# **BAB II KAJIAN PUSTAKA**

## **Kajian Topik**

Kajian topik adalah proses eksplorasi mendalam terhadap suatu subjek tertentu untuk memahami konsep, ruang lingkup, dan permasalahan yang ada di dalamnya. Eksplorasi mendalam terhadap literatur existing bertujuan memetakan lanskap permasalahan, melacak perkembangan riset terdahulu, dan menemukan celah pengetahuan yang potensial untuk dikembangkan melalui penelitian baru. Selain itu, kajian ini membantu merumuskan pertanyaan penelitian yang lebih terarah dan memastikan bahwa topik yang dipilih relevan serta memiliki kontribusi ilmiah maupun praktis. Proses ini juga memungkinkan peneliti untuk memahami pendekatan dan metode yang paling sesuai dalam menyelesaikan masalah yang diangkat, sehingga penelitian yang dilakukan menjadi lebih terstruktur dan berdampak.

### **Perusahan CV. Smartindo Telekom**

CV. Smartindo Telekom adalah sebuah perusahaan yang bergerak dibidang distributor telekomunikasi produk dari PT. Smartfren Telcom, dimana perusahaan menawarkan jasa untuk penjualan, retail ke outlet – outlet yang bekerja sama dengan perusahan. Perusahaan menawarkan produk – produk smartfen ke mitra outlet dan menjaga hubungan baik dengan mitra – mitra outlet tersebut.

Perusahan CV. Smartindo Telekom didirikan oleh Bapak Putra Yudha pada tahun 2015. Perusahan adalah anak perusahan dari PT. Wahana Putra Yudha yang merupakan pusat dari perusahan CV. Smartindo Telekom. CV. Smartindo Telekom didirikan atas dasar kerja sama antara CV. Smartfren Telecom dengan PT. Wahana Putra Yudha pada tahun 2015, dan diberikan kepercayaan untuk melayani wilayah distribusi antara lain Medan, Deli Serdang, Kabanjahe dan Sidikalang yang disebut area Cluster 1. Saat ini perusahaan sudah berkembang dan melakukan pengambilan area (take over) wilayah area Cluster 4 dan beberapa area cluster lainnya yang sekarang wilayah CV. Smartindo Telekom ampir mencakup keseluruhan pulau dari Sumatera.

### **Fraud**

Berdasarkan pernyataan dalam Statement on Auditing Standards (SAS) No. 99, kecurangan (fraud) diartikan sebagai tindakan yang dilakukan secara sengaja yang mengakibatkan penyajian informasi yang menyesatkan atau keliru secara material dalam laporan keuangan yang sedang diaudit. [REF CODE ASD14] Sementara itu, dalam Black’s Law Dictionary, kecurangan dijelaskan sebagai segala bentuk tindakan yang dirancang oleh seseorang dengan tujuan memperoleh keuntungan pribadi melalui penyesatan atau manipulasi terhadap orang lain. Tindakan ini mencakup berbagai cara yang tidak wajar, tersembunyi, licik, maupun yang tidak terduga, yang pada akhirnya menyesatkan atau merugikan pihak lain. [REF CODE ASD1]

Definisi kecurangan oleh Ikatan Akuntan Indonesia melalui Standar Profesional Akuntan Publik (SPAP) bagian 316 memberikan penjelasan mengenai bentuk-bentuk kecurangan dalam akuntansi sebagai berikut:

1. Kecurangan dalam pelaporan keuangan, yakni penyimpangan yang disengaja berupa manipulasi atau penghilangan informasi penting dalam laporan keuangan yang bertujuan menyesatkan para pengguna laporan tersebut. [REF CODE ASD2]
2. Penyalahgunaan aset, yaitu tindakan yang tidak semestinya terhadap aset perusahaan, seperti pencurian atau penggelapan, yang mengakibatkan laporan keuangan tidak merefleksikan kondisi sebenarnya sesuai prinsip akuntansi yang berlaku. [REF CODE ASD2]

Secara umum, tindakan kecurangan dapat diklasifikasikan ke dalam dua kategori utama, salah satunya adalah:

1. Kecurangan dari dalam organisasi (internal fraud), yaitu bentuk penyimpangan yang dilakukan oleh pihak internal perusahaan, seperti penyalahgunaan wewenang, penggelapan aset, atau pelanggaran terhadap kebijakan dan prosedur yang berlaku, yang berakibat pada timbulnya kerugian bagi perusahaan.. Dimana bentuk-bentuk kecurangan yang berasal dari dalam organisasi mencakup berbagai praktik seperti transaksi terselubung yang luput dari pencatatan resmi, penggelapan aset oleh pegawai internal, manipulasi kewajiban perpajakan, penyajian laporan keuangan yang menyimpang dari kondisi riil perusahaan, serta pemanfaatan informasi privileged untuk keuntungan pribadi dalam perdagangan saham.
2. Kecurangan eksternal (external fraud), yaitu kerugian yang ditimbulkan akibat tindakan curang yang dilakukan oleh pihak di luar perusahaan, seperti individu atau entitas yang tidak memiliki hubungan langsung dengan internal organisasi. Contoh dari jenis kecurangan ini mencakup aksi pencurian, pemalsuan dokumen (forgery), peretasan sistem (hacking), serta berbagai bentuk manipulasi lainnya yang dilakukan oleh pihak eksternal.. [REF CODE ASD3]

Dalam konteks yang lebih luas, tindakan kecurangan mengakomodasi beraneka ragam teknik dan strategi yang dirancang dengan level kepandaian dan perencanaan yang sistematis. Motivasi di balik aktivitas ini berpusat pada ambisi memperoleh advantage finansial dengan cara menyesatkan target melalui penyampaian data atau keterangan yang tidak akurat. Tidak terdapat definisi tunggal yang secara mutlak dapat menggambarkan makna dari kecurangan, karena tindakan ini mencakup berbagai bentuk manipulasi, tipu daya, maupun cara-cara tidak etis dan menyimpang. Secara umum, kecurangan merupakan bentuk penipuan yang memiliki unsur-unsur sebagai berikut:

1. Adanya suatu pernyataan atau penyampaian informasi,
2. Berkaitan dengan hal yang bersifat penting atau signifikan,
3. Informasi tersebut tidak sesuai dengan kenyataan,
4. Dilakukan secara sengaja atau tanpa kehati-hatian
5. Menyebabkan pihak lain bertindak berdasarkan informasi tersebut
6. Serta mengakibatkan kerugian bagi pihak yang menjadi korban **[REF CODE ASD10]**

Dalam konteks ini, tindakan yang dimaksud dapat mengarah pada perilaku curang. Di lingkungan kerja, praktik kecurangan kerap terjadi dalam berbagai bentuk. Menurut The Association of Certified Fraud Examiners (ACFE), terdapat tiga kategori utama kecurangan yang umum ditemukan, yaitu sebagai berikut:

1. Korupsi (*Corruption)*

Fenomena korupsi terwujud melalui ragam modus mulai dari transaksi penyuapan, adanya konflik interest yang tidak dikelola dengan baik, penerimaan kompensasi tidak resmi, hingga praktik ekstorsi untuk keuntungan finansial. Korupsi merupakan perilaku menyimpang yang dilakukan oleh seseorang yang memiliki wewenang, di mana ia secara tidak sah menggunakan posisi atau pengaruhnya untuk memperoleh keuntungan pribadi maupun untuk orang lain, dengan cara yang bertentangan dengan kewajiban dan etika jabatan. Menurut data dari Association of Certified Fraud Examiners (ACFE), korupsi menyumbang sekitar 10% dari seluruh kasus kecurangan yang terjadi di lingkungan kerja, dan sekitar 90% dari total kerugian akibat korupsi berasal dari praktik penyuapan.

1. Penyalahgunaan Aset (Asset Misappropriation)

Jenis kecurangan ini merupakan bentuk yang paling sering ditemukan dalam praktik, di mana pelaku memanfaatkan atau mengambil aset milik organisasi untuk kepentingan pribadi. Berdasarkan temuan dari Association of Certified Fraud Examiners (ACFE), sekitar 85% kasus kecurangan yang diteliti masuk dalam kategori ini. Penyalahgunaan dapat dilakukan secara langsung maupun tidak langsung, dan biasanya melibatkan elemen-elemen seperti uang tunai, rekening giro, persediaan barang, peralatan operasional, perlengkapan kantor, bahkan data atau informasi penting perusahaan.

1. Kecurangan Laporan Keuangan (Financial Statement Fraud)

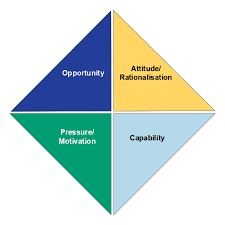
Jenis kecurangan ini umumnya dilakukan oleh pihak manajemen, di mana laporan keuangan sengaja dimanipulasi untuk memberikan gambaran yang menyesatkan mengenai kondisi keuangan perusahaan. Meskipun semua bentuk kecurangan pada akhirnya mempengaruhi pelaporan keuangan, suatu tindakan hanya dikategorikan sebagai skema kecurangan laporan keuangan apabila modifikasi tersebut memberikan keuntungan langsung maupun tidak langsung kepada pelakunya. Artinya, laporan keuangan tidak hanya digunakan untuk menutupi tindakan curang lainnya, tetapi menjadi alat utama dalam upaya manipulasi yang dilakukan. **[REF CODE ASD9]**

Kecurangan merupakan tindakan tidak etis yang dilakukan secara sengaja dan sistematis, dengan maksud memperoleh keuntungan pribadi melalui manipulasi atau penyimpangan, yang pada akhirnya merugikan pihak lain secara langsung maupun tidak langsung. Menurut teori Fraud Triangle dari Donald Cressey, fraud bisa terwujud dikarena kombinasi preassure (tekanan), opportunity (kesempatan), dan rationalization (rasionalisasi). Pencegahan fraud dapat dilakukan melalui penguatan pengendalian internal, audit berkala, penggunaan teknologi pemantauan, serta membangun budaya kerja yang menjunjung tinggi integritas dan transparansi.

