



## **Clustering Data Produksi Untuk Identifikasi Pola Perakitan Hardware Menggunakan K-Means**

**Zachol Fataha<sup>1\*</sup>, Kiki Setiawan Heri Ananda Putra<sup>2</sup>**

<sup>1</sup> Teknologi Informasi, Universitas Ibrahimy

<sup>2</sup> Sistem Informasi, Universitas Ibrahimy

<sup>1\*</sup> [zacholfataha@gmail.com](mailto:zacholfataha@gmail.com), <sup>2\*</sup> [kikisetiawan398@gmail.com](mailto:kikisetiawan398@gmail.com)

---

### **Abstrak**

Penelitian ini bertujuan mengidentifikasi pola pengelompokan komponen perakitan komputer menggunakan metode K-Means Clustering. Dataset yang digunakan mencakup CPU, motherboard, RAM, GPU, PSU, dan casing. Proses penelitian meliputi eksplorasi data, pembersihan data, imputasi nilai hilang, encoding kategorikal, normalisasi, serta rekayasa fitur seperti total TDP dan rasio daya PSU. Hasil clustering menunjukkan pembentukan empat kelompok utama, yaitu komponen low-end, mid-range, high-end, dan cluster RAM yang berdiri sendiri. Temuan ini membuktikan bahwa K-Means efektif dalam memetakan karakteristik hardware dan dapat digunakan sebagai dasar sistem rekomendasi perakitan PC.

**Kata Kunci:** K-Means, Clustering, Hardware, Algoritma

---

### **PENDAHULUAN**

Perkembangan teknologi informasi dalam dua dekade terakhir telah mendorong peningkatan kebutuhan perangkat komputasi di berbagai sektor kehidupan, mulai dari pendidikan, bisnis, industri kreatif, gaming, hingga penelitian ilmiah. Komputer pribadi (Personal Computer/PC) masih menjadi pilihan utama bagi banyak pengguna karena fleksibilitas dalam memilih dan meningkatkan (upgrade) komponen sesuai kebutuhan. Praktik merakit PC secara mandiri kini semakin populer karena memberikan kebebasan bagi pengguna untuk mendapatkan performa terbaik sesuai anggaran. Namun, di balik fleksibilitas tersebut, proses pemilihan komponen PC menimbulkan permasalahan baru, yaitu kompleksitas dalam memahami spesifikasi teknis, kompatibilitas antar komponen, kesesuaian daya, serta perbandingan harga dan performa. Keragaman jenis dan jumlah komponen PC yang beredar di pasaran menyebabkan proses pengambilan keputusan menjadi semakin sulit, terutama bagi pengguna yang tidak memiliki pemahaman teknis mendalam[1].

Komponen PC seperti CPU, GPU, RAM, Motherboard, PSU, dan Case masing-masing memiliki karakteristik teknis yang berbeda-beda, dan setiap produsen biasanya menggunakan standar spesifikasi yang tidak selalu sama satu sama lain. Sebagai contoh, CPU memiliki parameter seperti jumlah core, jumlah thread, frekuensi dasar (base clock), frekuensi boost (turbo clock), TDP, hingga jenis socket. GPU memiliki variabel berbeda seperti VRAM, TDP, panjang fisik, jumlah slot PCIe, serta kebutuhan daya minimum. RAM memiliki variasi kapasitas, kecepatan (MHz), timing, dan jumlah modul. Sementara motherboard memiliki perbedaan chipset, form factor, slot RAM, dukungan socket CPU, dan tipe memori. Perbedaan struktur data dan istilah antar kategori komponen ini menyebabkan dataset komponen PC bersifat heterogen, tidak seragam, dan sulit dianalisis tanpa proses standarisasi yang matang[2].

Selain itu, jumlah produk komponen PC yang tersedia di pasaran sangat banyak dan terus bertambah setiap tahun. Produsen seperti Intel, AMD, NVIDIA, MSI, ASUS, Corsair, dan lainnya merilis produk baru dengan rentang harga yang luas dan varian spesifikasi yang semakin kompleks. Akibatnya, pasar komponen PC menghasilkan data berukuran besar (big data) yang terus berkembang. Menganalisis ribuan produk secara manual untuk menemukan pola, hubungan, atau segmentasi pasar menjadi hampir mustahil dilakukan tanpa bantuan teknologi analisis data modern. Kondisi ini menimbulkan kebutuhan untuk mengembangkan suatu metode yang mampu mengelompokkan komponen PC berdasarkan kesamaan karakteristik sehingga informasi lebih mudah dipahami oleh pengguna maupun peneliti[3].

