

Data Mining Menggunakan Algoritma *K-Means* Untuk Menentukan Game Terpopuler Pada Platform Steam Dengan *RapidMiner*

¹Aryo Chandra Ray Hash, ²Deryl Iman Condro Baskoro, ³I Putu Ganesa Weda, ⁴Muhammad Fakhil,
⁵Muhammad Fauzan, ⁶Maulana Fansyuri

¹Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

²Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

³Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

⁴Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

⁵Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

⁶Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

¹aryochandra1209@gmail.com, ²derylbaskoro@gmail.com, ³iputuweda125@gmail.com, ⁴mfakhil288@gmail.com,
⁵muhammadfauzannn176@gmail.com, ⁶dosen02359@unpam.ac.id

Abstract

With the number of games increasing every year, it is a challenge to determine which games are the most popular on the Steam platform. This study uses the K-Means Clustering algorithm in RapidMiner to group games based on their popularity. Ratings and estimated number of game downloads are the variables used in this study. Data were collected from the top game sales Dataset on the Steam platform. Clustering produces two clusters: less dan most populer, indicate the level of game popularity. This study can help game developers and publishers understand what features users are most interested in in a game

Keywords: Data Mining, K-Means, RapidMiner, Steam, Popular Games

Abstrak

Di tengah jumlah game yang terus bertambah setiap tahunnya, menjadi tantangan tersendiri untuk menentukan game mana yang paling populer di platform Steam. Penelitian ini menggunakan algoritma *clustering* K-Means dalam *RapidMiner* untuk mengelompokkan game berdasarkan popularitasnya. Penilaian dan perkiraan jumlah unduhan game adalah variabel yang digunakan dalam penelitian ini. Data dikumpulkan dari *Dataset* penjualan game teratas di platform Steam. *Clustering* menghasilkan dua klaster: Kurang Populer dan Paling Populer, yang menunjukkan tingkat popularitas game. Penelitian ini dapat membantu pengembang dan penerbit game memahami fitur apa yang paling diminati pengguna dalam game.

Kata Kunci: *Data Mining, K-Means, RapidMiner, Steam, Game Populer*

A. PENDAHULUAN

Dengan ribuan judul game dari berbagai genre dan pengembang, Steam adalah salah satu distribusi digital game terbesar di dunia. Dengan jutaan pengguna yang aktif setiap hari, Steam juga menjadi pasar bagi para pengembang dan sumber data yang sangat kaya tentang perilaku dan preferensi pengguna. Mencari game mana yang benar-benar populer dan mendapatkan perhatian besar dari komunitas pengguna menjadi sulit di tengah banyaknya game yang tersedia.

Jumlah unduhan dan rating pengguna adalah beberapa cara untuk mengetahui seberapa populer suatu game. Namun, mengelompokkan game berdasarkan metrik tersebut secara manual sangat tidak efektif, terutama untuk data yang sangat besar. Akibatnya, pendekatan yang sistematis dan berbasis teknologi diperlukan untuk membantu proses analisis dan pengambilan keputusan. Data mining, khususnya metode *clustering*, adalah salah satu pendekatan yang relevan dalam situasi ini.

Untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan hal-hal yang serupa, *clustering* adalah metode pembelajaran tidak supervisi. *K-Means Clustering* adalah salah satu algoritma *clustering* yang paling banyak digunakan. Algoritma ini membagi data ke dalam beberapa kluster berdasarkan kedekatan jarak antar data terhadap *centroid*. Kemudian algoritma memperbarui posisi *centroid* untuk mencapai pengelompokan yang ideal (Han, Kamber, & Pei, 2012).

K-Means Clustering digunakan untuk membentuk dua kluster utama dalam penelitian ini:

- Game paling populer karena memiliki banyak unduhan dan rating tinggi.
- Game yang kurang populer karena memiliki rating yang lebih rendah dan jumlah unduhan yang lebih sedikit.

Pengelompokan ini bertujuan untuk menyederhanakan peta popularitas game dan membantu pengembang dan pengamat industri memahami game mana yang paling diminati oleh pengguna.

Perangkat lunak analitik data *RapidMiner* membantu proses analisis dengan mendukung berbagai teknik data mining, termasuk *K-Means Clustering*. Perangkat lunak ini juga menawarkan fitur preprocessing, pelatihan model, dan visualisasi hasil, serta memungkinkan pembuatan model dengan antarmuka visual tanpa harus menulis kode pemrograman.

Penelitian ini bertujuan untuk mengidentifikasi game-game terpopuler di platform Steam dengan menggunakan algoritma *K-Means* dalam *RapidMiner* dan membentuk dua kluster utama. Diharapkan metode ini dapat membantu pelaku industri dalam membuat strategi distribusi dan pengembangan game yang lebih tepat sasaran.

B. TEORI DAN METODELOGI

RapidMiner adalah perangkat lunak data science berbasis GUI yang dimaksudkan untuk membuat proses analisis data lebih mudah tanpa menulis kode. Mulai dari data preprocessing, modelling, evaluation, dan deployment, platform ini mendukung berbagai langkah dalam proses pengolahan data. Metode drag-and-drop memungkinkan pengguna membuat workflow analisis. Ini ideal untuk akademisi, analis bisnis, dan peneliti dari berbagai latar belakang (*RapidMiner* Documentation, 2023).

