Journal Paper Competition Accounting Festival 2026

FRAUDLENS: DETEKSI FRAUD LAPORAN KEUANGAN BERBASIS BENFORD DAN ISOLATION FOREST

Mohamad Zulfahmi Al Fareza[1] Nabila Agatha Parsa[2]
[1,2]Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya
mohamadzulfahmialfareza@gmail.com[1]

ARTICLE INFO

Article history:
Received:
Received in revised form:
Accepted:

Keywords: *Benford's Law, Financial Fraud Detection, Fraudlens, Forensic Audit, Hybrid Method, Isolation Forest*

Paper type
Jenis artikel (artikel penelitian)

ABSTRACT (in English)

Financial fraud detection is often hindered by the fundamental limitations of single method approaches, where statistical methods are prone to overlooking micro scale manipulation while Artificial Intelligence (AI) algorithms risk bias when facing massive data distortion. To address these security gaps, this study aims to design and build a fraud detection system named FraudLens to produce an effective and affordable computer assisted audit instrument for micro entities. Using a Research and Development (R&D) approach with a prototyping model, this research utilizes simulation data (synthetic data) divided into three dataset scenarios (normal, minor manipulation, and massive) to overcome real data privacy constraints. The system development uses Python 3.10 and the Pandas library with a Streamlit web based interface, applying a hybrid analysis tool that integrates Benford's Law statistical method and the Unsupervised Machine Learning Isolation Forest algorithm. Research findings reveal significant contradictory results: in the minor manipulation scenario, the statistical method failed to detect anomalies, yet the Isolation Forest algorithm achieved a 100% detection rate, conversely, in the massive manipulation scenario, the AI algorithm experienced a fatal detection failure with only a 5.3% success rate due to the Masking Effect phenomenon, while Benford's Law successfully identified manipulation indications with a MAD score of 0.0817. The novelty of this research lies in the empirical proof that without cross validation, AI is susceptible to Normalization of Deviance, making the integration of both methods absolutely necessary to create a double layered defense mechanism that covers the blind spots of each approach in maintaining audit integrity.

PENDAHULUAN

Stabilitas dan perkembangan ekonomi suatu negara sangat bergantung pada iklim investasi yang sehat, baik yang bersumber dari sektor publik maupun swasta. Investasi memegang peranan vital dalam mengakselerasi pertumbuhan ekonomi melalui peningkatan modal dan penciptaan lapangan kerja (Safitri et al., 2025). Dalam ekosistem ini, laporan keuangan menjadi instrumen fundamental bagi investor dan penyedia dana hibah untuk menilai kinerja serta kelayakan entitas penerima dana. Akurasi informasi dalam laporan keuangan tidak hanya menentukan kualitas keputusan pendanaan, tetapi juga menjadi pilar utama transparansi pembangunan ekonomi. Namun, tekanan untuk mempertahankan reputasi atau mendapatkan akses pendanaan sering kali mendorong manajemen melakukan manipulasi laporan keuangan, mulai dari pengakuan pendapatan fiktif hingga penyembunyian kewajiban, yang mendistorsi kondisi ekonomi riil perusahaan. Fenomena kecurangan laporan keuangan (*financial statement fraud*) masih menjadi ancaman serius global, termasuk di Indonesia. Berdasarkan laporan *Survei Fraud Indonesia* oleh ACFE Indonesia Chapter (2020), kecurangan laporan keuangan merupakan skema dengan dampak kerugian median terbesar meskipun frekuensinya lebih rendah dibandingkan korupsi aset. Kasus-kasus besar yang melibatkan entitas terkemuka seperti PT Garuda Indonesia dan Jiwasraya menunjukkan bahwa metode audit konvensional sering kali terlambat mendeteksi anomali yang tersembunyi. Audit manual, meskipun akurat, memerlukan sumber daya waktu dan biaya yang besar, sehingga kurang efisien jika diterapkan sebagai alat penapis awal (*screening*) pada populasi data yang masif, terutama dalam konteks seleksi penerima dana hibah atau kredit mikro.

Tantangan utama dalam deteksi dini kecurangan adalah keterbatasan metode tunggal. Pendekatan statistik populer seperti Hukum Benford (Benford's Law) terbukti efektif mendeteksi manipulasi pada tingkat makro atau struktural (Nigrini, 2012). Namun, metode ini memiliki celah penelitian berupa rendahnya sensitivitas terhadap manipulasi parsial yang jumlahnya sedikit namun bernilai material. Sebaliknya, pendekatan modern berbasis kecerdasan buatan seperti *Machine Learning* mampu mendeteksi pencilan (*outlier*) secara presisi, namun rentan mengalami bias jika data latih telah terdistorsi oleh manipulasi massal. Belum banyak penelitian yang mencoba menggabungkan kedua pendekatan ini, statistik deduktif dan pembelajaran mesin induktif dalam satu platform terintegrasi untuk menutupi kelemahan masing-masing metode.

Untuk menjawab kesenjangan tersebut, penelitian ini bertujuan mengembangkan sistem deteksi dini bernama *FraudLens* yang mengimplementasikan metode hibrida. Kebaruan penelitian ini terletak pada integrasi algoritma *Isolation Forest* untuk mendeteksi anomali individual dan Hukum Benford untuk memvalidasi integritas struktur data secara keseluruhan. Sistem ini dikembangkan berbasis web untuk memungkinkan pengguna mengunggah data keuangan dan memperoleh analisis risiko ganda secara instan. Melalui pendekatan ini, diharapkan tercipta instrumen audit berbantuan komputer yang tidak hanya cepat dan objektif, tetapi juga tangguh dalam mendeteksi berbagai modus manipulasi, sehingga dapat mendukung pengambilan keputusan investasi yang lebih akuntabel dan tepat sasaran.