Seiring berkembangnya ilmu data, pendekatan Machine Learning, khususnya Unsupervised Learning, telah menjadi solusi yang efektif dalam menganalisis dataset besar tanpa memerlukan label atau kategori awal. Salah satu algoritma yang paling banyak digunakan adalah K-Means Clustering, karena kemampuannya dalam mengelompokkan data ke dalam beberapa cluster berdasarkan kemiripan fitur[4]. Dengan menerapkan K-Means pada dataset komponen PC, ribuan produk dapat dikelompokkan menjadi beberapa segmen, misalnya produk low-end, mid-range, hingga high-end. Teknik ini juga memungkinkan peneliti untuk menemukan hubungan antar fitur teknis seperti pengaruh TDP terhadap

performa, segmentasi harga, kecenderungan spesifikasi berdasarkan kategori produk, serta pola yang tidak selalu terlihat secara kasat mata[5].

Penelitian ini berfokus pada proses penggabungan berbagai dataset komponen PC dari kategori CPU, GPU, RAM, Motherboard, PSU, dan Case kemudian melakukan standarisasi fitur, normalisasi, penanganan data hilang, hingga ekstraksi fitur numerik yang relevan seperti TDP, kapasitas memori, core/thread, VRAM, wattage, dan form factor. Data yang telah diproses kemudian dianalisis menggunakan K-Means untuk mendapatkan kelompok komponen PC yang memiliki karakteristik serupa. Proses ini dilengkapi dengan visualisasi seperti PCA (Principal Component Analysis), elbow method untuk menentukan jumlah cluster optimal, heatmap korelasi untuk mengetahui hubungan antar fitur, serta statistik deskriptif dari setiap cluster[6]. Dengan adanya analisis ini, pengguna dapat memperoleh gambaran lebih jelas mengenai segmentasi produk berdasarkan performa, efisiensi daya, kapasitas, dan harga. Selain bermanfaat bagi pengguna umum yang ingin merakit PC, penelitian ini juga memberikan nilai tambah bagi pelaku industri seperti toko komputer, reviewer hardware, maupun pengembang aplikasi rekomendasi komponen PC otomatis. Melalui pendekatan ini, kompleksitas dalam memilih komponen yang tepat dapat diminimalisir, proses perakitan PC menjadi lebih mudah, dan keputusan pembelian dapat dilakukan secara lebih objektif dan berbasis data.

Dengan demikian, penelitian mengenai klasterisasi komponen PC menggunakan metode K-Means ini sangat relevan dan penting dilakukan mengingat tingginya kompleksitas data komponen PC, kebutuhan analisis yang lebih efisien, serta tingginya permintaan pengguna terhadap sistem rekomendasi yang mampu memberikan hasil yang akurat. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem analisis data komponen PC serta membantu pengguna dalam memahami segmentasi produk secara lebih praktis dan ilmiah[7].

## METODE

Penelitian ini menggunakan metode kualitatif deskriptif dengan dukungan pendekatan komputasional untuk mengolah data spesifikasi komponen hardware PC menggunakan algoritma K-Means. Pendekatan kualitatif dipilih karena penelitian ini tidak hanya berfokus pada hasil angka atau perhitungan, tetapi juga menekankan pemahaman, penjelasan, serta interpretasi pola yang muncul dari hasil klasterisasi data numerik. Proses analisis melibatkan pemahaman mendalam mengenai hubungan antar komponen PC, seperti CPU, GPU, RAM, PSU, motherboard, dan casing, yang kemudian dianalisis menggunakan model data mining untuk menemukan pola-pola tertentu dalam perakitan komputer[8].

