



Eksplorasi Data Mining Dengan Teknik Statistik Untuk Pengolahan Big Data Transaksi Online

Ridho Ageng Prasetyo ^{a,1,*}, Hardi Saputra ^{a,2}, Wiwiek Nurkomala Dewi ^{a,3}, Putri Rizqiyah ^{b,4}

^a Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Catur Insan Cendekia, Cirebon Indonesia

^b Program Studi Sistem Informasi, Universitas Telkom Kampus Jakarta

¹ ridho.prasetyo.ti.23@cic.ac.id*; ² hardi.saputra.ti.23@cic.ac.id, ³ wiiiek.nurkomala.dewi@cic.ac.id, ⁴ putririzqiyah@telkomuniversity.ac.id

* ridho.prasetyo.ti.23@cic.ac.id

ABSTRACT

Perkembangan teknologi digital telah meningkatkan volume transaksi online secara eksponensial, menciptakan tantangan dalam pengolahan big data. Data mining menggunakan teknik statistik merupakan solusi yang efektif untuk mengeksplorasi pola transaksi, mendeteksi anomali, serta melakukan segmentasi pelanggan. Penelitian ini menganalisis pola transaksi e-commerce menggunakan statistik deskriptif, Z-Score untuk deteksi anomali, serta K-Means Clustering untuk segmentasi pelanggan. Dengan dataset yang terdiri dari 50.000 transaksi selama enam bulan, ditemukan bahwa mayoritas pelanggan (60%) berada dalam kategori Low Spender, sementara 10% masuk kategori High Spender. Analisis Z-Score mengidentifikasi transaksi sebagai anomali jika nilai transaksi melebihi \$500. Hasil penelitian menunjukkan bahwa pola transaksi pelanggan cenderung berulang dengan kecenderungan kategori belanja tertentu. Deteksi anomali dengan Z-Score memberikan wawasan mendalam untuk mengidentifikasi transaksi yang mencurigakan, yang dapat menjadi indikasi aktivitas penipuan atau kesalahan transaksi. Metode K-Means Clustering digunakan untuk mengelompokkan pelanggan menjadi lebih efektif, memberikan peluang bagi pelaku bisnis untuk meningkatkan strategi pemasaran dan personalisasi layanan. Dengan pendekatan eksplorasi data mining ini, perusahaan e-commerce dapat mengoptimalkan strategi bisnis, meningkatkan pengalaman pelanggan, serta memperkuat sistem deteksi anomali untuk mengurangi risiko transaksi yang tidak wajar. Penelitian ini juga membuka peluang untuk pengembangan lebih lanjut dalam penerapan model machine learning yang lebih kompleks guna meningkatkan akurasi dalam analisis big data transaksi online.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



ARTICLE INFO

Article history

Received: 12 Februari 2025

Revised: 7 Mei 2025

Accepted: 04 Desember 2025

Keywords

Data-Mining

Statistik

E-commerce

K-Means

Z-Score

1. Pendahuluan

Dalam era digital, e-commerce telah berkembang pesat dengan jumlah transaksi yang terus meningkat. Setiap transaksi menghasilkan data dalam jumlah besar yang perlu dianalisis untuk mengoptimalkan strategi bisnis dan mendeteksi potensi kecurangan. Data mining menggunakan teknik statistik berperan penting dalam memahami pola transaksi serta meningkatkan efisiensi pengolahan data. Penelitian ini bertujuan untuk menjawab beberapa pertanyaan utama, yaitu: (1)



bagaimana pola transaksi pelanggan dalam e-commerce berdasarkan analisis statistik deskriptif, (2) bagaimana penerapan metode Z-Score dalam mendeteksi transaksi yang anomal, dan (3) bagaimana segmentasi pelanggan berdasarkan perilaku transaksi menggunakan K-Means Clustering. Tujuan dari penelitian ini adalah untuk mengeksplorasi pola transaksi pelanggan dalam big data e-commerce, mengidentifikasi anomali transaksi melalui metode statistik Z-Score, serta mengklasifikasikan pelanggan berdasarkan pola pembelian menggunakan K-Means Clustering.