MATERI DAN METODE PENELITIAN

Penelitian ini menggunakan pendekatan *Research and Development* (R&D) dengan model pengembangan *prototyping* untuk merancang bangun sistem deteksi kecurangan laporan keuangan bernama *FraudLens*. Metode ini dipilih karena memungkinkan pengembangan iteratif yang berfokus pada penciptaan produk perangkat lunak yang teruji secara fungsional (Rindrayani et al, 2025). Prosedur penelitian ini dibatasi pada empat tahapan utama, yaitu: (1) **Analisis Kebutuhan** (*Communication*), di mana peneliti melakukan studi literatur terkait metode audit forensik dan kebutuhan sistem, (2) **Perancangan Cepat** (*Quick Plan*), yaitu mendesain alur algoritma hibrida dan antarmuka pengguna, (3) **Pembuatan Prototipe** (*Modeling & Construction*), yaitu tahap pengkodean sistem menggunakan bahasa Python dan Streamlit, dan (4) **Penyerahan & Umpan Balik** (*Deployment and Feedback*), yang dalam penelitian ini dibatasi pada uji coba simulasi menggunakan dataset uji (*black box testing*) untuk mengukur akurasi deteksi tanpa penyebaran massal. Tujuan utama dari desain penelitian ini adalah menghasilkan instrumen audit berbantuan komputer (*Computer Assisted Audit Techniques* / CAATs) yang dapat diakses oleh entitas mikro tanpa biaya infrastruktur yang tinggi. Pengembangan sistem dilakukan menggunakan bahasa pemrograman Python 3.10 yang mengintegrasikan pustaka analisis data Pandas, perhitungan numerik NumPy, dan antarmuka pengguna berbasis web menggunakan Streamlit.

Data yang digunakan dalam penelitian ini merupakan data simulasi (*synthetic data*) yang dibangkitkan secara komputasional untuk merepresentasikan struktur buku besar. Penggunaan data simulasi dalam penelitian deteksi kecurangan keuangan telah menjadi standar yang diterima secara luas untuk mengatasi kendala kerahasiaan (*confidentiality*) data keuangan riil dan ketiadaan label kecurangan yang valid pada data publik (FCA, 2024). Selain itu, penggunaan data sintetis memungkinkan peneliti untuk menetapkan *ground truth* (kunci jawaban) yang presisi guna mengukur performa deteksi algoritma secara objektif (Chaudhari & Charate, 2025). Penelitian ini merancang tiga skenario dataset uji untuk mengukur sensitivitas sistem secara komprehensif. Variabel fokus dalam penelitian ini adalah atribut nominal transaksi, yang akan dianalisis menggunakan dua pendekatan hibrida.

Landasan teoretis pertama yang diterapkan adalah Hukum Benford (*Benford's Law*). Fenomena matematika ini pertama kali diamati oleh Newcomb (1881) dan kemudian diformalkan oleh Frank Benford (1938), yang menyatakan bahwa dalam sekumpulan data numerik alami, distribusi digit pertama tidaklah acak, melainkan mengikuti pola logaritmik tertentu. Dalam konteks audit forensik, Nigrini (2012) membuktikan bahwa manipulasi data keuangan oleh manusia cenderung melanggar pola alami ini karena bias kognitif manusia dalam memilih angka acak. Probabilitas kemunculan digit pertama (d) dirumuskan secara matematis sebagai berikut:

$$P(d) = \log_{10}\left(1 + \frac{1}{d}\right)$$

Berdasarkan rumus tersebut, digit angka 1 memiliki peluang muncul terbesar, sedangkan digit 9 memiliki peluang terkecil. Rincian distribusi probabilitas teoretis untuk setiap digit pertama disajikan dalam Tabel 1.

Tabel 1. Distribusi Probabilitas Hukum Benford

Digit Awal	1	2	3	4	5	6	7	8	9
Probabilitas (P)	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

Sumber: Nigrini (2012)

Sistem *FraudLens* menghitung probabilitas kemunculan digit pertama data aktual dan membandingkannya dengan distribusi teoretis Benford. Tingkat penyimpangan antara kedua distribusi tersebut diukur menggunakan metrik *Mean Absolute Deviation* (MAD). Wiryadinata et al. (2023) dalam studi terbarunya menunjukkan bahwa penggunaan ambang batas MAD pada klaster transaksi merupakan metode *screening* awal yang efektif untuk memfilter anomali sebelum diproses lebih lanjut oleh algoritma AI. Jika nilai MAD melampaui batas toleransi statistik 0.05 sebagaimana disarankan oleh Nigrini (2012), maka populasi data diindikasikan memiliki risiko manipulasi struktural yang tinggi.

Sebagai pelengkap analisis statistik makro, penelitian ini menerapkan algoritma *Machine Learning* tipe *Unsupervised Learning* yaitu *Isolation Forest*. Algoritma yang diperkenalkan oleh Liu et al (2008) ini memiliki pendekatan yang berbeda dibandingkan metode deteksi pencilan tradisional berbasis jarak (*distance based*) atau kepadatan (*density based*) seperti *Local Outlier Factor* (Breunig et al., 2000). *Isolation Forest* bekerja dengan prinsip isolasi menggunakan struktur pohon keputusan acak (*iTrees*). Algoritma ini berasumsi bahwa data anomali (kecurangan) memiliki karakteristik "sedikit dan berbeda" (*few and different*), sehingga lebih mudah diisolasi di dekat akar pohon (Liu et al., 2012). Dalam sistem ini, setiap transaksi diberikan skor anomali, semakin pendek jalur (*path length*) yang dibutuhkan untuk mengisolasi sebuah titik data, semakin tinggi indikasi bahwa transaksi tersebut adalah *fraud*. Keunggulan utama metode ini adalah efisiensi komputasi yang linear dan kemampuannya menangani fenomena *masking* dan *swamping* yang sering terjadi pada dataset keuangan berdimensi tinggi.

Guna memvalidasi keandalan sistem, penelitian membandingkan hasil prediksi sistem terhadap label kebenaran (*ground truth*) yang telah ditetapkan saat pembangkitan data. Pendekatan ini mengacu pada standar validasi model deteksi *fraud* yang disarankan oleh Chaudhari & Charate (2025), di mana efektivitas sistem diukur dari kemampuannya meminimalkan kesalahan klasifikasi.

Pengujian keandalan dilakukan secara bertahap menggunakan tiga skenario dataset yang telah dirancang sebelumnya. Pada tahap pertama, sistem diuji menggunakan Dataset A (Normal) untuk mengukur tingkat kesalahan positif (*False Positive Rate*). Tujuannya adalah memastikan sistem tidak memberikan peringatan palsu secara berlebihan pada data yang bersih, yang dapat mengganggu efisiensi kerja auditor. Tahap kedua menggunakan Dataset B (Manipulasi Minor), yang bertujuan menguji sensitivitas algoritma *Isolation Forest* dalam mendeteksi anomali bernilai ekstrem yang jumlahnya minoritas. Tahap ketiga merupakan uji ketahanan (*robustness test*) menggunakan Dataset C (Manipulasi Masif). Skenario ini dirancang untuk menguji batas kemampuan algoritma AI ketika menghadapi distorsi data yang ekstrem. Pada tahap ini, keandalan sistem dinilai berdasarkan kemampuan metode hibrida (khususnya Hukum Benford) dalam mendeteksi kerusakan struktur distribusi data yang mungkin lolos dari deteksi *outlier* biasa.