Pengembangan sistem analisis dalam penelitian ini menggunakan metode Waterfall, karena struktur tahapannya yang sistematis dan berurutan. Tahap pertama adalah analisis kebutuhan, yaitu proses mengidentifikasi jenis data, fitur numerik, serta variabel-variabel yang akan digunakan dalam proses klasterisasi. Data yang dikumpulkan kemudian dirancang ulang dalam tahap perancangan, yaitu pembuatan desain alur preprocessing, penggabungan dataset, serta rancangan teknis model clustering. Tahap ini dilanjutkan dengan proses coding menggunakan bahasa pemrograman Python. Seluruh proses pembersihan data, normalisasi, encoding, penghitungan fitur turunan, serta penerapan algoritma K-Means dilakukan pada tahap ini[9].

Setelah sistem berjalan, dilakukan pengujian untuk memastikan bahwa algoritma K-Means bekerja secara optimal dan menghasilkan cluster yang stabil. Pengujian meliputi evaluasi jumlah cluster menggunakan Elbow Method, validasi hasil cluster, serta pengecekan konsistensi data melalui visualisasi. Apabila ditemukan kesalahan atau ketidaksesuaian data, dilakukan proses pemeliharaan yaitu penyesuaian terhadap dataset, penambahan fitur baru, atau perbaikan alur preprocessing untuk meningkatkan akurasi dan kualitas clustering[10].

## HASIL DAN PEMBAHASAN

### Analisis Struktur dan Karakteristik Dataset

Dataset yang digunakan dalam penelitian ini merupakan gabungan dari berbagai kategori komponen komputer, seperti CPU, GPU, RAM, PSU, motherboard, dan casing, yang kemudian disatukan dalam satu bentuk dataset terstruktur. Dari total 7.259 data yang berhasil dihimpun, sebagian besar merupakan komponen non-CPU seperti GPU dan PSU, sementara data CPU hanya berjumlah 328 entri. Hal ini menyebabkan distribusi dataset menjadi tidak merata, namun tetap dapat diproses dengan teknik clustering setelah melalui tahap normalisasi dan imputasi data. Banyaknya kolom yang bernilai kosong pada dataset asli, terutama pada kolom GPU, RAM, dan PSU dalam baris data CPU, menunjukkan bahwa dataset ini memang tidak dirancang sebagai tabel tunggal sejak awal, sehingga proses penggabungan dan pembersihan data menjadi bagian penting dari penelitian.

The screenshot displays the Kaggle interface for the 'Computer Hardware Dataset'. The main content area includes the dataset title, a description, and a 'Data Card' section. The 'About Dataset' section provides details about the data, including its scope and the fact that it is a work in progress. The 'Usability' score is 8.82, and the 'License' is CC0: Public Domain. The 'Expected update frequency' is Monthly. The sidebar on the left shows navigation options like Home, Competitions, Datasets, Models, Benchmarks, Game Arena, Code, Discussions, and Learn.

Gambar 1. Halaman Kaggle Dataset Computer Hardware

Seluruh komponen dinormalisasi ke dalam format numerik melalui transformasi fitur seperti CPU\_TDP, GPU\_TDP, VRAM, RAM\_GB, PSU\_Wattage, dan lainnya agar dapat digunakan oleh algoritma K-Means. Proses ini menghasilkan dataset yang lebih stabil dan siap untuk diolah secara matematis. Dengan demikian, tahap preprocessing menjadi fondasi penting sebelum dilakukan analisis clustering.

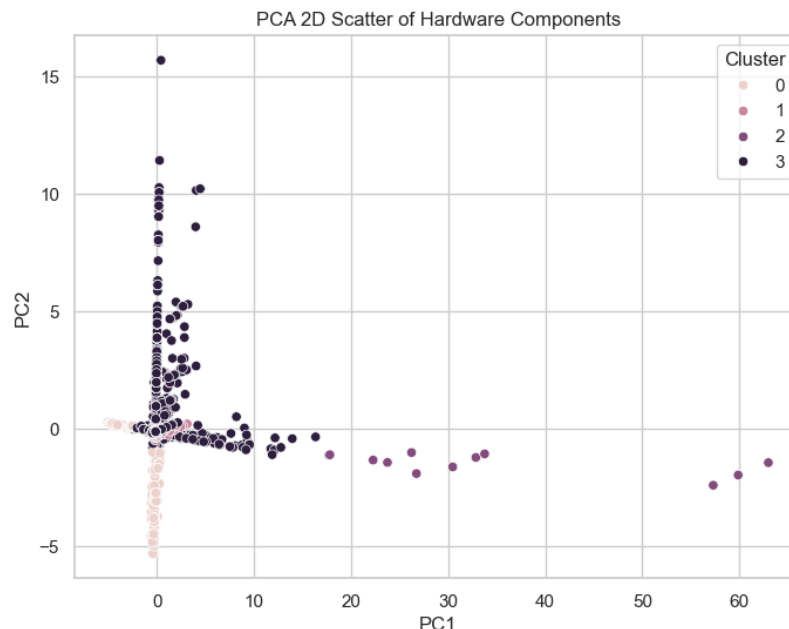
```
===== Unified Product Dataset (before processing) =====
```

