



Komparasi Metode *K-Means* dan *Hierarchical Clustering* untuk Pengelompokan Pola Bermain Pemain *Mobile Legends*

Savior Podung¹, Reynaldo Joshua Salaki²

^{1,2}Program Studi Teknik Informatika, Jurusan Teknik Elektro, Fakultas Teknik, Universitas Sam Ratulangi, Manado, Indonesia

Informasi Artikel

Sejarah Artikel:
Submit: 19 Juni 2026
Revisi: 16 Juli 2026
Diterima: 26 Juli 2026
Diterbitkan: 29 Juli 2026

Kata Kunci

Mobile Legends, Clustering, *K-Means*,
Hierarchical Clustering, Unsupervised Learning

Korespondensi

E-mail: saviorpodung026@student.unsrat.ac.id *

A B S T R A C T

The rapid growth of Mobile Legends: Bang Bang esports produces large volumes of gameplay statistics that remain underutilized beyond individual match summaries. This study compares K-Means and Hierarchical Clustering to group player playstyles using five gameplay variables (kill, death, assist, gold, and match duration) from 990 valid records in the MPL Cambodia Season 6 - BoxMatch dataset, following CRISP-DM preprocessing and Min-Max normalization. The Elbow Method identified K=3 as the optimal number of clusters. K-Means produced three interpretable groups - Carry/Damage Dealer, Late Game Fighter, and Sacrificial/Tank - with a Silhouette Score of 0.252 and a balanced distribution (42.6%, 18.5%, 38.9%), outperforming Hierarchical Clustering with Ward linkage (Silhouette Score 0.2203, less balanced distribution). The results were deployed into an interactive Flask-based web application for dataset upload, clustering visualization, and method comparison, which was validated through black box testing. This research demonstrates that combining clustering with web-based visualization can effectively reveal player playstyle patterns from competitive Mobile Legends gameplay statistics.

Abstrak

Perkembangan ekosistem e-sports Mobile Legends: Bang Bang menghasilkan data statistik gameplay dalam jumlah besar yang pemanfaatannya masih terbatas pada ringkasan hasil pertandingan individual. Penelitian ini membandingkan metode K-Means dan Hierarchical Clustering untuk mengelompokkan pola bermain pemain berdasarkan lima variabel gameplay (kill, death, assist, gold, durasi pertandingan) dari 990 data valid pada dataset MPL Cambodia Season 6 - BoxMatch, melalui preprocessing berbasis CRISP-DM dan normalisasi Min-Max. Elbow Method menentukan K=3 sebagai jumlah cluster optimal. K-Means menghasilkan tiga kelompok pola bermain yang dapat diinterpretasikan – Carry/Damage Dealer, Late Game Fighter, dan Sacrificial/Tank – dengan Silhouette Score 0,252 dan distribusi yang seimbang (42,6%, 18,5%, 38,9%), mengungguli Hierarchical Clustering dengan Ward linkage (Silhouette Score 0,2203, distribusi kurang seimbang). Hasil clustering diimplementasikan ke dalam aplikasi web berbasis Flask yang interaktif untuk unggah dataset, visualisasi clustering, dan perbandingan metode, yang telah divalidasi melalui pengujian black box. Penelitian ini membuktikan bahwa kombinasi clustering dengan visualisasi berbasis web mampu mengungkap pola bermain pemain secara efektif dari data statistik gameplay Mobile Legends yang kompetitif.



Pendahuluan

Perkembangan teknologi digital mendorong pertumbuhan industri game online dari sekadar hiburan menjadi bagian dari industri kreatif dan kompetitif yang dikenal sebagai electronic sports (e-sports) [1]. Salah satu genre yang berkembang pesat adalah Multiplayer Online Battle Arena (MOBA), yang menuntut kerja sama tim, strategi, dan kemampuan individu. Dalam konteks MOBA berbasis mobile, Mobile Legends: Bang Bang (MLBB) merupakan salah satu game dengan basis pemain aktif terbesar di Indonesia dan Asia Tenggara, dengan ekosistem kompetisi profesional seperti Mobile Legends Professional League (MPL) yang diselenggarakan secara rutin di berbagai negara pada tingkat nasional maupun internasional [2].

Sistem permainan MLBB secara otomatis menyediakan profil pemain dan statistik hasil pertandingan yang dapat diakses setelah permainan berakhir [3]. Statistik gameplay yang disajikan mencakup berbagai indikator performa pemain, seperti jumlah *kill*, *death*, *assist*, perolehan *gold*, serta durasi pertandingan. Data tersebut direkam secara otomatis oleh sistem permainan sebagai data numerik terstruktur yang merepresentasikan perilaku dan performa pemain [4]. Penyajian data statistik dalam bentuk antarmuka visual memungkinkan pengguna memahami performa dan pola permainan secara lebih jelas [5], sehingga MLBB sebenarnya telah memiliki sistem pencatatan data statistik yang cukup lengkap dan berpotensi untuk dianalisis lebih lanjut.

Meskipun data statistik gameplay tersedia dalam jumlah besar, pemanfaatannya masih cenderung terbatas pada penyajian hasil pertandingan atau performa individual tanpa analisis lanjutan untuk mengidentifikasi pola bermain secara mendalam [4]. Padahal, analisis terhadap data tersebut berpotensi menghasilkan informasi yang lebih komprehensif, seperti pemahaman terhadap perilaku pemain, kecenderungan strategi, serta perbedaan gaya bermain antar pemain. Untuk mengolah data berukuran besar tersebut, diperlukan teknologi komputasi yang mampu memproses data secara otomatis dan efisien. *Machine learning* dan *deep learning* berperan penting dalam mengekstraksi pola dan karakteristik perilaku dari data kompleks berskala besar [6], termasuk untuk analisis perilaku pemain, prediksi performa, deteksi kecurangan, dan penyeimbangan permainan. Penerapan *deep learning* pada data gameplay berbasis *multivariate time series* terbukti mampu mengidentifikasi pola aktivitas pemain secara akurat [7].

Dalam machine learning, terdapat dua pendekatan utama, yaitu *supervised learning* dan *unsupervised learning*. *Unsupervised learning* sangat sesuai untuk analisis data berskala besar yang tidak memiliki label karena mampu mengungkap pola berdasarkan kemiripan karakteristik data [8]. Data statistik gameplay MLBB umumnya tidak memiliki label kelas yang secara eksplisit menunjukkan pola bermain pemain, sehingga pendekatan ini menjadi relevan untuk digunakan, sejalan dengan temuan bahwa *unsupervised learning* efektif diterapkan pada data perilaku pengguna untuk mengidentifikasi pola tersembunyi tanpa memerlukan data berlabel [9].

Dua metode *clustering* yang umum digunakan dalam analisis data gameplay adalah K-Means dan Hierarchical Clustering. K-Means banyak diterapkan pada analisis data game karena kemudahan implementasi serta efisiensinya dalam mengelompokkan data numerik berskala besar berdasarkan tingkat kemiripan karakteristik [4], termasuk untuk mengelompokkan pemain MLBB berdasarkan statistik permainan dan karakteristik hero [11]. Di sisi lain, Hierarchical Clustering memiliki keunggulan dalam membentuk struktur pengelompokan secara bertingkat tanpa perlu menentukan jumlah cluster di awal, sehingga memungkinkan eksplorasi hubungan dan pola data yang lebih

mendalam [12]. Setiap algoritma clustering memiliki pendekatan pengelompokan yang berbeda sehingga dapat menghasilkan struktur cluster dan tingkat kualitas yang tidak sama pada data yang sama [13], sehingga diperlukan analisis perbandingan untuk menentukan metode yang paling sesuai dan efektif.