Untuk mengkuantifikasi tingkat keandalan tersebut, penelitian ini menggunakan metrik evaluasi utama yaitu Tingkat Deteksi (*Detection Rate*). Metrik ini dipilih karena dalam konteks audit forensik, risiko kegagalan mendeteksi kecurangan (*False Negative*) dianggap lebih fatal dibandingkan risiko mencurigai data valid (*False Positive*). Rumus perhitungan tingkat deteksi adalah sebagai berikut:

$$Dataction\ Rate = \frac{TP}{TP + FN} \times 100\%$$

Dimana:

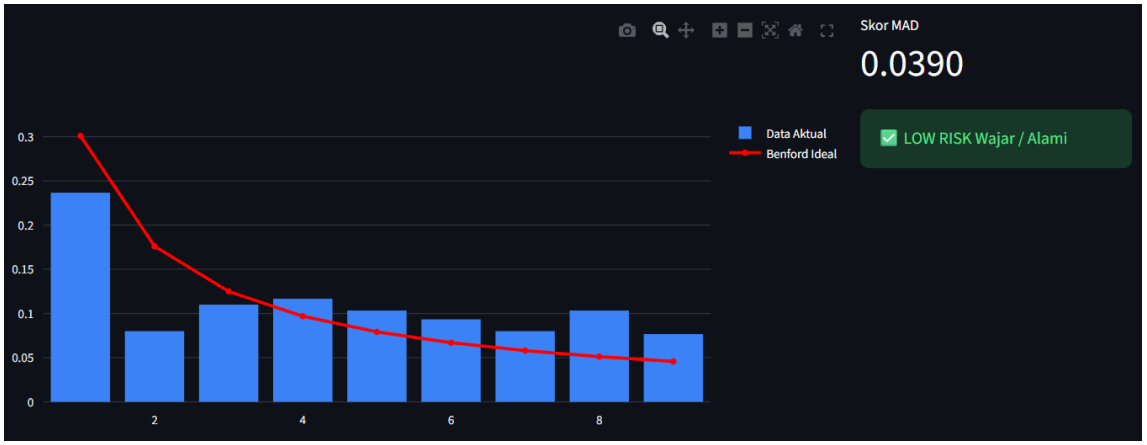
- TP (*True Positive*): Jumlah transaksi kecurangan yang berhasil dideteksi dengan tepat oleh sistem (status: *Anomali/High Risk*).
- FN (*False Negative*): Jumlah transaksi kecurangan yang gagal dideteksi atau lolos dari sistem (status: *Normal/Low Risk*).

Sistem dinyatakan andal apabila mampu mencapai tingkat deteksi di atas 90% pada kedua skenario manipulasi (Minor dan Masif), serta memberikan penjelasan visual yang dapat dipahami oleh pengguna.

HASIL PENELITIAN DAN PEMBAHASAN

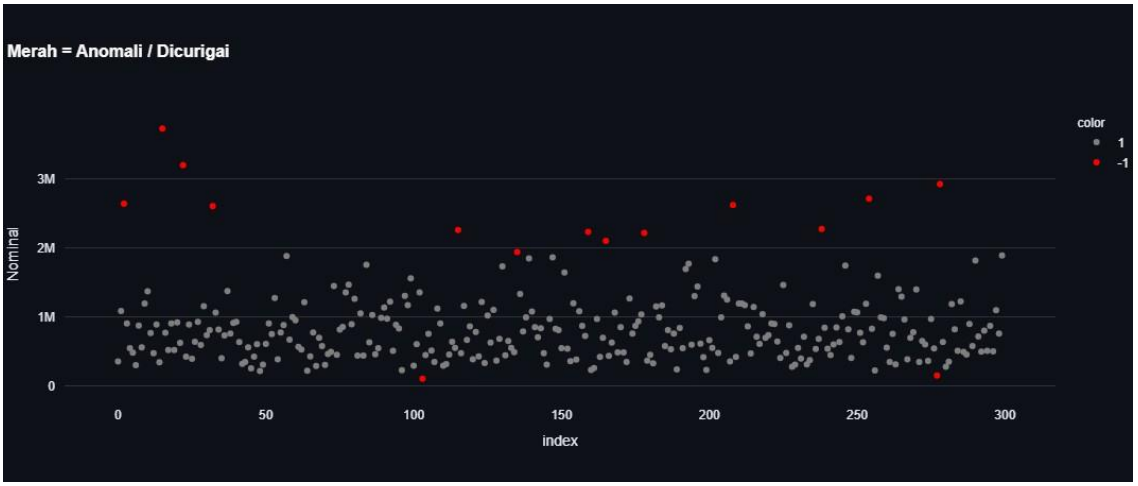
Analsis Dataset A Menggunakan Fraudlens

Pengujian tahap pertama difokuskan pada Dataset A yang merepresentasikan kondisi operasional normal tanpa adanya manipulasi data. Dataset ini memuat 300 entri transaksi valid yang didistribusikan mengikuti pola log normal. Tujuan utama dari skenario ini adalah untuk menetapkan garis dasar (*baseline*) kinerja sistem serta mengevaluasi tingkat kesalahan deteksi (*False Positive*) yang mungkin muncul saat sistem memproses data yang bersih. Berdasarkan hasil pemindaian menggunakan modul statistik, sistem menunjukkan harmonisasi yang tinggi antara distribusi data aktual dengan prediksi matematis Hukum Benford.



Gambar 1. Analisis Kepatuhan Hukum Benford (Dataset A)

Sebagaimana divisualisasikan pada Gambar 2, histogram frekuensi digit pertama data aktual memiliki pola penurunan logaritmik yang selaras dengan kurva referensi Benford. Secara kuantitatif, tingkat kesesuaian ini dikonfirmasi oleh nilai skor *Mean Absolute Deviation* (MAD) sebesar 0.0390. Mengingat nilai tersebut berada di bawah ambang batas kritis 0.05, sistem secara otomatis mengklasifikasikan dataset ini dalam status "LOW RISK" atau Wajar. Hal ini mengindikasikan bahwa variasi alami yang terjadi dalam transaksi, selama tidak ada intervensi struktural tetap akan dikenali oleh sistem sebagai data yang mematuhi hukum.