	Name	Price	CPU_TDP	CPU_Cores	CPU_Threads	Category	\
0	AMD Ryzen 5 5600X	\$158.86 USD	65.0	6.0	12.0	CPU	
1	AMD Athlon 3000G	\$53.22 USD	35.0	2.0	4.0	CPU	
2	AMD Ryzen 3 3300X	\$150.09 USD	65.0	4.0	8.0	CPU	
3	AMD Ryzen 5 5500	\$87.04 USD	65.0	6.0	12.0	CPU	
4	AMD Ryzen 5 5600	\$133.46 USD	65.0	6.0	12.0	CPU	

Gambar 2. Head Dataset Product

### Analisis Jumlah Cluster Menggunakan Elbow Method

Tahap penentuan jumlah cluster dilakukan menggunakan Elbow Method, yang memperlihatkan data inertia pada jumlah cluster 2 hingga 7. Nilai inertia mengalami penurunan signifikan pada penambahan cluster, terutama di titik K=4 dan K=5, sebelum kemudian melandai pada K=6 dan K=7. Pola ini menunjukkan bahwa titik optimum terdapat pada K=4 atau K=5. Berdasarkan hasil analisis, K=4 dipilih karena menawarkan keseimbangan terbaik antara pemisahan cluster dan efisiensi komputasi.



Gambar 3. Visualisasi PCA 2D Scatter

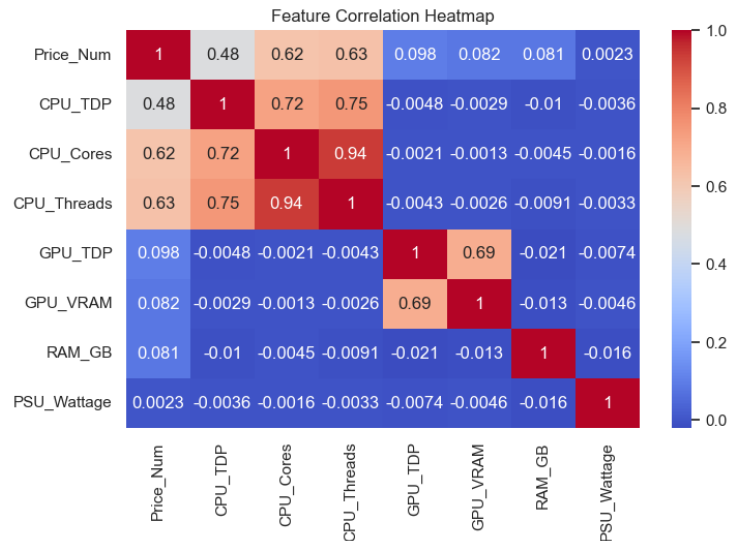
Penentuan cluster yang optimal ini sangat penting, mengingat dataset memuat jenis komponen yang berbeda-beda. Dengan memilih jumlah cluster yang tepat, pola kategori perangkat keras dapat dipisahkan secara akurat berdasarkan kesamaan spesifikasi.

