Asimetri Harga dan Spesifikasi pada Laptop *Gaming* 2026: Segmentasi *Tier* Performa Menggunakan K-Means Clustering

Rizky Amelia[1,*], Dicha Wijaya Kusuma[1], Dimas Ramadhani[1], Muhammad Iqbal Alexandre Saputra[1], Raditya Yusma Nata[1], Shabrina Alya Aureliya[1], Zakaria Fattawari[1]
[1]Informatika, Institut Teknologi Kalimantan;
11231020@student.itk.ac.id
11231021@student.itk.ac.id
11231056@student.itk.ac.id
11231080@student.itk.ac.id
11231090@student.itk.ac.id
11231092@student.itk.ac.id
[*]Korespondensi: rizky.amelia@lecturer.itk.ac.id;

Abstract – The 2026 gaming laptop market exhibits information asymmetry between price and technical specifications, making it difficult for consumers to assess a laptop's performance tier, particularly in the absence of official tier labels. (1) Background: important specifications are often embedded in unstructured product titles and descriptions, while price alone may not fully represent gaming performance. (2) Method: this study applies K-Means clustering to segment gaming laptops based on price and technical specifications. The Gaming Laptops 2026 dataset from Kaggle contained 614 records and was cleaned into 594 laptop products. GPU models were extracted from product text and mapped to numerical benchmark scores; because missing GPU information was not assumed to represent integrated graphics, the main clustering analysis used 377 observations with complete GPU benchmark information. Nine numerical features were transformed and MinMax-scaled. Cluster-number evaluation showed that k = 2 achieved better geometric quality, whereas k = 3 was selected as the final configuration to provide a more interpretable Entry, Mid-Range, and High-End segmentation. (3) Results: the final K-Means model with k = 3 achieved a Silhouette score of 0.224, a Davies–Bouldin Index of 1.530, and a Calinski–Harabasz Index of 113.6. The model showed relatively high stability across initialization seeds (ARI = 0.9549 ± 0.0352) and data sub-sampling (ARI = 0.8506 ± 0.1622). GPU benchmark produced the highest ANOVA F-score (575.8), followed by price and refresh rate. Ablation analysis further showed that GPU alone could not reproduce the full clustering structure. (4) Conclusion: GPU performance is an important discriminator of laptop tiers, but the resulting segmentation is determined by a combination of GPU benchmark, price, refresh rate, memory, storage, and other technical characteristics. The three-tier solution therefore represents a practical trade-off between geometric clustering quality and interpretability rather than a mathematically optimal partition.

Keywords – K-Means clustering, product segmentation, gaming laptops, performance tier, Silhouette Coefficient

Intisari – Pasar laptop *gaming* 2026 menghadirkan asimetri informasi antara harga dan spesifikasi teknis sehingga konsumen dapat mengalami kesulitan dalam menilai *tier* performa suatu laptop, terutama karena tidak tersedia label *tier* resmi. (1) Latar Belakang: spesifikasi penting sering tersimpan di dalam judul dan deskripsi produk yang tidak terstruktur, sedangkan harga saja belum tentu mencerminkan kemampuan *gaming* secara menyeluruh. (2) Metode: penelitian ini menerapkan algoritma K-Means untuk melakukan segmentasi laptop berdasarkan kombinasi harga dan spesifikasi teknis. Dataset Gaming Laptops 2026 dari Kaggle terdiri atas 614 observasi dan dibersihkan menjadi 594 produk laptop. Model GPU diekstraksi dari teks produk kemudian dipetakan menjadi skor *benchmark* numerik. Karena informasi GPU yang kosong tidak diasumsikan sebagai GPU terintegrasi, analisis *clustering* utama dilakukan pada 377 observasi yang memiliki informasi *benchmark* GPU lengkap. Sembilan fitur numerik kemudian ditransformasi dan diskalakan menggunakan MinMaxScaler. Evaluasi jumlah klaster menunjukkan bahwa k = 2 memiliki kualitas geometris yang lebih baik, sedangkan k = 3 dipilih sebagai konfigurasi akhir karena menghasilkan segmentasi *Entry, Mid-Range*, dan *High-End* yang lebih mudah diinterpretasikan. (3) Hasil: model K-Means final dengan k = 3 memperoleh *Silhouette* sebesar 0,224, *Davies–Bouldin Index* sebesar 1,530, dan *Calinski–Harabasz Index* sebesar 113,6. Model menunjukkan kestabilan yang relatif tinggi terhadap perubahan inisialisasi (ARI = 0,9549 ± 0,0352) dan sub-sampel data (ARI = 0,8506 ± 0,1622). GPU benchmark memiliki skor F ANOVA tertinggi sebesar 575,8, diikuti harga dan *refresh rate*. Hasil *ablation study* juga menunjukkan bahwa GPU saja belum dapat mereproduksi struktur *clustering* dari model lengkap. (4) Simpulan: performa GPU merupakan salah satu pembeda utama *antar-tier*, tetapi segmentasi yang terbentuk merupakan hasil kombinasi GPU *benchmark*, harga, *refresh rate*, memori, penyimpanan, dan karakteristik teknis lainnya. Dengan demikian, penggunaan tiga *tier* merupakan *trade-off* praktis antara kualitas geometris *clustering* dan interpretabilitas, bukan pembagian yang diklaim optimal secara matematis.