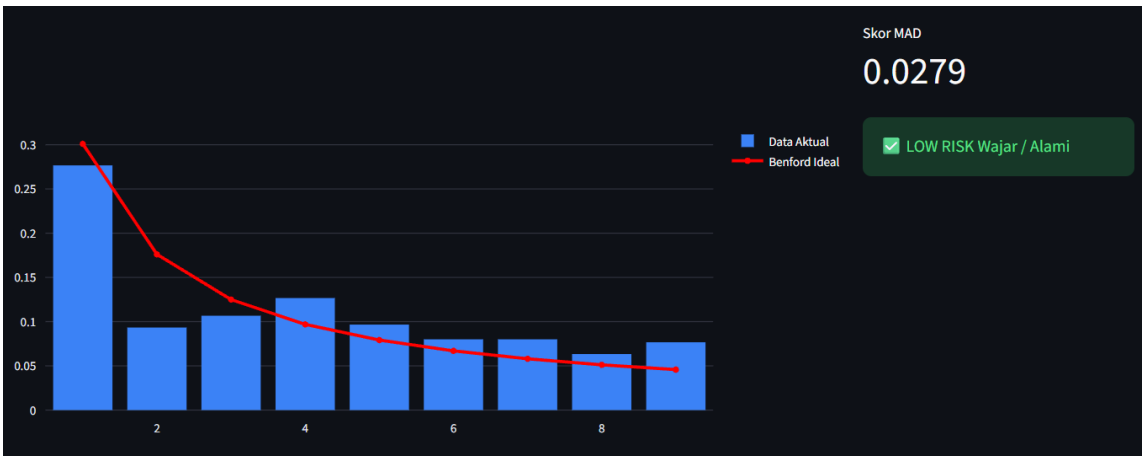


Gambar 2. Visualisasi Sebaran Outlier (Dataset A)

Evaluasi dilanjutkan pada modul deteksi anomali berbasis *Machine Learning*. Pada tahap ini, algoritma *Isolation Forest* mengidentifikasi adanya 15 transaksi yang ditandai sebagai *outlier* (titik merah pada *Scatter Plot*). Mengingat dataset ini tidak mengandung kecurangan (*Zero True Positive*), maka ke-15 deteksi tersebut dikategorikan sebagai peringatan palsu (*False Positive*). Berdasarkan formulasi evaluasi kinerja, tingkat kesalahan sistem pada kondisi normal dihitung dengan membagi jumlah deteksi anomali (15) dengan total populasi data (300), yang menghasilkan persentase sebesar 5%. Dengan kata lain, sistem terbukti memiliki tingkat spesifisitas sebesar 95% dalam mengenali data yang valid. Dalam konteks audit forensik, tingkat kesalahan 5% ini merupakan batas toleransi yang wajar, di mana sistem berfungsi efektif sebagai filter prioritas untuk mengarahkan auditor pada transaksi bernilai ekstrem tanpa membebani proses pemeriksaan dengan terlalu banyak peringatan yang tidak relevan.

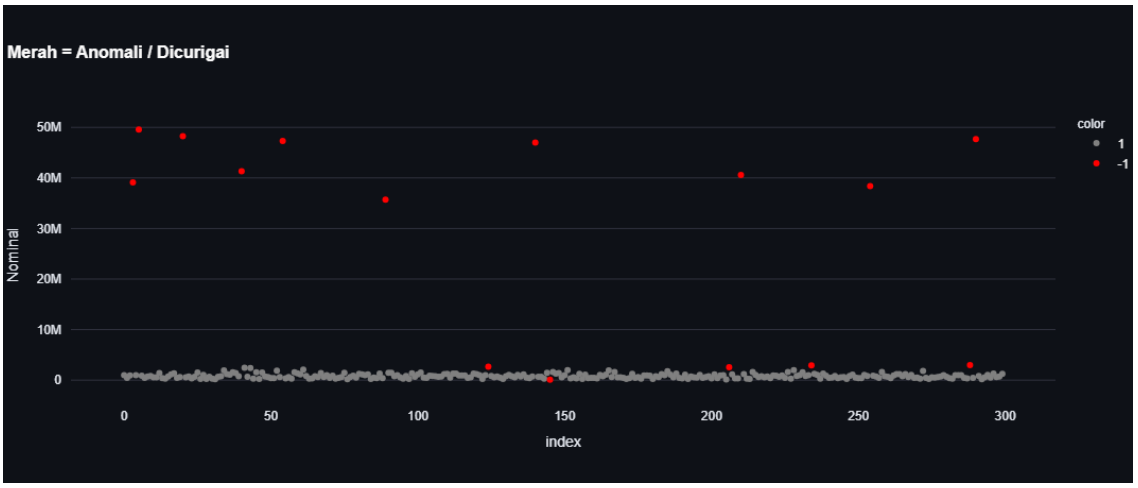
Analsis Dataset B Menggunakan Fraudlens

Setelah menetapkan garis dasar pada data normal, pengujian dilanjutkan menggunakan Dataset B. Skenario ini dirancang untuk mensimulasikan modus kecurangan "penyelundupan nilai ekstrem", di mana disisipkan 10 transaksi fiktif (3,3% dari populasi) dengan nominal fantastis (Rp 35.000.000 – Rp 50.000.000) di antara ratusan transaksi operasional yang wajar. Hasil pemindaian menggunakan Hukum Benford pada skenario ini mengungkap keterbatasan fundamental metode statistik.



Gambar 3. Analisis Kepatuhan Hukum Benford (Dataset B)

Sebagaimana ditampilkan pada Gambar 3, distribusi digit pertama data aktual masih terlihat selaras dengan kurva prediksi Benford. Sistem mencatatkan nilai skor MAD sebesar 0.0279. Karena nilai ini masih berada di bawah ambang batas 0.05, sistem menyimpulkan status "LOW RISK / Wajar". Kegagalan statistik dalam mendeteksi keberadaan 10 data manipulasi ini dikategorikan sebagai *False Negative*. Hal ini terjadi karena volume data manipulasi terlalu kecil dibandingkan total populasi, sehingga dominasi data normal berhasil "menutupi" penyimpangan tersebut dalam kalkulasi frekuensi agregat. Temuan ini mengonfirmasi bahwa audit yang hanya mengandalkan uji Benford rentan meloloskan *fraud* yang bersifat individual.



