**Dokumen Software Design pada Aplikasi Fraud Detection**

**Dylan Ciu**

Goal Planning Project Magang

Tabel Isi

[Pendahuluan 2](#_Toc204842211)

[Masalah 2](#_Toc204842212)

[Penjelasan Fitur Deteksi 3](#_Toc204842213)

[Tujuan 5](#_Toc204842214)

[Ruang Lingkup (Scope) 5](#_Toc204842215)

[System Overview 5](#_Toc204842216)

[Output & Perbandingan Hasil Deteksi Fraud 7](#_Toc204842217)

[Tabel Perbandingan: Sebelum dan Sesudah Fraud Detection 8](#_Toc204842218)

[Timeline Project 9](#_Toc204842219)

[Penutup 11](#_Toc204842220)

**Abstrak** - Meningkatnya volume transaksi digital telah memperbesar risiko terjadinya penipuan (fraud), khususnya dalam sistem keuangan berbasis teknologi. Sistem deteksi fraud konvensional yang bersifat rule-based sering tidak efektif menghadapi pola penipuan yang dinamis dan kompleks (Bolton & Hand, 2002). Oleh karena itu, dalam proyek ini dirancang sebuah sistem deteksi fraud berbasis pembelajaran mesin (machine learning) sederhana, dengan algoritma seperti Decision Tree dan Logistic Regression yang cocok untuk implementasi berskala kecil dan edukatif (Nguyen et al., 2022). Sistem dirancang secara modular, meliputi pemrosesan data, analisis eksploratif, pelatihan model, serta pemberian skor dan peringatan terhadap transaksi mencurigakan. Tidak seperti sistem komersial berbasis cloud, sistem ini dikembangkan secara lokal tanpa ketergantungan terhadap layanan pihak ketiga. Prototipe ini diharapkan dapat digunakan sebagai dasar pembelajaran dan pengembangan awal sistem keamanan finansial berbasis AI yang lebih luas.

# Pendahuluan

Fraud Detection merupakan sistem yang dirancang untuk mengenali transaksi mencurigakan secara otomatis melalui analisis pola data. Dengan meningkatnya penggunaan sistem pembayaran digital, risiko penipuan semakin besar, dan sistem pendeteksi berbasis machine learning menjadi kebutuhan krusial. Proyek ini bertujuan mengembangkan prototipe sistem Fraud Detection yang sederhana namun fungsional, berbasis data simulasi, dan menggunakan algoritma machine learning ringan seperti Decision Tree.

# Masalah

|  |  |  |
| --- | --- | --- |
| **Jenis Fraud** | **Pola Umum** | **Fitur Deteksi** |
| Fraud E-commerce | - Banyak transaksi ke merchant yang sama dengan nominal identik - Refund/pembatalan cepat setelah transaksi - Penyalahgunaan kode promo | Merchant\_id, refund\_time, promo\_code\_used, transaction\_amount |
| Pinjaman Online | - Pengajuan dari device/IP yang sama untuk banyak akun - Data pribadi tidak konsisten atau mencurigakan - Pengajuan nominal maksimum terus | Device\_id, loan\_amount, KYC\_status, account\_age, submission\_pattern |
| Card Not Present (CNP) | |  | | --- | |  |  |  | | --- | | - Transaksi tanpa kehadiran fisik kartu - Lokasi/perangkat transaksi baru - Banyak transaksi cepat dalam waktu singkat | | Card\_id, device\_id, location, transaction\_interval, amount |
| Cyber Fraud / Cyber Crime | - Akses cepat dan masif (burst) - Navigasi sistem tidak wajar (tidak melalui login/verifikasi) - Banyak request dari 1 device/IP | device\_id, user\_id, request\_frequency, access\_pattern, referral\_flow |

## Penjelasan Fitur Deteksi

1. Fraud E-commerce

Fitur Utama:

* Merchant\_id (identitas merchant)
* Refund\_time (waktu antara pembelian dan pembatalan)
* Promo\_code\_used (kode promo yang dipakai)
* Transaction\_amount (nilai belanja)

Cara kerja deteksi:

* Sistem menganalisis apakah user sering melakukan pembelian → lalu langsung meminta **refund dalam waktu sangat cepat** (refund\_time kecil).
* Jika merchant yang digunakan sering muncul dalam kasus refund cepat, maka merchant\_id dianggap memiliki **tingkat risiko tinggi**.
* Jika promo\_code\_used sama digunakan berkali-kali dari akun berbeda atau pada waktu tertentu → sistem bisa mendeteksi abuse promo.
* Pola pembelanjaan dengan nilai **identik** dan berulang juga menjadi sinyal kuat untuk fraud e-commerce.