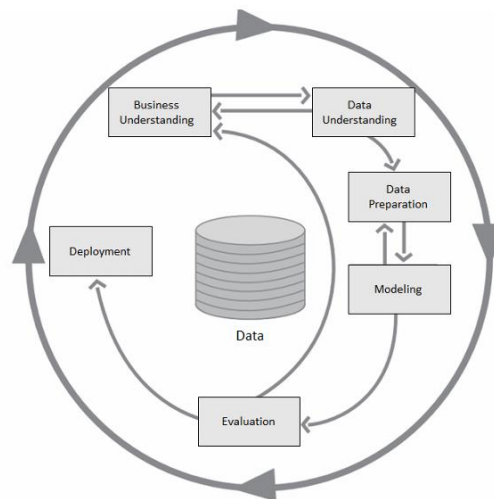
Beberapa penelitian terdahulu telah menerapkan pendekatan clustering pada data perilaku pemain game. Evaluasi metode K-Means dan Hierarchical Clustering pada segmentasi pemain MOBA berbasis pendekatan XGBoost menunjukkan bahwa pemilihan metode clustering perlu disesuaikan dengan karakteristik data, namun masih terkendala ukuran sampel kecil dan distribusi cluster yang tidak seimbang [13]. Penerapan clustering pada data statistik gameplay game mobile mampu mengidentifikasi kelompok pemain dengan karakteristik berbeda, namun mengabaikan aspek pergerakan spasio-temporal pemain [3]. Penerapan K-Means pada data statistik gameplay MLBB untuk mengelompokkan pemain berdasarkan kecenderungan hero dan role menunjukkan hasil yang relevan, tetapi masih bergantung pada data self-report yang subjektif dan terbatas pada metrik akumulasi akhir [11]. Sebagian besar penelitian tersebut hanya menerapkan satu metode clustering tanpa membandingkannya dengan metode lain, serta belum banyak yang mengimplementasikan hasil clustering ke dalam media visualisasi interaktif berbasis web yang dirancang khusus untuk konteks gameplay.

Kesenjangan penelitian pada bidang ini terletak pada tiga aspek. Pertama, sebagian besar penelitian terdahulu hanya menerapkan satu metode clustering tanpa membandingkannya dengan metode lain [3, 11, 13]. Kedua, evaluasi metode yang ada masih banyak menggunakan data pemain kasual atau data self-report yang subjektif, berbeda dengan penelitian ini yang menggunakan data statistik gameplay profesional dari liga kompetitif MPL Cambodia Season 6. Ketiga, belum banyak penelitian yang mengimplementasikan hasil clustering ke dalam media visualisasi interaktif berbasis web yang dirancang khusus untuk konteks gameplay Mobile Legends.

Metode Penelitian

2.1 Tahapan Penelitian

Penelitian ini menggunakan metodologi *CRISP-DM* (Cross-Industry Standard Process for Data Mining) yang terdiri dari enam fase, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* [16]. Tahapan dimulai dari identifikasi masalah dan tujuan penelitian, kajian teori serta kebutuhan sistem, pengumpulan dataset statistik gameplay, eksplorasi dan analisis data (EDA), pra-pemrosesan data, penerapan metode clustering, evaluasi hasil clustering, analisis dan interpretasi hasil, hingga implementasi visualisasi berbasis web. Pendekatan *CRISP-DM* dipilih karena bersifat iteratif dan fleksibel untuk proyek analisis data berskala kecil hingga menengah, sebagaimana digunakan secara luas pada penelitian penerapan metode data mining [17].



Gambar 1. Diagram Alur Metodologi CRISP-DM

2.2 Dataset dan Variabel Penelitian

Dataset yang digunakan adalah data statistik pertandingan Mobile Legends Professional League (MPL) Cambodia Season 6 – BoxMatch, berupa file CSV yang memuat rekaman statistik performa individu pemain pada 100 pertandingan kompetitif tingkat profesional, dengan total 1.010 baris data. Dataset diperoleh dari platform berbagi data terbuka Kaggle (<https://www.kaggle.com/datasets/bcakra/mobile-legend-tournament-match-mpl-cambodia-s6>) dan dipilih karena merepresentasikan permainan kompetitif tingkat tinggi sehingga relevan untuk menganalisis pola performa dan gaya bermain pemain berbasis data statistik gameplay profesional, berbeda dengan sebagian penelitian terdahulu yang menggunakan data self-report atau data pemain kasual. Lima variabel utama digunakan sebagai fitur numerik dalam proses clustering, yaitu Kill, Death, Assist, Gold, dan Time (durasi pertandingan), sebagaimana dirangkum pada Tabel 1.

Tabel 1. Variabel Utama Penelitian

Variabel	Penjelasan	Bentuk Data
Kill	Jumlah lawan yang berhasil dieliminasi pemain	Total per match (integer)
Death	Jumlah kematian pemain selama pertandingan	Total per match (integer)
Assist	Jumlah kontribusi membantu eliminasi lawan	Total per match (integer)
Gold	Total gold dari farming, kill, assist, dan objektif	Total per match (numerik)
Time	Lama waktu pertandingan berlangsung	Menit (hasil konversi MM:SS)

Sumber: Data Penelitian

Seluruh variabel bersifat numerik dan tidak memiliki label kelas yang secara eksplisit menunjukkan kategori atau gaya bermain pemain, sehingga dataset dianalisis menggunakan pendekatan *unsupervised learning*, khususnya metode clustering, untuk mengelompokkan pemain berdasarkan kemiripan karakteristik statistik gameplay yang dimiliki.

2.3 Pra-Pemrosesan dan Normalisasi Data

Tahap *Data Preparation* mencakup dua proses utama. Pertama, konversi format kolom Time dari teks MM:SS menjadi satuan menit desimal menggunakan persamaan (1), agar dapat diproses dalam perhitungan jarak Euclidean pada algoritma clustering berbasis jarak.

$$\text{Timemenit} = \text{menit} + \text{detik} / 60 \quad (1)$$

Kedua, penghapusan baris data tidak valid. Hasil eksplorasi data (EDA) menemukan 20 baris dengan nilai Gold = 0 yang secara logika gameplay tidak valid, karena mekanisme *passive gold* berjalan otomatis sepanjang pertandingan sehingga nilai Gold = 0 tidak mungkin terjadi pada

pertandingan yang tercatat sempurna. Baris-baris tersebut dihapus sehingga dataset akhir terdiri dari 990 baris data valid (98,02% dari total data awal). Selanjutnya, seluruh variabel dinormalisasi menggunakan *MinMaxScaler* dari pustaka Scikit-learn ke rentang 0–1, untuk mencegah variabel Gold (rentang 362–117.666) mendominasi perhitungan jarak Euclidean dibandingkan variabel lain dengan rentang yang jauh lebih kecil [18].