#### **Fraud Diamond**

**Fraud Diamond merupakan pengembangan lanjutan dari konsep Fraud Triangle yang awalnya diperkenalkan oleh Donald Cressey. Pada tahun 2004, David T. Wolfe dan Dana R. Hermanson memperluas model ini dengan menambahkan elemen keempat, yaitu Capability (Kemampuan). Model Fraud Diamond terdiri dari empat komponen utama yang saling berhubungan, yaitu: Pressure (Tekanan), Opportunity (Kesempatan), Rationalization (Rasionalisasi), dan Capability (Kemampuan). Tekanan mengacu pada dorongan internal yang membuat seseorang terdorong untuk melakukan kecurangan, yang biasanya berakar dari masalah finansial, beban pekerjaan yang terlalu tinggi, atau kebutuhan pribadi yang mendesak.** [REF CODE ASD4]. Selain tekanan, kesempatan juga menjadi faktor kunci. Kesempatan muncul ketika ada kelemahan dalam pengendalian internal, kurangnya pengawasan, atau akses berlebih yang dimiliki individu tertentu. "Opportunity allows the fraudster to commit the act without being caught, often due to weak controls or lack of oversight" [REF CODE ASD4].

Lebih jauh, pelaku kecurangan biasanya mencari pembenaran moral melalui rasionalisasi, agar tindakan mereka terasa sah atau setidaknya dapat diterima secara etika. Misalnya, seseorang mungkin merasa bahwa mereka hanya "meminjam sementara" atau "perusahaan sudah cukup kaya dan ini tidak akan merugikan siapa pun." "Rationalization helps the fraudster justify the unethical behavior, making it seem acceptable in their mind" [REF CODE ASD4]. Namun, elemen yang membuat Fraud Diamond lebih komprehensif dibandingkan model sebelumnya adalah **Capability (Kemampuan)**. Elemen ini menekankan bahwa meskipun seseorang memiliki tekanan, kesempatan, dan rasionalisasi, kecurangan besar tidak akan terjadi tanpa kemampuan tertentu. Kemampuan ini mencakup posisi strategis, pengetahuan teknis, kepercayaan diri tinggi, dan kemampuan memengaruhi orang lain agar turut membantu atau menutupi kecurangan. "Even if the first three elements are present, the fraud will not occur unless the person has the capability to commit and conceal it" [REF CODE ASD4].



Fraud Diamond juga dapat digunakan sebagai kerangka kerja untuk mendeteksi potensi kecurangan. Berikut beberapa pendekatan yang bisa dilakukan:

1. **Mendeteksi Tekanan (Pressure)**
   1. **Analisis gaya hidup karyawan**: Perubahan gaya hidup drastis yang tidak sesuai dengan penghasilan bisa menjadi tanda tekanan finansial.
   2. **Pantau target kerja yang tidak realistis**: Tekanan dari target yang terlalu tinggi dapat mendorong karyawan mencari jalan pintas dengan melakukan kecurangan.
   3. **Tinjau konflik pribadi atau profesional**: Karyawan yang mengalami masalah pribadi atau konflik dengan perusahaan bisa lebih rentan melakukan kecurangan.
   4. Cabang mungkin merasa tertekan untuk memenuhi target penjualan atau keuntungan tertentu yang ditetapkan oleh manajemen pusat. Hal ini dapat meningkatkan kemungkinan adanya perilaku kecurangan untuk mencapai target tersebut.
2. **Menganalisis Kesempatan (Opportunity)**
   1. **Evaluasi pengendalian internal**: Perusahaan perlu memastikan sistem pengendalian internal kuat dan meminimalkan celah yang bisa dieksploitasi.
   2. **Rotasi kerja dan audit mendadak**: Melakukan rotasi posisi karyawan dan audit mendadak dapat mengurangi kesempatan untuk menyembunyikan kecurangan.
   3. **Pemantauan akses dan otorisasi**: Pastikan hanya orang dengan otorisasi yang memiliki akses ke sistem keuangan atau data penting.
   4. Jika kebijakan internal lemah atau laporan keuangan tidak diawasi dengan baik, maka kesempatan untuk melakukan kecurangan meningkat, seperti memanipulasi transaksi atau laporan keuangan.
3. **Mengidentifikasi Rasionalisasi (Rationalization)**
   1. **Kaji budaya etika di perusahaan**: Budaya yang mendukung integritas dan kejujuran akan mengurangi kemungkinan karyawan membenarkan tindakan curang.
   2. Pembenaran yang diberikan oleh cabang seperti "untuk memenuhi target" atau "persaingan bisnis" dapat menjadi alasan bagi manajer untuk melakukan kecurangan atau penyalahgunaan.
   3. **Lakukan wawancara atau survei kepuasan kerja**: Ketidakpuasan karyawan bisa memicu rasionalisasi. Perusahaan harus memahami alasan di balik ketidakpuasan ini.
4. **Mengukur Kemampuan (Capability)**
   1. **Identifikasi posisi kunci yang rentan**: Orang dengan posisi strategis, seperti manajer keuangan atau IT, memiliki potensi lebih besar melakukan kecurangan karena akses dan pengetahuannya.
   2. **Pantau individu dengan pengaruh besar**: Karyawan yang memiliki pengaruh besar dan karisma tinggi bisa memanipulasi rekan kerja untuk ikut serta atau menutupi tindakan mereka.
   3. **Perhatikan karyawan dengan keterampilan teknis tinggi**: Individu dengan kemampuan teknologi canggih lebih mampu mengeksploitasi celah sistem.
   4. Kemampuan manajer cabang untuk mengelola operasi dan keuangan cabang dapat menjadi faktor penting dalam menentukan potensi terjadinya kecurangan. Cabang yang memiliki manajer dengan kemampuan rendah atau kurang diawasi lebih rentan terhadap potensi fraud.

### **Level Fraud pada Jabatan**

Fraud (kecurangan) dalam konteks perusahaan cabang bisa terjadi di berbagai level jabatan antara lain sales, admin, dan manajer. Berikut penjelasannya:

#### **Fraud oleh Sales**

Sales biasanya bertanggung jawab pada penjualan dan hubungan dengan pelanggan. Beberapa jenis kecurangan yang umum dilakukan sales:

1. **Penggelapan hasil penjualan,** dimana sales menerima uang dari pelanggan tetapi hanya melaporkan sebagian atau bahkan tidak melaporkan sama sekali.
2. **Mark-up harga,** dimana menaikkan harga jual di luar ketentuan perusahaan lalu mengambil selisihnya.
3. **Pemalsuan order,** dimana membuat pesanan fiktif untuk mencapai target dan mendapatkan komisi lebih besar.
4. **Manipulasi diskon atau promo, dimana** memberikan diskon tidak sah ke pelanggan dengan imbalan pribadi.

#### **2.1.3.2. Fraud oleh Admin**

Admin punya akses ke data dan dokumen penting. Kecurangan yang bisa terjadi di posisi ini:

1. **Pemalsuan laporan keuangan**, memanipulasi data penjualan, pengeluaran, atau stok agar terlihat sesuai target atau menutupi kerugian.
2. **Penggelapan uang kas kecil** dengan Mengambil uang kas dengan membuat kwitansi palsu atau mencatat biaya yang tidak ada.
3. **Manipulasi stok** mencatat barang sudah dikirim padahal belum, lalu menjual barang secara pribadi.

#### **Fraud Manajer Cabang**

Manajer cabang biasanya punya otoritas lebih besar. Jenis kecurangan yang mungkin dilakukan:

1. ***Kickback* (suap balik) yaitu** Kerja sama gelap dengan supplier atau pihak ketiga untuk mendapatkan komisi ilegal.
2. **Manipulasi target** dengan Menggelembungkan angka penjualan agar cabang terlihat sukses dan mendapatkan bonus lebih besar.
3. **Rekayasa pengeluaran** dengan melaporkan biaya operasional lebih besar dari kenyataan lalu mengambil selisih dana.
4. **Nepotisme atau favoritisme** dengan mempekerjakan orang terdekat yang tidak kompeten demi keuntungan pribadi.

### **Machine Learning**

Machine Learning (ML) adalah salah satu cabang dari kecerdasan buatan (Artificial Intelligence) yang memungkinkan sistem komputer untuk mempelajari pola dari data secara otomatis dan membuat prediksi atau keputusan tanpa harus diprogram secara rinci untuk setiap tugasnya. ML menggunakan berbagai algoritma statistik dan matematika untuk menemukan pola-pola tersembunyi dalam data, yang kemudian diaplikasikan dalam pengambilan keputusan atau prediksi pada data baru. Proses ini biasanya melibatkan tahap pelatihan model dengan menggunakan kumpulan data yang besar agar model dapat memahami hubungan yang kompleks dan pola yang ada.