Kata kunci – K-Means *clustering*, segmentasi produk, laptop *gaming*, tingkatan performa, *Silhouette Coefficient*

I. PENDAHULUAN

Pasar laptop *gaming* pada tahun 2026 berkembang sangat cepat sehingga menjadi sulit dipahami baik oleh pembeli maupun penjual. Pada satu daftar produk dapat terdapat ratusan varian dengan rentang harga yang lebar, dari unit *entry-level* hingga *workstation*, namun tidak tersedia label resmi yang menyatakan sebuah laptop tergolong kelas pemula, menengah, atau kelas atas. Kehadiran komponen terspesialisasi seperti NPU untuk AI *on-device* dan GPU generasi baru yang lebih efisien membuat pembeli tidak

hanya dihadapkan pada pilihan murah atau mahal, tetapi juga pada spesifikasi yang kerap tidak sebanding dengan harganya. Informasi penting seperti kapasitas RAM, jenis GPU, penyimpanan, dan *refresh rate* layar justru tersembunyi di dalam judul produk yang tidak terstruktur sehingga sulit dibandingkan secara objektif.

Kondisi tersebut menimbulkan ketidakseimbangan informasi antara harga pasar dan spesifikasi teknis. Pembeli sering kesulitan menilai apakah sebuah laptop tergolong *overpriced* atau *best-value* berdasarkan tren perangkat keras terkini, persoalan yang sejalan dengan temuan mengenai kesenjangan informasi harga–atribut produk pada pasar daring [1]. Di sisi lain, siklus pengembangan perangkat keras yang sangat cepat menyebabkan penurunan harga yang kurang menentu sehingga diperlukan sistem yang mampu mengelompokkan produk secara otomatis untuk mempermudah pemasaran maupun pengambilan keputusan pembelian. Karena data yang tersedia sama sekali tidak memiliki label *tier*, pendekatan yang sesuai bukanlah klasifikasi, melainkan *clustering*.

Sejumlah penelitian terdahulu telah menerapkan metode *clustering* untuk segmentasi pasar dan produk karena kemampuannya mengidentifikasi pola pada data tanpa memerlukan label kelas [2]. Pada konteks perangkat komputasi, K-Means dan metode *clustering* lainnya telah digunakan untuk mengelompokkan laptop berdasarkan harga dan spesifikasi [3], [4], serta untuk menganalisis pola pada data produksi dan harga perangkat keras [5], [6]. Namun, penelitian-penelitian tersebut pada umumnya berfokus pada penggunaan atribut numerik yang telah tersedia secara terstruktur. Pada data produk daring, sebagian informasi teknis justru tersimpan dalam teks judul yang tidak terstruktur sehingga memerlukan proses ekstraksi sebelum dapat digunakan dalam pemodelan. Selain itu, representasi GPU menggunakan nomor model secara langsung belum tentu mencerminkan perbedaan performa aktual antar-generasi dan varian GPU. Kondisi tersebut menunjukkan adanya ruang penelitian untuk membentuk segmentasi laptop dengan menggabungkan ekstraksi spesifikasi dari teks produk dan representasi performa GPU yang lebih informatif.

Berdasarkan uraian tersebut, penelitian ini menerapkan algoritma K-Means untuk membentuk segmentasi performa laptop *gaming* berdasarkan kombinasi harga dan spesifikasi teknis. Penelitian ini bertujuan untuk (1) membangun model *clustering* K-Means untuk mengelompokkan laptop ke dalam *tier* performa yang dapat diinterpretasikan, (2) mengevaluasi pemilihan jumlah klaster dengan mempertimbangkan kualitas geometris dan interpretabilitas praktis, (3) mengidentifikasi fitur yang paling berperan dalam membedakan karakteristik antar-klaster, serta (4) mengevaluasi kontribusi fitur GPU melalui *ablation study* dan pemeriksaan kestabilan hasil *clustering*. Penelitian ini tidak menawarkan metode *clustering* baru, melainkan memberikan kontribusi berupa penerapan segmentasi produk yang mengintegrasikan ekstraksi spesifikasi dari judul produk tidak terstruktur, pemetaan model GPU ke skor *benchmark* performa, serta evaluasi kontribusi fitur dan kestabilan klaster. Dengan pendekatan tersebut, penelitian ini menganalisis sejauh mana kombinasi harga dan spesifikasi teknis dapat menghasilkan segmentasi *Entry*, Mid-Range, dan High-End yang bermakna pada dataset yang digunakan.

II. DASAR TEORI

Clustering merupakan teknik *unsupervised learning* yang bertujuan mengelompokkan objek ke dalam beberapa kelompok (klaster) sedemikian rupa sehingga objek dalam satu klaster memiliki kemiripan tinggi, sementara objek antar-klaster berbeda. Berbeda dengan klasifikasi, *clustering* tidak memerlukan label sehingga sesuai untuk menemukan struktur pengelompokan yang belum diketahui sebelumnya. K-Means adalah salah satu algoritma *clustering* berbasis partisi yang paling banyak digunakan karena efisien pada data numerik berdimensi besar [2].