Tabel 2. Data Sebelum dan Sesudah Pra-Pemrosesan (Baris Pertama)

Variabel	Sebelum (Raw)	Sesudah Konversi Time	Sesudah Normalisasi (0–1)
Kill	6	6	0,50
Death	4	4	0,50
Assist	8	8	0,36
Gold	13.798	13.798	0,1145
Time	24:20	24,33	0,63

Sumber: Hasil Pengolahan Data

2.4 Penerapan Metode Clustering

Penentuan jumlah cluster optimal (K) dilakukan menggunakan Elbow Method dengan menghitung nilai Within-Cluster Sum of Squares (WCSS) untuk K=2 hingga K=10, kemudian mengidentifikasi titik elbow sebagai acuan jumlah cluster paling optimal [19]. Rentang K=2 hingga K=10 dipilih pada tahap Elbow Method agar pola penurunan WCSS dapat diamati secara menyeluruh sebelum kurva melandai. Setelah titik elbow teridentifikasi di sekitar K=3, evaluasi perbandingan Silhouette Score selanjutnya difokuskan pada rentang K=2 hingga K=6 agar analisis tetap relevan dengan jumlah cluster yang realistis untuk diinterpretasikan sebagai pola bermain, karena nilai K yang terlalu besar akan menghasilkan cluster yang sulit dimaknai secara praktis. Algoritma K-Means diterapkan dengan parameter *random_state* = 42 dan *n_init* = 10 untuk memastikan konsistensi hasil [10]. Algoritma Hierarchical Clustering diterapkan menggunakan pendekatan agglomerative dengan metode linkage Ward, Complete, dan Average untuk perbandingan, di mana Ward dipilih sebagai metode utama karena meminimalkan varians di dalam setiap cluster [12].

2.5 Evaluasi Clustering

Evaluasi kualitas hasil clustering dilakukan menggunakan Silhouette Score sebagai metrik utama, yaitu metrik yang mengukur seberapa baik suatu data tergabung dalam clusternya sendiri dibandingkan dengan cluster lain, dengan nilai berkisar antara -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan struktur cluster yang lebih jelas [20]. Sebagai metrik pendukung, evaluasi juga mempertimbangkan Davies-Bouldin Index (semakin rendah semakin baik) dan Calinski-Harabasz Index (semakin tinggi semakin baik) agar penilaian kualitas cluster lebih komprehensif. Perbandingan dilakukan pada seluruh kombinasi metode clustering dan nilai K dari 2 hingga 6, dengan mempertimbangkan pula keseimbangan distribusi data antar cluster sebagai indikator kebermaknaan hasil pengelompokan.

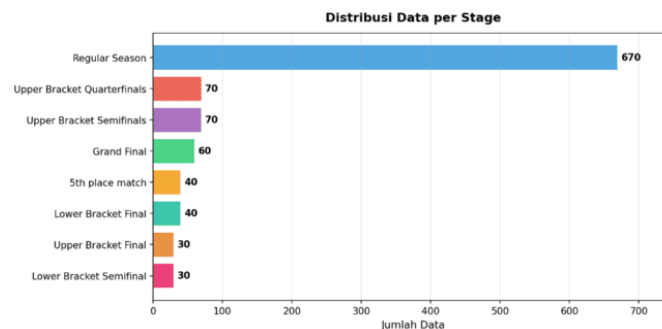
2.6 Implementasi Aplikasi Web

Hasil clustering diimplementasikan ke dalam aplikasi web menggunakan *framework Flask* (Python) untuk backend, serta HTML, CSS, dan JavaScript untuk frontend. Aplikasi terdiri dari enam halaman utama, yaitu *Dashboard*, *Upload Dataset*, *Elbow Method*, *K-Means*, *Hierarchical*, dan *Perbandingan Metode*, yang mengintegrasikan seluruh tahapan penelitian mulai dari unggah dataset, pra-pemrosesan otomatis, proses clustering interaktif, hingga visualisasi dan evaluasi hasil. Implementasi ini merepresentasikan fase *Deployment* pada metodologi CRISP-DM.

Hasil dan Pembahasan

3.1 Hasil Pengumpulan dan Eksplorasi Data

Dataset MPL Cambodia Season 6 – BoxMatch terdiri dari 1.010 baris dan 27 kolom, mencakup delapan tahap kompetisi dari *Regular Season* hingga *Grand Final*. Data terbanyak berasal dari tahap *Regular Season* yaitu 670 data (66,3%), sebagaimana disajikan pada Gambar 1, karena pada tahap ini jumlah pertandingan yang dimainkan jauh lebih banyak dibandingkan tahap *playoff*.



Gambar 2. Distribusi Data per Stage Kompetisi

Dari 27 kolom yang tersedia, hanya lima variabel numerik (Kill, Death, Assist, Gold, Time) yang digunakan sebagai fitur utama clustering. Hasil pengecekan *missing values* menunjukkan tidak ada nilai kosong pada kelima variabel tersebut, namun ditemukan 20 baris dengan nilai Gold = 0 yang secara logika gameplay tidak valid. Statistik deskriptif kelima variabel utama disajikan pada Tabel 2.

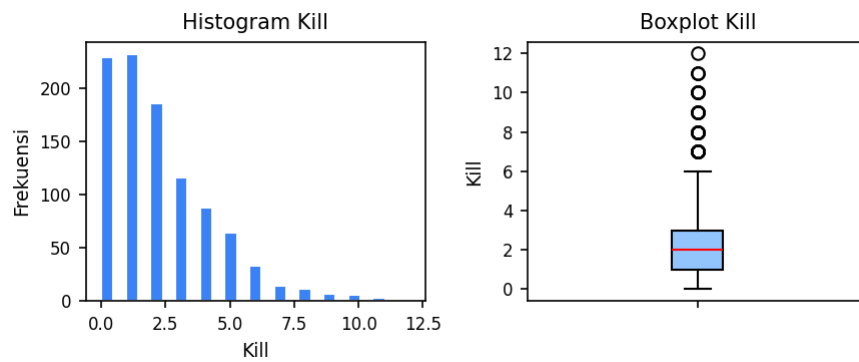
Tabel 2. Statistik Deskriptif Variabel Utama

Statistik	Kill	Death	Assist	Gold
Mean	2,19	2,19	5,54	9.676,14
Std	2,14	1,66	3,71	4.883,91
Min	0	0	0	0
Median	2	2	5	9.113,50
Max	12	8	22	117.666

