

Studi Perbandingan Pola Transaksi Nasabah Normal dan Fraudulent dengan Pendekatan EDA pada Data Perbankan

Luthfatul ika zahro^{1*},

¹Institut Teknologi Nasional Malang

¹2318082@scholar.itn.ac.id



Histori Artikel:

Diajukan: 27 October 2025

Disetujui: 28 October 2025

Dipublikasi: 29 October 2025

Kata Kunci:

Analisis Data Penjualan

Menggunakan Metode

Exploratory Data Analysis

(EDA) pada Dataset Transaksi

Barang

Digital Transformation

Technology (Digitech) is an

Creative Commons License This

work is licensed under a

Creative Commons Attribution-

NonCommercial 4.0 International

(CC BY-NC 4.0).

Abstrak

Aktivitas transaksi keuangan perbankan semakin meningkat seiring berkembangnya layanan digital, namun hal ini juga diikuti oleh meningkatnya risiko transaksi fraud. Penelitian ini bertujuan untuk menganalisis perbedaan perilaku transaksi antara nasabah normal dan nasabah fraudulent dengan menggunakan pendekatan Exploratory Data Analysis (EDA). Dataset yang digunakan berasal dari data perbankan yang berisi atribut seperti jumlah transaksi, saldo sebelum dan sesudah transaksi, serta status transaksi (normal atau fraud).

Melalui tahapan EDA, dilakukan analisis statistik deskriptif, visualisasi distribusi variabel, dan korelasi antar fitur untuk mengungkap pola yang membedakan kedua kelompok nasabah. Hasil analisis menunjukkan bahwa nasabah dengan aktivitas fraud cenderung memiliki nominal transaksi yang lebih besar, frekuensi transaksi tinggi dalam waktu singkat, serta perubahan saldo yang signifikan dibandingkan dengan nasabah normal. Temuan ini memberikan wawasan awal dalam memahami karakteristik perilaku finansial yang berisiko serta dapat dijadikan acuan bagi sistem pendeteksian fraud berbasis data

PENDAHULUAN

Perkembangan layanan perbankan digital memudahkan transaksi nasabah, namun juga meningkatkan risiko penipuan (*fraud*). Deteksi dini transaksi mencurigakan penting untuk mencegah kerugian finansial dan reputasi bagi bank. Analisis Data Eksploratif (EDA) membantu memahami pola transaksi melalui visualisasi dan statistik, sehingga perbedaan antara transaksi normal dan fraudulent dapat diidentifikasi.

Penelitian ini bertujuan membandingkan pola transaksi nasabah normal dan fraudulent menggunakan EDA, untuk memberikan wawasan yang berguna dalam pengembangan sistem deteksi penipuan yang lebih efektif

STUDI LITERATUR

Penelitian terkait deteksi transaksi fraud telah banyak dilakukan dengan berbagai pendekatan analisis data. Menurut Han et al. (2021), *fraud detection* merupakan proses penting dalam mengidentifikasi aktivitas tidak sah yang melibatkan penyalahgunaan sistem keuangan. Pendekatan berbasis *machine learning* sering digunakan untuk memprediksi kemungkinan terjadinya fraud, namun sebelum model diterapkan, tahap eksplorasi awal melalui *Exploratory Data Analysis (EDA)* sangat penting untuk memahami distribusi data dan hubungan antar variabel.

Menurut Tukey (1977), EDA merupakan pendekatan untuk memahami struktur data melalui metodestatistik dan visualisasi guna menemukan pola, anomali, atau hubungan antar fitur. Dalam konteks transaksi perbankan, EDA dapat membantu mendeteksi perbedaan perilaku antara kelompok pengguna. Studi yang dilakukan oleh Carcillo et al. (2019) menunjukkan bahwa transaksi fraud sering kali memiliki karakteristik ekstrem, seperti nominal besar, pola berulang dalam waktu singkat, serta dilakukan di luar jam operasional. Selain itu, penelitian oleh Dal Pozzolo et al. (2018) juga menggarisbawahi bahwa ketidakseimbangan jumlah data antara transaksi fraud dan normal merupakan tantangan umum yang perlu ditangani dengan analisis cermat.