Algoritma K-Means menjadi model yang cukup populer sebagai pendekatan *clustering* berbasis partisi karena dikenal lebih efisien dan mudah dalam proses implementasinya [7]. K-Means membagi data menjadi k klaster dengan cara meminimalkan *within-cluster sum of squares* (WCSS), yaitu total kuadrat jarak *Euclidean* antara tiap objek dengan pusat (*centroid*) klasternya. Prosesnya dimulai dengan inisialisasi *centroid*, penetapan tiap objek ke *centroid* terdekat, lalu pembaruan posisi *centroid* sebagai rata-rata anggota klaster, dan diulang hingga konvergen. Inisialisasi *k-means++* digunakan untuk memilih *centroid* awal yang tersebar sehingga mempercepat konvergensi dan mengurangi risiko terjebak pada optimum lokal [3]. Karena berbasis jarak *Euclidean*, seluruh fitur perlu diskalakan agar fitur berskala besar tidak mendominasi perhitungan jarak.

Pemilihan jumlah klaster merupakan salah satu tahapan penting dalam penerapan K-Means. Metode Elbow mengevaluasi perubahan nilai WCSS terhadap peningkatan jumlah klaster dan mengidentifikasi titik ketika penurunan WCSS mulai berkurang secara siginfikan [8]. Sementara itu, *Silhouette Coefficient* mengukur tingkat kesesuaian suatu objek terhadap klasternya dibandingkan dengan klaster terdekat lainnya, dengan rentang nilai -1 hingga 1; nilai yang lebih tinggi menunjukkan pemisahan klaster yang lebih baik [9]. Salah satu tantangan utama penerapan K-Means adalah penentuan jumlah klaster yang optimal, maka dari itu Metode *Elbow* dan *Silhouette Coefficient* umum digunakan sebagai dasar pemilihan jumlah klaster yang paling representatif [10]. Dalam praktiknya, pemilihan jumlah klaster tidak hanya dapat mempertimbangkan nilai metrik internal, tetapi juga interpretabilitas hasil sesuai konteks permasalahan.

Pengukuran validitas klaster menggunakan pendekatan *Silhouette Coefficient* dan *Davies-Bouldin Index* menjadi syarat yang cukup krusial untuk memastikan bahwa segmentasi *tier* performa memiliki tingkat kohesi internal yang tinggi serta pemisahan antarklaster yang optimal [11]. Pada studinya, pendekatan berbasis algoritma ekstraksi fitur mampu mengurai atribut spesifikasi teknis guna mendeteksi

penyimpangan atau asimetri antara harga pasar dan nilai intrinsik suatu produk komputasi [12].

Untuk mengevaluasi kualitas klaster secara internal, penelitian ini juga menggunakan *Davies–Bouldin Index* dan *Calinski–Harabasz Index* [13]. *Davies–Bouldin* membandingkan dispersi dalam klaster dengan jarak antarklaster, sehingga nilai yang lebih rendah menunjukkan pemisahan yang lebih baik [14]. Sebaliknya, *Calinski–Harabasz* membandingkan variasi antarklaster dengan variasi dalam klaster, sehingga nilai yang lebih tinggi menunjukkan struktur klaster yang lebih baik [15]. Kestabilan hasil *clustering* dapat dievaluasi menggunakan *Adjusted Rand Index* (ARI), yaitu ukuran kesamaan antara dua hasil penugasan klaster yang telah disesuaikan terhadap kemungkinan kesamaan yang terjadi secara acak [16].

GPU merupakan salah satu komponen penting yang memengaruhi kemampuan grafis sebuah laptop, khususnya pada aplikasi dengan kebutuhan komputasi visual tinggi. Namun, nomor model GPU tidak selalu memiliki hubungan linear dengan performa aktual karena dipengaruhi oleh perbedaan generasi, arsitektur, kelas produk, dan variasi implementasi perangkat. Oleh karena itu, penelitian ini memetakan model GPU ke skor *benchmark* G3D sebagai representasi numerik kemampuan grafis relatif. Skor *benchmark* tersebut tetap memiliki keterbatasan karena performa laptop juga dipengaruhi oleh TGP, sistem pendingin, konfigurasi memori, dan batas daya; oleh karena itu, skor digunakan sebagai proksi performa relatif, bukan sebagai prediksi *frame rate* spesifik pada tiap laptop.

Dalam konteks pemasaran, *clustering* telah dimanfaatkan untuk segmentasi, penargetan, dan pemosisian produk, termasuk pada pasar laptop di Indonesia [4]. Pendekatan K-Means maupun K-Medoids juga telah digunakan untuk mengelompokkan laptop berdasarkan harga dan spesifikasi [3], sedangkan penerapan serupa ditemukan pada pengelompokan data produksi dan harga perangkat keras [5], [6]. Penelitian ini menggunakan dasar tersebut untuk membentuk *tier* performa laptop, tetapi dengan menambahkan proses ekstraksi spesifikasi dari teks produk yang tidak terstruktur serta representasi performa GPU berbasis skor *benchmark*. Dengan demikian, hasil *clustering* tidak hanya didasarkan pada harga, tetapi pada kombinasi beberapa karakteristik teknis yang telah diproses menjadi fitur numerik.