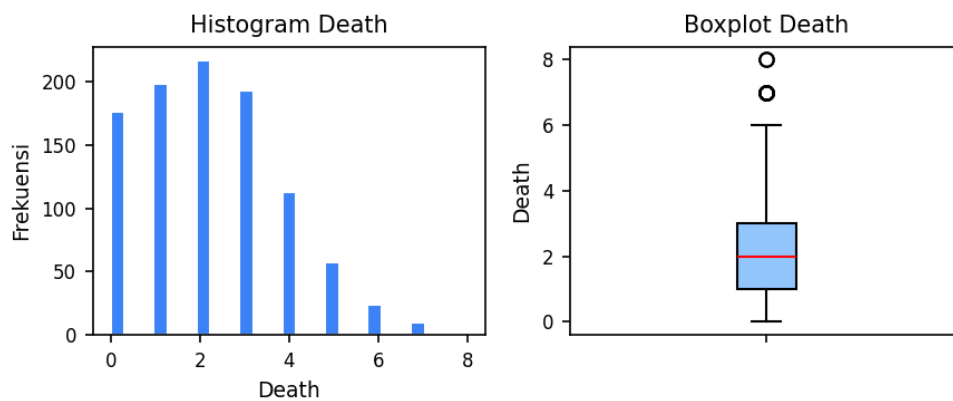
Sumber: Hasil Pengolahan Data

Berdasarkan Tabel 2, Kill dan Death memiliki rata-rata yang hampir sama (2,19), menunjukkan keseimbangan antara eliminasi dan kematian secara keseluruhan, sedangkan Assist memiliki rata-rata lebih tinggi (5,54), menandakan kontribusi assist yang besar dalam pertandingan kompetitif. Variabel Gold memiliki rentang nilai yang sangat lebar (0-117.666) dengan standar deviasi 4.883,91, jauh lebih besar dibandingkan variabel lainnya. Perbedaan skala ini menegaskan perlunya normalisasi data sebelum clustering berbasis jarak dijalankan, karena variabel dengan rentang nilai besar dapat mendominasi perhitungan jarak Euclidean dan menghasilkan cluster yang bias [18].

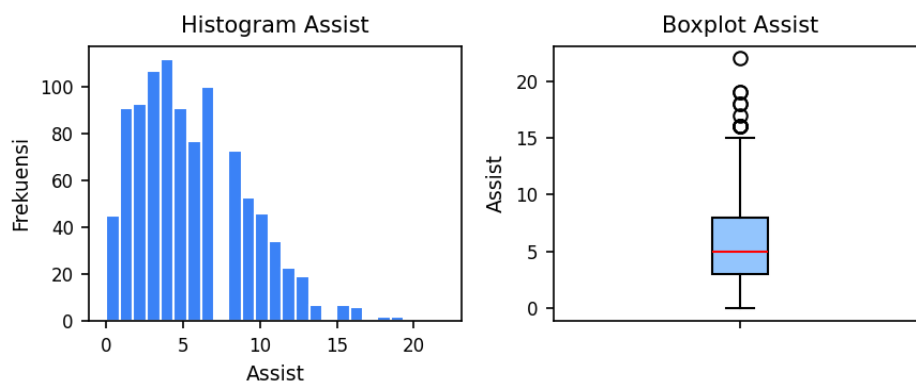
Untuk melihat sebaran data dan mendeteksi keberadaan outlier pada tiap variabel sebelum proses clustering dijalankan, Gambar 1a hingga Gambar 1e menyajikan histogram dan boxplot untuk masing-masing dari kelima variabel utama.



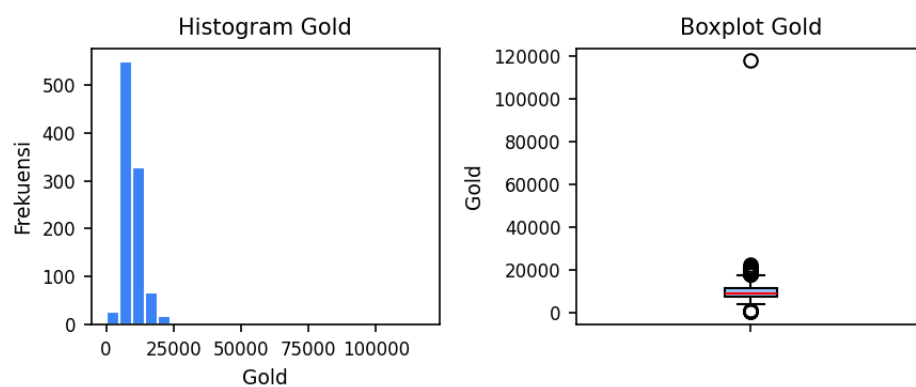
Gambar 3. Histogram dan Boxplot Variabel Kill



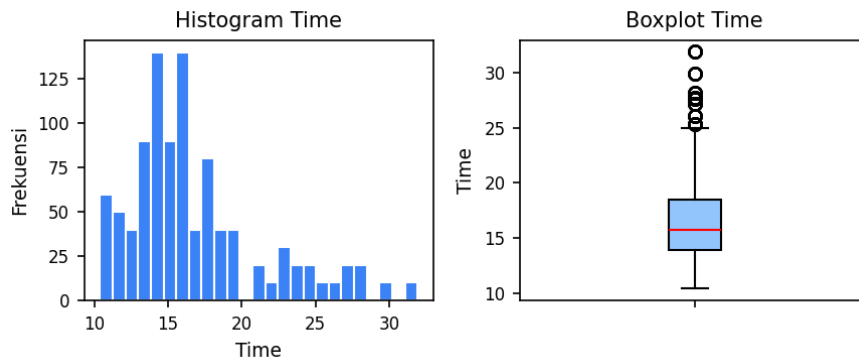
Gambar 4. Histogram dan Boxplot Variabel Death



Gambar 5. Histogram dan Boxplot Variabel Assist



Gambar 6. Histogram dan Boxplot Variabel Gold



Gambar 7. Histogram dan Boxplot Variabel Time

Histogram pada Gambar 1a, 1b, dan 1e (Kill, Death, dan Time) memperlihatkan pola right-skewed dengan mayoritas nilai rendah dan ekor panjang ke nilai tinggi, Gambar 1c (Assist) mendekati distribusi normal dengan sedikit kemiringan, sedangkan Gambar 1d (Gold) sangat dipengaruhi oleh nilai ekstrem sehingga sebagian besar data tampak menumpuk pada rentang nilai rendah.

Tabel 3. Ringkasan Deteksi Outlier Berbasis IQR

Variabel	Q1	Q3	Jumlah Outlier	% dari Data	Skewness
Kill	1	3	42	4,2%	1,33
Death	1	3	12	1,2%	0,58
Assist	3	8	12	1,2%	0,73
Gold	7.559	11.474	37	3,7%	12,22
Time	13,87	18,43	80	8,1%	1,19

Sumber: Hasil Pengolahan Data

Tabel di atas menunjukkan bahwa variabel Gold memiliki tingkat skewness yang sangat ekstrem (12,22) akibat satu nilai anomali sebesar 117.666, jauh melampaui nilai tertinggi kedua (22.481), atau hampir lima kali lipatnya. Nilai tersebut kemungkinan besar merupakan kesalahan pencatatan data dan bukan performa gameplay yang wajar, namun tetap dipertahankan dalam dataset karena tidak melanggar aturan validasi $\text{Gold} > 0$ yang diterapkan pada tahap pra-pemrosesan. Variabel Time juga memiliki proporsi outlier yang cukup besar (8,1%), yang mengindikasikan adanya pertandingan dengan durasi sangat singkat maupun sangat panjang dibandingkan mayoritas data. Temuan ini memperkuat urgensi normalisasi MinMaxScaler yang diterapkan pada tahap Data Preparation, karena tanpa normalisasi, nilai ekstrem pada variabel Gold berpotensi mendominasi perhitungan jarak Euclidean dan menghasilkan cluster yang bias.