Gambar 4. Visualisasi Sebaran Outlier (Dataset B)

Berbanding terbalik dengan hasil statistik, modul *Machine Learning* menunjukkan sensitivitas yang superior. Algoritma *Isolation Forest* mendeteksi total 15 transaksi sebagai *outlier*. Berdasarkan visualisasi pada Gambar 4, terlihat jelas adanya pemisahan klaster yang ekstrem. Titik-titik merah pada bagian atas grafik merepresentasikan transaksi manipulasi dengan nilai di atas Rp 35 Juta yang berhasil diisolasi sepenuhnya oleh algoritma.

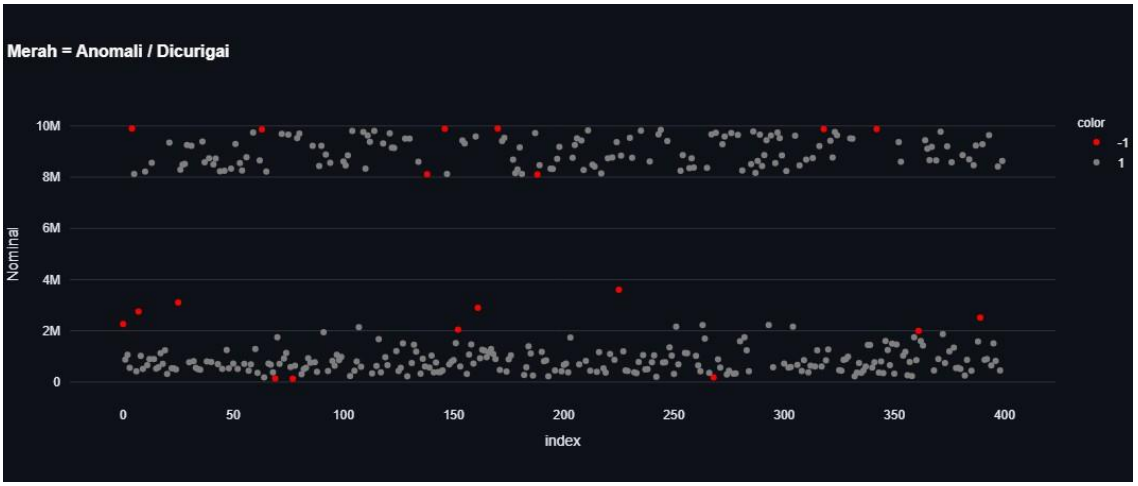
Dari 15 deteksi tersebut, 10 di antaranya adalah transaksi manipulasi yang disuntikkan, sementara 5 lainnya merupakan variasi data normal (*noise*) yang terbawa. Mengacu pada rumus evaluasi, sistem mencatatkan skor *Detection Rate* sebesar 100%. Hal ini membuktikan bahwa pada kasus manipulasi berskala kecil, algoritma AI memiliki sensitivitas sempurna untuk memisahkan data curang dari data normal.

Keberhasilan AI ini disebabkan oleh mekanisme kerjanya yang tidak bergantung pada frekuensi digit, melainkan pada jarak isolasi data. Transaksi bernilai puluhan juta rupiah tersebut memiliki densitas yang sangat rendah dan jarak yang jauh dari klaster data mayoritas, sehingga sistem dengan mudah menandainya sebagai anomali. Perbandingan hasil pada Skenario B ini membuktikan secara empiris bahwa integrasi AI mutlak diperlukan untuk menutupi *blind spot* metode statistik dalam mendeteksi kecurangan berskala mikro.

Analsis Dataset C Menggunakan Fraudlens

Skenario pengujian terakhir merupakan uji ketahanan (*robustness test*) menggunakan Dataset C, yang dirancang untuk mensimulasikan kondisi korupsi terstruktur atau penggelembungan (*mark up*) anggaran secara sistematis. Pada dataset ini, disuntikkan sebanyak 150 data transaksi manipulasi (setara dengan 37,5% dari total 400 populasi data) dengan pola nominal berulang pada kisaran Rp 8.000.000 hingga Rp 9.900.000. Tujuan utama dari eksperimen ini adalah untuk mengevaluasi respons sistem ketika proporsi data kecurangan cukup besar untuk mendominasi profil data normal. Hasil pemindaian algoritma *Isolation Forest* pada skenario ini mengungkap kerentanan fundamental *Machine Learning* ketika menghadapi serangan data berskala masif. Sebagaimana divisualisasikan pada Gambar 5 , terbentuk dua klaster data yang sangat padat: klaster bawah yang merupakan transaksi valid dan klaster atas yang merupakan transaksi manipulasi. Namun, bertentangan dengan ekspektasi, algoritma hanya mengidentifikasi 19 transaksi sebagai *outlier*, sementara 131 transaksi manipulasi lainnya lolos dari deteksi. Lebih spesifik lagi, dari 19 temuan tersebut, hanya 8 transaksi yang benar-benar merupakan data manipulasi (*True Positive*), sementara sisanya adalah peringatan palsu.

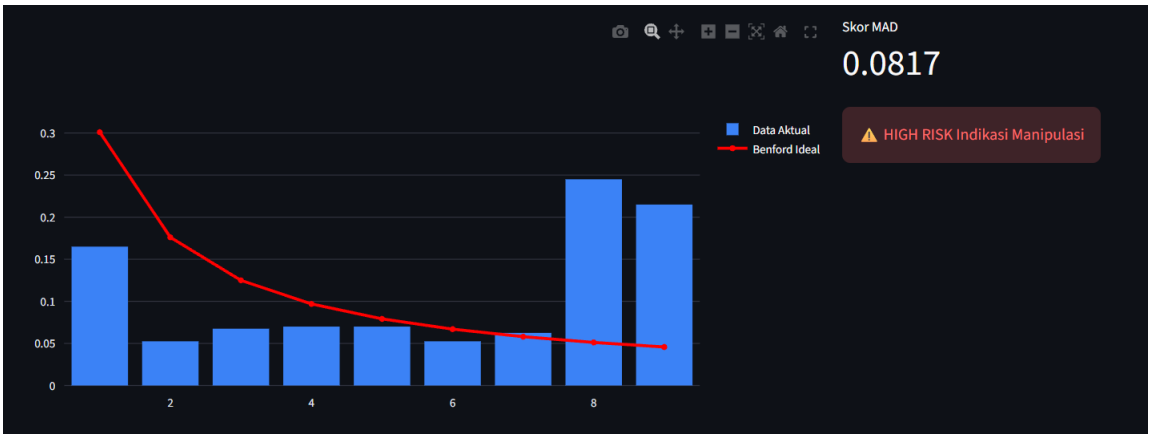
Secara matematis, hal ini menyebabkan skor *Detection Rate* anjlok menjadi hanya 5,3%. Artinya, sebanyak 142 transaksi manipulasi (94,7%) gagal dideteksi dan dianggap wajar oleh sistem.