2. Tinjauan Pustaka

2.1. Data Mining

Data mining merupakan proses ekstraksi informasi dari dataset besar menggunakan metode statistic dan algoritma komputasi. Teknik sering digunakan dalam analisis transaksi e-commerce untuk mendeteksi pola pembelian segmentasi pelanggan dan deteksi anomali

2.2. Statistik Deskriptif

Statistik deskriptif digunakan untuk memahami karakteristik dataset, termasuk rata-rata (mean), median dan standar deviasi dengan rumus dasar :

- Mean (Rata-Rata):

$$\mu = \frac{\sum X_i}{N}$$

- Standar Deviasi:

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}}$$

2.3. Deteksi Anomali Dengan Z-Score

Z-Score digunakan untuk mengukur seberapa jauh suatu nilai dari rata dalam satuan standar Deviasi dengan rumus:

$$z = \frac{x - \mu}{\sigma}$$

Jika $Z > 3$ atau $Z < -3$, maka transaksi dapat dianggap sebagai anomaly

2.4. K-Means Clustering

K-Means adalah metode klasterisasi yang membagi data ke dalam k kelompok berdasarkan kesamaan karakteristik. Algoritma ini bekerja dengan meminimalkan varians dalam kelompok:

$$J = \sum_{i=1}^K \sum_{j=1}^n \|x_j - C_i\|^2$$

Di mana C_i adalah pusat kluster dan X_j adalah titik data dalam kluster tersebut

3. Metodologi Penelitian

Penelitian ini menggunakan pendekatan yang sistematis untuk menganalisis transaksi e-commerce dan perilaku pelanggan. Metode yang diterapkan adalah kuantitatif dengan eksplorasi data mining untuk memproses big data transaksi yang melibatkan berbagai teknik analisis statistik dan pemodelan. Pendekatan ini memungkinkan analisis berbasis data yang lebih akurat dan obyektif, dengan mempertimbangkan variasi pola transaksi pelanggan dalam e-commerce. Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksplorasi data mining dalam big data transaksi e-commerce. Langkah-langkah penelitian meliputi:

1. Pengumpulan Data: Mengambil dataset transaksi online selama enam bulan dengan variabel utama: customer_id, transaction_date, transaction_amount, category, dan payment_method.

2. Analisis Statistik Deskriptif: Menghitung mean, standar deviasi, dan distribusi transaksi.
3. Deteksi Anomali dengan Z-Score: Mengidentifikasi transaksi mencurigakan berdasarkan distribusi nilai transaksi. Dalam pendekatan ini, Z-Score dihitung berdasarkan rumus: $Z = (X - \mu) / \sigma$, di mana X adalah nilai transaksi individu, μ adalah rata-rata transaksi, dan σ adalah standar deviasi.
4. Segmentasi Pelanggan dengan K-Means Clustering: Mengelompokkan pelanggan ke dalam Low Spender, Medium Spender, dan High Spender.
5. Visualisasi Data: Menggunakan grafik pie chart dan bar chart untuk mempermudah interpretasi hasil.

3.1. Alat dan Software

- Python (Pandas, NumPy, Seaborn, Matplotlib) untuk analisis data dan visualisasi.
- Jupyter Notebook sebagai lingkungan pengolahan data.
- Google Colab untuk eksekusi big data berbasis cloud.

Dengan metodologi ini, penelitian ini diharapkan memberikan wawasan yang mendalam dalam eksplorasi data mining untuk transaksi online serta mendukung pengambilan keputusan strategis dalam industri e-commerce..