Dengan demikian, EDA menjadi fondasi penting dalam memahami perilaku transaksi finansial dan dapat membantu perbankan mengembangkan sistem deteksi dini berbasis data. Penelitian ini mengadopsi pendekatan serupa dengan fokus utama pada identifikasi perbedaan karakteristik perilaku transaksi antara nasabah normal dan nasabah fraudulent menggunakan teknik eksplorasi visual dan statistik deskriptif.

METODE

Langkah-langkah dalam metode penelitian yang dilakukan adalah sebagai berikut:

1. Pengumpulan Data

Data yang digunakan berasal dari dataset publik Kaggle berisi transaksi nasabah perbankan, baik normal maupun fraudulent. Variabel utama meliputi amount, type, oldbalanceOrg, newbalanceOrig, oldbalanceDest, newbalanceDest, dan isFraud sebagai indikator penipuan.

2. Integrasi Data

Data diimpor ke Python menggunakan pustaka pandas. Dilakukan pengecekan tipe data, nilai hilang, dan pemisahan data menjadi dua kelompok: transaksi normal (isFraud=0) dan fraudulent (isFraud=1) untuk analisis komparatif.

3. Pembersihan Data

Proses ini meliputi penanganan missing values, penghapusan outlier, penyesuaian tipe data, serta validasi logika antara saldo awal, jumlah transaksi, dan saldo akhir agar data siap dianalisis.

4. Eksplorasi dan Visualisasi Data

Tahap EDA digunakan untuk menggali perbedaan pola antara transaksi normal dan fraud melalui analisis deskriptif dan visualisasi seperti countplot, bar chart, heatmap, dan scatter plot.

Hasilnya menunjukkan transaksi fraud lebih sering terjadi pada jenis TRANSFER dan CASH_OUT dengan nominal tinggi.

5. Analisis Data

Dilakukan perbandingan statistik sederhana antara kedua kelompok. Korelasi antar variabel dianalisis untuk melihat hubungan antar fitur. Hasil menunjukkan perbedaan signifikan dalam pola transaksi serta ketidakseimbangan data antara kelas normal dan fraud.

6. Kesimpulan

Pendekatan EDA berhasil mengidentifikasi ciri utama transaksi fraud: nominal besar, jenis transaksi tertentu (TRANSFER/CASH_OUT), dan ketidaksesuaian saldo.

HASIL

A. Hasil penggabungan dua dataset transaksi menghasilkan ribuan entri nasabah dengan data yang telah dibersihkan. Proses pembersihan mencakup konversi tanggal ke format datetime, kolom numerik ke tipe data angka, penghapusan duplikat, penambahan label (Normal dan Fraud), serta penggabungan kedua dataset menjadi satu transaksi

Dataset 1 (setelah dibersihkan):														
company name	model name	mobile weight	ram	front camera	back camera	processor	battery capacity	screen size	screen size	launched price (pakistan)	launched price (india)	launched price (china)	launched price (usa)	launched price
0	Apple	iPhone 16 128GB	174g	5.9	12MP	48.0 A17 Bionic	3.0 6.1 inches	3.0 6.1 inches	3.0 6.1 inches	PKR 224,999	79999.0	CNY 5,799	USD 799	
1	Apple	iPhone 16 256GB	174g	5.9	12MP	48.0 A17 Bionic	3.0 6.1 inches	3.0 6.1 inches	3.0 6.1 inches	PKR 234,999	84999.0	CNY 6,499	USD 849	
2	Apple	iPhone 16 512GB	174g	5.9	12MP	48.0 A17 Bionic	3.0 6.1 inches	3.0 6.1 inches	3.0 6.1 inches	PKR 244,999	89999.0	CNY 6,499	USD 899	
3	Apple	iPhone 16 Plus 128GB	203g	5.9	12MP	48.0 A17 Bionic	4.0 6.7 inches	4.0 6.7 inches	4.0 6.7 inches	PKR 249,999	89999.0	CNY 6,499	USD 899	
4	Apple	iPhone 16 Plus 256GB	203g	5.9	12MP	48.0 A17 Bionic	4.0 6.7 inches	4.0 6.7 inches	4.0 6.7 inches	PKR 259,999	94999.0	CNY 6,499	USD 949	