Proses ML dimulai dengan pengumpulan data yang relevan, kemudian data tersebut diproses dan dibersihkan agar siap digunakan dalam pelatihan model. Selanjutnya, algoritma ML dipilih dan diterapkan untuk melatih model menggunakan data tersebut. Model yang sudah dilatih kemudian diuji dengan data baru untuk mengevaluasi performanya. Jika hasilnya memuaskan, model dapat digunakan untuk prediksi atau klasifikasi dalam aplikasi nyata.

Machine Learning beroperasi dengan cara menganalisis data untuk mengidentifikasi pola serta hubungan matematis antara variabel input dan output. Proses ini diawali dengan pelatihan model menggunakan data yang sudah diberi label (supervised learning) ataupun data yang belum diberi label (unsupervised learning). Setelah pelatihan, model mampu menggeneralisasi pola-pola tersebut sehingga dapat melakukan prediksi atau pengambilan keputusan pada data baru yang belum pernah diproses sebelumnya. Secara sederhana, algoritma ML dilatih dengan contoh input-output, misalnya pasangan data (2,10), (5,19), dan (9,31). Algoritma akan mencari fungsi matematis yang menghubungkan input dan output, seperti o = 3 × i + 4 o=3×i+4. Setelah model terlatih, jika diberikan input baru, misalnya 7, model dapat memprediksi outputnya, yaitu 25. Keakuratan prediksi ini bergantung pada kualitas dan kuantitas data yang digunakan selama pelatihan

Jenis - Jenis Machine Learning yang biasa digunakan adalah sebagai berikut:

1. Model Supervised Learning dilatih dengan menggunakan data yang sudah memiliki label, yang berarti setiap data input disertai dengan output yang telah diketahui sebelumnya. Melalui proses ini, model mempelajari hubungan antara input dan output agar dapat memprediksi hasil pada data yang belum pernah ditemui. Contoh penerapannya antara lain adalah klasifikasi email spam dan estimasi harga properti.
2. Model Unsupervised Learning menggunakan data tanpa label, dengan tujuan utama mengidentifikasi pola, struktur tersembunyi, atau mengelompokkan data menjadi beberapa kategori. Contohnya termasuk segmentasi pelanggan dan identifikasi anomali dalam data.
3. Model Reinforcement Learning belajar melalui proses interaksi dengan lingkungan di mana model menerima umpan balik berupa hadiah (reward) atau hukuman (penalti). Tujuannya adalah untuk mengoptimalkan pengambilan keputusan agar memperoleh reward sebesar mungkin, dan metode ini sering diterapkan dalam bidang robotika serta pengembangan permainan komputer.

Beberapa algoritma yang sering dipakai dalam Machine Learning adalah berikut ini:

1. Decision Tree yaitu Algoritma yang memanfaatkan struktur berbentuk pohon untuk mengambil keputusan dengan menganalisis fitur-fitur yang ada pada data.
2. Random Forest yaitu Kombinasi dari banyak decision tree yang bekerja secara paralel untuk meningkatkan akurasi dan mengurangi overfitting.
3. Logistic Regression yaitu Sebuah metode statistik yang umum digunakan untuk melakukan klasifikasi dengan dua kategori atau kelas.
4. Support Vector Machine (SVM) yaitu Algoritma yang berfungsi menentukan garis pemisah (hyperplane) optimal guna membedakan kelompok data yang berbeda.
5. XGBoost: Algoritma boosting yang efisien dan sering digunakan dalam kompetisi ML karena performanya yang tinggi.

Machine Learning telah banyak diterapkan di berbagai bidang, seperti kesehatan, keuangan, pemasaran, dan lain-lain. Contohnya, dalam bidang kesehatan, ML digunakan untuk memprediksi tingkat kasus penyakit menular, membantu pengambilan keputusan dalam program vaksinasi dan pelayanan kesehatan. Studi yang dilakukan oleh **[REF CODE ASD6]** mengaplikasikan beragam algoritma machine learning, termasuk decision tree, random forest, logistic regression, SVM, dan XGBoost, untuk memprediksi tingkat kejadian penyakit di Indonesia. Hasil penelitian ini memberikan dukungan penting bagi para pembuat kebijakan agar dapat merancang kebijakan kesehatan dengan lebih cepat dan tepat.

Selain itu, ML juga digunakan untuk prediksi diagnosis penyakit seperti diabetes menggunakan algoritma neural network yang dioptimasi dengan algoritma evolusi untuk meningkatkan akurasi prediksi **[REF CODE ASD7]**. Dalam bidang lain, ML digunakan untuk prediksi ketepatan penempatan karir dengan model klasifikasi seperti Random Forest dan SVM, yang dievaluasi berdasarkan akurasi prediksi **[REF CODE ASD8]**. Oleh karena itu machine Learning adalah teknologi yang sangat penting dalam era digital saat ini karena kemampuannya dalam mengolah data besar dan memberikan insight prediktif yang dapat mendukung pengambilan keputusan otomatis dan cepat. Dengan berbagai algoritma yang tersedia, ML dapat disesuaikan untuk berbagai kebutuhan aplikasi, mulai dari klasifikasi, regresi, hingga clustering.

### **Random Forest**

Random Forest merupakan sebuah algoritma machine learning yang termasuk dalam kategori ensemble learning, dimana banyak pohon keputusan (decision tree) digabungkan untuk meningkatkan ketepatan prediksi sekaligus mengurangi kemungkinan overfitting. Algoritma ini dikembangkan oleh Leo Breiman dan Adele Cutler, dan dapat diterapkan pada berbagai masalah baik klasifikasi maupun regresi.

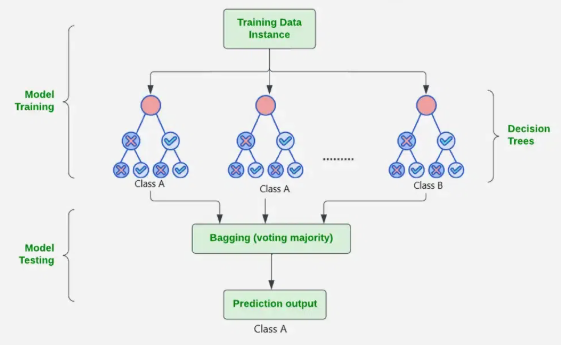
Random Forest adalah kumpulan pohon keputusan yang masing-masing dibangun dari data dan fitur acak, sehingga menghasilkan model yang kuat dan stabil. Random Forest juga telah digunakan untuk meningkatkan hasil klasifikasi pada deteksi pasien diabetes, dengan akurasi mencapai 95,45% setelah penerapan normalisasi data. **[REF CODE ASD13]** Struktur gambar Random Forest adalah representasi visual dari banyak pohon keputusan yang berdiri sendiri, dengan proses agregasi hasil sebagai komponen penting dalam menghasilkan prediksi akhir. Random Forest merupakan salah satu algoritma klasifikasi yang banyak digunakan karena mampu menghasilkan prediksi dengan akurasi tinggi, seperti yang dibuktikan pada penelitian klasifikasi cuaca dan banjir di Indonesia. **[REF CODE ASD12]**

Random Forest menciptakan sejumlah besar pohon keputusan dengan menggunakan teknik bootstrap sampling dan pemilihan fitur secara acak untuk meningkatkan presisi sekaligus mengurangi risiko overfitting. Setiap pohon terdiri dari struktur decision tree yang meliputi root node, node internal, cabang, dan daun yang menghasilkan prediksi. Hasil akhir ditentukan melalui mekanisme voting mayoritas pada kasus klasifikasi atau rata-rata pada kasus regresi dari semua pohon yang tergabung dalam model.

Pada Random Forest, setiap pohon dibangun dengan data bootstrap sampling dan pada setiap split hanya dipertimbangkan subset acak fitur, sehingga pohon-pohon yang terbentuk berbeda dan tidak berkorelasi tinggi. Namun, proses pembentukan pohon pada masing-masing decision tree tetap mengikuti prinsip di atas, yaitu mencari fitur terbaik berdasarkan kriteria seperti Gini atau Entropy untuk memisahkan data secara rekursif hingga mencapai kondisi berhenti.

Selain bootstrap sampling, Random Forest juga menerapkan random feature selection pada setiap node split dalam decision tree. Alih-alih mempertimbangkan semua fitur, algoritma hanya memilih subset fitur secara acak untuk menentukan split terbaik. Hal ini menurunkan korelasi antar pohon dan meningkatkan keragaman model sehingga mengurangi varians dan risiko overfitting.