Gambar 5. Visualisasi Sebaran Outlier (Dataset C)

Kegagalan deteksi ini terjadi akibat fenomena *Masking Effect* atau penyamaran data. Karena volume data manipulasi sangat besar (150 entri) dan menggerombol secara rapat pada rentang nilai yang sama, algoritma menginterpretasikan kepadatan tersebut sebagai pola yang wajar (*normal cluster*). Hal ini dikonfirmasi oleh fitur analisis wawasan AI sistem yang secara otomatis memperluas batas rentang wajar (*normal boundaries*) menjadi Rp 200.638 hingga Rp 9.834.417. Akibatnya, transaksi *mark up* bernilai 8 hingga 9 juta rupiah dianggap masuk dalam toleransi kewajaran sistem. Temuan ini membuktikan bahwa tanpa validasi silang, AI berisiko mengalami *Normalization of Deviance*, di mana anomali yang terjadi berulang-ulang dianggap sebagai standar normal yang baru.

Di saat algoritma AI mengalami bias akibat volume data, metode statistik Hukum Benford menunjukkan peran vitalnya sebagai jaring pengaman terakhir. Grafik distribusi digit pada **Gambar 6** memperlihatkan kerusakan struktural data yang sangat ekstrem.



Gambar 6. Analisis Kepatuhan Hukum Benford (Dataset C)

Terjadi lonjakan frekuensi yang drastis pada digit awal 8 dan 9, yang visualisasinya jauh melampaui garis prediksi Benford. Secara matematis, penyimpangan ini menghasilkan skor *Mean Absolute Deviation* (MAD) sebesar 0.0817. Mengingat nilai ini melampaui ambang batas kritis 0.05, sistem secara tegas memberikan status "HIGH RISK / Indikasi Manipulasi". Berbeda dengan AI yang "terkecoh" oleh densitas klaster data, Hukum Benford bersifat objektif terhadap probabilitas kemunculan digit, sehingga manipulasi sebesar apa pun tidak dapat bersembunyi di balik volume data.

Temuan kontradiktif antara Skenario B (di mana AI unggul) dan Skenario C (di mana Statistik unggul) menyempurnakan pembuktian hipotesis penelitian ini. Eksperimen membuktikan bahwa penggunaan metode tunggal memiliki celah keamanan yang fatal: metode statistik lemah terhadap *fraud* mikro, sedangkan metode AI rentan bias terhadap *fraud* makro. Oleh karena itu, integrasi kedua metode dalam sistem *FraudLens* terbukti menjadi solusi yang komprehensif, menciptakan mekanisme pertahanan lapis ganda yang mampu menutupi kelemahan masing-masing pendekatan untuk menjamin integritas hasil audit.

KESIMPULAN

Penelitian ini berhasil mengembangkan *FraudLens*, sebuah sistem deteksi kecurangan hibrida yang mengintegrasikan Hukum Benford dan algoritma *Isolation Forest*. Temuan empiris membuktikan bahwa penggunaan metode tunggal memiliki celah keamanan fatal, di mana metode statistik terbukti lemah dalam mendeteksi manipulasi berskala mikro, sedangkan *Machine Learning* rentan mengalami bias (*normalization of deviance*) saat menghadapi manipulasi berskala masif. Oleh karena itu, integrasi kedua metode terbukti menjadi solusi pertahanan lapis ganda yang komprehensif, di mana statistik mengamankan integritas struktural data sementara AI efektif mengisolasi anomali pencilan. Secara praktis, penelitian ini menghadirkan instrumen audit digital yang efisien dan terjangkau bagi entitas mikro, dengan kemampuan spesifisitas tinggi dalam memilah transaksi wajar. Kontribusi teoretis ini menegaskan perlunya validasi silang antara pendekatan parametrik dan non parametrik dalam audit forensik. Untuk pengembangan selanjutnya, disarankan agar pengujian diperluas menggunakan data keuangan riil guna mengukur ketahanan sistem terhadap variasi modus kecurangan yang lebih kompleks di lingkungan operasional nyata.

DAFTAR PUSTAKA

BUKU

Financial Conduct Authority (FCA). (2024). *Report: Using Synthetic Data in Financial Services*. London: FCA.

Nigrini, M. J. (2012). *Benford's Law: Applications for forensic accounting, auditing, and fraud detection*. John Wiley & Sons.

Rindrayani, S. R., Rustiyana, R., Judijanto, L., Abdullah, G., & Ardiyanti, A. D. (2025). *Metode Penelitian dan Pengembangan: R&D Research and Development*. PT. Sonpedia Publishing Indonesia.

JURNAL

Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American philosophical society*, 551-572.

Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000). LOF: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data* (pp. 93-104). https://doi.org/10.1145/342009.335388

Chaudhari, A. V., & Charate, P. A. (2025). Synthetic Financial Document Generation and Fraud Detection Using Generative AI and Explainable ML. *Journal of Recent Trends in Computer Science and Engineering*, *13*(2), 45-59. https://doi.org/10.70589/JRTCSE.2025.13.2.6

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. In *2008 eighth ieee international conference on data mining* (pp. 413-422). IEEE.

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2012). Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *6*(1), 1-39. https://doi.org/10.1145/2133360.2133363

Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of mathematics*, *4*(1), 39-40. https://doi.org/10.2307/2369148

Safitri, N., Saputri, Y. O., & Sanjaya, I. M. (2025). Analisis peran investasi dalam mendorong pertumbuhan ekonomi nasional di era globalisasi. Jurnal Media Akademik (JMA), 3(11). https://doi.org/10.62281/sgz38g41

Wiryadinata, D., Sugiharto, A., & Tarno. (2023). The Use of Machine Learning to Detect Financial Transaction Fraud: Multiple Benford Law Model for Auditors. *Journal of Information Systems Engineering and Business Intelligence*, 9(1). http://dx.doi.org/10.20473/jisebi.9.2.239-252