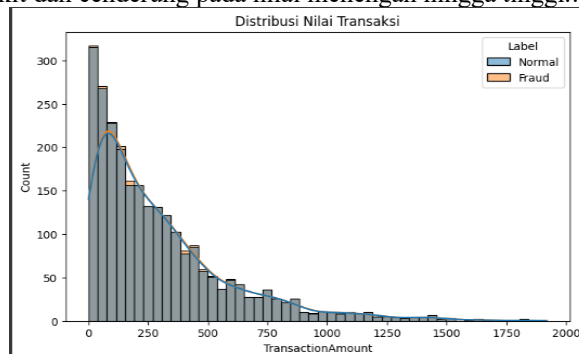
Dataset 2 (setelah dibersihkan):														
unnamed: 0	name	brand	model	battery capacity (mAh)	screen size (inches)	touchscreen	resolution x	resolution y	processor	ram (mb)	internal storage (gb)	rear camera	front	from
0	OnePlus 7T Pro McLaren Edition	OnePlus	7T Pro McLaren Edition	4085.0	6.67	Yes	1440	3120	8	12000	256.0	48.0	48.0	
1	Realme X2 Pro	Realme	X2 Pro	4000.0	6.59	Yes	1080	2400	8	6000	64.0	64.0	64.0	
2	iPhone 11 Pro Max	Apple	iPhone 11 Pro Max	3669.0	6.59	Yes	1242	2688	6	4000	64.0	12.0	12.0	
3	iPhone 11	Apple	iPhone 11	3110.0	6.19	Yes	828	1792	6	4000	64.0	12.0	12.0	
4	LG G8X ThinQ	LG	G8X ThinQ	4000.0	6.49	Yes	1080	2340	8	6000	128.0	12.0	12.0	

B. Eksplorasi dan Visualisasi Data

Beberapa grafik utama yang dihasilkan antara lain:

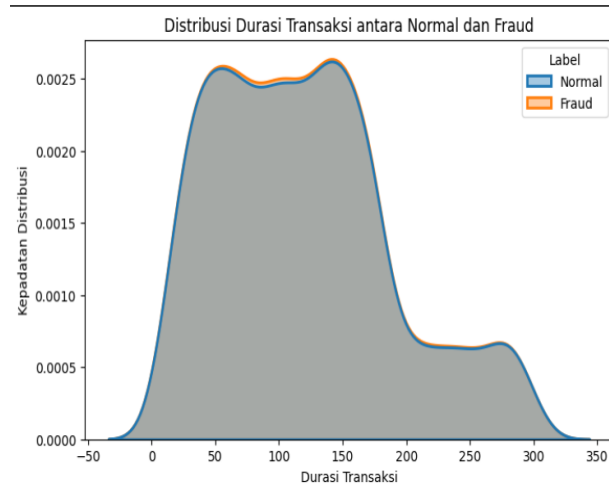
1. Distribusi nilai transaksi normal dan fraud

Grafik tersebut menunjukkan bahwa Sebagian besar transaksi pada grafik tersebut adalah transaksi normal, karena jumlahnya jauh lebih banyak dan mendominasi di semua rentang nilai, terutama pada nilai transaksi kecil. Transaksi fraud hanya muncul sedikit dan cenderung pada nilai menengah hingga tinggi..