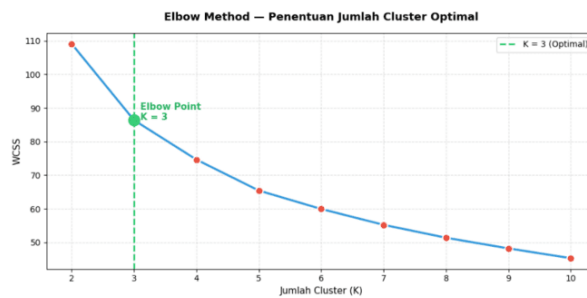
3.2 Hasil Pra-Pemrosesan dan Normalisasi

Setelah konversi format kolom Time ke satuan menit desimal dan penghapusan 20 baris data dengan $\text{Gold} = 0$, dataset akhir terdiri dari 990 baris data valid (98,02%). Seluruh variabel kemudian dinormalisasi menggunakan *MinMaxScaler* ke rentang 0–1. Sebagai contoh, nilai Gold pada baris pertama yang semula 13.798 menjadi 0,1145 setelah normalisasi, sejalan dengan nilai Kill, Death, Assist, dan Time yang juga berada pada rentang 0–1. Hasil ini memastikan setiap variabel memberikan kontribusi yang proporsional dalam pembentukan cluster pada tahap *Modeling* selanjutnya.

3.3 Hasil Elbow Method

Perhitungan nilai WCSS dilakukan untuk $K=2$ hingga $K=10$ pada data yang telah dinormalisasi. Penurunan WCSS paling signifikan terjadi dari $K=2$ ke $K=3$ dengan selisih 22,7330, kemudian melambat secara konsisten pada nilai K berikutnya ($K=4$ hanya 11,7433; $K=10$ hanya 2,9071). Pola ini,

yang divisualisasikan pada Gambar 2, mengindikasikan bahwa penambahan cluster di atas K=3 tidak memberikan peningkatan kualitas pengelompokan yang berarti.



Gambar 8. Grafik Elbow Method

Berdasarkan titik elbow pada K=3 dan interpretasi yang bermakna terhadap tiga pola bermain yang dapat dibedakan secara jelas, jumlah cluster optimal yang ditetapkan dalam penelitian ini adalah K = 3.

Tabel 4. Nilai WCSS untuk Setiap Jumlah Cluster (K)

K	2	3	4	5	6	7	8	9	10
WCSS	109,1269	86,3939	74,6506	65,4387	59,9564	55,2373	51,3935	48,2203	45,3132

Sumber: Hasil Pengolahan Data

3.4 Hasil K-Means Clustering

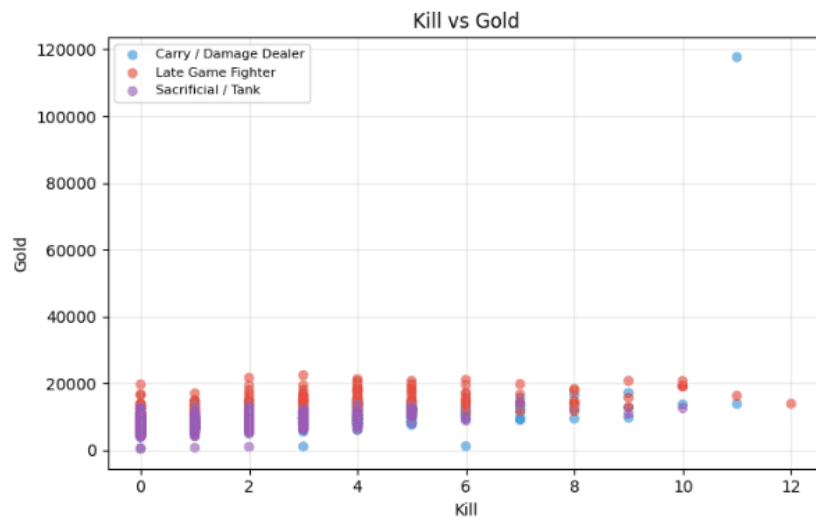
Penerapan K-Means dengan K=3, random_state = 42, dan n_init = 10 menghasilkan Silhouette Score sebesar 0,252, yang menunjukkan bahwa struktur cluster mulai terbentuk, tetapi pemisahnya belum kuat. Ketiga cluster tetap diberi label berdasarkan karakteristik dominannya, yaitu Carry/Damage Dealer, Late Game Fighter, dan Sacrificial/Tank, sebagaimana dirangkum pada Tabel 3.

Tabel 5. Ringkasan Hasil K-Means Clustering (K=3)

Cluster	Kill	Death	Assist	Gold	n
0 – Carry/DD	2,61	0,82	6,03	9.659,37	422
1 – Late Fighter	3,28	2,89	7,83	14.365,71	183
2 – Tank	1,23	3,39	3,94	7.968,11	385

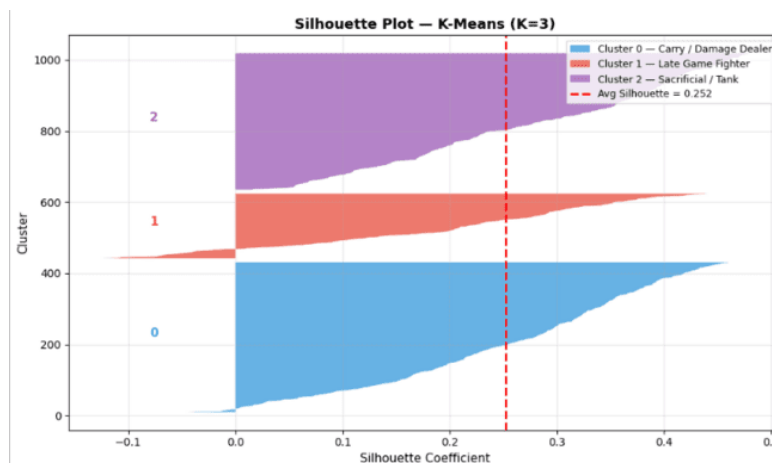
Sumber: Hasil Pengolahan Data

Cluster 0 (Carry/Damage Dealer, 422 pemain atau 42,6%) dicirikan oleh tingkat kematian terendah (0,82 death/match) dan durasi pertandingan terpendek (14,81 menit), merepresentasikan pemain yang mendominasi pertandingan secara efisien sejak awal. Cluster 1 (Late Game Fighter, 183 pemain atau 18,5%) memiliki nilai Kill, Assist, dan Gold tertinggi namun durasi pertandingan terpanjang (24,88 menit), merepresentasikan pemain yang berkontribusi besar dalam pertandingan yang berlangsung lama. Cluster 2 (Sacrificial/Tank, 385 pemain atau 38,9%) memiliki tingkat kematian tertinggi (3,39 death/match) dengan Kill dan Gold terendah, merepresentasikan peran defensif yang bertugas menarik perhatian lawan sebagai tameng tim.



Gambar 9. Scatter Plot K-Means (Kill vs Gold)

Gambar 10 memperlihatkan bahwa ketiga cluster memiliki distribusi yang cukup terpisah, terutama pada sumbu Gold, dengan Cluster 1 berada pada posisi Gold tertinggi dan Cluster 2 pada posisi terendah. Kualitas pemisahan masing-masing cluster divisualisasikan pada *silhouette plot* Gambar 4, dengan nilai silhouette rata-rata Cluster 0 sebesar 0,2543, Cluster 1 sebesar 0,1929, dan Cluster 2 sebesar 0,2775.