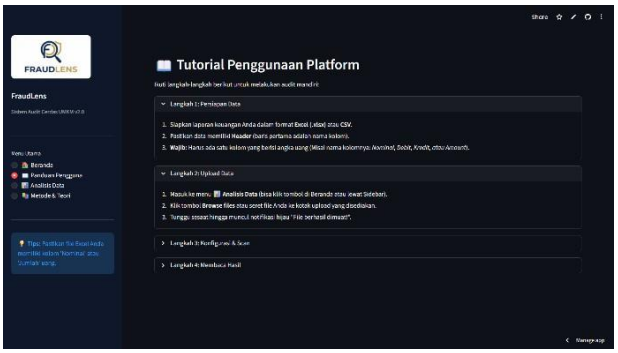
INTERNET

Association of Certified Fraud Examiners (ACFE) Indonesia Chapter. (2020). Survei fraud Indonesia 2019. ACFE Indonesia Chapter. https://acfe-indonesia.or.id/survei-fraud-indonesia/

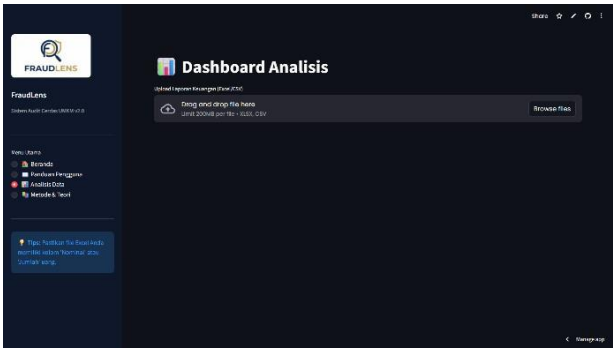
LAMPIRAN

Lampiran 1. Dokumentasi Antarmuka Pengguna (UI/UX)

Gambar 7. Halaman Beranda



Gambar 8. Halaman Tutorial



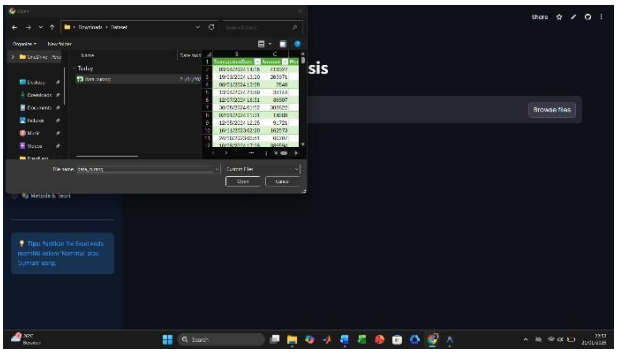
Gambar 9. Halaman Dashboard Analisis

Gambar 10. Halaman Landasan Ilmiah

Antarmuka sistem *FraudLens* dirancang secara terintegrasi, dimulai dari halaman ***Beranda*** yang berfungsi sebagai gerbang interaksi utama berisi proposisi nilai dan navigasi cepat. Untuk memastikan akseptabilitas pengguna awam serta transparansi ilmiah, sistem dilengkapi halaman ***Tutorial*** yang memandu operasional teknis langkah demi langkah serta halaman ***Landasan Ilmiah*** yang menjabarkan dasar teoritis metode hibrida (Benford dan *Isolation Forest*) agar hasil analisis dapat dipertanggungjawabkan. Seluruh proses eksekusi kemudian dipusatkan pada ***Dashboard Analisis***, sebuah modul inti yang menangani input data pengguna, validasi format otomatis, hingga penyajian visualisasi deteksi kecurangan secara *real-time*.

Alur Penggunaan *FraudLens*

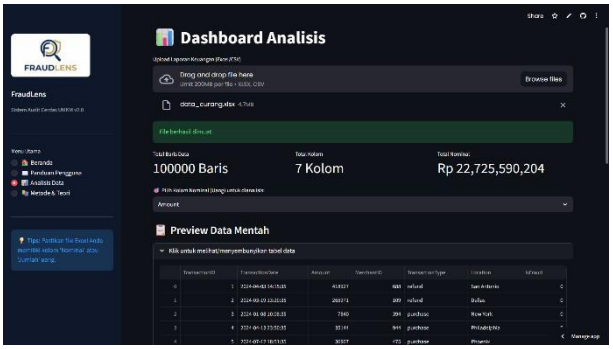
1. Input File Laporan



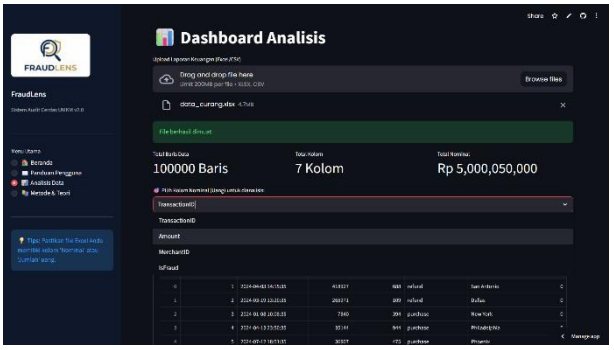
Gambar 11. Proses Upload File

Untuk memulai analisis kecurangan laporan dilakukan dengan mengupload file laporan dengan format .xlsx atau .csv pada halaman Dashboard Analisis.

2. Tampilan Awal Dashboard Analisis Setelah Dinput Data



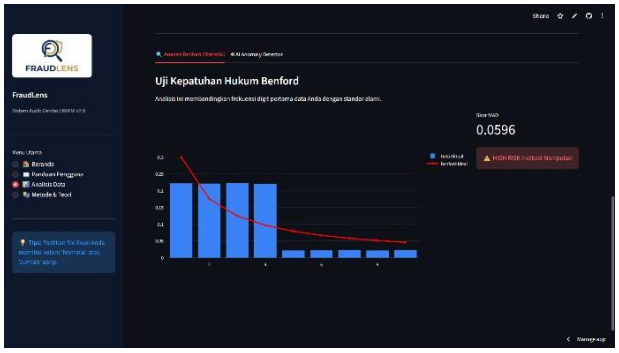
Gambar 12. Tampilan Setelah Diinput Data



Gambar 13. Pemilihan Kolom untuk Dianalisis

Setelah dokumen berhasil diunggah, sistem secara otomatis melakukan pemindaian awal (*parsing*) yang hasilnya ditampilkan pada **Gambar 12**. Antarmuka secara responsif menyajikan metadata statistik vital meliputi total baris, jumlah kolom, dan akumulasi nominal transaksi serta tabel pratinjau (*raw data preview*) untuk memverifikasi bahwa integritas data tetap terjaga selama proses transfer. Selanjutnya, **Gambar 13** mendemonstrasikan mekanisme konfigurasi analisis, di mana pengguna diwajibkan melakukan pemetaan variabel dengan memilih kolom target (misalnya 'Amount') melalui menu *dropdown*. Fitur seleksi kolom ini krusial untuk memastikan algoritma *Benford* dan *Isolation Forest* hanya memproses atribut numerik yang relevan, sehingga akurasi deteksi anomali dapat terjamin sebelum proses pemindaian mendalam dimulai.