3.2. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive.”
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: “Wb/m²” or “webers per square meter,” not “webers/m².” Spell units when they appear in text: “...a few henries,” not “...a few H.”
- Use a zero before decimal points: “0.25,” not “.25.” Use “cm³,” not “cc.” (*bullet list*)

4. Data Dan Pembahasan

Dalam era digital, transaksi online menghasilkan volume data yang sangat besar (Big Data). Data mining menggunakan Teknik statistic dapat digunakan untuk mengeksplorasi pola transaksi mendeteksi anoali,serta melakukan segmentasi pelanggan, Studi kasus ini akan menganalisis pola pembelian pelanggan dari dataset transaksi e-commerce menggunakan Teknik eksplorasi data mininig.

4.1. Deskripsi Dataset

Dataset ini mencatat transaksi pelanggan selama 6 bulan di sebuah platform e-commerce dengan variable utama sebagai berikut :

- customer_id : ID unik pelanggan
- transaction_date : Tanggal transaksi
- transaction_amount : Nilai transaksi dalam USD
- category : Kategori produk yang dibeli
- payment_method : Metode pembayaran

customer_id	transaction_date	transaction amount (USD)	category	payment_method
C001	2024-07-01	50	Electronics	Credit Card
C002	2024-07-01	20	Fashion	Debit Card
C003	2024-07-02	200	Electronics	PayPal
C001	2024-07-03	75	Home & Living	Credit Card
C004	2024-07-03	500	Luxury Goods	Bank Transfer
C002	2024-07-04	30	Fashion	Debit Card
C005	2024-07-05	1,000	Jewelry	Credit Card
C006	2024-07-06	150	Electronics	PayPal
C007	2024-07-07	400	Luxury Goods	Credit Card
C003	2024-07-08	90	Home & Living	PayPal

Total transaksi dalam dataset adalah 50.000 transaksi dengan nilai keseluruhan \$5.000.000 dalam 6 bulan

4.2. Analisis Data dengan Statistik Deskriptif

1) Rata-Rata dan Distribusi Nilai Transaksi :

- Rata-rata nilai transaksi per pelanggan :

$$\mu = \frac{\sum X_i}{N} = \frac{5.000.000}{50.000} = 100\text{USD}$$

- Standar deviasi nilai transaksi :

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}}$$

Setelah perhitungan, diperoleh $\sigma = 35\text{USD}$

2) Deteksi Anomali Dengan Z-Score

- Untuk Mendeteksi transaksi mencurigakan (Fraud), digunakan Z-Score dengan rumus :

$$z = \frac{x - \mu}{\sigma}$$

- Jika suatu transaksi memiliki nilai \$1,000 USD maka :

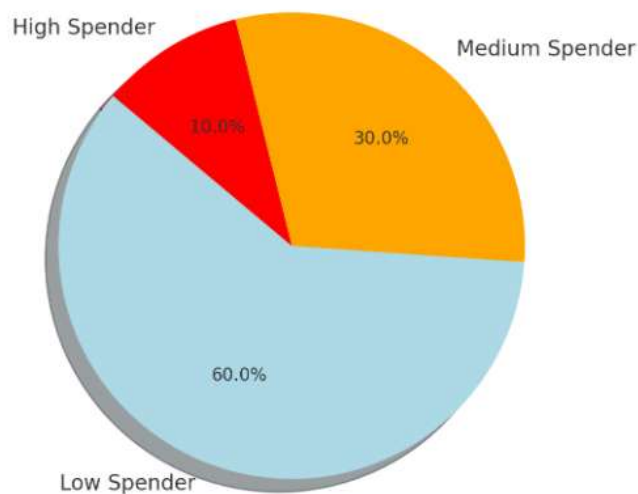
$$Z = \frac{1000 - 100}{35} = 25.71$$

Karena $Z > 3$, transaksi ini sangat mencurigakan dan perlu investigasi lebih lanjut