Gambar 10. Silhouette Plot K-Means

3.5 Hasil Hierarchical Clustering

Penerapan Hierarchical Clustering dengan K=3 dan *linkage* Ward menghasilkan *Silhouette Score* sebesar 0,2203. Tabel 4 merangkum karakteristik ketiga cluster yang terbentuk.

Tabel 6. Ringkasan Hasil Hierarchical Clustering (K=3, Ward)

Cluster	Kill	Death	Assist	Gold	n
0	2,22	1,30	4,86	9.054,58	589
1	1,64	3,81	5,92	9.148,19	264
2	3,18	2,97	7,80	14.778,35	137

Sumber: Hasil Pengolahan Data

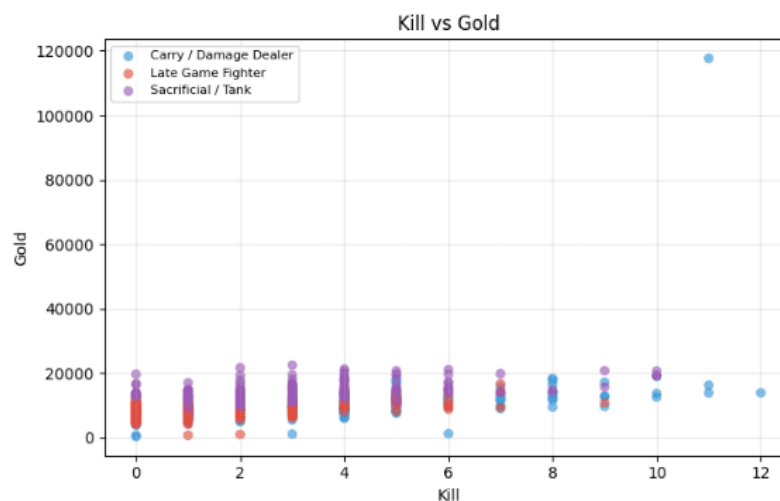
Cluster 0 merupakan kelompok terbesar (589 pemain, 59,5%) dengan Death terendah (1,30) dan durasi terpendek (14,57 menit), mencerminkan gaya bermain efisien. Cluster 1 (264 pemain, 26,7%) memiliki Death tertinggi (3,81) dengan Kill terendah (1,64), mencerminkan peran defensif dengan kontribusi ofensif rendah. Cluster 2 merupakan kelompok terkecil (137 pemain, 13,8%) dengan Kill,

Assist, dan Gold tertinggi serta durasi terpanjang (25,80 menit), mencerminkan performa statistik terbaik secara keseluruhan.

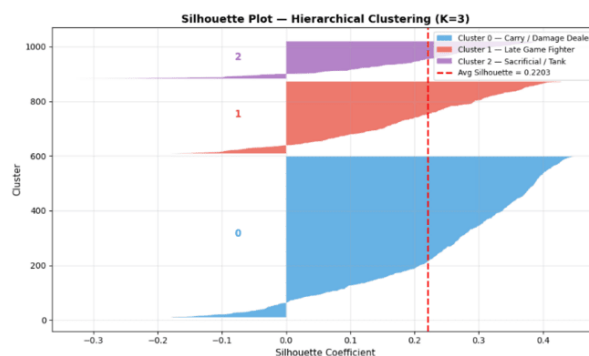


Gambar 11. Dendrogram Hierarchical Clustering (Ward, Sampel 200 Data)

Dendrogram pada Gambar 10, yang dibuat menggunakan sampel acak 200 data karena keterbatasan komputasi untuk memvisualisasikan seluruh 990 data, bersifat representatif dan bukan merupakan gambaran penuh dari seluruh proses penggabungan cluster pada keseluruhan data. Visualisasi ini memperlihatkan struktur pengelompokan bertingkat dari proses agglomerative clustering. Garis merah horizontal menunjukkan threshold pemotongan yang menghasilkan tiga cluster, sedangkan jarak vertikal antar penggabungan merepresentasikan tingkat ketidakmiripan antar cluster.



Gambar 11. Scatter Plot Hierarchical Clustering (Kill vs Gold)



Gambar 12. Silhouette Plot Hierarchical Clustering

Berdasarkan Gambar 12, nilai silhouette rata-rata Cluster 0 sebesar 0,2430, Cluster 1 sebesar 0,1890, dan Cluster 2 sebesar 0,1831. Cluster 0 memiliki nilai silhouette tertinggi yang mengindikasikan pemisahan paling jelas di antara ketiga cluster yang terbentuk. Perlu dicatat bahwa distribusi cluster pada Hierarchical Clustering (59,5%, 26,7%, 13,8%) kurang seimbang dibandingkan

K-Means, sehingga dalam penelitian ini Hierarchical Clustering berperan sebagai metode pembanding untuk memperkuat validitas analisis, bukan sebagai metode utama dalam interpretasi pola bermain.

3.6 Perbandingan K-Means dan Hierarchical Clustering

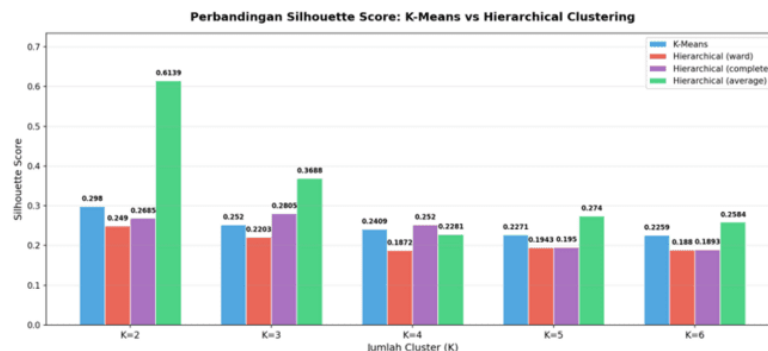
Perbandingan dilakukan dengan menghitung *Silhouette Score* untuk seluruh kombinasi metode dan nilai K dari 2 hingga 6, sebagaimana dirangkum pada Tabel 5.

Tabel 7. Perbandingan Silhouette Score K-Means dan Hierarchical Clustering (K=2-6)

K	K-Means	Ward	Complete	Average
2	0,2980	0,2490	0,2685	0,6139
3	0,2520	0,2203	0,2805	0,3688
4	0,2409	0,1872	0,2520	0,2281
5	0,2271	0,1943	0,1950	0,2740
6	0,2259	0,1880	0,1893	0,2584

Sumber: Hasil Pengolahan Data

Nilai *Silhouette Score* tertinggi secara keseluruhan diperoleh Hierarchical *Average Linkage* pada K=2 (0,6139). Namun, pemeriksaan distribusi data menunjukkan bahwa nilai tinggi tersebut tidak merepresentasikan pengelompokan yang bermakna, karena *Average Linkage* menghasilkan distribusi 989 berbanding 1 data pada K=2, dan 974, 1, 15 data pada K=3. Pola ini merupakan ciri umum *chaining effect* pada *Average Linkage*, yang cenderung membentuk satu cluster besar dan memisahkan beberapa *outlier* ke dalam cluster yang sangat kecil [12]. Pola serupa, meskipun tidak seekstrem itu, juga teramati pada *Complete Linkage* dengan distribusi 781, 179, dan 30 data pada K=3.