### Interpretasi Hasil Clustering K-Means

Hasil clustering menunjukkan pembagian dataset ke dalam empat kelompok besar yang memiliki karakteristik berbeda. Cluster pertama merupakan cluster terbesar dengan 6.257 item, yang didominasi oleh data PSU, GPU, motherboard, dan casing. Cluster ini cenderung menampung komponen-komponen dengan nilai fitur menengah dan variatif, seperti watt PSU, ukuran casing, atau kapasitas daya GPU. Dominasi jumlah data di cluster ini menunjukkan bahwa sebagian besar dataset terdiri dari perangkat keras non-prosesor yang memiliki variasi spesifikasi yang luas. Cluster kedua berisi 729 item yang didominasi oleh GPU dan PSU. Cluster ini menampilkan karakteristik perangkat keras yang memiliki konsumsi daya relatif lebih tinggi dibandingkan komponen umum lainnya. Nilai rata-rata harga pada cluster ini juga lebih tinggi dibandingkan beberapa cluster lain, menunjukkan bahwa GPU dan PSU yang masuk dalam kelompok ini cenderung merupakan perangkat kelas menengah hingga tinggi.

Cluster ketiga berisi 261 item dan hampir seluruhnya merupakan RAM. Ini menunjukkan bahwa algoritma K-Means mampu memisahkan perangkat RAM dengan baik karena karakteristik numeriknya yang relatif homogen, seperti ukuran kapasitas (8 GB, 16 GB, 32 GB), kecepatan clock, dan struktur modul. Dengan demikian, RAM membentuk cluster tersendiri karena memiliki pola nilai fitur yang konsisten satu sama lain. Cluster keempat merupakan cluster paling kecil, hanya berisi 12 item, dan seluruhnya merupakan CPU dengan spesifikasi tinggi. CPU dalam kelompok ini umumnya memiliki jumlah core, thread, dan TDP yang jauh lebih besar dibandingkan CPU biasa.

Harga rata-rata cluster ini juga paling tinggi dibandingkan cluster lainnya, menunjukkan bahwa ini adalah kelompok

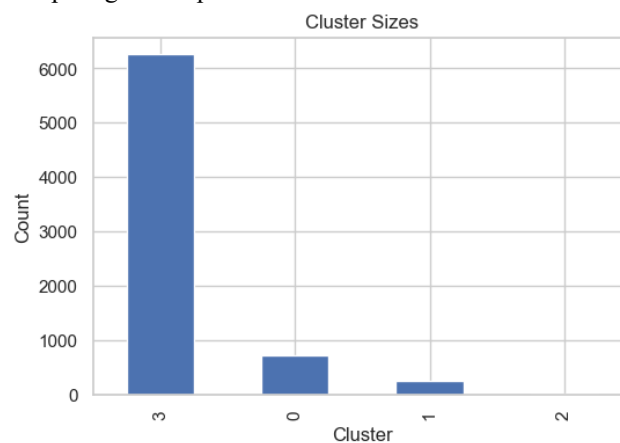


prosesor kelas high-end.

Gambar 4. Visualisasi Feature Correlation Heatmap

### Analisis Harga Rata-rata per Cluster

Hasil perhitungan harga rata-rata menunjukkan variasi nilai yang cukup signifikan antar cluster. Cluster RAM memiliki harga rata-rata tertinggi, yaitu sekitar 36.735, yang dipengaruhi oleh banyaknya produk RAM premium berkapasitas besar. Sementara itu, cluster CPU high-end mencatat harga rata-rata lebih dari 250.000, menjadikannya cluster dengan nilai komponen paling mahal per item.



Gambar 5. Grafik Hasil Clustering

Cluster produk GPU-PSU memiliki rata-rata harga menengah yaitu sekitar 18.000. Nilai ini selaras dengan kondisi pasar, di mana GPU kelas entry hingga mid-range mendominasi dataset. Cluster besar yang memuat PSU, casing, dan motherboard mencatat harga rata-rata sekitar 20.398, sehingga menunjukkan bahwa komponen-komponen pendukung sistem juga memiliki variasi harga yang sangat luas.