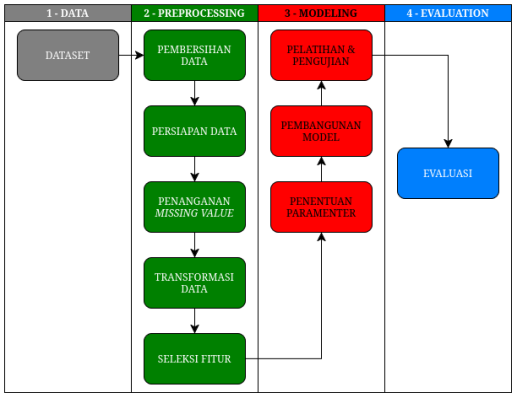
III. METODOLOGI

Penelitian ini menggunakan dataset *Gaming Laptops 2026* yang bersumber dari Kaggle dan terdiri atas 614 baris dengan 10 atribut, yaitu judul produk (*title*), merek (*brand*), harga (*price*), harga awal (*list_price*), persentase diskon (*discount_pct*), mata uang, *rating* (*stars*), jumlah ulasan (*reviews_count*), kategori produk (*breadCrumbs*), dan deskripsi (*description*). Sebagian besar spesifikasi teknis laptop tidak tersedia sebagai atribut terstruktur, tetapi tertanam di dalam teks judul dan deskripsi produk.

Tahap pembersihan data dimulai dengan menghapus kolom *price_currency* karena hanya memiliki satu nilai unik, melakukan normalisasi penulisan pada atribut *brand*, serta menyaring produk non-laptop berdasarkan informasi pada *breadCrumbs* dan kata kunci pada judul. Sebanyak 20 observasi teridentifikasi sebagai produk non-laptop dan dikeluarkan sehingga tersisa 594 observasi. Kolom dengan tingkat kekosongan tinggi seperti *list_price* tidak digunakan sebagai fitur pemodelan. Sementara itu, kolom *title* dan *description* dipertahankan sementara untuk proses ekstraksi spesifikasi sebelum dikeluarkan pada tahap seleksi fitur.

Karena spesifikasi teknis tersedia pada teks bebas, kapasitas RAM (ram_*gb), kapasitas penyimpanan (storage_gb), ukuran layar (screen_in), dan refresh rate (refresh*_hz) diekstraksi dari judul (*title*) menggunakan *regular expression*. Model GPU diekstraksi dari judul dan ketika pola yang relevan tersedia, dilengkapi dari deskripsi. Audit ekstraksi menemukan model GPU spesifik pada 410 dari 594 observasi (69,0%). Model tersebut kemudian dipetakan ke skor PassMark D3D yang digunakan dalam eksperimen; sebanyak 377 observasi memiliki skor *benchmark* yang dapat dipetakan. Sebanyak 217 observasi tanpa skor *benchmark* tidak diberi nilai 0 dan tidak diasumsikan menggunakan GPU terintegrasi, melainkan dikeluarkan dari analisis utama yang melibatkan GPU. Untuk fitur lain, *refresh_hz* yang kosong diisi 60 Hz, sedangkan *ram_gb, storage_gb, screen_in*, dan *rating* menggunakan median.

Fitur *price* dan *reviews_count* yang berdistribusi menceng ditransformasi menggunakan log1p. Sembilan fitur numerik yang digunakan pada model utama adalah *price_log*, *reviews_log, discount_pct, stars, ram_gb, storage_gb*, *screen_in, refresh_hz*, dan *gpu_benchmark*. Seluruh fitur numerik diskalakan menggunakan MinMaxScaler ke rentang [0, 1] karena K-Means menggunakan jarak Euclidean. Setelah baris tanpa *gpu_benchmark* dikeluarkan, matriks model utama berisi 377 observasi dan sembilan fitur. Brand tidak digunakan pada model ginal agar segmentasi berfokus pada karakteristik harga dan spesifikasi teknis; pengaruh algoritma diperiksa secara keseluruhan menggunakan *Algglomerative Clustering*. Keseluruhan tahapan *pipeline* ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

Algoritma utama yang digunakan adalah K-Means (k = 3, inisialisasi *k-means++*, *n_init* = 10, *random_state* = 42) yang dijalankan pada sembilan fitur numerik tanpa menyertakan fitur *brand*. Sebagai pembanding digunakan (1) K-Means dengan tambahan fitur *brand* hasil *one-hot encoding* (18 fitur) untuk menguji pengaruh fitur kategorikal, dan (2) *Agglomerative Clustering* (*linkage = ward*, k = 3) sebagai algoritma independen untuk memverifikasi bahwa struktur *tier* yang ditemukan bukan sekadar artefak satu algoritma.

