



Comparative Performance Analysis of K-NN and K-Means Algorithms for Gaming Mouse Recommendation System

Analisis Perbandingan Kinerja Algoritma K-NN dan K-Means untuk Sistem Rekomendasi Mouse Gaming

Suhardi Aras^{1*}, Agniel Lorensyus Malino²,
Yuchiyo Heizkia Reenhard Paliyama³, Ridho Bintang Ramadhan⁴

^{1,2,3,4}Faculty of Engineering, Universitas Muhammadiyah Sorong, Indonesia

E-Mail: ¹suhardi.aras@um-sorong.ac.id, ²nielmano1405@gmail.com,
³reenhard07@gmail.com, ⁴ridho2005bintang@gmail.com

Received Jul 08th 2025; Revised Oct 15th 2025; Accepted Oct 24th 2025; Available Online Oct 31th 2025

Corresponding Author: Suhardi Aras

Copyright © 2025 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

The demand for high-performance gaming peripherals, such as gaming mice, has increased alongside the rapid growth of the gaming industry and digital technology. Features like high DPI, macro buttons, precise sensors, and ergonomic design have become key considerations for gamers. However, the vast variety of available products often confuses consumers in choosing the right mouse for their preferences. This study aims to develop a gaming mouse recommendation system using a machine learning approach by comparing two algorithms: K-Nearest Neighbor (K-NN) and K-Means. The research methodology involves data collection from Kaggle, data preprocessing (cleaning and normalization), dimensionality reduction using PCA, clustering with K-Means, and classification with K-NN, enhanced by the SMOTE technique to address class imbalance. K-NN, as a supervised learning algorithm, classifies data based on similarity to its nearest neighbors, while K-Means, an unsupervised method, groups data based on feature similarity. Evaluation metrics include accuracy, precision, recall, and F1-score. The results show that K-Means performs effectively in product segmentation, while K-NN achieves high accuracy in classification. Combining both algorithms leads to a more accurate and user-oriented recommendation system for gaming peripherals.

Keyword: Algorithm, Gaming Mouse, K-Means, K-NN, System

Abstrak

Kebutuhan akan perangkat keras seperti *mouse* untuk bermain gim yang mendukung performa meningkat seiring dengan pertumbuhan industri gim dan teknologi digital. Fitur seperti *DPI* tinggi, tombol makro, sensor presisi, dan desain ergonomis menjadi daya tarik utama bagi pengguna. Namun, banyaknya variasi produk membuat konsumen kesulitan memilih *mouse* yang sesuai dengan preferensi mereka. Penelitian ini bertujuan untuk membangun sistem rekomendasi *mouse gaming* berbasis pembelajaran mesin dengan membandingkan dua algoritma, yaitu *K-Nearest Neighbor (K-NN)* dan *K-Means*. Metode penelitian meliputi pengumpulan data dari *Kaggle*, *preprocessing data* (pembersihan, normalisasi), reduksi dimensi dengan *PCA*, pengelompokan menggunakan *K-Means*, serta klasifikasi menggunakan *K-NN* dengan penerapan *SMOTE* untuk mengatasi ketidakseimbangan kelas. *K-NN*, sebagai algoritma *supervised learning*, memanfaatkan kedekatan antar data untuk menentukan label, sedangkan *K-Means* merupakan algoritma *unsupervised* yang mengelompokkan data berdasarkan kemiripan fitur. Evaluasi dilakukan menggunakan metrik *akurasi*, *presisi*, *recall*, dan *F1-score*. Hasil menunjukkan bahwa *K-Means* efektif dalam segmentasi produk, sementara *K-NN* memberikan akurasi tinggi dalam klasifikasi. Kombinasi keduanya menghasilkan sistem rekomendasi yang lebih akurat dan relevan bagi pengguna.

Kata Kunci: Algoritma, *K-Means*, *K-NN*, *Mouse Gaming*, Sistem

1. PENDAHULUAN

Kebutuhan akan *hardware* yang mendukung kinerja pengguna telah meningkat secara signifikan sebagai akibat dari pertumbuhan industri gim dan teknologi digital. *Mouse gaming* adalah salah satu perangkat yang melihat peningkatan permintaan. Sekarang *mouse gaming* dapat digunakan tidak hanya



sebagai alat penunjuk. Mereka memiliki fitur khusus seperti *Dots Per Inch (DPI)* tinggi, tombol makro, sensor optik yang presisi, dan desain ergonomis yang membuatnya nyaman untuk digunakan selama waktu yang lama [1]. Pelanggan dihadapkan pada banyak pilihan yang sering kali membingungkan seiring bertambahnya variasi produk di pasaran. Ini terutama berlaku bagi mereka yang tidak memahami spesifikasi teknis yang sesuai dengan kebutuhannya.

Pemilihan *mouse gaming* yang tepat dalam hal ini sangat penting agar pengalaman bermain gim dapat disesuaikan dengan preferensi setiap pengguna. Namun, banyaknya produk yang tersedia membuat proses pengambilan keputusan tidak efektif. Sementara beberapa pengguna mungkin lebih memperhatikan desain, yang lain mungkin memperhatikan kinerja atau harga. Akibatnya, diperlukan sebuah sistem pintar yang dapat membantu pengguna memilih produk *mouse gaming* sesuai dengan preferensi dan kebutuhan mereka [2].