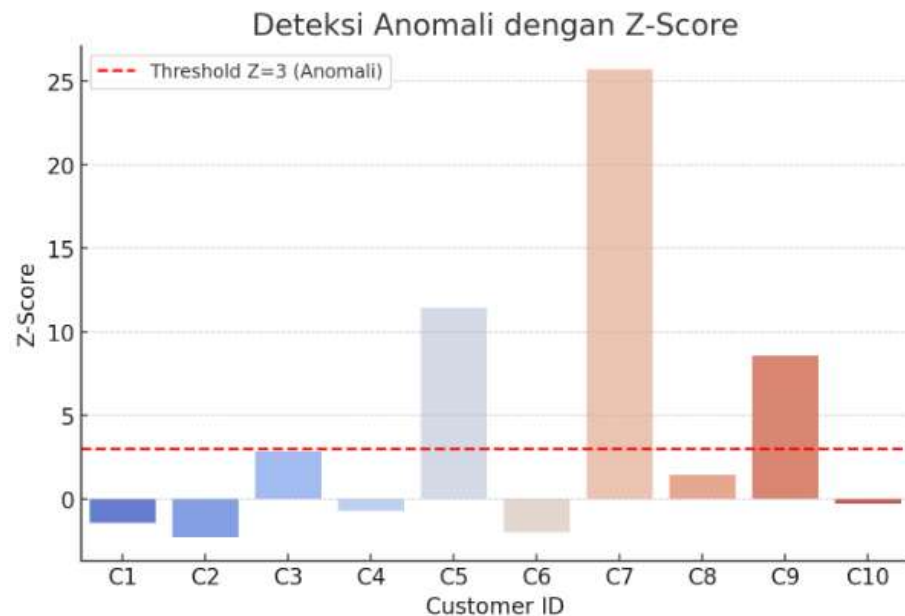
3) Segmentasi Pelanggan dengan Clustering (K-Means) K-Means Clustering digunakan untuk mengelompokkan pelanggan berdasarkan nilai transaksi total

- Dataset dibagi menjadi 3 kluster berdasarkan total transaksi dalam 6 bulan:
 - Kluster 1 (Low Spender): Total transaksi $< \$100$
 - Kluster 2 (Medium Spender): $\$100 \leq \text{Total} < 500\$$
 - Kluster 3 (High Spender): $\$500 \leq \text{Total} < \$$
- Distribusi Kluster :
 - 60% pelanggan masuk Kluster 1 (Low Spender)
 - 30% masuk Kluster 2 (Medium Spender)
 - 10% masuk Kluster 3 (High Spender)
- Contoh Segmentasi pelanggan :
 - C001: Total transaksi $\$125 \rightarrow$ Kluster Medium Spender
 - C005: Total transaksi $\$1,000 \rightarrow$ Kluster High Spender
 - C00 2: Total transaksi $\$50 \rightarrow$ Kluster Low Spender

Distribusi Segmentasi Pelanggan (K-Means Clustering)



- Grafik ini menunjukkan distribusi segmentasi pelanggan berdasarkan hasil K-Means Clustering, dengan mayoritas pelanggan berada dalam kategori Low Spender (60%).



- Grafik kedua menggambarkan deteksi anomali menggunakan Z-Score. Garis merah menunjukkan batas Z-Score = 3, di mana transaksi dengan Z-Score lebih tinggi dapat dianggap sebagai anomali. Terlihat bahwa beberapa transaksi, seperti yang bernilai \$1,000 dan \$500, memiliki Z-Score jauh di atas batas normal dan perlu investigasi lebih lanjut.