Karena *clustering* tidak memiliki label kebenaran, evaluasi internal menggunakan *Silhouette Coefficient* sebagai acuan utama, *Davies–Bouldin Index*, dan *Calinski–Harabasz Index*. Kestabilan inisialisasi diukur menggunakan ARI antar-hasil dari 20 *random seed*, sedangkan kestabilan terhadap data diperiksa melalui 30 sub-sampel acak berukuran 80% data. *Ablation study* membandingkan empat konfigurasi, yaitu *full features*, GPU-only, tanpa GPU, dan *baseline rule-based* berbasis kuantil *gpu_benchmark*. Agar hasil *ablation* dapat dibandingkan secara langsung, penugasan klaster dari setiap konfigurasi dievaluasi pada ruang sembilan fitur penuh yang sama dan dibandingkan terhadap label *full features* menggunakan ARI. Skor F-ANOVA (*f_classif*) digunakan sebagai ukuran kepentingan kausal. Validasi praktis dilakukan dengan memeriksa median profil antar-*tier* dan contoh produk yang paling dekat dengan *centroid*; pemeriksaan ini tidak diperlakukan sebagai validasi *ground truth* eksternal.

IV. HASIL DAN PEMBAHASAN

A. Penentuan Jumlah Klaster

Evaluasi jumlah klaster dilakukan pada 377 observasi dengan *benchmark* GPU lengkap. Pada 20 *random seed*, k – 2 memberikan kualitas geometris yang lebih baik dengan *Silhouette* 0,2560 ± 0,0002 dan *Davies-Bouldin* 1,4708 ± 0,0004, dibandingkan k = 3 dengan *Silhouette* 0,2223 ± 0,0013 dan *Davies-Bouldin* 1,5321 ± 0,0022. Stabilitas inisiasi juga lebih tinggi daripada k = 2 (ARI 0,9936 ± 0,0092) daripada k = 3 (ARI 0,9549 ± 0,0352). Meskipun demikian, k = 3 dipilih sebagai konfigurasi karena menghasilkan tiga profil yang lebih informatif untuk konteks segmentasi: median *gpu_benchmark* meningkat dari 5.139 pada *Entry* menjadi 16.724 pada *Mid-Range* dan 22.448 pada *High-End*, sementara median harga meningkat dari USD 1.099,5 menjadi USD 1.449,5 dan USD 2.599. Dengan demikian, pemilihan k = 3 merupakan *trade-off* antara kualitas geometris dan interpretabilitas praktis, bukan klaim bahwa tiga klaster merupakan solusi matematis optimal.

B. Perbandingan Model dan Konfigurasi

Model final (K-Means tanpa *brand*, k = 3) dibandingkan dengan *Agglomerative Clustering* pada ruang fitur yang sama. Hasil pada Tabel 1 menunjukkan bahwa K-Means memperoleh *Silhouette* lebih tinggi (0,224 vs 0,181), *Davies-Bouldin* lebih rendah (1,530 vs 1,617), dan *Calinski-Harabasz* lebih tinggi (113,6 vs 90,6). Nilai ARI antara kedua hasil *clustering* sebesar 0,595 menunjukkan kesesuaian sedang, sehingga pembanding ini diperlakukan sebagai pemeriksaan sensivitas terhadap pilihan algoritma, bukan bukti adanya *ground truth* klaster. Perbandingan model dan konfigurasi ditunjukkan pada Tabel 1.

Tabel 1. Perbandingan model *clustering* pada 337 observasi

Model	*Silhouette*	*Davies-Bouldin*	*Calinski-Harabasz*
K-Means (k = 3)	0,224	1,530	113,6
Agglomerative (k = 3)	0,181	1,617	90,9

C. Ablation Study

Ablation study pada tabel 2 menguji apakah struktur segmentasi terutama merupakan hasil satu fitur GPU. Ketika seluruh *assignment* dievaluasi pada ruang sembilan fitur penuh yang sama, model *full features* memperoleh *Silhouette* 0,224 dan menjadi acuan ARI. GPU-only menghasilkan ARI 0,500 terhadap model penuh, sedangkan model tanpa GPU menghasilkan ARI 0,415. *Baseline rule-based* yang membagikan *gpu_benchmark* berdasarkan dua kuantil memiliki ARI 0,260. Hasil ini menunjukkan bahwa GPU merupakan pembeda kuat, tetapi GPU saja tidak cukup untuk mereproduksi struktur yang terbentuk dari kombinasi harga dan spesifikasi lain. Evaluasi pada ruang fitur penuh ditunjukkan pada Tabel 2.

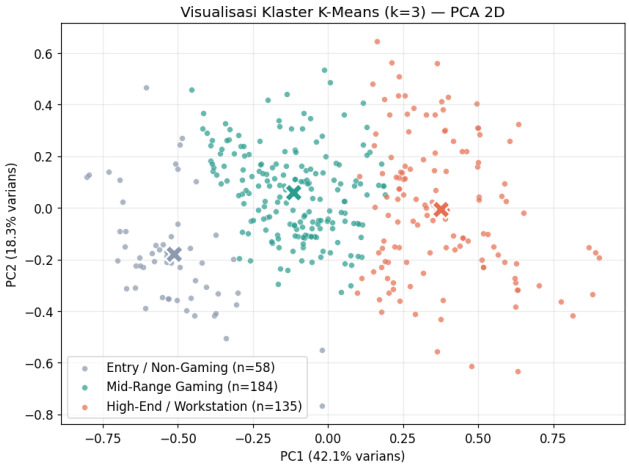
Tabel 2. *Ablation Study* (evaluasi pada ruang fitur penuh)