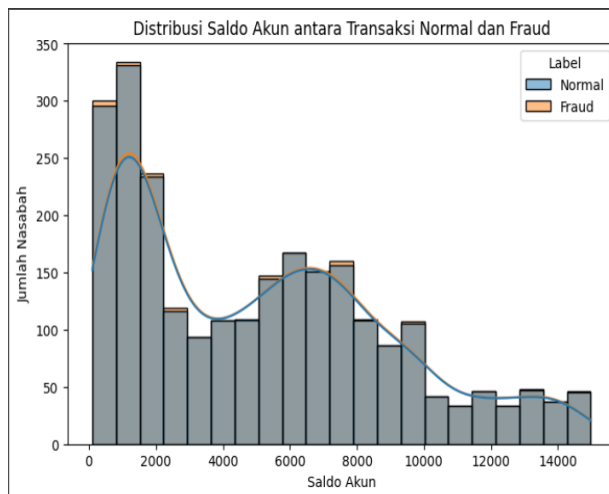
2. Perbandingan Durasi Transaksi (TransactionDuration)

Grafik tersebut menunjukkan bahwa sebagian besar transaksi, baik normal maupun fraud, memiliki durasi antara 50 hingga 200 detik. Pola distribusinya hampir sama, menandakan bahwa durasi transaksi tidak terlalu membedakan kedua jenis transaksi. Namun, transaksi normal tetap sedikit lebih banyak di seluruh rentang durasi..



3. Distribusi Saldo Akun (AccountBalance)

Grafik tersebut menunjukkan distribusi saldo akun antara transaksi normal dan fraud dengan pola yang naik turun pada beberapa rentang saldo. Terlihat bahwa jumlah nasabah cukup tinggi pada saldo di bawah 2000, kemudian menurun dan kembali meningkat di sekitar saldo 6000–8000 sebelum akhirnya turun lagi. Pola distribusi antara transaksi normal dan fraud hampir sama, namun transaksi normal tetap lebih banyak di seluruh rentang saldo

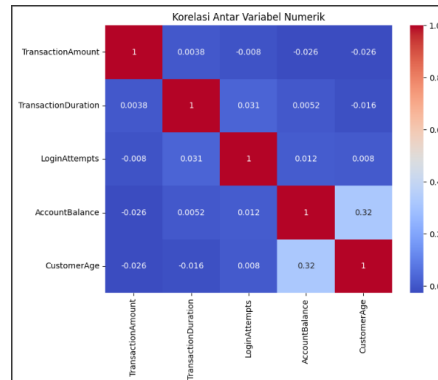


PEMBAHASAN

Berdasarkan hasil visualisasi heatmap korelasi, terlihat hubungan antar variabel numerik pada data nasabah. Nilai korelasi berada pada rentang -1 hingga 1, di mana semakin mendekati 1 menunjukkan hubungan positif yang kuat. Dari hasil korelasi Pearson, Kendall, dan Spearman, dapat dilihat bahwa beberapa variabel memiliki hubungan positif cukup kuat, sedangkan lainnya lemah atau hampir tidak berkorelasi. Visualisasi ini membantu memahami keterkaitan antar fitur seperti saldo, durasi, dan jumlah transaksi. Selain itu, analisis tren waktu menunjukkan perbandingan jumlah transaksi normal dan fraud setiap bulan, sehingga dapat terlihat pola terjadinya aktivitas penipuan pada data nasabah dari waktu ke waktu.