Struktur dan komponen gambar Random Forest dapat dipahami sebagai kumpulan (ensemble) dari banyak pohon keputusan (decision trees) yang bekerja secara paralel dan independen, Selanjutnya, hasil dari setiap pohon tersebut digabungkan untuk menghasilkan prediksi akhir yang lebih tepat dan konsisten. Berikut ini adalah ilustrasi struktur dari algoritma Random Forest:



Berikut ini adalah penjabaran secara sederhana dan singkat dengan flowchart sederhana pada algoritma random forest ialah:

1. Mulai
2. Ambil dataset asli
3. Lakukan bootstrap sampling (ambil sampel acak dengan pengembalian dari dataset asli untuk membuat beberapa subset data)
4. Bangun decision tree untuk setiap subset data
5. Pada setiap node, pilih subset fitur secara acak
6. Tentukan split terbaik berdasarkan fitur yang dipilih
7. Bangun pohon sampai kriteria penghentian tercapai (misal max depth atau node minimum)
8. Ulangi langkah 3-4 untuk membentuk banyak pohon (n\_trees)
9. Untuk data baru, lakukan prediksi dengan setiap pohon
10. Gabungkan hasil prediksi dari semua pohon
11. Untuk klasifikasi: voting mayoritas
12. Untuk regresi: rata-rata prediksi
13. Output hasil prediksi akhir
14. Selesai

Berikut adalah langkah – langkah algoritma Random Forest adalah sebagai berikut:

#### **Bootstrap Sampling**

Bootstrap sampling dalam Random Forest adalah proses pengambilan sampel acak dari dataset asli dengan pengembalian (sampling dengan replacement) untuk membentuk beberapa subset data yang masing-masing digunakan untuk melatih satu pohon keputusan secara independen. Ini berarti beberapa data dapat terpilih lebih dari satu kali dalam satu subset, sementara beberapa data lain mungkin tidak terpilih sama sekali. Cara kerja dari bootstrap sampling dalam random forest adalah sebagai berikut:

1. Dari dataset pelatihan yang tersedia, dilakukan pengambilan sampel secara acak dengan pengembalian untuk membentuk subset data baru yang berukuran sama dengan dataset asli.
2. Setiap subset ini menjadi data latih untuk satu pohon keputusan dalam hutan acak. Karena adanya pengembalian, setiap pohon mendapatkan data yang sedikit berbeda, menciptakan variasi antar pohon.
3. Variasi ini penting untuk mengurangi korelasi antar pohon dan menghindari overfitting, sehingga hasil agregasi prediksi menjadi lebih stabil dan akurat.
4. Selain itu, sekitar sepertiga data yang tidak terpilih dalam bootstrap disebut out-of-bag (OOB), yang bisa digunakan untuk validasi model tanpa perlu data terpisah

Misal dataset asli berisi 1000 sampel. Dengan bootstrap sampling, untuk setiap pohon diambil 1000 sampel secara acak dengan pengembalian. Jadi, satu pohon mungkin memiliki 700 sampel unik, dan 300 sampel terduplikasi, sementara 300 lainnya tidak terpilih. Pohon lain akan memiliki subset yang berbeda, sehingga menghasilkan model yang lebih robust saat hasilnya digabungkan. Dengan demikian, bootstrap sampling adalah fondasi penting dalam Random Forest yang memungkinkan pembentukan banyak pohon keputusan yang berbeda dan meningkatkan performa model secara keseluruhan

#### **Pembentukan Decision Tree**

Pembentukan Decision Tree adalah proses konstruksi pohon keputusan yang dimulai dari root node (akar) yang berisi seluruh dataset, kemudian secara rekursif membagi data menjadi subset-subset yang lebih homogen berdasarkan fitur terbaik yang dipilih pada setiap langkah pemisahan (splitting). Proses ini terus berlangsung sampai memenuhi kondisi penghentian tertentu, seperti ketika semua data dalam subset tergolong ke dalam kelas yang seragam, kedalaman pohon telah mencapai batas maksimal, atau jumlah data dalam subset menjadi terlalu sedikit untuk diproses lebih lanjut. Berikut ini adalah langkah – langkah pembentukan pohon keputusan secara detail yaitu:

1. Mulai dari Root Node

Seluruh data pelatihan berada pada root node

1. Pemilihan Fitur Terbaik untuk Memisahkan Data (Splitting)

Pada setiap node, algoritma mencari fitur dan nilai ambang (threshold) yang paling efektif untuk membagi data menjadi subset yang lebih homogen terhadap target kelas. Kriteria pemilihan fitur biasanya menggunakan metrik:

1. Gini Impurity

di mana merepresentasikan proporsi kelas ke- *i* pada node *t*. Semakin rendah nilai Gini, maka tingkat kemurnian node tersebut semakin tinggi.

1. Entropy dan Information Gain:

Information Gain adalah pengurangan entropy setelah pemisahan:

di mana *k* adalah subset hasil pemisahan, jumlah sampel di subset, dan *N* jumlah sampel di parent node. Algoritma memilih fitur dan threshold dengan nilai Gini terendah atau Information Gain tertinggi.

1. Membagi Data ke Subset Berdasarkan Fitur Terpilih

Data dibagi ke cabang-cabang baru sesuai dengan hasil pengujian fitur (misalnya, nilai fitur ≤ threshold ke cabang kiri, > threshold ke cabang kanan).

1. Rekursi pada Subset Data

Proses pemilihan fitur dan pembagian data diulang pada setiap node anak yang baru terbentuk.

1. Kriteria Berhenti

Proses pembentukan pohon berhenti jika:

1. Seluruh data pada node tersebut termasuk dalam satu kelas yang identik (node yang homogen) .
2. Pohon telah mencapai tingkat kedalaman maksimum yang telah ditetapkan sebelumnya.
3. Jumlah data dalam node kurang dari batas minimum.
4. Tidak ada peningkatan signifikan dalam pemisahan data.
5. Penentuan Label pada Node Daun (Leaf Node)

Pada node daun, kelas yang dipilih biasanya adalah kelas mayoritas dari data yang berada di node tersebut (untuk klasifikasi), atau rata-rata nilai target (untuk regresi).

Decision tree dimulai dari root dan membandingkan nilai fitur untuk menelusuri cabang hingga mencapai simpul daun. Algoritma memilih fitur terbaik berdasarkan metrik seperti Gini Impurity, Entropy, dan Information Gain untuk memisahkan data secara rekursif. Setiap node internal berisi kondisi uji atribut, dan daun merepresentasikan kelas target [REF CODE ASD5]. Random Forest membangun banyak decision tree dengan subset data dan fitur acak untuk mengurangi korelasi antar pohon. Dengan demikian, pembentukan decision tree adalah proses rekursif pemilihan fitur terbaik untuk memisahkan data secara bertingkat, menghasilkan struktur pohon yang dapat digunakan untuk klasifikasi atau regresi.

#### **Prediksi Setiap Pohon**

Setelah pohon terbentuk, setiap pohon memberikan prediksi untuk data baru. Untuk klasifikasi, pohon memberikan kelas; untuk regresi, pohon memberikan nilai numerik.

#### **Voting atau Averaging untuk Prediksi Akhir**

Setelah itu, untuk klasifikasi, prediksi akhir diambil berdasarkan voting mayoritas dari hasil prediksi semua pohon (kelas yang paling banyak dipilih). Pada regresi, prediksi akhir adalah rata – rata dari semua pohon. Rumus prediksi regresi Random Forest antara lain:

Dimana:

= prediksi akhir

*N* = total pohon yang ada dalam hutan

= prediksi yang dihasilkan oleh pohon ke-*i* untuk input

### **Metrik Evaluasi Klasifikasi**

Ada empat metrik evaluasi utama yang sering digunakan dalam klasifikasi, termasuk pada model seperti Random Forest: Accuracy, Precision, Recall, dan F1-Score, berikut adalah empat metrik evaluasi:

#### **Accuracy**

Accuracy merupakan salah satu metrik yang paling sering dipakai untuk menilai kemampuan sebuah model dalam mengklasifikasikan data secara keseluruhan. Nilai accuracy dihitung dari perbandingan antara jumlah prediksi yang tepat (true positive ditambah true negative) dengan total data yang dianalisis. Dengan kata lain, akurasi menggambarkan seberapa besar persentase prediksi yang benar dari seluruh prediksi yang dibuat oleh model.

Accuracy cocok digunakan jika distribusi kelas positif dan negatif pada dataset seimbang. Namun, jika dataset tidak seimbang (misal, data positif jauh lebih sedikit dari negatif), accuracy bisa menyesatkan karena model bisa saja mendapatkan nilai accuracy tinggi hanya dengan memprediksi mayoritas kelas saja.

#### **Precision**

Precision menunjukkan tingkat ketepatan model dalam memprediksi kelas positif. Nilai precision dihitung sebagai perbandingan antara jumlah prediksi positif yang benar (true positive) dengan keseluruhan prediksi positif yang dilakukan (true positive ditambah false positive).

Precision penting untuk meminimalkan prediksi positif yang salah (false positive), misalnya pada deteksi spam, di mana kita ingin memastikan email yang diklasifikasikan sebagai spam benar-benar spam.

#### **Recall**

Recall, yang juga dikenal sebagai sensitivity, menilai kemampuan model dalam mengidentifikasi seluruh data positif dengan tepat. Recall dihitung sebagai perbandingan antara jumlah prediksi positif yang benar (true positive) dengan total data yang memang benar positif (true positive ditambah false negative). Tujuannya adalah untuk meminimalkan jumlah kesalahan berupa false negative..