Konfigurasi	*Silhouett*	*Davies-Bouldin*	ARI vs Full
Full features	0,224	1,530	1,000
GPU-only	0,178	1,720	0,500
Tanpa GPU	0,223	1,560	0,415
GPU *rule baseline*	0,122	2,181	0,260

D. Visualisasi dan Karakteristik Klaster

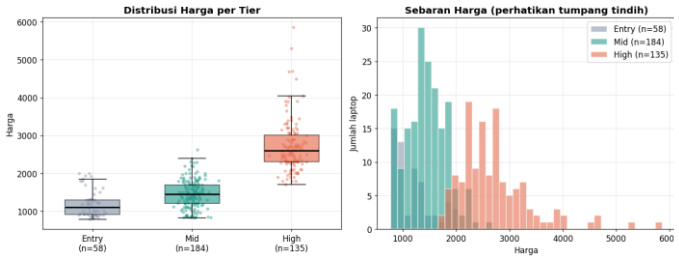
Hasil *clustering* divisualisasikan dalam ruang dua dimensi menggunakan PCA seperti ditunjukkan pada Gambar 2. Dua komponen utama menjelaskan sekitar 60,4% variansi total (PC1 = 42,1%, PC2 = 18,3%). *Loading* PC1 paling besar berasal dari *gpu_benchmark* (0,57) dan *price_log* (0,53), diikuti *refresh_hz* (0,41), sedangkan PC2 lebih dipengaruhi oleh *reviews_log, storage_gb*, dan *refresh_hz*. Visualisasi memperlihatkan kecenderungan pemisahan antar-*tier*, tetapi masih terdapat tumpang tindih sehingga PCA digunakan sebagai alat interpretasi, bukan bukti validitas klaster. Model

final menghasilkan 58 laptop pada *Entry*, 184 pada *Mid-Range*, dan 135 pada *High-End* (total 377). Visualisasi klaster K-Means disajikan pada Gambar 2.



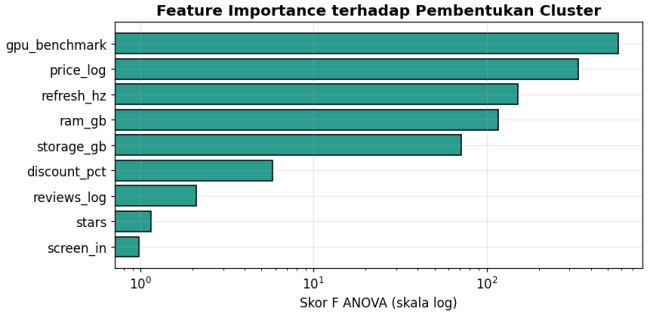
Gambar 2. Visualisasi Klaster K-Means (k = 3) Pada Ruang PCA 2D

Distribusi harga pada Gambar 3 menunjukkan bahwa *tier High-End* memiliki median harga lebih tinggi, sedangkan *Entry* dan *Mid-Range* masih mengalami tumpang tindih. Median harga masing-masing *tier* adalah sekitar USD 1.099,5, USD 1.449,5, dan USD 2.599. Tumpang tindih tersebut menunjukkan bahwa harga memiliki daya pembeda, tetapi tidak memadai sebagai satu-satunya dasar pembentukan *tier*. Distribusi harga per *tier* disajikan pada Gambar 3.



Gambar 3. Distribusi harga per *tier*

Skor F ANOVA pada gambar 4 menunjukkan bahwa *gpu_benchmark* memiliki separasi univariat tertinggi terhadap label klaster (F = 575,8), diikuti dengan *price_log* (38,6), *refresh_hz* (152,2), *ram_gb* (116,3), dan *storage_gb* (71,0). Nilai ini mendukung peran GPU sebagai pembeda penting, tetapi tidak boleh ditafsirkan sebagai kontribusi kausal atau sebagai bukti bahwa satu fitur membentuk klaster secara mandiri; hasil *ablation* pada Tabel 2 menunjukkan bahwa struktur model penuh tetap bergantung pada kombinasi beberapa fitur. Feature importance vs pembentukan kluster disajikan pada Gambar 4.



Gambar 4. *Feature Importance* vs Pembentukan Klaster

Profil rata-rata tiap *tier* dirangkum pada Tabel 3 menggunakan *mean* ± *standard deviation* (SD). SD pada tabel menggambarkan variasi antar-laptop di dalam masing-masing *tier*, bukan variasi antar-*random seed*.

Tabel 2. Profil tiap *tier* (mean ± SD)

Tier	n	Harga	RAM	Storage	GPU-rank	Refresh	Layar
Entry	58	1188,9 ± 341,8	27,7 ± 11,4	1182,5 ± 671,9	5996,1 ± 2418,0	74,1 ± 27,7	15,7 ± 0,9
Mid-Range	184	1449,9 ± 364,8	25,9 ± 9,3	1069,6 ± 418,7	16431,2 ± 2164,4	153,6 ± 40,6	16,6 ± 5,7
High-End	135	2.710,5 ± 681,4	48,7 ± 18,7	2096,3 ± 1132,1	21502,8 ± 3846,9	203,0 ± 60,9	16,4 ± 0,9

Satuan: harga (USD), RAM & *storage* (GB), *refresh* (Hz), layar (inci).