Gambar 13. Grafik Perbandingan Silhouette Score Antarmetode

Pada metode dengan distribusi cluster yang seimbang dan dapat diinterpretasikan secara bermakna, yaitu K-Means dan Hierarchical Ward Linkage, K-Means menghasilkan *Silhouette Score* lebih tinggi (0,2520) dibandingkan Hierarchical Ward Linkage (0,2203) pada K=3, dengan distribusi yang lebih proporsional (42,6%, 18,5%, 38,9% berbanding 59,5%, 26,7%, 13,8%). Sebagaimana terlihat pada Gambar 8, K-Means juga menunjukkan nilai *Silhouette Score* yang relatif stabil pada seluruh nilai K yang diuji, sedangkan Hierarchical Clustering cenderung tidak stabil dan rentan terhadap distribusi cluster yang tidak seimbang, khususnya pada Complete dan Average Linkage. Tabel 7 merangkum distribusi jumlah data pada setiap cluster untuk seluruh metode dan nilai K, sebagai dasar verifikasi bahwa pemilihan metode terbaik tidak hanya didasarkan pada skor *Silhouette* tertinggi, tetapi juga pada interpretabilitas dan keseimbangan distribusi cluster yang dihasilkan. Dengan demikian, K-Means Clustering pada K=3 ditetapkan sebagai hasil clustering utama yang digunakan sebagai dasar interpretasi pola bermain pemain Mobile Legends dalam penelitian ini, sedangkan hasil Hierarchical Clustering digunakan sebagai pembanding untuk memperkuat validitas analisis, bukan sebagai metode utama.

Tabel 8. Distribusi Jumlah Data per Cluster untuk Setiap Metode dan Nilai K

K	Metode	Distribusi Jumlah Data per Cluster (n)
2	K-Means	672 / 318
2	Ward	401 / 589
2	Complete	960 / 30
2	Average	989 / 1
3	K-Means	422 / 183 / 385
3	Ward	589 / 264 / 137
3	Complete	179 / 30 / 781
3	Average	974 / 1 / 15
4	K-Means	390 / 155 / 277 / 168
4	Ward	274 / 264 / 137 / 315
4	Complete	781 / 30 / 55 / 124
4	Average	955 / 19 / 15 / 1
5	K-Means	258 / 153 / 204 / 235 / 140
5	Ward	264 / 315 / 137 / 215 / 59
5	Complete	30 / 124 / 55 / 624 / 157
5	Average	170 / 785 / 15 / 1 / 19
6	K-Means	204 / 76 / 232 / 94 / 131 / 253
6	Ward	137 / 315 / 181 / 215 / 59 / 83
6	Complete	55 / 124 / 29 / 624 / 157 / 1
6	Average	157 / 785 / 13 / 1 / 19 / 15

Sumber: Hasil Pengolahan Data

Tabel 7 memperlihatkan bahwa Complete dan Average Linkage secara konsisten menghasilkan distribusi yang sangat timpang pada hampir seluruh nilai K, sehingga meskipun beberapa kombinasi menghasilkan Silhouette Score yang tinggi, hasil tersebut tidak bermakna secara praktis untuk interpretasi pola bermain.

Sebagai metrik pendukung, perhitungan pada K=3 menunjukkan K-Means memperoleh Davies-Bouldin Index sebesar 1,3291 dan Calinski-Harabasz Index sebesar 359,31, sedangkan Hierarchical Ward Linkage memperoleh Davies-Bouldin Index sebesar 1,5521 dan Calinski-Harabasz Index sebesar 277,00. Nilai Davies-Bouldin yang lebih rendah dan Calinski-Harabasz yang lebih tinggi pada K-Means mengonfirmasi bahwa metode ini menghasilkan pemisahan cluster yang secara konsisten lebih baik dibandingkan Hierarchical Ward Linkage.

3.7 Implementasi Aplikasi Web

Hasil clustering diimplementasikan ke dalam aplikasi web berbasis *framework Flask* yang terdiri dari enam halaman utama. Halaman *Dashboard* menyajikan ringkasan dataset dan statistik deskriptif beserta tombol pemuatan dataset secara otomatis. Halaman *Upload Dataset* memungkinkan pengguna mengunggah file CSV melalui klik atau *drag and drop*, yang kemudian diproses secara otomatis pada tahap pra-pemrosesan dan normalisasi di *backend*. Halaman *Elbow Method* menampilkan grafik WCSS beserta tabel nilai untuk setiap K. Halaman K-Means dan Hierarchical masing-masing menampilkan *Silhouette Score*, tabel ringkasan cluster, *scatter plot*, *silhouette plot*, distribusi pemain per cluster, indikator performa terbaik, tabel perbandingan antar cluster, serta rekomendasi gaya bermain yang dihasilkan secara otomatis, dengan tambahan dendrogram pada halaman Hierarchical. Halaman Perbandingan Metode menampilkan grafik batang dan tabel lengkap perbandingan *Silhouette Score* seluruh kombinasi metode dan nilai K, disertai rekomendasi metode terbaik secara otomatis.

Untuk memvalidasi bahwa seluruh fitur berjalan sesuai fungsinya, dilakukan pengujian black box terhadap aplikasi web dengan 15 skenario pengujian yang mencakup jalur normal maupun penanganan input tidak valid, sebagaimana dirangkum pada Tabel 6.

Tabel 10. Hasil Pengujian Black Box Aplikasi Web

ID Uji	Fitur	Skenario Pengujian	Hasil Aktual	Status
BB-01	Load Default Dataset	Memuat dataset bawaan sistem	Berhasil, 990 baris data termuat dengan kolom sesuai	Valid
BB-02	Upload Dataset	Mengunggah CSV dengan format kolom sesuai	Berhasil, pratinjau data dan statistik deskriptif tampil	Valid
BB-03	Upload Dataset	Mengunggah tanpa memilih berkas	Pesan "Tidak ada file yang diunggah." ditampilkan	Valid
BB-04	Upload Dataset	Mengunggah CSV dengan kolom tidak sesuai	Pesan "Kolom 'Kill' tidak ditemukan di dataset." ditampilkan	Valid
BB-05	Upload Dataset	Mengunggah berkas CSV kosong	Kesalahan tertangani, pesan teknis pustaka ditampilkan	Valid dengan catatan
BB-06	Elbow Method	Menjalankan tanpa dataset dimuat	Pesan "Upload dataset terlebih dahulu." ditampilkan	Valid
BB-07	Elbow Method	Menjalankan dengan dataset valid	Grafik WCSS K=2-10 tampil, titik elbow di K=3	Valid
BB-08	K-Means Clustering	Menjalankan tanpa dataset dimuat	Pesan peringatan ditampilkan	Valid
BB-09	K-Means Clustering	Menjalankan dengan K=3 (skenario utama)	Silhouette Score 0,252; distribusi 422/183/385 sesuai hasil penelitian	Valid
BB-10	K-Means Clustering	Menjalankan dengan K=1 (nilai tidak valid)	Kesalahan tertangani tanpa aplikasi berhenti berfungsi	Valid dengan catatan
BB-11	K-Means Clustering	Menjalankan dengan K besar (K=10)	Berhasil, Silhouette Score 0,225	Valid
BB-12	K-Means Clustering	Menjalankan tanpa parameter K	Berhasil, sistem otomatis menggunakan K=3	Valid
BB-13	Hierarchical Clustering	Menjalankan Ward Linkage, K=3	Silhouette Score 0,2203; distribusi 589/264/137 sesuai hasil penelitian	Valid
BB-14	Hierarchical Clustering	Menjalankan dengan linkage tidak valid	Kesalahan tertangani tanpa aplikasi berhenti berfungsi	Valid dengan catatan
BB-15 <i>Sumber: Hasil</i>	Perbandingan Metode	Menjalankan perbandingan	Rekomendasi K-Means, K=3,	Valid