#### **F1-Score**

F1-score merupakan ukuran evaluasi yang menggabungkan precision dan recall dalam satu nilai dengan menggunakan rata-rata harmonis. Metrik ini sangat berguna untuk mengevaluasi performa model pada data yang tidak seimbang, karena memberikan penilaian yang seimbang antara ketepatan dan kemampuan model dalam menangkap seluruh kasus positif.

F1-score akan memberikan gambaran yang lebih baik tentang performa model jika ingin memperhatikan baik precision maupun recall secara bersamaan.

### **Google Form**

Google Forms merupakan alat berbasis web yang sangat efektif untuk membuat dan menyebarkan kuisioner secara online. Ketika responden mengisi kuisioner, data secara otomatis tersimpan dan terorganisir dalam Google Sheets, yaitu spreadsheet online yang terintegrasi langsung dengan Google Forms. Data yang masuk dari Google Forms langsung tersimpan rapi dalam Google Sheets tanpa perlu input manual, menghemat waktu dan mengurangi kesalahan. Proses pengumpulan dan pengolahan data hasil pengisian yaitu.

1. Penyusunan instrumen dilakukan melalui platform Google Forms, di mana peneliti merancang berbagai jenis pertanyaan seperti pilihan ganda, jawaban singkat, serta skala Likert untuk mengakomodasi variasi data yang dibutuhkan.
2. Distribusi Kuisioner disebarkan melalui link yang bisa dibagikan via email, media sosial, atau platform lainnya.
3. Pengumpulan Data Responden dengan mengisi kuisioner secara online, dan data otomatis tersimpan di Google Sheets.
4. Pengelolaan Data di Google Sheets dibersihkan dan diorganisir menggunakan fitur spreadsheet, siap untuk dianalisis lebih lanjut.

### **Google Sheet**

Google Sheets merupakan layanan spreadsheet berbasis daring yang memfasilitasi pengguna dalam menyusun, memodifikasi, serta mengatur data dalam bentuk tabel terdiri atas baris dan kolom, dengan akses yang dapat dilakukan secara real-time melalui internet. Berbeda dengan software spreadsheet tradisional seperti Microsoft Excel yang berbasis offline, Google Sheets menyimpan data secara otomatis di cloud sehingga data dapat diakses dan dikerjakan secara real-time dari berbagai perangkat yang terhubung internet. Fungsi utama dari google sheet adalah:

1. Pengolahan Data: Memungkinkan pengurutan, manipulasi, dan kalkulasi data menggunakan rumus matematika dan fungsi statistik
2. Visualisasi Data: Mendukung pembuatan grafik dan diagram yang membantu dalam analisis data dan pelaporan
3. Integrasi dengan Google Forms: Data hasil pengisian kuisioner Google Forms otomatis tersimpan di Google Sheets, memudahkan pengelolaan dan analisis data survei.
4. Penyimpanan Otomatis: Data tersimpan secara otomatis di cloud sehingga mengurangi risiko kehilangan data akibat gangguan perangkat

### **VS Code**

Visual Studio Code (VS Code) adalah editor kode sumber buatan Microsoft yang ringan namun kaya fitur, mendukung berbagai bahasa pemrograman, termasuk Dart dan Python. VS Code dilengkapi dengan IntelliSense untuk memberikan saran kode yang cerdas, debugger bawaan untuk membantu menemukan dan memperbaiki error, serta integrasi Git agar lebih mudah mengelola versi kode. Selain itu, tersedia banyak ekstensi yang bisa diinstal, seperti Flutter untuk pengembangan aplikasi Dart dan Python extension yang mendukung analisis kode, virtual environment, hingga Jupyter Notebook. Kombinasi fitur ini menjadikan VS Code pilihan populer bagi pengembang aplikasi.

### **Aplikasi Android**

Aplikasi Android adalah perangkat lunak yang dirancang untuk berjalan di sistem operasi Android, yang dikembangkan oleh Google. Aplikasi Android biasanya dibuat menggunakan bahasa pemrograman seperti Java, Kotlin, atau Dart (melalui Flutter), dan bisa diunduh melalui Google Play Store atau sumber lainnya. Aplikasi Android dalam konteks laporan audit adalah perangkat lunak yang dirancang untuk berjalan pada sistem operasi Android dan dapat digunakan untuk mendukung proses audit, baik internal maupun eksternal. Aplikasi ini dapat membantu auditor dalam mengumpulkan data, mencatat temuan, dan membuat laporan secara real-time langsung dari lapangan. Selain itu, aplikasi Android juga memungkinkan integrasi dengan berbagai sistem lain, seperti database perusahaan atau cloud storage, untuk mempercepat akses informasi dan memperkuat akurasi hasil audit. Dengan fitur seperti GPS, kamera, dan notifikasi, aplikasi audit berbasis Android juga dapat mempermudah pelacakan lokasi, dokumentasi bukti visual, serta pengingat jadwal audit. Penggunaan aplikasi semacam ini meningkatkan efisiensi, mengurangi kesalahan manual, dan mempercepat proses pelaporan hasil audit.

### **Bahasa Pemrograman Pyhton**

Python merupakan salah satu bahasa pemrograman yang banyak diminati karena struktur sintaksnya yang ringkas dan intuitif, sehingga cocok digunakan baik oleh pemula maupun programmer berpengalaman. Bahasa ini mendukung beragam pendekatan pemrograman, mulai dari paradigma objek, prosedural, hingga fungsional. Penerapannya sangat luas, mencakup pengembangan aplikasi web, analisis data, kecerdasan buatan, hingga proses otomasi berbagai tugas digital. Python memiliki pustaka (library) yang sangat kaya, seperti Django dan Flask untuk web, Pandas dan NumPy untuk analisis data, serta TensorFlow untuk machine learning. Kemampuan Python yang fleksibel serta dukungan dari komunitas global yang aktif menjadikannya sebagai salah satu bahasa pemrograman yang paling banyak digunakan di seluruh dunia.

### **Bahasa Pada Pemrograman Dart**

Dart merupakan bahasa pemrograman yang dikembangkan oleh Google dan dirancang khusus untuk mendukung pengembangan aplikasi lintas platform secara efisien, terutama pada sisi antarmuka pengguna seperti aplikasi mobile dan web. Bahasa ini memperoleh popularitas yang signifikan melalui integrasinya dengan framework Flutter, yang memungkinkan pengembang untuk membangun aplikasi Android, iOS, web, dan desktop dengan satu basis kode terpadu. Dart mengadopsi paradigma pemrograman berorientasi objek serta menawarkan sintaksis yang ringkas dan mudah dibaca, sehingga memfasilitasi proses pembelajaran dan pengembangan. Dart mendukung dua mode kompilasi, yakni Just-in-Time (JIT) yang dioptimalkan untuk proses pengembangan, serta Ahead-of-Time (AOT) yang meningkatkan performa eksekusi pada saat aplikasi dijalankan dalam lingkungan produksi.

### **Framework Flutter**

Flutter merupakan sebuah framework bersifat open-source yang dikembangkan oleh Google dengan tujuan memfasilitasi pengembangan aplikasi multiplatform menggunakan satu kode sumber yang sama. Melalui penggunaan Flutter, pengembang dapat membangun aplikasi yang dapat dijalankan pada berbagai sistem operasi seperti Android, iOS, web, serta platform desktop termasuk Windows, macOS, dan Linux. Flutter mengadopsi bahasa pemrograman Dart sebagai dasar pengembangannya dan dilengkapi dengan sejumlah fitur unggulan.

Salah satu fitur utama adalah Hot Reload, yang memungkinkan pengembang untuk mengimplementasikan perubahan pada kode secara real-time tanpa perlu melakukan kompilasi ulang atau memulai ulang aplikasi, sehingga sangat mendukung efisiensi dalam proses pengembangan perangkat lunak.. Selain itu, Flutter menyediakan widget yang kaya dan dapat dikustomisasi, sehingga memudahkan pembuatan antarmuka yang menarik dan responsif. Kecepatan, efisiensi, serta kemampuannya dalam menghasilkan performa tinggi menjadikan Flutter pilihan populer di kalangan pengembang aplikasi modern.