===== CLUSTER SUMMARY =====

```
Cluster 0 (729 items):
Category
GPU      444
PSU      209
CPU       76
Name: count, dtype: int64
Avg price: 18031.95145631068

Cluster 1 (261 items):
Category
RAM      261
Name: count, dtype: int64
Avg price: 36735.42156862745

Cluster 2 (12 items):
Category
CPU       12
Name: count, dtype: int64
Avg price: 252212.08333333334

Cluster 3 (6257 items):
Category
...
PSU       728
CPU       240
```

## Gambar 6. Kesimpulan Cluster

Perbedaan harga antar cluster mempertegas kemampuan K-Means dalam mengelompokkan perangkat keras menurut karakteristik nilai dan fungsinya, sehingga proses clustering tidak hanya berhasil secara matematis, tetapi juga logis jika dikaitkan dengan realitas pasar perangkat komputer.

### KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan menggunakan metode K-Means Clustering terhadap data komponen komputer, dapat disimpulkan bahwa proses eksplorasi, pembersihan, dan rekayasa fitur berhasil membentuk dataset yang representatif untuk dianalisis. Algoritma K-Means mampu mengelompokkan komponen hardware ke dalam beberapa cluster yang mencerminkan karakteristik performa, kebutuhan daya, dan kategori pasar, seperti low-end, mid-range, high-end, serta cluster khusus RAM. Hasil ini menunjukkan bahwa teknik clustering efektif dalam memahami pola spesifikasi hardware dan dapat digunakan sebagai dasar sistem rekomendasi perakitan PC maupun analisis segmentasi produk.

Adapun saran dari penelitian ini, penggunaan dataset dengan cakupan lebih luas, penambahan fitur teknis yang lebih spesifik, serta penerapan algoritma clustering lain seperti DBSCAN atau Hierarchical Clustering dapat meningkatkan akurasi dan kedalaman analisis. Integrasi metode ini ke dalam aplikasi rekomendasi rakitan PC juga berpotensi memberikan manfaat praktis bagi pengguna, teknisi, maupun penyedia layanan e-commerce komponen komputer.

### DAFTAR PUSTAKA

- [1] R. F. Daud, I. Komunikasi, U. M. Kotabumi, and L. Utara, "Dampak Perkembangan Teknologi Komunikasi Terhadap Bahasa Indonesia," vol. 5, no. 2, pp. 252–269, 2021, doi: 10.30596/interaksi.v5i2.7539.
- [2] W. Yunanri, A. Fauzan, A. Yani, and M. A. Aziz, "Analisis Performance Central Prosessing Unit ( CPU ) Realtime Menggunakan Metode Benchmarking An Analysis of Performance Central Processing Unit ( CPU ) for Real Time Using Benchmarking Method," vol. 20, no. 2, pp. 237–248, 2021, doi: 10.30812/matrik.
- [3] F. Montoya, Z. Aulia, A. Nugroho, and D. Aribowo, "Optimasi Performa Komputer melalui Teknik Overlocking pada Prosessor AMD Ryzen 5 5600x," 2023.
- [4] A. A. Zulyani, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Menentukan Tingkat Vaksinasi Pada Kecamatan Tambun Selatan," vol. 3, pp. 7037–7050, 2023.
- [5] R. S. Nurhalizah and R. Ardianto, "Analisis Supervised dan Unsupervised Learning pada Machine Learning : Systematic Literature Review," vol. 4, no. 1, pp. 61–72, 2024.
- [6] P. Component, "Pengembangan Aplikasi Berbasis Web dengan R Shiny untuk Analisis Data Menggunakan Algoritma".
- [7] A. Sujiana, U. Budiyanto, and F. T. Informasi, "Cogito Smart Journal | VOL. 9 - NO.2, DECEMBER 2023 ■ 252," vol. 9, no. 2, pp. 252–265, 2023.
- [8] S. Agustinus, B. Telaumbanua, F. Setiadi, and S. Nurjanah, "Analisis Clustering Menggunakan Metode Enhanced Fuzzy C-Means Clustering Dengan Algoritma Rock Pada Student Performance Dataset," vol. 7, no. 3, 2025, doi: 10.32877/bt.v7i3.2287.
- [9] A. K. Clustering and G. Belajar, "Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar," vol. 12, 2022, doi: 10.34010/jati.v12i1.
- [10] I. N. Abrar and A. Abdullah, "Klasifikasi Penyakit Liver Menggunakan Metode Elbow Untuk Menentukan K Optimal pada Algoritma K-Nearest Neighbor ( K-NN )," vol. 12, pp. 218–228, 2023.