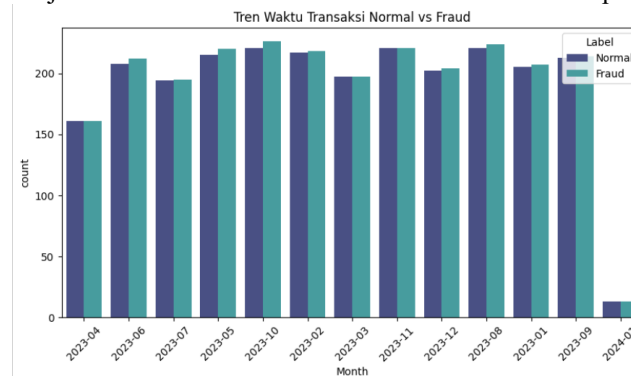
4. Heatmap korelasi antar variabel numerik

Grafik heatmap di atas menunjukkan korelasi antar variabel numerik pada data nasabah, seperti TransactionAmount, TransactionDuration, LoginAttempts, AccountBalance, dan CustomerAge. Dari hasil visualisasi, tampak bahwa sebagian besar korelasi antar variabel bernilai mendekati nol, yang berarti hubungan antar variabel relatif lemah atau hampir tidak ada. Satu-satunya hubungan yang cukup terlihat adalah antara AccountBalance dan CustomerAge dengan nilai korelasi sekitar 0.32, menunjukkan bahwa nasabah yang lebih tua cenderung memiliki saldo akun lebih tinggi.



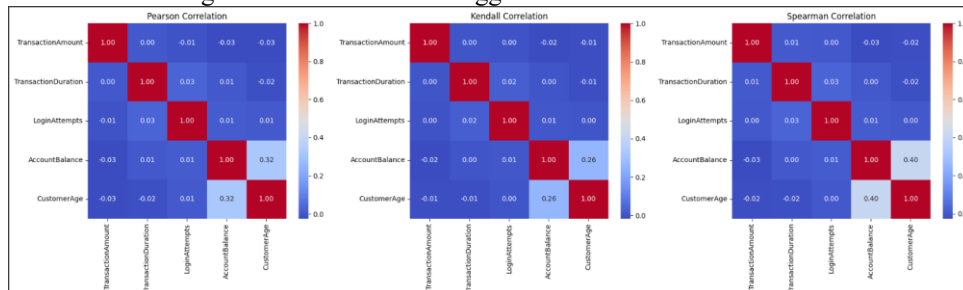
5. Analisis tren waktu (jika ada kolom tanggal)

Grafik tersebut menunjukkan tren waktu transaksi normal dan fraud pada data nasabah dari April 2023 hingga Januari 2024. Terlihat bahwa jumlah transaksi, baik normal maupun fraud, cenderung stabil setiap bulan, dengan perbedaan jumlah yang tidak terlalu signifikan. Pada beberapa bulan seperti Mei hingga Desember 2023, transaksi fraud sedikit lebih banyak dibanding normal, sementara pada bulan lain jumlahnya hampir seimbang. Namun, terjadi penurunan tajam jumlah transaksi pada Januari 2024, baik untuk transaksi normal maupun fraud. Secara keseluruhan, grafik ini menunjukkan bahwa aktivitas transaksi relatif konsisten sepanjang tahun.



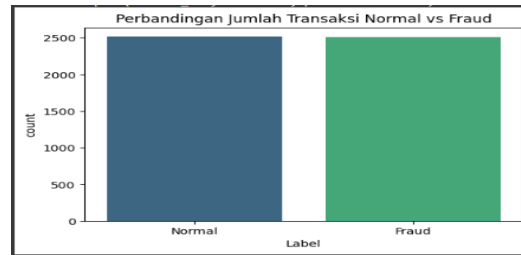
6. Tampilkan tabel korelasi

Sementara itu diagram tersebut menunjukkan hasil korelasi Pearson, Kendall, dan Spearman antar variabel nasabah. Secara umum, korelasi antar variabel sangat lemah (mendekati nol), artinya tiap variabel saling independen. Hanya AccountBalance dan CustomerAge yang menunjukkan korelasi positif sedang, artinya nasabah yang lebih tua cenderung memiliki saldo lebih tinggi..