### **Unified Modeling Language (UML)**

Unified Modeling Language (UML) merupakan bahasa standar yang digunakan untuk menggambarkan sistem secara visual, terutama pada sistem yang berorientasi objek. UML berperan sebagai alat bantu penting dalam proses perancangan, visualisasi, dokumentasi, serta pengembangan sistem perangkat lunak maupun sistem kompleks lainnya. UML membantu menggambarkan bagaimana sebuah sistem bekerja melalui berbagai diagram yang menjelaskan struktur dan perilaku sistem tersebut secara terperinci. **UML REF**

UML pertama kali dikembangkan oleh Object Management Group (OMG) dan versi awalnya dirilis pada tahun 1997. UML menjadi bahasa pemodelan yang sangat penting dalam rekayasa perangkat lunak karena dapat mempermudah komunikasi antara pengembang sistem dan pengguna, serta memfasilitasi proses pengembangan perangkat lunak yang berkelanjutan. Fungsi utama dan tujuan penggunaan UML adalah sebagai berikut:

1. Memvisualisasikan sistem dari UML menyediakan diagram yang memudahkan pemahaman sistem secara keseluruhan dan bagian-bagiannya.
2. Mendokumentasikan sistem dengan UML membantu mendokumentasikan kebutuhan, desain, dan arsitektur sistem sehingga memudahkan pengembangan dan pemeliharaan.
3. Membantu analisis dan perancangan, UML digunakan untuk menganalisis kebutuhan sistem dan merancang solusi yang tepat sebelum implementasi coding.
4. Memfasilitasi komunikasi dan menjadi jembatan antara pengembang dan pengguna sistem, sehingga kebutuhan pengguna dapat diterjemahkan dengan jelas ke dalam desain sistem **UML REF2**

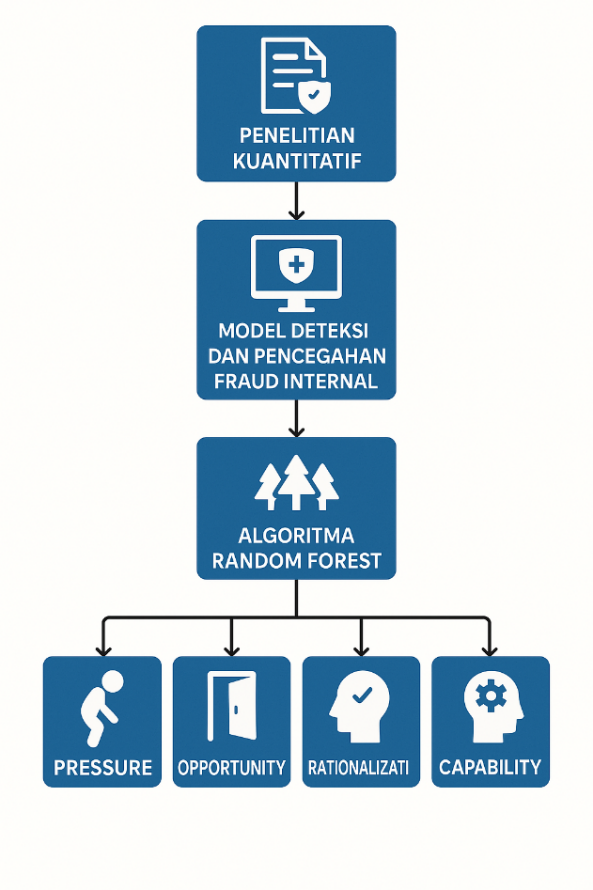
Komponen – komponen y:ang terdapat pada diagram UML adalah sebagai berikut ini:

1. Use Case Diagram dapat menjelaskan hubungan antara pengguna (aktor) dengan sistem, serta menggambarkan fungsi-fungsi utama yang disediakan oleh sistem tersebut.
2. Class Diagram dapat memvisualisasikan struktur kelas dalam sistem, lengkap dengan atribut dan metode yang melekat pada masing-masing kelas.
3. Activity Diagram dapat mengilustrasikan alur kerja atau proses bisnis yang terjadi dalam sistem secara rinci.
4. Sequence Diagram dapat menggambarkan interaksi antar objek dalam urutan kronologis selama proses berlangsung,
5. Diagram lainnya seperti State Diagram, Component Diagram, dan Deployment Diagram juga sering digunakan sesuai kebutuhan sistem **UML REF2**

UML adalah bahasa pemodelan visual yang esensial dalam perancangan sistem berorientasi objek. Dengan UML mempermudah proses analisis, desain, dokumentasi, dan komunikasi dalam pengembangan perangkat lunak. Diagram UML seperti Use Case, Class, Activity, dan Sequence sangat berguna untuk menggambarkan kebutuhan dan desain sistem secara jelas.

## **Metode dan Tahapannya**

Dalam penelitian ini, digunakan metode kuantitatif, yaitu pendekatan yang memanfaatkan data numerik dan teknik statistik untuk mengumpulkan serta mengolah data yang dapat diukur secara objektif. Metode ini bertujuan untuk memberikan penjelasan, melakukan prediksi, atau mengendalikan fenomena tertentu melalui pendekatan eksperimen. Fokus utama adalah membangun model deteksi dan pencegahan fraud internal di CV. Smartindo Telekom menggunakan algoritma Random Forest, dengan pendekatan analisis berdasarkan kerangka Fraud Diamond. Berikut ini adalah alur kerja metodologis penelitian, yaitu:



Fraud Diamond mencakup empat elemen utama yang menjadi dasar variabel penelitian:

1. ***Pressure* (Tekanan) adalah dorongan yang memotivasi seseorang melakukan tindakan kecurangan, yang dapat berasal dari masalah keuangan pribadi atau tekanan yang diberikan oleh pihak atasan.**
2. ***Opportunity* (Kesempatan) adalah c**elah dalam sistem yang memungkinkan individu melakukan kecurangan tanpa terdeteksi.
3. ***Rationalization* (Rasionalisasi) adalah** justifikasi moral yang memungkinkan pelaku merasa tindakannya dapat dibenarkan.
4. ***Capability* (Kemampuan) adalah k**apasitas individu, baik dari sisi keahlian maupun posisi strategis, yang memungkinkan mereka melancarkan kecurangan.

### **Pengumpulan Data**

Proses pengumpulan data dalam penelitian ini bertujuan untuk mengumpulkan informasi yang relevan mengenai faktor-faktor dalam model Fraud Diamond, serta data historis yang berkaitan dengan aktivitas internal di CV. Smartindo Telekom. Data yang dikumpulkan terdiri dari dua kategori, yaitu:

1. Data Primer adalah data yang diperoleh melalui penyebaran kuesioner kepada karyawan perusahaan untuk mengukur aspek tekanan (pressure), peluang (opportunity), rasionalisasi (rationalization), dan kemampuan (capability). Kuesioner tersebut disusun berdasarkan indikator yang mewakili keempat komponen utama Fraud Diamond.
2. Data Sekunder dimana meliputi data historis aktivitas keuangan, catatan transaksi penjualan, laporan audit internal, serta catatan pelanggaran atau dugaan fraud yang pernah terjadi. Data ini digunakan untuk membangun model prediksi dan pelatihan algoritma Random Forest.

Proses pengumpulan data dilakukan secara sistematis agar hasil analisis dapat merepresentasikan kondisi aktual perusahaan dan mendukung pengembangan model deteksi serta pencegahan fraud secara akurat.

### **Variabel Pada Penelitian dan Indikator**

Penelitian ini mengadopsi variabel-variabel yang dirancang berdasarkan kerangka Fraud Diamond, yang mencakup empat komponen utama, yaitu Pressure, Opportunity, Rationalization, dan Capability. Masing-masing komponen tersebut diwakili oleh beberapa indikator yang dikembangkan menjadi pertanyaan dalam kuesioner serta fitur pada dataset yang digunakan dalam model algoritma Random Forest. Berikut ini adalah rincian variabel beserta indikatornya:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Variabel** | **Indikator Penelitian** | **Jenis Variabel** | **Skala pada Pengukuran** |
| 1 | Pressure | Tanggung jawab keuangan tinggi | Independen | Likert |
|  |  | Tekanan dari atasan untuk mencapai target | Independen | Likert |
|  |  | Masalah pribadi (utang, biaya hidup) | Independen | Likert |
|  |  | Ketidakpuasan terhadap gaji | Independen | Likert |
| 2 | Opportunity | Kelemahan dalam pengendalian internal | Independen | Likert |
|  |  | Akses bebas terhadap data/sistem keuangan | Independen | Likert |
|  |  | Tidak adanya audit internal yang rutin | Independen | Likert |
|  |  | Kurangnya pengawasan langsung dari atasan | Independen | Likert |
| 3 | Rationalization | Anggapan bahwa tindakan tidak merugikan perusahaan secara langsung | Independen | Likert |
|  |  | Pembenaran tindakan sebagai balas jasa atas kerja keras | Independen | Likert |
|  |  | Lingkungan kerja yang permisif | Independen | Likert |
|  |  | Merasa bahwa semua orang juga melakukannya | Independen | Likert |
| 4 | Capability | Jabatan atau posisi strategis di perusahaan | Independen | Likert |
|  |  | Pengalaman kerja yang panjang | Independen | Likert |
|  |  | Pemahaman terhadap sistem dan proses internal | Independen | Likert |
|  |  | Kemampuan menyembunyikan tindakan | Independen | Likert |
| 5 | Fraud | Terjadi atau tidaknya tindakan fraud (berdasarkan audit/laporan valid) | Dependen (Label) | Nominal (Biner) |

Tabel ini berfungsi untuk menguraikan tipe serta susunan variabel yang akan dipakai dalam tahapan pengolahan data dan pembangunan model Random Forest. Selain variabel bebas (fitur), penelitian ini juga membutuhkan variabel dependen atau label sebagai acuan dalam supervised learning: Fraud: Label biner (0 = tidak ada fraud, 1 = terdapat fraud) berdasarkan temuan audit atau laporan valid dari manajemen.

Skala Likert merupakan teknik pengukuran yang sering dipakai dalam penelitian untuk menilai sikap, opini, atau pandangan individu terhadap suatu pernyataan atau topik tertentu. Metode ini menggunakan rangkaian pernyataan yang disertai pilihan jawaban berjenjang, contohnya seperti "Sangat Setuju", "Setuju", "Netral", "Tidak Setuju", hingga "Sangat Tidak Setuju". Berdasarkan **[REF CODE ASD11]** skala Likert digunakan untuk mengukur sikap, pandangan, serta persepsi individu atau kelompok terhadap suatu fenomena sosial. Pilihan jawaban dalam skala ini disusun secara bertingkat untuk menangkap intensitas respon yang diberikan.