Membantu pelanggan memilih produk *mouse gaming* yang tepat dari banyaknya pilihan yang ada adalah masalah utama yang muncul [3]. Jika tidak ditangani dengan baik, konsumen dapat salah membeli produk, yang dapat menyebabkan kenyamanan, kinerja gim, atau bahkan pemborosan biaya. Di sisi lain, produsen dan platform *e-commerce* menghadapi masalah dalam memberikan rekomendasi yang tepat bagi calon pembeli.

Mengembangkan sistem rekomendasi adalah salah satu solusi strategis untuk masalah ini. Data historis, preferensi, dan kesamaan antarproduk dapat membantu pelanggan memilih produk melalui sistem rekomendasi [4]. Dua metode algoritma, *K-Nearest Neighbor (K-NN)* dan *K-Means*, akan digunakan untuk membangun sistem rekomendasi dalam penelitian ini. Dengan melakukan perbandingan, Anda dapat mengetahui metode mana yang lebih efektif untuk memberikan rekomendasi kepada pengguna tentang produk *mouse gaming*.

Algoritma *K-NN* merupakan metode *supervised learning* yang banyak digunakan dalam sistem rekomendasi karena mampu mengidentifikasi kemiripan antar data untuk memberikan hasil yang relevan. Penelitian terdahulu menunjukkan bahwa *K-NN* berbasis *item-based collaborative filtering* berhasil memberikan rekomendasi produk *e-commerce* dengan tingkat akurasi yang baik (*MAE* 1,05 dan *RMSE* 1,36) [5]. Kelebihan utama dari *K-NN* adalah kemampuannya dalam mempersonalisasi rekomendasi berdasarkan kesamaan preferensi pengguna, menjadikannya cocok untuk kebutuhan seperti pemilihan *mouse gaming* yang sangat tergantung pada preferensi individual.

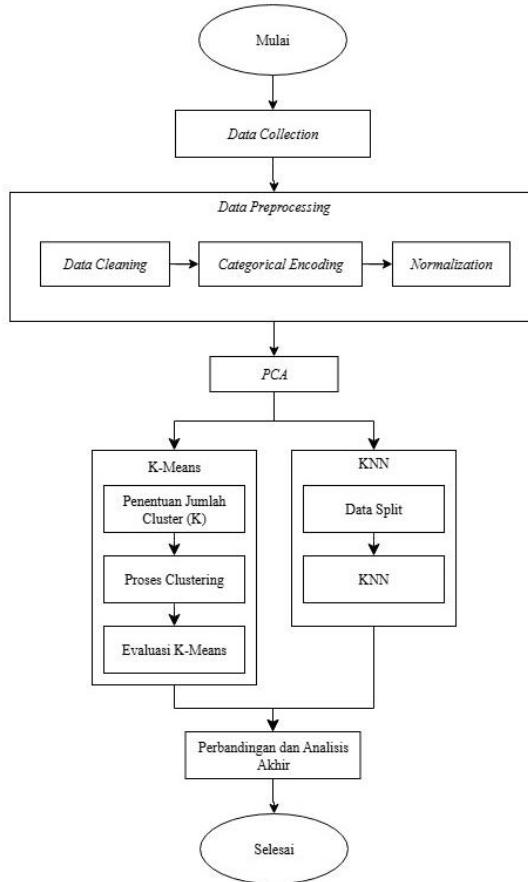
Clustering merupakan salah satu teknik dalam *unsupervised learning* yang bertujuan untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan tingkat kemiripan antar data. Setiap kelompok atau *cluster* berisi data dengan karakteristik yang serupa, sehingga dapat mempermudah proses analisis dan rekomendasi produk. Di sisi lain, *K-Means* adalah algoritma *unsupervised learning* yang efektif dalam mengelompokkan data ke dalam beberapa *cluster* berdasarkan fitur tertentu. Algoritma ini juga dapat mempercepat proses rekomendasi serta mengurangi kompleksitas data tanpa mengorbankan akurasi hasilnya [6], [7]. Kombinasi *K-Means* dan *K-NN* juga telah digunakan dalam pengembangan sistem rekomendasi, yang menunjukkan peningkatan efisiensi dan relevansi rekomendasi [8]. Oleh karena itu, membandingkan kedua algoritma ini dalam konteks sistem rekomendasi *mouse gaming* menjadi langkah penting untuk menentukan pendekatan yang paling optimal.

Tujuan penelitian ini adalah untuk membuat dan menerapkan sistem rekomendasi untuk produk *mouse gaming* yang menggunakan algoritma *K-NN* dan *K-Means*. Selain itu, mereka juga akan melakukan analisis perbandingan terhadap kinerja kedua algoritma tersebut. Diharapkan hasil penelitian ini akan membantu pengembang sistem mengembangkan sistem rekomendasi *produk* yang lebih akurat dan sesuai dengan kebutuhan pengguna. Mereka juga akan menunjukkan pengembang sistem metode algoritma yang lebih optimal untuk rekomendasi perangkat keras permainan.