2. Pinjaman Online

Fitur Utama:

* Device\_id (ID perangkat)
* Loan\_amount (jumlah pinjaman yang diajukan)
* KYC\_status (status verifikasi pengguna)
* Submission\_pattern (pola pengajuan)

Cara kerja deteksi:

* Sistem memonitor pengajuan pinjaman dari device\_id yang sama untuk beberapa akun → ini menunjukkan adanya **akun ganda atau pemalsuan identitas**.
* Pengajuan terus-menerus dengan **jumlah maksimal** juga akan dianggap mencurigakan (loan\_amount ekstrem).
* Jika pengguna tidak menyelesaikan verifikasi (KYC) namun tetap mengajukan pinjaman, maka KYC\_status = Unverified akan meningkatkan skor risiko.
* Sistem juga bisa mendeteksi pola submission\_pattern yang tidak wajar — misalnya pengajuan dilakukan di waktu larut malam atau dalam interval terlalu cepat.

3. Card Not Present (CNP)

Fitur Utama:

* Card\_id (ID kartu)
* Device\_id (Perangkat yang digunakan)
* Location (Lokasi transaksi)
* Transaction\_interval (Selang waktu antar transaksi)
* Amount (Jumlah transaksi)

Cara kerja deteksi:

* Jika kartu digunakan dari lokasi/IP baru yang belum pernah tercatat → naikkan skor risiko.
* Deteksi burst transaction → banyak transaksi dalam waktu sangat singkat.
* Transaksi dari akun yang belum pernah pakai kartu tersebut sebelumnya = anomali.

4. Cyber Fraud / Cyber Crime

Fitur Utama:

* Device\_id (ID perangkat)
* User\_id (ID pengguna)
* Request\_frequency (Frekuensi permintaan akses)
* Access\_pattern (Pola akses pengguna)
* Referral\_flow (Alur rujukan system)

Cara kerja deteksi:

* Sistem memonitor **jumlah request per user/device** dalam periode waktu.
* Jika satu device menghasilkan terlalu banyak request/login/trx dalam waktu singkat → dianggap sebagai serangan otomatis (bot).
* Pola navigasi yang tidak biasa (e.g. langsung ke transaksi tanpa login atau verifikasi) → indikator fraud berbasis bot/script injection.