### **Teknik Pada Analisis Data**

Proses analisis data pada penelitian ini dilaksanakan secara sistematis dengan tujuan mengembangkan model deteksi serta pencegahan kecurangan internal menggunakan algoritma Random Forest, yang didasarkan pada konsep Fraud Diamond. Tahapan analisis data meliputi beberapa proses berikut:

1. **Pra-Pemrosesan Data**

Sebelum dilakukan analisis, data yang dikumpulkan melalui kuesioner dan data sekunder terlebih dahulu diproses agar siap digunakan dalam pelatihan model. Tahapan pra-pemrosesan data meliputi:

1. Data cleaning, yaitu mengeliminasi data yang duplikat, mengisi nilai yang hilang (missing value), serta menghapus data pencilan (outlier) bila diperlukan
2. Transformasi data, yaitu mengkonversi data kualitatif seperti skala Likert menjadi format numerik.
3. Normalisasi/standardisasi: Menyesuaikan skala nilai antar variabel agar seragam, jika diperlukan untuk meningkatkan performa algoritma.
4. Labeling: Memberikan label fraud (1) atau tidak fraud (0) berdasarkan data audit atau hasil validasi dari manajemen.
5. **Pembentukan Dataset**

Setelah melewati proses pra-pemrosesan, data kemudian dipisah menjadi dua bagian yaitu:

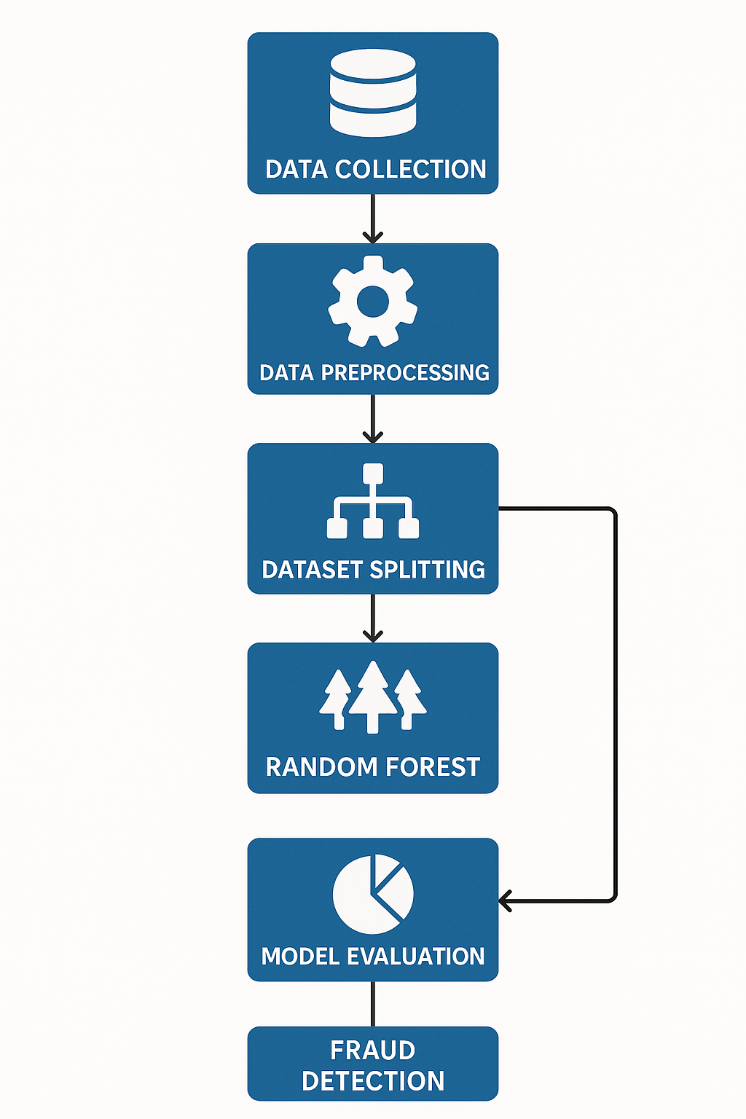
1. Data Latih (Training Set), yang berfungsi untuk melatih model Random Forest.
2. Data Uji (Testing Set), yang digunakan untuk mengukur dan menilai kinerja model dalam mendeteksi fraud.
3. Implementasi Algoritma Random Forest

Random Forest merupakan metode klasifikasi yang mengandalkan prinsip ensemble learning dengan menggabungkan banyak pohon keputusan untuk memperoleh hasil prediksi yang lebih akurat. Algoritma ini dipilih karena efektivitasnya dalam mengelola data yang kompleks dan multivariat serta kemampuannya dalam mengidentifikasi fitur-fitur kunci yang berperan signifikan dalam mendeteksi fraud. Tahapan implementasi model ini dijelaskan sebagai berikut:

1. Melatih model dengan parameter yang telah ditentukan.
2. Melakukan validasi model menggunakan cross-validation atau confusion matrix.
3. Mengukur performa model dengan metrik seperti:
4. Accuracy (akurasi)
5. Precision (ketepatan)
6. Recall (sensitivitas)
7. F1-Score
8. ROC-AUC Curve
9. Analisis Feature Importance

Setelah model terbentuk, dilakukan analisis terhadap fitur (variabel) yang paling berpengaruh dalam mendeteksi fraud. Ini berguna untuk mengidentifikasi faktor-faktor internal yang paling berkontribusi terhadap risiko terjadinya fraud di perusahaan.

Berikut adalah gambar flowchart analisis data atau visual alur kerja dari mulai proses input data hingga output model dari analisi data dimana adalah sebagai berikut ini:



## **Penelitian Terkait**

Untuk lebih dalam memahami dan mempelajari tentang random forest, penulis akan memberikan lima penelitian terkait sebagai bahan referensi untuk penulisan skripsi ini antara lain adalah:

1. Penelitian yang dilakukan oleh **[REF CODE Penlitian1]**, berfokus pada pengembangan dan evaluasi model klasifikasi menggunakan algoritma Random Forest untuk memprediksi risiko stroke. Studi ini memanfaatkan data klinis dan faktor risiko yang telah dikumpulkan dari pasien dalam penelitian stroke sebelumnya. Model Random Forest dilatih dengan cermat menggunakan data pelatihan, kemudian diuji untuk menghasilkan prediksi pada data yang terpisah. Evaluasi performa model dilakukan dengan beberapa metrik penting, yaitu akurasi, presisi, recall, dan skor F1. Hasilnya menunjukkan bahwa model Random Forest menunjukkan kinerja yang sangat baik dalam memprediksi stroke. Model ini mencapai akurasi sebesar 93,6%, presisi 91,4%, recall 96,1%, dan skor F1 sebesar 93,7%. Temuan ini mengindikasikan potensi besar algoritma Random Forest sebagai alat yang efektif untuk mendukung diagnosis dan pencegahan stroke.
2. Penelitian dari **[REF CODE Penlitian2].** bertujuan untuk mengidentifikasi Pokémon Legendaris menggunakan Algoritma SF-Random Forest. Data yang digunakan meliputi 800 sampel Pokémon dengan berbagai karakteristik. Sebelum analisis, data diproses dengan membersihkan data yang tidak relevan, mengisi kekosongan, menormalisasi skala nilai, dan menyeimbangkan jumlah data untuk setiap kategori. Algoritma Random Forest digunakan untuk mengidentifikasi fitur-fitur yang paling relevan dalam menentukan apakah suatu Pokémon termasuk kategori legendaris atau tidak. Hasilnya menunjukkan bahwa model SF-Random Forest sangat akurat. Model ini mencapai tingkat akurasi, presisi, recall, dan skor F1 sebesar 100%. Ini berarti algoritma tersebut sangat baik dalam mengenali ciri-ciri penting dan mengatasi masalah ketidakseimbangan data. Dengan demikian, SF-Random Forest terbukti efektif untuk tugas pengenalan pola yang rumit.
3. Penelitian oleh **[REF CODE Penlitian3].** mengulas secara mendalam optimasi hyperparameter pada algoritma Random Forest dengan tujuan untuk meningkatkan akurasi prediksi cakupan jaringan 5G. Studi ini menekankan pentingnya proses optimasi hyperparameter untuk memaksimalkan kinerja model. Serangkaian eksperimen dilakukan dengan menguji berbagai kombinasi hyperparameter, termasuk 'max\_depth', 'max\_features', 'min\_samples\_leaf', 'min\_samples\_split', dan 'n\_estimators', menggunakan teknik optimasi yang relevan. Hasil penelitian menunjukkan bahwa pemilihan kombinasi hyperparameter yang optimal secara signifikan memperbaiki performa model. Model yang telah dioptimasi berhasil mencapai Root Mean Squared Error (RMSE) minimum sebesar 0,6, yang jauh lebih baik dibandingkan model Random Forest tanpa optimasi hyperparameter yang memiliki RMSE 1,14. Studi ini secara tegas menegaskan bahwa optimasi hyperparameter memegang peranan krusial dalam meningkatkan keakuratan dan konsistensi model Random Forest untuk prediksi cakupan 5G.
4. Penelitian oleh **[REF CODE Penlitian4].** meneliti penggunaan algoritma Random Forest untuk mengatasi masalah data tak seimbang dalam klasifikasi, khususnya pada data stunting. Penelitian ini mengintegrasikan metode Synthetic Minority Over-sampling Technique (SMOTE) dan SMOTE-Edited Nearest Neighbor (SMOTE-ENN) untuk menyeimbangkan jumlah data antar kelas. Tujuannya adalah meningkatkan akurasi model dalam memprediksi stunting, sebuah kondisi yang sering kali memiliki jumlah kasus minoritas yang jauh lebih sedikit dibandingkan kasus non-stunting. Hasil penelitian menunjukkan bahwa algoritma Random Forest, ketika dikombinasikan dengan teknik penyeimbangan data seperti SMOTE dan SMOTE-ENN, mampu mengklasifikasikan data tak seimbang dengan performa yang lebih baik. Ini menunjukkan efektivitas Random Forest dalam menangani dataset yang kompleks dan kemampuannya untuk tetap memberikan hasil yang akurat bahkan ketika menghadapi distribusi kelas yang tidak merata. Penerapan ini sangat relevan untuk masalah kesehatan publik di mana deteksi dini kondisi seperti stunting menjadi krusial.
5. Penelitian **[REF CODE Penlitian5].** oleh memaparkan penggunaan model Algoritma Random Forest Classifier (RFC) sebagai metode diagnosis terhadap potensi kerusakan atau penurunan kinerja fungsional pada komponen-komponen sistem pembangkit daya, dengan fokus utama pada indikator efisiensi energi. Berdasarkan temuan studi ini, dengan menggunakan label indikator Net Plant Heat Rate (NPHR) serta fitur prediktor dari sepuluh variabel terkontrol teratas, ditemukan bahwa fitur-fitur yang mengalami penurunan fungsi operasional terkait efisiensi energi, dan memerlukan perhatian prioritas, meliputi: Suhu Air Umpan Akhir, Suhu Gas Keluar Pemanas Udara Utama, serta Kandungan Oksigen dalam Gas Buang. Hasil ini memberikan panduan berharga untuk strategi pemeliharaan yang lebih efektif dan peningkatan efisiensi operasional secara keseluruhan pada sistem pembangkit daya..