2. METODOLOGI PENELITIAN

Metodologi penelitian ini dibagi dalam lima tahap dapat dilihat pada Gambar 1. Adapun penjelasannya yaitu :

1. *Data Collection*, Data spesifikasi lengkap *mouse gaming* dikumpulkan dari website *Kaggle*.
2. *Data Preprocessing*, Data yang telah dianalisis terlebih dahulu lalu melakukan penghapusan kolom yang tidak relevan, *encoding* data kategorikal, serta normalisasi fitur menarik untuk menyamakan skala data sebelum digunakan dalam proses klasifikasi.
3. *PCA*, Setelah data bersih dan sudah normal, maka akan dilakukan reduksi dimensi menggunakan *PCA* untuk menyederhanakan fitur sebelum *modeling*.
4. Proses *Modeling*, Tahap ini menggunakan dua jalur paralel, yaitu *K-Means* dan *K-NN*. *K-Means* digunakan untuk proses pengelompokan data, sedangkan *K-NN* digunakan untuk klasifikasi data. Keduanya menggunakan hasil reduksi dimensi dari *PCA* sebagai *input* dan dievaluasi untuk dibandingkan efektivitasnya.
5. Perbandingan dan Analisis Hasil, Hasil dari kedua algoritma kemudian dibandingkan dan dianalisis untuk menentukan metode yang paling efektif dalam sistem rekomendasi *mouse gaming*.



Gambar 1. Metode Penelitian

2.1. *Data Collection*

Data collection merupakan proses sistematis untuk mengumpulkan, mengukur, dan merekam data yang akurat dan relevan dari berbagai sumber guna mendukung proses analisis dan pengambilan keputusan yang tepat [9]. Dalam konteks penelitian ini, data collection bertujuan untuk memperoleh informasi teknis mengenai produk *mouse gaming* yang digunakan sebagai dasar pengembangan sistem rekomendasi. Tujuan pengumpulan data ini adalah untuk menyediakan basis data yang lengkap dan representatif mengenai karakteristik *mouse gaming*, seperti ukuran, berat, jenis sensor, serta fitur ergonomis yang memengaruhi pengalaman pengguna.

Metode pengumpulan data yang digunakan dalam penelitian ini adalah data sekunder, yaitu data yang diperoleh dari pihak ketiga. *Dataset* diunduh dari situs web Kaggle melalui tautan "<https://www.kaggle.com/datasets/ellimaaac/gaming-mouse-specs/data>". *Dataset* tersebut berisi 1.057 entri yang mewakili berbagai merek dan tipe *mouse gaming*, dengan total 24 variabel independen. Adapun variabel-variabel tersebut mencakup informasi seperti: dimensi fisik (panjang, lebar, tinggi), berat, bentuk, teknologi sensor, posisi sensor, *polling rate*, *DPI*, akselerasi, jumlah tombol tambahan, serta fitur-fitur ergonomis lainnya seperti *thumb rest* dan material permukaan. Data yang terkumpul ini selanjutnya digunakan dalam proses *preprocessing* dan modeling sebagai *input* utama dalam sistem rekomendasi *mouse gaming*.

2.2. *Data Preprocessing*

Data preprocessing adalah proses yang melibatkan transformasi dan pembersihan data agar sesuai dengan kebutuhan model komputasional untuk digunakan atau dianalisis oleh algoritma pembelajaran mesin. *Preprocessing* data bertujuan untuk meningkatkan kualitas, memastikan struktur dan skala data konsisten, dan mengurangi suara atau inkonsistensi. Agar model seperti *K-Means* dan *K-NN* dapat belajar dengan lebih akurat, membuat prediksi yang lebih akurat, dan dapat digeneralisasi dengan baik, hal ini sangat penting.

Preprocessing mencakup beberapa tahap penting dalam penelitian ini. Pertama, penanganan nilai hilang (*missing value*) untuk menghindari bias analisis. Kedua, pengubahan data kategorikal menjadi numerik melalui *encoding* agar dapat diproses oleh algoritma. Ketiga, dilakukan normalisasi atau *scaling* data numerik guna menyamakan skala antar fitur dan meningkatkan performa algoritma yang sensitif terhadap jarak antar data. Ini berkaitan dengan hasil Hakim (2020), yang menunjukkan bahwa *preprocessing*

sangat memengaruhi akurasi model, terutama dalam klasifikasi, seperti model *Naive Bayes* untuk analisis sentiment [10].

2.3. Principal Component Analysis (PCA)

Untuk menyederhanakan data berukuran besar menjadi dimensi yang lebih sederhana, analisis komponen utama *PCA* adalah metode reduksi dimensi. Kumpulan fitur asli diubah menjadi kumpulan komponen utama (*principal components*) yang tidak berkorelasi satu sama lain dan disusun berdasarkan jumlah variasi (*variance*) yang dijelaskan. Mengatasi masalah multikolinearitas dan mempercepat proses komputasi pada data besar adalah dua contoh hasil yang baik dari penggunaan metode ini. Sebelum digunakan dalam proses *clustering* dan klasifikasi, *PCA* digunakan untuk mengurangi kompleksitas data dalam penelitian ini. Ini meningkatkan efisiensi dan kinerja model [11].

Studi ini menggunakan *PCA* untuk membagi dimensi data hasil normalisasi menjadi dua komponen utama (*PCA1* dan *PCA2*). Tujuan dari pengurangan ini adalah untuk membuat hasil pengelompokan *K-Means* dan *K-NN* lebih mudah dilihat dalam grafik dua dimensi, sehingga distribusi dan pemisahan antar klaster dapat dilihat dengan lebih jelas.