D. Validasi dan Pemeriksaan *Overfitting*

Sebagai validasi praktis, median profil diperiksa menurut urutan *Entry-Mid-High*. Median *gpu_benchmark* meningkat 5.139 → 16.724 → 22.448, *refresh rate* 60 → 165 → 240 Hz, dan harga USD 1.099,5 → 1.449,5 → 2.599; median RAM dan *storage* juga tidak menurut. Selain itu, lima produk yang paling dekat dengan *centroid* tiap *tier* diperiksa, Contoh representatifnya mencakup ASUS Zenbook S 16 dengan Radeon 880M pada *Entry*, Lenovo Legion 5 dengan RTX 5060 pada *Mid-Range*, dan ASUS ROG Strix G16 dengan RTX 5070 Ti pada *High-End*. Pemeriksaan ini mendukung keterbacaan profil *tier* pada dataset, tetapi tetap merupakan validasi praktis/internal karena tidak menggunakan label pakar atau *ground truth* eksternal.

Kestabilan model dinilai menggunakan ARI. Pada 20 *random seed*, hasil k = 3 memiliki rata-rata pairwise ARI 0,9549 ± 0,0352. Pada 30 sub-sampel acak sebesar 80% data, kesesuaian terhadap label model penuh menghasilkan ARI 0,8506 ± 0,1622. Nilai tersebut menunjukkan bahwa solusi relatif dapat direproduksi terhadap perubahan inisialisasi dan pengurangan sebagian data, tetapi tidak membuktikan bahwa klaster merupakan kategori performa yang benar secara eksternal.

E. Pembahasan

Temuan paling penting adalah bahwa *gpu_benchmark* memberikan separasi univariat terbesar antar-klaster (F = 575,8), diikuti dengan *price_log* dan *refresh_hz*. Namun, dominasi skor F tidak berarti hasil *clustering* hanya ditentukan GPU. Pada *ablation study*, GPU-*only* hanya memiliki ARI sebesar 0,500 terhadap model penuh, sedangkan *baseline rule-based* berbasis GPU memiliki ARI 0,260. Dengan demikian, GPU berfungsi sebagai pembeda penting dalam dataset yang dianalisis, tetapi harga, *refresh rate*, RAM, *storage*, dan atribut lain tetap berkontribusi terhadap struktur segmentasi.

Perbandingan antar-*tier* pada Tabel 3 mengungkap bahwa harga rata-rata *Entry* dan *Mid-Range* relatif berdekatan (USD 1.188,9 ± 341,8 vs USD 1.449,9 ± 364,8), tetapi *benchmark* GPU dan *refresh rate* berbeda lebih jelas. *Entry* memiliki *gpu_benchmark* rata-rata 5.996,1 ± 2.418,0 dan *refresh rate* 74,1 ± 27,7 Hz, sedangkan *Mid-Range* mencapai 16.431,2 ± 2.164,4 dan 153,6 ± 40,6 Hz. *High-End* memiliki harga rata-rata USD 2.710,5 ± 681,4, RAM 48,7 ± 18,7 GB, *storage* 2.096,3 ± 1.132,1 GB dan *gpu_benchmark* 21.502,8 ± 3.846,9. Pola tersebut menunjukkan bahwa dataset ini, laptop dengan harga yang sebagian tumpang tindih dapat memiliki profil grafis dan layar yang berbeda, sehingga harga saja tidak cukup menggambarkan *tier* yang terbentuk

Label *Entry* pada penelitian ini merupakan label operasional untuk klaster dengan profil *benchmark* GPU dan *refresh rate* terendah; klaster tersebut masih dapat memuat sebagian laptop dengan GPU *diskret entry-level* sehingga tidak disamakan secara mutlak dengan kategori *non-gaming*. Generalisasi hasil juga dibatasi oleh beberapa faktor: data berasal dari satu dataset Kaggle yang merepresentasikan *snapshot* pasar, hanya 377 dari 594 laptop memiliki *benchmark* GPU yang dapat dipetakan, beberapa spesifikasi lain masih berasal dari ekstraksi teks dan imputasi, serta skor G3D tidak menangkap seluruh variasi TGP dan implementasi GPU pada setiap laptop. Oleh karena itu, batas *tier* yang diperoleh berlaku terutama pada observasi yang dianalisis dan dapat berubah ketika sumber data, cakupan GPU, atau periode pasar berbeda.

V. SIMPULAN

Penelitian ini membangun model *clustering* K-Means pada 377 observasi dengan *benchmark* GPU lengkap, yang berasal dari 594 laptop hasil pembersihan terhadap 614 baris dataset *Gaming* Laptops 2026. Model ginal menggunakan k = 3 dan menghasilkan *tier Entry, Mid-Range*, dan *High-End* dengan *Silhouette* 0,224, *Davies-Bouldin* 1,530, dan *Calinski-Harabasz* 113,6. Evaluasi k = 2 dan k = 3 menunjukkan bahwa k =2 memiliki kualitas geometris dan kestabilan yang lebih baik, sedangkan k = 3 dipilih karena memberikan tiga profil yang lebih mudah diinterpretasikan dalam konteks segmentasi. Dengan demikian, tiga klaster diposisikan sebagai *trade-off* praktis, bukan jumlah klaster yang diklaim optimal secara matematis.