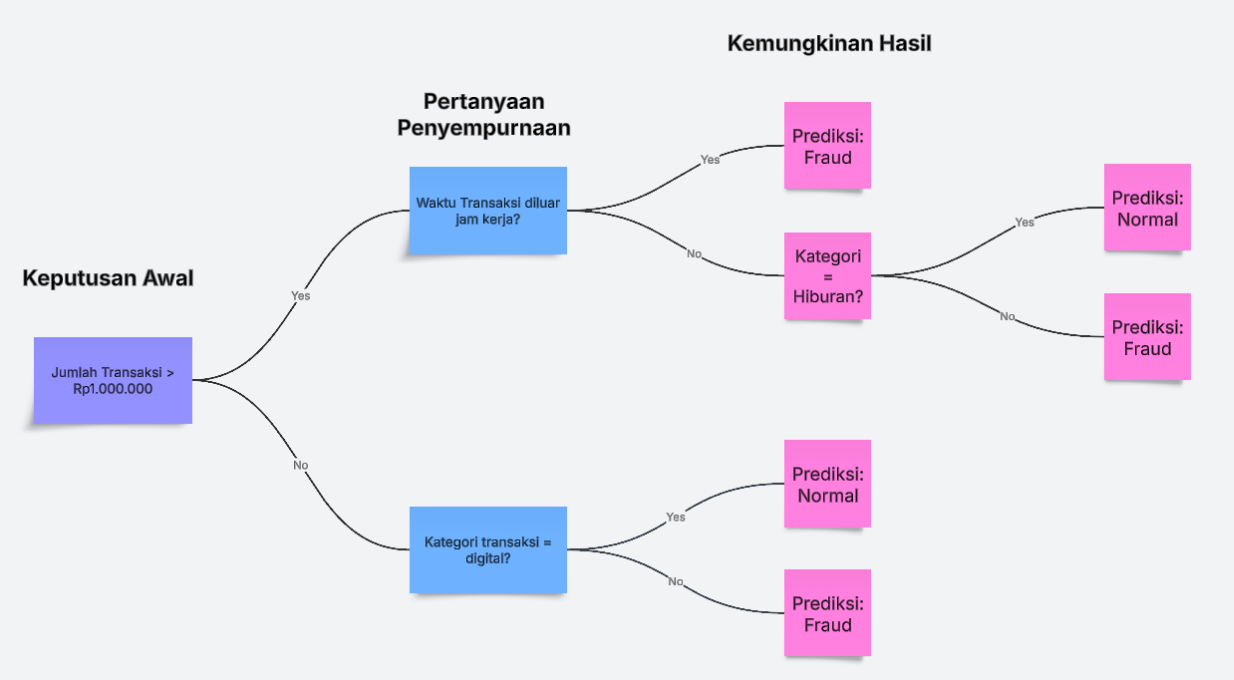
# Tujuan

* Mendeteksi aktivitas transaksi yang mencurigakan secara real-time atau batch.
* Mengembangkan sistem scoring dan alert otomatis.
* Menyediakan arsitektur modular untuk integrasi dengan sistem keuangan digital skala kecil.

# Ruang Lingkup (Scope)

* Bahasa pemrograman: Python
* Dataset: Public dataset transaksi (format CSV)
* Algoritma: Decision Tree, Logistic Regression
* Tools: Pandas, Matplotlib, Scikit-learn
* Output: Terminal alert atau tabel flag transaksi

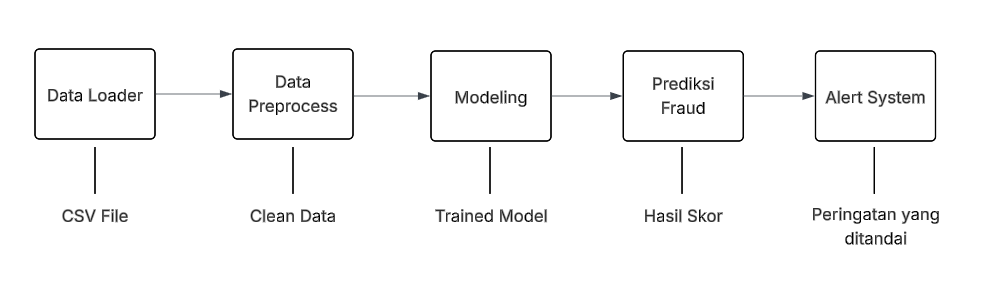
# System Overview



*Gambar 1: Struktur Decision Tree untuk Fraud Detection*

**Penjelasan Node:**

* **Jumlah Transaksi:** nilai threshold bisa didapat dari EDA (misal outlier > 1 juta rupiah)
* **Waktu Transaksi:** fraud kadang terjadi pada jam tidak biasa (misal tengah malam)
* **Kategori:** beberapa kategori seperti top-up digital, elektronik mahal, lebih rentan



*Gambar 2: Diagram Arsitektur Sistem untuk Fraud Detection*

**Alur Sistem**

**Data Loader**

* Memuat dataset transaksi dalam format CSV. Dataset mencakup atribut seperti user\_id, timestamp, amount, merchant, category, location, dan label is\_fraud.

**Data Preprocessing**

* Menangani missing value, mengonversi data waktu, melakukan encoding terhadap data kategorikal, dan normalisasi terhadap nilai jumlah transaksi (amount).
* Menghasilkan data bersih siap pakai untuk model machine learning.

**Exploratory Data Analysis (EDA)**

* Menyajikan visualisasi dan statistik seperti jumlah transaksi per kategori, distribusi nilai transaksi, serta deteksi outlier.
* Tujuannya untuk menemukan pola dasar yang dapat mengindikasikan adanya transaksi abnormal.

**Model Training & Inference**

* Menggunakan algoritma Decision Tree dan/atau Logistic Regression untuk membangun model prediksi apakah suatu transaksi tergolong fraud atau tidak.
* Model dilatih menggunakan data historis berlabel dan divalidasi dengan metrik seperti akurasi, precision, recall, dan F1-score.