2.4. Clustering

Clustering merupakan salah satu metode dalam *unsupervised learning* yang bertujuan untuk mengelompokkan data ke dalam beberapa klaster berdasarkan kemiripan karakteristik antar data. Tidak seperti metode *supervised learning* yang membutuhkan label, *clustering* bekerja tanpa informasi target dan hanya mengandalkan pola alami yang terdapat pada data. Algoritma *clustering* seperti *K-Means* mengelompokkan data dengan meminimalkan jarak antar data dalam satu klaster dan memaksimalkan jarak antar klaster yang berbeda. Proses ini sangat berguna dalam eksplorasi data awal maupun dalam sistem rekomendasi, karena memungkinkan pengelompokan produk atau item berdasarkan kemiripan fitur [12].

Dalam konteks sistem rekomendasi, *clustering* digunakan untuk mengidentifikasi kelompok item atau pengguna yang memiliki preferensi serupa, sehingga sistem dapat memberikan rekomendasi yang lebih relevan tanpa perlu informasi eksplisit dari pengguna. Salah satu algoritma yang umum digunakan adalah *K-Means*, yang bekerja dengan menentukan pusat klaster (*centroid*) secara iteratif hingga hasil pengelompokan stabil [13].

2.5. K-Means

Untuk pengelompokan (*clustering*) data, algoritma *K-Means* adalah salah satu pendekatan pembelajaran tanpa pengawasan yang paling banyak digunakan. *K-Means* membagi sekumpulan data ke dalam sejumlah klaster (*k*) berdasarkan kedekatan jarak terhadap *centroid* (titik pusat) masing-masing klaster. Proses iteratif algoritma ini dimulai dengan pemilihan *centroid* secara acak, kemudian pengelompokan data berdasarkan kedekatan ke *centroid*, dan pembaruan *centroid* hingga proses konvergen atau stabil. Algoritma ini sangat efektif ketika diterapkan pada data yang memiliki pemisahan kelompok yang jelas. Studi oleh Fahmiyah dan Ningrum (2023) menggunakan *K-Means* untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan nilai Indeks Pembangunan Manusia (*IPM*). Metode *Elbow* dan indeks *Calinski-Harabasz* digunakan untuk memilih jumlah klaster. *K-Means* dapat mengelompokkan wilayah berdasarkan kemiripan karakteristik sosial ekonomi [14].

Selain itu, telah terbukti bahwa penggunaan *K-Means* memiliki efek positif pada bidang kesehatan dan demografi masyarakat. Studi Pramudiyanti dkk. (2024) menggunakan *K-Means* untuk menentukan alasan mengapa kebutuhan ber-KB di Kabupaten Banyuwangi tidak terpenuhi. Hasil *clustering* menunjukkan bahwa alasan mengapa kebutuhan Keluarga Berencana (KB) tidak terpenuhi dapat dikategorikan secara khusus, seperti karena alasan ekonomi, ketidaktahuan, atau kesulitan layanan kesehatan. Oleh karena itu, ini dapat digunakan oleh pemerintah daerah untuk membuat kebijakan yang lebih fokus [15]. Selain itu, Irfiani dan Rani (2021) menggunakan algoritma yang sama untuk mengklasifikasikan status gizi balita di Posyandu dengan menggunakan data berat badan dan tinggi badan. Mereka menemukan bahwa teknik *K-Means* cukup efektif untuk mengelompokkan balita ke dalam kategori gizi buruk, kurang, normal, dan obesitas. Ini juga dapat digunakan untuk membantu petugas kesehatan memantau lebih baik kesehatan anak-anak [16].

Algoritma *K-Means* sangat digunakan di bidang bisnis dan pendidikan, selain di bidang sosial dan kesehatan. Handayani dkk. (2022) membuat sistem berbasis web menggunakan Laravel yang memanfaatkan *K-Means* untuk mengkategorikan gaya belajar siswa. Ini memungkinkan guru menyesuaikan metode pengajaran mereka sesuai dengan masing-masing kelompok siswa. Metode ini meningkatkan interaksi antara guru dan siswa dan mendukung personalisasi pembelajaran [17]. Dalam industri ritel, Abid dkk. menggunakan algoritma *K-Means* pada data penjualan komoditas pangan di Toko Tani Indonesia. Tujuan dari penerapan algoritma ini adalah untuk mengidentifikasi pola pembelian konsumen dan mengoptimalkan pengelolaan stok barang. Hasil klasterisasi ini memberikan pemilik toko informasi penting untuk menentukan strategi penjualan yang lebih baik dan untuk merencanakan distribusi barang yang lebih baik [18].

2.6. K-NN

Untuk satu metode klasifikasi *supervised learning* adalah algoritma *K-NN*, yang bekerja berdasarkan prinsip kesamaan antar data. Pada dasarnya, *K-NN* menentukan kelas data baru dengan menghitung jarak ke seluruh data pelatihan, kemudian menggunakan tetangga terdekat untuk membuat keputusan. Algoritma ini sangat populer di banyak industri karena kesederhanaannya, seperti pengambilan keputusan berbasis data, sistem rekomendasi, dan klasifikasi gambar. Dalam penelitian yang dilakukan oleh Khairi dkk. (2021), *K-NN* berhasil digunakan untuk mengklasifikasi status ekonomi masyarakat Desa Sapikerep dengan akurasi 98,68%. Ini menunjukkan bahwa *K-NN* efektif dalam pengambilan keputusan berbasis data sosial [19].