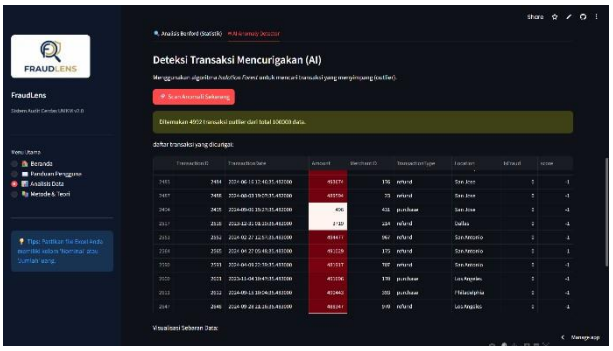
3. Uji kepatuhan Hukum Benford



Gambar 14. Dokumentasi Hasil Uji Kepatuhan Hukum Benford

Setelah konfigurasi variabel selesai, sistem menyajikan hasil analisis statistik makro pada modul Uji Kepatuhan Hukum Benford (**Gambar 14**). Antarmuka ini menampilkan visualisasi grafik komparatif yang menyandingkan distribusi frekuensi digit pertama dari data aktual pengguna (direpresentasikan oleh histogram batang biru) dengan kurva distribusi probabilitas ideal Benford (garis merah). Untuk memudahkan pengambilan keputusan, sisi kanan panel dilengkapi dengan metrik kuantitatif berupa nilai skor *Mean Absolute Deviation* (MAD) serta indikator status risiko otomatis (label peringatan berwarna) yang menyimpulkan apakah data tergolong wajar atau terindikasi manipulasi.

4. Deteksi Transaksi Kecurangan



Gambar 15. Dokumentasi Transaksi Mencurigakan

Gambar 16. Pemilihan Kolom untuk Dianalisis

Tahap akhir dari proses audit digital difasilitasi oleh modul Deteksi Transaksi Kecurangan (AI), seperti yang terlihat pada **Gambar 15**. Modul ini merupakan antarmuka bagi algoritma *Machine Learning* (*Isolation Forest*) untuk melakukan pemindaian mendalam terhadap setiap baris transaksi. Antarmuka menyajikan hasil deteksi dalam dua format utama:

1. **Tabel Rincian Anomali**
   Daftar interaktif yang secara spesifik menampilkan baris data yang teridentifikasi sebagai *outlier* (pencilan), lengkap dengan atribut kunci (seperti ID transaksi, tanggal, dan nominal) serta label skor anomali (-1) yang ditandai dengan warna merah untuk memudahkan prioritas pemeriksaan.
2. **Visualisasi Sebaran Data**
   Grafik *scatter plot* yang memetakan posisi seluruh transaksi, di mana titik data berwarna merah merepresentasikan anomali yang terisolasi dari klaster data normal (titik abu-abu), memberikan gambaran visual mengenai pola sebaran kecurangan terhadap populasi data secara keseluruhan.

Lampiran 2. Tautan Repositori Kode dan Akses Aplikasi

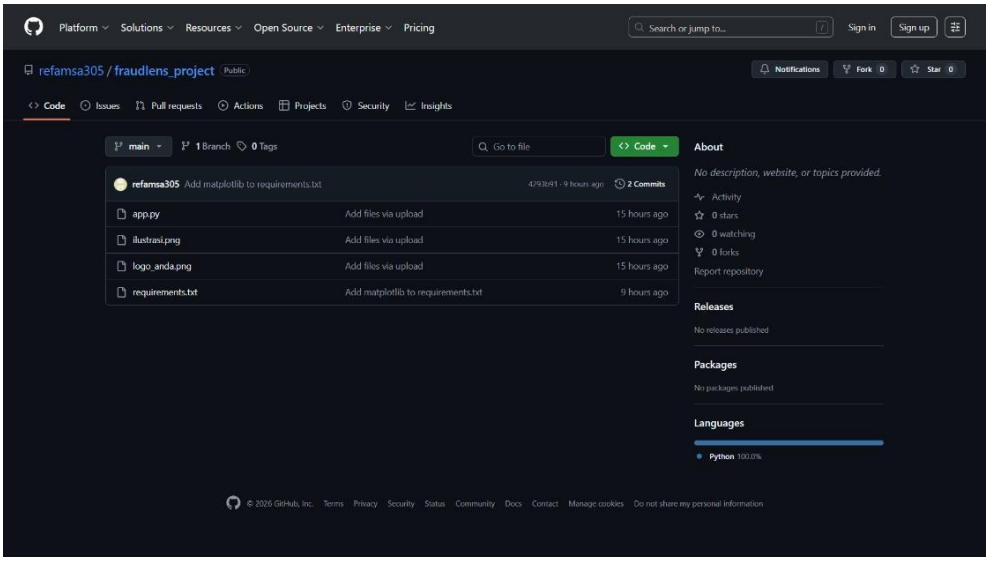
Penelitian ini menjunjung tinggi prinsip transparansi dan reproduksibilitas ilmiah (*scientific reproducibility*). Seluruh kode sumber (*source code*), dataset simulasi, dan aset pendukung yang digunakan dalam pengembangan sistem *FraudLens* dapat diakses secara terbuka oleh publik. Selain itu, prototipe aplikasi telah diunggah ke layanan *cloud hosting* agar dapat diuji coba secara langsung tanpa perlu instalasi lokal.

Berikut adalah rincian akses untuk repositori dan demo aplikasi:

A. Repositori Kode Sumber (GitHub)

Repositori ini memuat seluruh skrip pemrograman Python (termasuk algoritma *Isolation Forest* dan Hukum Benford), file konfigurasi sistem, serta dataset yang digunakan dalam pengujian.

Platform : Github
Pemilik : refamsa305
Nama Repositori : fraudlens_project
Tautan URL : https://github.com/refamsa305/fraudlens_project.git



Gambar 17. Dokumentasi Github

B. Demo Aplikasi (Streamlit Community Cloud)

Aplikasi *FraudLens* dapat diakses melalui peramban web (*web browser*) modern (seperti Google Chrome, Microsoft Edge, atau Safari) melalui tautan berikut:
https://fraudlensproject.streamlit.app/

FRAUDLENS

Gambar 18. Logo Fraudlens