**Scoring & Alerting**

* Setiap transaksi akan diberi skor probabilitas (misalnya 0.0–1.0) untuk mengukur kemungkinan fraud.
* Transaksi yang melebihi ambang batas akan ditandai dan dicetak sebagai alert dalam terminal atau disimpan dalam log file.

# Output & Perbandingan Hasil Deteksi Fraud

**Studi Kasus & Temuan**

* Algoritma Random Forest mencapai akurasi pelatihan hingga 99,5%, dan akurasi pengujian sekitar 74,5 %

Sumber: <https://medium.com/%40prasetyow938/fraud-detection-in-online-transactions-using-machine-learning-2489fe0f8e48>

* Analisis dari literatur Indonesia secara sistematis (Diponegoro Journal of Accounting) menemukan bahwa sistem fraud detection berbasis AI memiliki efektivitas rata-rata sebesar **76,7 %** dalam meningkatkan efektivitas dan kecepatan kontrol fraud di lembaga keuangan local.

Sumber:<https://ejournal3.undip.ac.id/index.php/accounting/article/view/46107?utm_source=chatgpt.com>

* Sedangkan platform pinjaman tunai (cash loan) di Indonesia yang menggunakan sistem fraud detection real-time mencatat peningkatan akurasi deteksi fraud hingga **300 %**, efisiensi operasional meningkat **30 %**, dan potensi kerugian finansial yang dicegah mencapai **USD 2 juta** dalam dua bulan operasi.

Sumber: <https://trustdecision.com/resources/blog/case-study-detecting-application-fraud-on-an-indonesian-cash-loan-platform?utm_source=chatgpt.com>

# Tabel Perbandingan: Sebelum dan Sesudah Fraud Detection

|  |  |  |
| --- | --- | --- |
| **Aspek** | **Sebelum Penerapan Fraud Detection** | **Setelah Penerapan Fraud Detection** |
| Akurasi Prediksi Fraud | Rule-based, akurasi kurang lebih atau sama dengan 50% (cenderung error tinggi). | Random Forest / ML: akurasi hingga ~74 %—99 % tergantung dataset & model.  Sumber:  1.) <https://medium.com/%40prasetyow938/fraud-detection-in-online-transactions-using-machine-learning-2489fe0f8e48>  2.)  <https://jurnal.polgan.ac.id/index.php/sinkron/article/view/14645?utm_source=chatgpt.com> |
| Efektivitas Deteksi & Keamanan | Sistem manual/statistic rentan slip. | Efektivitas rata-rata ~76,7 % menurut SLR bisnis Indonesia.  Sumber:  <https://ejournal3.undip.ac.id/index.php/accounting/article/view/46107?utm_source=chatgpt.com> |
| Operasional & Efisiensi | Banyak review manual, lambat | Otomatisasi mempercepat ~30 % di platform fintech Indonesia  Sumber:  <https://trustdecision.com/resources/blog/case-study-detecting-application-fraud-on-an-indonesian-cash-loan-platform?utm_source=chatgpt.com> |
| Kerugian Finansial Preventif | Tidak ada system prediktif. | Potensi kerugian dicegah sebesar USD 2 juta dalam 2 bulan (platform loan)  Sumber:  <https://trustdecision.com/resources/blog/case-study-detecting-application-fraud-on-an-indonesian-cash-loan-platform?utm_source=chatgpt.com> |