Selain itu, *K-NN* banyak digunakan dalam pengolahan data teks seperti klasifikasi berita dan deteksi hoaks. Dalam penelitian yang dilakukan oleh Pratiwi dkk. (2022), algoritma *K-NN* digunakan untuk mendeteksi hoaks berita dengan teknik representasi teks *TF-IDF* dan *preprocessing* seperti *stemming* dan penghapusan *stopword*. Hasilnya menunjukkan bahwa *K-NN* sangat relevan untuk digunakan dalam *text mining* dan klasifikasi konten digital, dengan akurasi yang diperoleh sebesar 93,33% dengan nilai $k = 3$ [20].

Selain itu, *K-NN* dapat digunakan untuk klasifikasi dalam bidang pertanian dan industri. Dengan menggunakan algoritma *K-NN*, Lestari dkk. (2023) dapat mengevaluasi kualitas biji kopi yang kuat dengan menggunakan fitur warna *RGB* dari gambar biji. Penelitian tersebut menunjukkan bahwa dengan pemrosesan data visual yang tepat, *K-NN* dapat memberikan akurasi klasifikasi sekitar 70%, yang cukup baik untuk kebutuhan awal sistem kontrol mutu berbasis citra digital [21].

3. HASIL DAN PEMBAHASAN

3.1. Data Collection

Dataset mouse gaming sebanyak 1.057 baris yang diperoleh dari platform Kaggle digunakan dalam penelitian ini. *Dataset* ini berisi informasi tentang spesifikasi teknis dari berbagai jenis *mouse gaming* yang saat ini tersedia untuk dibeli. Adapun variabel yang dikumpulkan, termasuk yang berkaitan dengan desain ergonomis *mouse*. Keseluruhan fitur ini dipilih karena dinilai dapat memengaruhi kenyamanan dan preferensi pengguna saat memilih *mouse gaming*. Ini juga digunakan sebagai dasar untuk analisis dan klasifikasi lebih lanjut. Tabel 1 menunjukkan contoh data yang digunakan dalam penelitian ini.

Tabel 1. Data Spesifikasi *Mouse Gaming*

Brand	Name	Length (mm)	Width (mm)	Height (mm)	Weight (g)	Shape	Hump Placement	...	Middle Buttons
Rapoo	VT7 Max	127.4	64.3	39.8	53.0	Symmetrical	Center	...	2
Endgame Gear	OP1 8K v2	118.2	60.5	37.2	49.5	Symmetrical	Back - moderate	...	0
Endgame Gear	OP1 4K v2	118.2	60.5	37.2	58.5	Symmetrical	Back – moderate	...	0
Attack Shark	G3	116.5	62.9	38.2	59.0	Symmetrical	Back - moderate	...	0
Valkyrie	VK M1 Pro	119.5	62.0	40.0	58.0	Symmetrical	Center	...	0
...
Zowie	EC1-C	130.0	69.0	42.0	80.0	Ergonomic	Center	...	0

3.2. Data Preprocessing

Sebelum proses pelatihan model, tahap *preprocessing* data sangat penting untuk membuat data sesuai dan ideal untuk dianalisis oleh algoritma pembelajaran mesin. *Dataset* awal terdiri dari 24 fitur yang menunjukkan berbagai aspek dari *mouse gaming*. Fitur-fitur tersebut termasuk dimensi fisik *mouse* (panjang, lebar, tinggi), bobot, posisi *thumb rest*, teknologi sensor, nilai DPI, frekuensi polling, kecepatan pelacakan, dan akselerasi. Namun, proses pemodelan tidak menggunakan semua fitur ini. Oleh karena itu, seleksi fitur dilakukan untuk memilih atribut yang paling relevan untuk sistem rekomendasi yang dibuat.

Beberapa fitur yang dipertahankan adalah panjang (mm), berat, sentuhan jari, bahan, teknologi sensor, posisi sensor, DPI, frekuensi *polling*, kecepatan pengawasan (IPS), dan percepatan (G). Pemilihan ini didasarkan pada manfaat fitur *mouse* untuk kenyamanan, presisi, dan performa saat digunakan untuk bermain game. Selain itu, karena sebagian besar algoritma pembelajaran mesin hanya dapat memproses data numerik, fitur kategorikal seperti teknologi sensor dan material diubah ke bentuk numerik untuk memastikan kompatibilitas dengan algoritma yang digunakan. Berikut *dataset* setelah dilakukan *preprocessing* data ditunjukkan pada Tabel 2.

3.3. PCA

Setelah normalisasi selesai, *PCA* digunakan untuk menurunkan dimensi. Tujuan utama penggunaan *PCA* adalah untuk mengurangi kompleksitas data dengan menggabungkan informasi dari banyak fitur ke dalam dua komponen utama, *PCA1* dan *PCA2*. Ini juga bertujuan untuk membuat hasil *clustering* algoritma *K-Means* lebih mudah dilihat dalam ruang dua dimensi, yang menunjukkan variasi terbesar dalam data. Selain itu, komponen *PCA1* dan *PCA2* digunakan sebagai komponen *input* dalam proses klasifikasi menggunakan algoritma *K-NN*, sehingga struktur data menjadi lebih teliti. Sebagai sampel representatif dari hasil transformasi, Tabel 3 berikut menunjukkan lima data teratas setelah menggunakan proses *PCA*.