Pengujian Aplikasi		seluruh metode dan K	Silhouette 0,252	Score	
--------------------	--	----------------------	---------------------	-------	--

Kesimpulan

Penelitian ini berhasil menerapkan metode K-Means dan Hierarchical Clustering untuk mengelompokkan pola bermain pemain Mobile Legends berdasarkan variabel Kill, Death, Assist, Gold, dan durasi pertandingan pada 990 data MPL Cambodia Season 6. Berdasarkan Elbow Method, jumlah cluster optimal adalah tiga. K-Means menghasilkan tiga pola bermain, yaitu Carry/Damage Dealer, Late Game Fighter, dan Sacrificial/Tank, dengan Silhouette Score 0,2520 serta distribusi cluster yang lebih seimbang dibandingkan Hierarchical Clustering yang memperoleh skor 0,2203. Evaluasi menggunakan Davies-Bouldin Index dan Calinski-Harabasz Index juga menunjukkan bahwa K-Means lebih stabil dan proporsional sehingga dipilih sebagai metode utama. Hasil clustering telah diterapkan pada aplikasi web berbasis Flask dan seluruh fiturnya berjalan dengan baik berdasarkan pengujian black box. Meskipun demikian, penelitian masih terbatas pada satu liga dan label cluster belum divalidasi menggunakan data role atau hero. Penelitian selanjutnya disarankan menggunakan dataset yang lebih luas serta mengembangkan hasil clustering menjadi sistem rekomendasi strategi dan pembentukan tim e-sports.

Daftar Pustaka

- [1] W. K. Chap, A. C. Rao, and P. K. Pandey, "A Technological Review on Rise of Esports in World Economy," in ACM Int. Conf. Proc. Series, Dec. 2022. doi: 10.1145/3590837.3590916.
- [2] F. Arianto, H. M. Az-Zahra, and B. S. Prakoso, "User Experience Analysis on Mobile-Based MOBA Games Using UX Curve Method," J. Games, Game Art, Gamification, vol. 9, no. 2, pp. 43-51, 2024. doi: 10.21512/jggag.v9i2.10372.
- [3] Y. Qiu, Y. Gong, and G. Liu, "User Behavior Analysis and Clustering in a MMO Mobile Game: Insights and Recommendations," arXiv preprint, pp. 1-16, 2024. [Online]. Available: <http://arxiv.org/abs/2407.11772>
- [4] S. Yu, M. Yang, L. H. Wei, J. S. Hu, H. W. Tseng, and T. H. Meen, "Combination of self-organizing map and k-means methods of clustering for online games marketing," Sensors Mater., vol. 32, no. 8, pp. 2697-2707, 2020. doi: 10.18494/SAM.2020.2800.
- [5] E. Karatzas et al., "VICTOR: A visual analytics web application for comparing cluster sets," Comput. Biol. Med., vol. 135, pp. 1-18, 2021. doi: 10.1016/j.compbiomed.2021.104557.
- [6] Z. Hu et al., "Deep learning applications in games: a survey from a data perspective," Appl. Intell., vol. 53, no. 24, pp. 31129-31164, 2023. doi: 10.1007/s10489-023-05094-2.
- [7] J. P. Pinto, A. Pimenta, and P. Novais, "Deep Learning and Multivariate Time Series for Cheat Detection in Video Games," in Proc. 2021 IEEE 8th Int. Conf. Data Sci. Adv. Anal. (DSAA), 2021, pp. 3037-3057. doi: 10.1109/DSAA53316.2021.9564219.
- [8] M. T. Almuqati, F. Sidi, S. N. M. Rum, M. Zolkepli, and I. Ishak, "Challenges in Supervised and Unsupervised Learning: A Comprehensive Overview," Int. J. Adv. Sci. Eng. Inf. Technol., vol. 14, no. 4, pp. 1449-1455, 2024. doi: 10.18517/ijaseit.14.4.20191.
- [9] S. Naeem, A. Ali, S. Anam, and M. M. Ahmed, "An Unsupervised Machine Learning Algorithms: Comprehensive Review," Int. J. Comput. Digit. Syst., vol. 13, no. 1, pp. 911-921, 2023. doi: 10.12785/ijcds/130172.
- [10] A. Ashabi, S. Bin Sahibuddin, and M. Salkhordeh Haghighi, "The systematic review of K-means clustering algorithm," in ACM Int. Conf. Proc. Series, Feb. 2020, pp. 13-18. doi: 10.1145/3447654.3447657.
- [11] Danendra Athallah et al., "Implementasi algoritma K-Means untuk menganalisa pemain video game Mobile Legends untuk mengetahui tipe hero dan role yang sering digunakan pada setiap kalangan," J. Tek. Inform., 2022.

- [12] P. Shetty and S. Singh, "Hierarchical Clustering: A Survey," *Int. J. Appl. Res.*, vol. 7, no. 4, pp. 178–181, 2021. doi: 10.22271/allresearch.2021.v7.i4c.8484.
- [13] S. Yildirim, "XGBoost-Driven Evaluation of Clustering Methods for MOBA Player Segmentation," *J. Emerg. Comput. Technol.*, vol. 5, no. 1, pp. 84–95, 2025. doi: 10.57020/ject.1770052.
- [14] G. Zhou, J. Ewald, and J. Xia, "OmicsAnalyst: A comprehensive web-based platform for visual analytics of multi-omics data," *Nucleic Acids Res.*, vol. 49, no. W1, pp. W476–W482, 2021. doi: 10.1093/nar/gkab394.
- [15] X. Wu, A. Poorthuis, R. Zurita-Milla, and M. J. Kraak, "An interactive web-based geovisual analytics platform for co-clustering spatio-temporal data," *Comput. Geosci.*, vol. 137, pp. 1–25, 2020. doi: 10.1016/j.cageo.2020.104420.
- [16] C. Schröer, F. Kruse, and J. Marx, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021. doi: 10.1016/j.procs.2021.01.199.
- [17] A. Rotondo and F. Quilligan, "Evolution Paths for Knowledge Discovery and Data Mining Process Models," *SN Comput. Sci.*, vol. 1, no. 2, pp. 1–19, 2020. doi: 10.1007/s42979-020-0117-6.
- [18] C. Wongoutong, "The impact of neglecting feature scaling in k-means clustering," *PLoS ONE*, vol. 19, no. 12, pp. 1–19, 2024. doi: 10.1371/journal.pone.0310839.
- [19] N. I. Boyko and O. A. Tkachyk, "Hierarchical clustering algorithm for dendrogram construction and cluster counting," *Informatics Math. Methods Simul.*, vol. 13, no. 1–2, pp. 5–15, 2023. doi: 10.15276/imms.v13.no1-2.5.