7. Perbandingan jumlah transaksi normal vc fraud

Grafik di atas memperlihatkan perbandingan jumlah transaksi normal dan fraud yang terlihat hampir seimbang, masing-masing berada di kisaran 2500 transaksi. Keseimbangan ini menunjukkan bahwa distribusi data antara kedua kategori cukup proporsional, sehingga analisis maupun pelatihan model deteksi kecurangan (fraud detection) dapat dilakukan dengan lebih objektif. Dengan proporsi data yang seimbang, model tidak akan condong memprediksi hanya pada salah satu kelas (misalnya hanya normal saja), dan hasil evaluasi akurasi nantinya bisa lebih mencerminkan performa sebenarnya dalam membedakan transaksi normal dan transaksi fraud.



KESIMPULAN

Berdasarkan hasil visualisasi dan analisis korelasi yang telah dilakukan, dapat disimpulkan bahwa hubungan antar variabel numerik seperti TransactionAmount, TransactionDuration, LoginAttempts, AccountBalance, dan CustomerAge umumnya memiliki korelasi yang lemah. Hanya terdapat sedikit hubungan positif antara AccountBalance dan CustomerAge, menunjukkan bahwa nasabah yang lebih tua cenderung memiliki saldo lebih tinggi. Analisis tren waktu juga menunjukkan bahwa jumlah transaksi normal dan fraud relatif seimbang setiap bulannya, tanpa fluktuasi ekstrem. Selain itu, grafik perbandingan jumlah transaksi memperlihatkan bahwa proporsi data normal dan fraud hampir sama, sehingga dataset ini cukup ideal untuk analisis lebih lanjut atau pelatihan model deteksi kecurangan karena tidak ada ketimpangan kelas yang signifikan.

UCAPAN TERIMA KASIH

Sebagai penulis, saya mengucapkan terima kasih kepada semua pihak yang telah membantu dalam proses penelitian dan penyusunan analisis ini. Terima kasih kepada dosen pembimbing atas arahan dan bimbingannya, serta kepada rekan-rekan yang telah memberikan dukungan dan masukan yang berarti. Berkat bantuan dan kerja sama tersebut, penelitian mengenai analisis transaksi nasabah dan deteksi potensi kecurangan ini dapat terselesaikan dengan baik.

REFERENSI

- Anderson, T. W., & Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18(1), 47-82.
- Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons.
- Carcillo, F., Dal Pozzolo, A., Le Borgne, Y. A., Caelen, O., Mazzer, Y., & Bontempi, G. (2019). Scarff: A scalable framework for streaming credit card fraud detection with Spark. *Information Fusion*, 41, 182-194.
- Chen, Y., Zhang, X., & Liu, Y. (2020). Machine learning approaches for house price prediction: A comparative study. *Journal of Real Estate Research*, 42(3), 345-367.
- Cleveland, W. S. (1993). *Visualizing data*. Hobart Press.
- Dal Pozzolo, A., Caelen, O., Le Borgne, Y. A., Waterschoot, S., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. *Expert Systems with Applications*, 41(10), 4915-4928.
- Fadillah, A., & Sari, R. P. (2023). Exploratory data analysis for real estate valuation using Python: A case study on Indonesian housing data. *International Journal of Data Science and Analytics*, 15(2), 112-128.
- Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Elsevier.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.
- Khairina, N., & Harahap, M. K. (2018). Menjaga Kerahasiaan Data dengan Steganografi Kombinasi LSB-2 dengan LSB-3. *Sinkron - Jurnal & Penelitian Teknik Informatika*, 3(1), 286-288.
- Khairina, N., Harahap, M. K., & Lubis, J. H. (2018). The Authenticity of Image using Hash MD5 and Steganography Least Significant Bit. *International Journal Of Information System & Technology*, 2(1), 1-6.
- Kumar, S., Singh, R., & Gupta, A. (2022). Location-based factors influencing real estate prices: An empirical analysis. *Urban Studies*, 59(4), 678-695.
- Li, S., & Wang, H. (2017). Big data analytics in real estate: Opportunities and challenges. *Journal of Property Research*, 34(2), 123-145.
- Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.