Tabel 2. Preprocessing Data

N o	Length (mm)	Weight	Thumb Rest	Material	Sensor Technol- ogy	Sensor Position	DPI	...	Acceleration (G)
1	0.806375	0.039683	0.0	0.000000	0.0	0.273276	0.220299	...	0.0
2	0.128485	0.154531	0.0	0.000000	0.0	-0.273276	0.220299	...	0.0
3	0.270147	-0.373796	0.0	-0.223144	0.0	-0.177164	0.076749	...	0.0
4	-2.391518	-0.599392	0.0	0.000000	0.0	0.565778	0.076749	...	0.0
5	0.688380	1.419206	0.0	0.000000	0.0	0.504626	-1.224045	...	0.0

Tabel 3. Hasil PCA

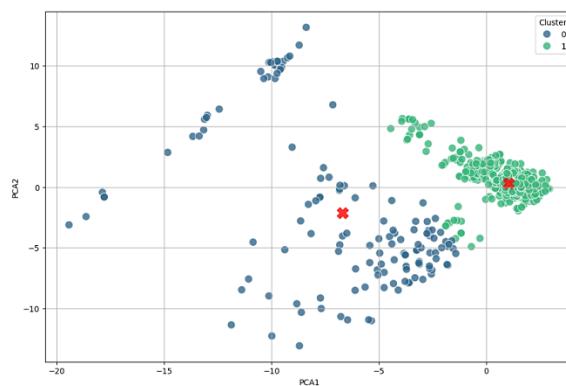
No	PC1	PC2
0	0.424997	0.179110
1	0.421974	0.185718
2	0.416729	0.165157
3	0.481864	0.121558
4	0.099412	-0.050933

3.4. Modeling

3.4.1 K-Means

Untuk melakukan *clustering*, algoritma *K-Means* menggunakan proses normalisasi menggunakan *RobustScaler* untuk mengatasi *outlier*. Selanjutnya, untuk mempertahankan variansi data sebesar 95%, proses *PCA* digunakan untuk mengurangi dimensi. Metode *Elbow* dan Skor *Silhouette* digunakan untuk menentukan jumlah *cluster* terbaik; hasilnya menunjukkan bahwa dua *cluster* adalah pilihan terbaik, dengan skor *Silhouette* 0,524.

Setelah model dijalankan, hasil evaluasi tambahan ditemukan, yang menunjukkan pemisahan *cluster* yang cukup baik. Nilai *Davies-Bouldin* sebesar 1.426, yang menunjukkan peningkatan yang lebih kecil, dan Nilai *Calinski-Harabasz* sebesar 335.118, yang menunjukkan peningkatan yang lebih besar. Gambar 2 menunjukkan dua kelompok data yang sangat berbeda. Titik *centroid* ditunjukkan dengan "X" merah. Menurut analisis distribusi rekomendasi, *Cluster 0* mengandung 28% lebih banyak data yang direkomendasikan daripada *Cluster 1*, yang hanya 8%. Ini menunjukkan kecenderungan karakteristik produk yang berbeda antar *cluster*.



Gambar 2. Visualisasi Clustering

3.4.2 K-NN

Algoritma *K-NN* digunakan untuk menentukan apakah sebuah *mouse* direkomendasikan atau tidak setelah data dibagi dan diseimbangkan. Proses ini menghadapi masalah karena ada ketidakseimbangan jumlah data di antara kelas. Akibatnya, diperlukan penyesuaian untuk memastikan bahwa model tidak bisa

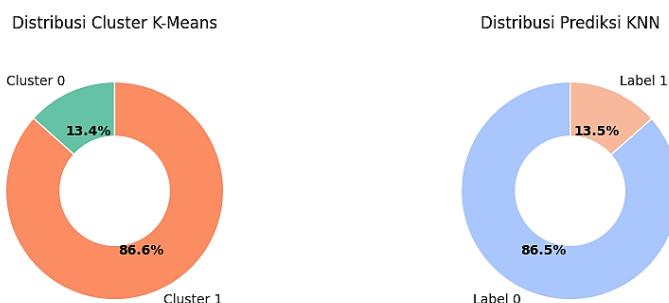
terhadap salah satu kelas. Setelah pelatihan dan pengujian, *K-NN* memberikan hasil yang sangat baik dengan akurasi sebesar 95%. Model juga dapat mengidentifikasi *mouse* yang disarankan dengan cukup baik, yang ditunjukkan oleh tingkat keberhasilan dalam mengklasifikasikan data. Hasil ini menunjukkan bahwa *K-NN* mungkin menjadi algoritma yang dapat diandalkan untuk membantu sistem rekomendasi produk berbasis data.

3.5. Evaluasi

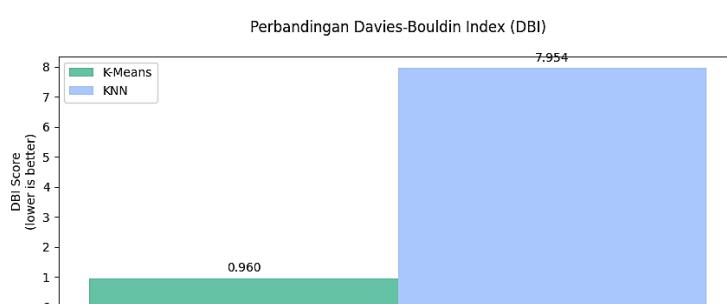
Proses evaluasi dilakukan untuk mengevaluasi kinerja kedua algoritma: *K-Means* untuk *clustering* dan *K-NN* untuk klasifikasi. Gambar 2 menunjukkan visualisasi *clustering*, menunjukkan bahwa hasil *K-Means* menunjukkan kemampuan model untuk memisahkan data ke dalam dua *cluster* yang sangat berbeda. Nilai *Silhouette Score* sebesar 0,524 menunjukkan bahwa pemisahan antar *cluster* cukup jelas, sementara nilai *Davies-Bouldin Index (DBI)* sebesar 0,960, yang menunjukkan kualitas pemisahan *cluster* yang baik (nilai yang lebih rendah menunjukkan nilai yang lebih tinggi). Gambar 3 menunjukkan visualisasi distribusi, yang menunjukkan bahwa *Cluster 1* menguasai 86,6% data, sedangkan *Cluster 0* hanya memiliki 13,4%.