GPU *benchmark* merupakan pembeda penting antar *tier*, tetapi *ablation study* menunjukkan bahwa GPU saja tidak mereproduksi struktur model penuh. Pemeriksaan praktis terhadap median produl dan produk yang dekat dengan *centroid* juga mendukung interpretasi *tier*, meskipun belum menggantikan validasi eksternal. Untuk pengembangan berikutnya, disarankan (1) meningkatkan cakupan ekstraksi dan pemetaan *benchmark* GPU agar lebih banyak observasi dapat digunakan, (2) memasukkan informasi TGP, varian *mobile* GPU, CPU, dan karakteristik termal jika tersedia, (3) memperoleh label *tier* dari pakar atau *benchmark gaming* terstandar untuk validasi eksternal, serta (4) memperluas data lintas sumber dan kuartal agar kestabilan batas tier terhadap perubahan pasar dapat diuji.

REFERENSI

[1] Y. Tang, Y. Zhang, and X. Ning, “Uncertainty in the platform market: The information asymmetry perspective,” *Computers in Human Behavior*, vol. 148, art. 107918, doi: 10.1016/j.chb.2023.107918, 2023.

[2] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, “K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data,” *Information Sciences*, vol. 622, pp. 178–210, doi: 10.1016/j.ins.2022.11.139, 2023.

[3] I. Fajar & A. Setiawan. “Cluster analysis on laptop sales data and specifications using K-means and K-medoids methods,” *Journal of Data Science Insights*, vol. 2, no. 2, pp. 105–113, 2024.

[4] T. C. Saputra, S. M. Fadhilah, S. U. Mangkuto, and J. . Heikal, “Segmentation, targeting and positioning analysis using k-means clustering model: A case study of the laptop market in Indonesia”, ijafibs, vol. 12, no. 2, pp. 195–203, Sep. 2024.

[5] Z. Fataha and K. S. H. A. Putra, “Clustering data produksi untuk identifikasi pola perakitan hardware menggunakan K-means,” *Gudang Jurnal Multidisiplin Ilmu*, vol. 3, no. 11, pp. 250–254, doi: 10.59435/gjmi.v3i11.1897, 2025.

[6] R. Marmay, R. Luthvina, and O. Ulandari, “Implementasi algoritma K-means dalam pengelompokan data harga laptop,” *Journal Scientific of Mandalika (JSM)*, 2026.

[7] M. Alves Gomes and T. Meisen, “A review on customer segmentation methods for personalized customer targeting in e-commerce use cases,” *Information Systems and e-Business Management*, vol. 21, no. 3, pp. 527–570, doi: 10.1007/s10257-023-00640-4, 2023.

[8] M. Cui, “Introduction to the K-Means Clustering Algorithm Based on the Elbow Method,” *Accounting, Auditing and Finance*, vol. 1, pp. 5–8, 2020, doi: 10.23977/accaf.2020.010102.

[9] G. Ogbuabor and F. N. Ugwoke, “Clustering Algorithm for a Healthcare Dataset Using Silhouette

Score Value," *International Journal of Computer Science & Information Technology (IJCSIT)*, vol. 10, no. 2, pp. 27–37, 2018, doi: 10.5121/ijcsit.2018.10203.

[10] M. Shutaywi and N. N. Kachouie, "Silhouette analysis for performance evaluation in machine learning with applications to clustering," *Entropy*, vol. 23, no. 6, art. 759, doi: 10.3390/e23060759, 2021.

[11] A. Ramadhan, D. Kurniawan, & T. Haryanto. "Evaluation of K-Means, DBSCAN, and hierarchical clustering for performance segmentation". *Jurnal Teknik Informatika (JUTIF)*, 6(3), 1605–1614, 2025.

[12] Z. John, N. Anand, D. Surjeet & A. Zainab. "House price prediction using hedonic pricing model and machine learning techniques. Concurrency and Computation: Practice and Experience". 34. 10.1002/cpe.7342, 2022.

[13] A. A. Wani. "Comprehensive analysis of clustering algorithms: Exploring limitations and innovative solutions". PeerJ Computer Science, 10, e2286. https://doi.org/10.7717/peerj-cs.2286, 2024.

[14] G. Liu. "A New Index for Clustering Evaluation Based on Density Estimation." 10.48550/arXiv.2207.01294, 2022.

[15] F. A. Ilham, D. E. Eko, B. Randi, N. Y. Ilham & L. Ridho. "Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index evaluation on K-Means algorithm for classifying flood-affected areas in Jakarta". *Journal of Applied Informatics and Computing (JAIC)*, 7(1), 89–97. https://doi.org/10.30871/jaic.v7i1.4947, 2023.

[16] BD. Zoltan. "A new approach for evaluating internal cluster validation indices". *arXiv preprint arXiv:2308.03894*. https://doi.org/10.48550/arXiv.2308.03894, 2023.