# **REFERENSI**

**Referensi code**

**ASD1**   
Agus Defri Yando, Kecenderungan Kecurangan Akuntansi, (Batam: CV. Batam Publisher, 2020), hal 81.

**ASD2**  
Arum Ardianingsih, Audit Laporan Keuangan, (Jakarta: PT. Bumi Aksara, 2018), hal 74.

**ASD3**Dewi Hanggraeni, Manajemen Risiko dan Environmental, Social, and governance (ESG) Teori dan hasil penelitian, (Bogor: PT Penerbit IPB Press, 2021), hal. 51.

**ASD4**  
Wolfe, D. T., & Hermanson, D. R. (2004). *The Fraud Diamond: Considering the Four Elements of Fraud*. **The CPA Journal**, 74(12), 38-42.

**ASD5**M. Fahrul Rizki Aditya, Nuril Lutvi Azizah dan Uce Indahyanti (2024). Prediksi Penyakit Hipertensi Menggunakan metode Decision Tree dan Random Forest. Jurnal Ilmiah KOMPUTASI, Volume 23 No :1, Maret 2024, p-ISSN 1412-9434/e-ISSN 2549-7227  
 **ASD6**  
R. G. Wardhana, G. Wang, dan F. Sibuea, "Penerapan Machine Learning dalam Prediksi Tingkat Kasus Penyakit di Indonesia," Jurnal Online Informatika Sistem dan Manajemen, vol. 5, no. 1, pp. 40-50, 2023. [Online]. Available: <https://jurnal.amikom.ac.id/index.php/joism/article/download/1136/401/5841>  
  
**ASD7**  
A. P. Sari dan S. Suhardi, "Implementasi Metode Machine Learning Menggunakan Algoritma Evolving Artificial Neural Network Pada Kasus Prediksi Diagnosis Diabetes," JATIKOM, vol. 3, no. 2, 2020. DOI: 10.17509/jatikom.v3i2.27885. [Online]. Available: <https://ejournal.upi.edu/index.php/JATIKOM/article/view/27885>

**ASD8**  
M. Nawawi, A. Sihombing, dan Y. Yuliati, "Model Klasifikasi Machine Learning untuk Prediksi Ketepatan Penempatan Karir," Jurnal Saintekom, vol. 14, no. 1, pp. 13-25, 2024. DOI: 10.33020/saintekom.v14i1.512. [Online]. Available: <https://ojs.stmikplk.ac.id/index.php/saintekom/article/download/512/187/3607>  
  
**ASD9**  
Nur Lazimatul Hima Solehah, et al., Kecurangan Akuntansi Ditinjau dari Pengendalian Internal, Moralitas, dan Personal Culture, (Banten: CV. AA. RIZKY, 2020), hal 24.

**ASD10**  
Arum Ardianingsih, Audit Laporan Keuangan (Jakarta: PT. Bumi Aksara, 2018), hal 78.

**ASD11**  
S. Sugiyono, "Skala Likert digunakan untuk mengukur sikap, pendapat, dan persepsi seseorang atau sekelompok orang tentang fenomena sosial," Al-Kharaj: Jurnal Ekonomi, Keuangan & Bisnis Syariah, vol. 6, no. 7, pp. 5490–5500, 2019.  
  
**ASD12**  
F. D. Rahman, M. I. Zulfa, and A. Taryana, “Clustering dan Klasifikasi Data Cuaca Cilacap dengan Menggunakan Metode K-Means dan Random Forest,” Jurnal SINTA: Sistem Informasi dan Teknologi Komputasi, vol. 1, no. 2, pp. 90–97, 2024, doi: 10.61124/sinta.v1i2.15  
https://jurnalsinta.id/index.php/sinta/article/view/15  
  
**ASD13**  
A. Fauziah and J. Hernadi, “Klasifikasi Data Tak Seimbang Menggunakan Algoritma Random Forest dengan SMOTE dan SMOTE-ENN (Studi Kasus pada Data Stunting),” Jurnal Teknomatika, vol. 17, no. 2, pp. 38-47, Des. 2024, doi: 10.30989/teknomatika.v17i2.1530. Sumber  
https://ejournal.unjaya.ac.id/index.php/teknomatika/article/view/1530/906  
  
**ASD14**  
<https://ejournal.uki.ac.id/index.php/beuki/article/download/595/458>  
  
**Penelitian1**  
M. Fadli and R. A. Saputra, “KLASIFIKASI DAN EVALUASI PERFORMA MODEL RANDOM FOREST UNTUK PREDIKSI STROKE,” JT Jurnal Teknik, vol. 12, no. 2, pp. 72–80, 2023, doi: 10.31000/jt.v12i2.9099.  
Ref <https://jurnal.umt.ac.id/index.php/jt/article/view/9099>

**Penelitian2**  
A. Prayoga, Y. V. Via, and I. G. S. M. Diyasa, “Classifying Legendary Pokémon with SF-Random Forest Algorithm,” Journal of Information Systems and Informatics, vol. 6, no. 3, 2024, doi: 10.51519/journalisi.v6i3.859.  
ref https://www.journal-isi.org/index.php/isi/article/view/859

**Penelitian3**H. Yuliana, Iskandar, Hendrawan, S. Basuki, M. R. Hidayat, A. Charisma, and H. Vidyaningtyas, “Hyperparameter Optimization of Random Forest Algorithm to Enhance Performance Metric Evaluation of 5G Coverage Prediction,” Buletin Pos dan Telekomunikasi, vol. 22, no. 1, 2024, doi: 10.17933/bpostel.v22i1.390.  
Ref <https://bpostel.komdigi.go.id/index.php/bpostel/article/view/390>  
  
**Penelitian4**A. Fauziah and J. Hernadi, “Klasifikasi Data Tak Seimbang Menggunakan Algoritma Random Forest dengan SMOTE dan SMOTE-ENN (Studi Kasus pada Data Stunting),” Jurnal Teknomatika, vol. 17, no. 2, pp. 38-47, Des. 2024, doi: 10.30989/teknomatika.v17i2.1530.  
ref https://ejournal.unjaya.ac.id/index.php/teknomatika/article/view/1530 **Penelitian5**G. A. M. Ashfania, T. Prahasto, A. Widodo, and T. Warsokusumo, “Penggunaan Algoritma Random Forest untuk Klasifikasi berbasis Kinerja Efisiensi Energi pada Sistem Pembangkit Daya,” ROTASI, vol. 24, no. 3, pp. 14–21, Jul. 2022, doi: 10.14710/rotasi.24.3.14-21.  
Ref https://ejournal.undip.ac.id/index.php/rotasi/article/view/47375  
  
Referensi bacaan  
Santosa, R. B. (2018). *Analisis Fraud Diamond Theory terhadap Terjadinya Fraud*. **Jurnal Ekonomi dan Bisnis**, Universitas Jember.