Algoritma *K-NN* sangat baik untuk klasifikasi, dengan akurasi mencapai 95%, dan dapat menemukan kelas minoritas dengan mudah. Produk ini direkomendasikan. Namun, dengan menggunakan metrik *DBI*, model *K-NN* menunjukkan nilai *DBI* sebesar 7,954, Gambar 4 menunjukkan bahwa struktur pemisahan antarkelas tidak jelas seperti model *K-Means* meskipun klasifikasi akurat. Prediksi *K-NN* kurang berbeda dari hasil *clustering K-Means*, yang lebih terorganisir. Visualisasi persebaran pada Gambar 5 menunjukkan perbedaan ini.

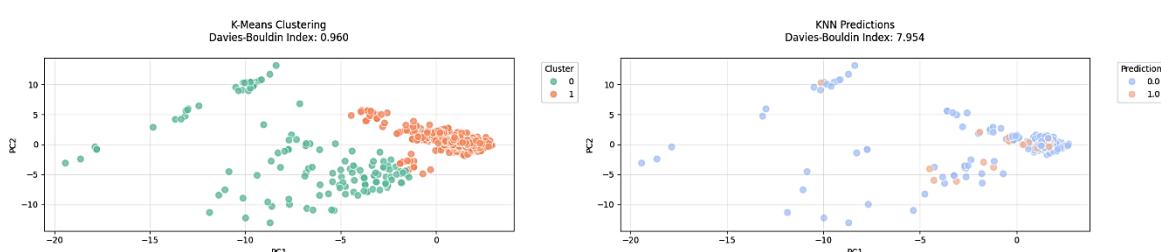
Secara keseluruhan, *K-Means* lebih unggul dalam memetakan struktur data dan menemukan pola tersembunyi, sedangkan *K-NN* lebih unggul dalam memberikan prediksi akhir terhadap data baru. Kombinasi keduanya dapat meningkatkan sistem rekomendasi, dengan *K-Means* memahami karakteristik kelompok produk dan *K-NN* membuat keputusan akhir yang unik untuk setiap produk.



Gambar 3. Distribusi *Cluster K-Means* dan Prediksi *K-NN*



Gambar 4. Perbandingan DBI Algoritma *K-Means* dan *K-NN*



Gambar 5. Plotting Hasil Clustering

4. DISKUSI

Hasil analisis menunjukkan bahwa *K-Means* dapat membentuk kelompok data yang konsisten dan terpisah dengan lebih baik daripada *K-NN* dengan nilai *DBI* sebesar 0,960, *K-Means* jauh lebih baik daripada *K-NN* dengan 7,954, yang menunjukkan kemampuan pemisahan data yang lebih baik. Ini terutama berlaku untuk kelompokan *mouse gaming* berdasarkan spesifikasi teknis seperti bobot, bentuk, dan *DPI*. Namun, *K-NN* unggul dalam prediksi dengan akurasi 95%, tetapi struktur kelasnya tidak jelas. Hasil menunjukkan bahwa *K-Means* memiliki kemampuan yang lebih baik untuk menemukan pola alami dalam data, sedangkan *K-NN* memiliki kemampuan yang lebih baik untuk membuat keputusan klasifikasi. Kombinasi keduanya dapat berfungsi sebagai pendekatan yang saling melengkapi untuk sistem rekomendasi. Namun, faktor-faktor kontekstual seperti preferensi pengguna yang tidak tercakup dalam data teknis harus dipertimbangkan.

5. KESIMPULAN

Dengan menggunakan pendekatan gabungan algoritma *K-Means* dan *K-NN*, penelitian ini berhasil membangun sistem rekomendasi untuk permainan *mouse*. Hasil evaluasi menunjukkan bahwa algoritma *K-Means* dapat mengelompokkan data ke dalam dua *cluster* dengan struktur yang jelas, seperti yang ditunjukkan oleh nilai *DBI* 0,960. Sementara itu, algoritma *K-NN* menunjukkan performa klasifikasi yang tinggi dengan akurasi 95%. Dalam kombinasi kedua algoritma ini, *K-Means* membantu menemukan pola tersembunyi dalam data, sedangkan *K-NN* memberikan keputusan klasifikasi yang akurat. Hasil ini menunjukkan bahwa sistem rekomendasi produk yang efektif dapat dibangun dengan menggunakan pendekatan berbasis spesifikasi teknis.