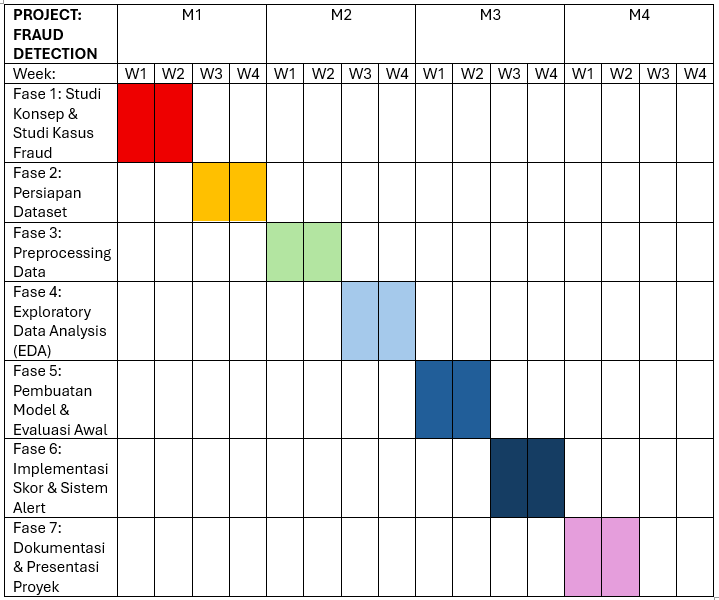
**Interpretasi & Relevansi**

* **Sebelum:** Sistem tradisional rule-based kurang adaptif terhadap pola baru → sering menimbulkan false negative maupun false positive.
* **Setelah:** Model ML (seperti Decision Tree, Random Forest) yang dilatih pada data historis dapat secara signifikan meningkatkan akurasi dan memperingatkan transaksi abnormal lebih cepat.
* Efisiensi operasional meningkat, mengurangi beban manual.

**Penerapan ke Project**

* **Target akurasi awal:** minimal mencapai ~70–80 % (dengan model sederhana).
* **Keunggulan tambahan:** meskipun tidak sebesar platform besar, efek positif berupa penghematan waktu dan deteksi dini tetap signifikan**.**
* Dengan peningkatan algoritma (ensemble seperti Random Forest atau XGBoost), akurasi dapat meningkat lebih jauh.

# Timeline Project



**Fase 1: Studi Konsep & Studi Kasus Fraud *(Minggu 1–2)***

* Mempelajari definisi dan karakteristik fraud dalam transaksi digital.
* Mempelajari referensi dan/atau studi kasus dari fintech local, contoh seperti Akulaku, Kredivo, dan e-wallet Indonesia.
* Mengidentifikasi jenis-jenis transaksi mencurigakan dan pola umum penipuan yang sering terjadi.

**Fase 2: Persiapan Dataset *(Minggu 3–4)***

* Mengumpulkan dan mengevaluasi dataset publik yang relevan, seperti data transaksi digital dari Kaggle.
* Menyesuaikan struktur data dengan kebutuhan sistem (user\_id, amount, waktu, kategori, is\_fraud).
* Melakukan pembersihan awal dan penyusunan format CSV sebagai input sistem.

**Fase 3: Preprocessing Data *(Minggu 5–6)***

* Membersihkan data dari duplikasi, missing values, dan data tidak valid.
* Melakukan encoding terhadap fitur kategorikal seperti merchant atau kategori transaksi.
* Normalisasi nilai transaksi (amount) dan penambahan fitur baru seperti waktu transaksi (jam), lokasi, dan frekuensi user.

**Fase 4: Exploratory Data Analysis (EDA) *(Minggu 7–8)***

* Membuat visualisasi statistik dasar: distribusi jumlah transaksi, kategori, waktu, dan lokasi.
* Mengidentifikasi anomali dan outlier yang berpotensi menunjukkan fraud.
* Menyusun insight dari pola transaksi sebagai dasar pemilihan fitur penting dalam model.

**Fase 5: Pembuatan Model & Evaluasi Awal *(Minggu 9–10)***

* Membangun model klasifikasi menggunakan algoritma Decision Tree dan Logistic Regression.
* Melatih dan menguji model dengan data yang telah diproses.
* Mengukur performa model menggunakan metrik seperti akurasi, precision, recall, dan F1-score.

**Fase 6: Implementasi Skor & Sistem Alert *(Minggu 11–12)***

* Menerapkan sistem skor probabilitas untuk setiap transaksi berdasarkan output model.
* Menetapkan threshold skor untuk menentukan level risiko (misalnya: skor > 0.8 = high risk).
* Menampilkan alert pada terminal dan/atau menyimpan hasil deteksi ke dalam file log CSV.

**Fase 7: Dokumentasi & Presentasi Proyek *(Minggu 13–14)***

* Menyusun laporan akhir proyek: latar belakang, metode, arsitektur sistem, model, dan hasil evaluasi.
* Membuat visualisasi tambahan seperti diagram sistem, decision tree, dan grafik hasil.
* Menyiapkan slide presentasi dan demo sistem sederhana untuk finalisasi proyek.

# Penutup

Sistem Fraud Detection ini menjadi langkah awal dalam membangun solusi keamanan transaksi berbasis AI. Dengan pendekatan modular dan model sederhana, sistem ini dapat dikembangkan lebih lanjut untuk kebutuhan industri maupun akademik.