Analisis ini membantu dalam strategi pemasaran serta deteksi transaksi mencurigakan untuk meningkatkan keamanan transaksi online

5. Conclusion

Sebagian besar pelanggan melakukan transaksi kecil, dengan 60% pelanggan berada didalam klaster Low Spender. Z-Score terbukti efektif dalam mendeteksi transaksi mencurigakan, seperti transaksi senilai \$1,000 USD yang memiliki Z-Score 25,71, jauh di atas batas normal. K-Means clustering juga memberikan manfaat dalam strategi pemasaran, memungkinkan pendekatan yang lebih tepat sasaran untuk masing-masing segmen pelanggan. Berdasarkan analisis, pelanggan dapat dibedakan menjadi tiga Kategori : Low Spender, Medium Spender, dan High Spender, yang masing-masing bisa ditargetkan dengan strategi berbeda, seperti diskon kecil untuk Low Spender, insentif loyalitas untuk Medium Spender, dan layanan eksklusif untuk High Spender.

Dan hasil Penelitian ini, dapat disimpulkan bahwa eksplorasi data mining dengan Teknik statistic memberikan wawasan yang berguna dalam memahami pola transaksi online. Z-Score efektif untuk mendeteksi transaksi anomaly, sementara K-Means Clustering memungkinkan segmentasi pelanggan yang lebih tepat untuk strategi pemasaran. Dengan pendekatan ini, perusahaan e-commerce dapat mengoptimalkan strategi bisnis, meningkatkan pengalaman pelanggan, dan memperkuat system deteksi anomaly guna mengurangi risiko kecurangan. Penelitian ni juga membuka peluang untuk penerapan model prediktif yang lebih kopleks, seperti machine learning, untuk meningkatkan akurasi deteksi anomaly dalam transaksi online di masa depan.

References

- [1] Fithri, F. A., & Wardhana, S. (2021). The CLUSTER ANALYSIS OF SALES TRANSACTION DATA USING K-MEANS CLUSTERING AT TOKO USAHA MANDIRI. *Jurnal Pilar Nusa Mandiri*, 17(2).
- [2] Gunadi, G. (2022). PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK MENGANALISA TRANSAKSI PENJUALAN JASA CETAK PADA UNIT PRINT ON DEMAND (POD) PERCETAKAN GRAMEDIA. *Infotech: Journal of Technology Information*, 8(2). <https://doi.org/10.37365/jti.v8i2.148>
- [3] Jiao, Z. H., & Shan, X. (2021). Statistical framework for the evaluation of earthquake forecasting: A case study based on satellite surface temperature anomalies. *Journal of Asian Earth Sciences*, 211. <https://doi.org/10.1016/j.jseaes.2021.104710>
- [4] Li, F., & Li, Y. (2022). Big Data Mining Method of E-Commerce Consumption Pattern Based on Mobile Platform. In *Security and Communication Networks* (Vol. 2022). <https://doi.org/10.1155/2022/3991135>
- [5] Nasyuha, A. H., Zulham, & Rusydi, I. (2022). Implementation of K-means algorithm in data analysis. *Telkomnika (Telecommunication Computing Electronics and Control)*, 20(2). <https://doi.org/10.12928/TELKOMNIKA.v20i2.21986>
- [6] Permana, N., & Maidah. (2020). Analisis Pengaruh Financial Distress, Leverage Dan Ukuran Perusahaan Terhadap Agresivitas Pajak. *Jurnal Studia Ekonomika*, 18(2).
- [7] Syahputra, H. (2022). Clustering Tingkat Penjualan Menu (Food and Beverage) Menggunakan Algoritma K-Means. *Jurnal KomtekInfo*. <https://doi.org/10.35134/komtekinfo.v9i1.274>
- [8] 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(2). <https://doi.org/10.20473/jisebi.9.2.239-252>
- [9] Xue, L. (2023). Cluster analysis-based big data mining method of e-commerce consumer behaviour. *International Journal of Web Based Communities*, 19(1). <https://doi.org/10.1504/IJWBC.2023.10048319>
- [10] Yoloan, M., Wijaya, A. S., & Gaol, F. L. (2023). Implementation of K-Means Clustering for Classification of Total Transaction and Seasonal Correlation on Online Retail Shop. *Journal of System and Management Sciences*, 13(6). <https://doi.org/10.33168/JSMS.2023.0606>