REFERENSI

- [1] R. A. Zafs *et al.*, “Ergonomic analysis of gaming mouse using electromyography and subjective assessment,” *Journal of Mechatronics, Electrical Power, and Vehicular Technology*, vol. 15, no. 2, pp. 177–185, 2024, doi: 10.55981/j.mev.2024.1110.
- [2] Fahmi Sulaiman, Nana Suarna, and Iin, “Pengukuran Kualitas Perangkat Lunak Sistem Informasi Pengarsipan Dokumen Laporan Jalan Tol Menggunakan Metode Mccall,” *INFOTECH journal*, vol. 8, no. 1, pp. 34–40, Mar. 2022, doi: 10.31949/infotech.v8i1.2234.
- [3] Y. Pan, “The role of recommendation algorithms in driving the development of e-commerce platforms,” *Applied and Computational Engineering*, vol. 92, no. 1, pp. 206–211, Oct. 2024, doi: 10.54254/2755-2721/92/20241754.
- [4] J. Yu and A. Wang, “Adaptive recommendation systems: A comparative analysis of K-NN-based algorithms and hybrid models,” *Applied and Computational Engineering*, vol. 73, no. 1, pp. 24–32, Jul. 2024, doi: 10.54254/2755-2721/73/20240356.
- [5] K. Ramadhan Putra and I. Fathur Rahman, “MIND (Multimedia Artificial Intelligent Networking Database Pemanfaatan Metode Collaborative Filtering dengan Algoritma K-NN pada Sistem Rekomendasi Produk,” *Journal MIND Journal / ISSN*, vol. 9, no. 1, pp. 113–123, 2024, doi: 10.26760/mindjournal.v9i1.113-123.
- [6] M. Mustika Dewi, L. Dwi Farida, and dan Akhmad Dahlan, “Edu Komputika Journal Implementasi K-Means dan Collaborative Filtering untuk Sistem Rekomendasi,” *Edu Komputika*, vol. 10, no. 2, 2023, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/edukom>
- [7] A. N. D. Andra and Z. K. A. Baizal, “E-commerce Recommender System Using PCA and K-Means Clustering,” *Jurnal RESTI*, vol. 6, no. 1, pp. 57–63, Feb. 2022, doi: 10.29207/resti.v6i1.3782.
- [8] H. Mutiasari, T. W. Purboyo, and R. A. Nugrahaeni, “Sistem Rekomendasi Film Menggunakan Metode K-Means Clustering (Movie Recommendation System Using K-Means Clustering Method.”
- [9] A. Bachir, A. Sultan, and S. S. Abu-Naser, “Predictive Modeling of Breast Cancer Diagnosis Using Neural Networks:A Kaggle Dataset Analysis,” 2023. [Online]. Available: www.ijeaais.org/ijaer
- [10] B. Hakim, “Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning,” *JBASE - Journal of Business and Audit Information Systems*, vol. 4, no. 2, Aug. 2021, doi: 10.30813/jbase.v4i2.3000.
- [11] I. T. Jolliffe and J. Cadima, “Principal component analysis: A review and recent developments,” Apr. 13, 2016, *Royal Society of London*. doi: 10.1098/rsta.2015.0202.
- [12] A. K. John, “Data Mining Third Edition.”, Springer, 2007
- [13] A. K. Jain, “Data Clustering: 50 Years Beyond K means Data Clustering: 50 Years Beyond K means 50 Years Beyond K-means 50 Years Beyond K-means.”
- [14] H. Yansah, T. Tri Wulansari, and F. Alameka, “Penerapan Algoritma K-Means Dalam Clustering Produk Terlaris Pada Fr Parfum,” 2022.
- [15] A. P. Pramudiyanti, M. F. Shafiro, L. A. Salim, and Wasyik, “K-Means Cluster Analysis Related To Unmet Need For Family Planning In Banyuwangi, Indonesia: A Case Study,” *Journal of Public Health Research and Community Health Development*, vol. 7, no. 2, pp. 136–142, Mar. 2024, doi: 10.20473/jphrecode.v7i2.39691.

- [16] E. Irfiani, S. Sulistia Rani, S. Nusa Mandiri Jl Kramat Raya No, and J. Pusat, "Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita," vol. 6, no. 4, pp. 17–27, 2018.
- [17] F. Handayani, "Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar," *Jurnal Teknologi dan Informasi*, doi: 10.34010/jati.v12i1.
- [18] D. Abid, R. Wirya Adikusuma, A. Mufti, A. Fikri, D. Rinci, and K. Hapsari, "Penerapan Metode K-Means Clustering Untuk Analisa Penjualan Komoditas Toko Tani Indonesia Article History ABSTRAK," *Jurnal Riset Inovasi Bidang Informatika Dan Pendidikan Informatika (KERNEL)*, vol. 3, no. 2, 2022.
- [19] A. Khairi, A. Fais Ghazali, and A. Darul Nur Hidayah, "Implementasi K-Nearest Neighbor (K-Nn) Untuk Klasifikasi Masyarakat Pra Sejahtera Desa Sapikerep Kecamatan Sukapura," vol. 2, no. 3, pp. 319–323.
- [20] M. Diki Hendriyanto and N. Sari, "Muhammad Diki Hendriyanto Penerapan Algoritma K-Nearest Neighbor Penerapan Algoritma K-Nearest Neighbor dalam Klasifikasi Judul Berita Hoax."
- [21] P. Ayu Lestari, D. Puspita, and S. Aminah, "Penerapan Algoritma K-Nearest Neighbor (K-NN) Pada Klasifikasi Kualitas Biji Kopi Robusta," 2024.