RapidMiner dapat digunakan untuk mengolah data seperti jumlah unduhan dan rating pengguna dalam analisis game di platform Steam. *RapidMiner* memungkinkan pembuatan model yang akurat secara visual dan efisien dan memungkinkan visualisasi hasil pengelompokan secara langsung dengan dukungan terhadap berbagai algoritma machine learning, termasuk *K-Means Clustering*.

Data mining adalah proses eksplorasi data menyeluruh untuk menemukan pola tersembunyi, tren, atau pengetahuan baru dari kumpulan data yang sangat besar. Untuk mengekstrak informasi berharga dari data mentah, proses ini menggunakan statistik, pembelajaran mesin, dan basis data.

Data mining mencakup banyak langkah saat diterapkan, mulai dari pengumpulan data, pembersihan, transformasi, eksplorasi, pemodelan, dan akhirnya evaluasi hasilnya. Dengan jumlah unduhan dan rating pengguna sebagai indikator utama, data mining digunakan dalam penelitian ini untuk menentukan permainan Steam terpopuler. Teknik ini sangat bermanfaat untuk membantu proses pengambilan keputusan berbasis data, khususnya dalam industri game digital yang sangat kompetitif.

Kaggle adalah sebuah platform untuk komunitas ilmu data dan pembelajaran mesin yang membantu peneliti, akademisi, dan praktisi dalam membuat dan menguji model analisis data dengan menyediakan banyak *Dataset* publik, kompetisi, dan forum diskusi. Platform ini juga memungkinkan pengguna berbagi dan mengakses berbagai jenis data dari berbagai bidang seperti kesehatan, ekonomi, transportasi, dan kepuasan hidup masyarakat. Selain itu, *Kaggle* menawarkan notebook interaksi yang memungkinkan peneliti berbagi dan menguji model analisis data.

Dataset "Game Terpopuler" yang digunakan dalam penelitian ini diperoleh dari *Kaggle.com*, karena platform ini menyediakan *Dataset* yang lengkap, terkurasi, dan dapat diakses secara publik. Hal ini memungkinkan analisis yang lebih cepat dan dengan kualitas yang lebih baik. Pengguna dapat memahami struktur data dan melakukan preprocessing dengan benar jika ada metadata dan dokumentasi yang baik tersedia. Dengan algoritma *K-Means Clustering*, *Kaggle* menjadi sumber data yang sah

dan dapat diandalkan untuk penelitian berbasis data mining seperti klasifikasi.

Salah satu teknik utama dalam pengolahan data adalah *clustering*, yang bertujuan untuk mengelompokkan data ke dalam beberapa grup atau kluster berdasarkan hal-hal yang serupa. Karena tidak memerlukan label kelas pada data, teknik ini termasuk dalam kategori pembelajaran yang tidak diawasi (Tan, Steinbach, & Kumar, 2019).

Untuk analisis game, *clustering* memungkinkan pengelompokan game ke dalam kategori seperti "paling populer" dan "kurang populer" berdasarkan karakteristik yang sebanding. Hal ini sangat membantu dalam pemetaan popularitas game dan memberikan gambaran umum tentang tren pasar. Hasil *clustering* juga dapat digunakan sebagai dasar untuk strategi pengembangan game dan pemasaran yang lebih tepat sasaran.

Algoritma pengclustoran K-Means membagi data ke dalam kluster K berdasarkan jarak terdekat antara *centroid* kluster dan data. Ini terdiri dari inisialisasi *centroid* secara acak, pembagian data berdasarkan seberapa dekat *centroid* berada, dan iterasi ulang sampai posisi *centroid* konvergen atau stabil (Jain, 2010).

Karena sifatnya yang sederhana, efisien, dan cocok untuk data berukuran besar, algoritma ini banyak digunakan. Studi ini menggunakan K-Means untuk membagi game-game di Steam menjadi dua kluster utama: game paling populer dan game paling kurang populer, berdasarkan jumlah unduhan dan rating pengguna. Algoritma ini memiliki beberapa keterbatasan meskipun efektif. Salah satunya adalah kebutuhan untuk menentukan jumlah kluster awal dan sensitivitas terhadap posisi awal *centroid* inisialisasi. Oleh karena itu, metode inisialisasi *centroid* dan nilai K yang tepat sangat mempengaruhi kualitas hasil akhir.

C. HASIL DAN PEMBAHASAN

Penerapan penelitian ini menggunakan perhitungan Algoritma *K-Means Clustering* yang dilakukan dengan menggunakan alat *RapidMiners* sebagai perangkat yang dapat membantu pengguna dalam menjalankan perhitungan dengan menggunakan operator. Operator ini berfungsi untuk mengubah data dan kemudian menghubungkannya dengan node-node Operator lainnya. Tugas pengguna adalah menghubungkan node tersebut untuk memeriksa hasil tampilan grafik dari perhitungan.

1. *T*

Dataset ini mencakup informasi mengenai sejumlah judul permainan, dengan total sebanyak 2.380 data entri. Setiap entri terdiri dari tiga atribut penting, yaitu nama_game, penilaian, dan perkiraan_unduhan. Kolom game_name mencakup daftar nama-nama permainan yang ada, dituliskan dalam format teks, seperti Counter-Strike 2,

PUBG: BATTLEGROUNDS, dan Dota 2. Kolom penilaian adalah nilai yang diberikan oleh pengguna untuk game, disajikan dalam bentuk angka desimal (real), biasanya dalam rentang 1 sampai 5. Sementara itu, kolom perkiraan_unduhan menunjukkan taksiran jumlah unduhan dari setiap permainan, disajikan dalam angka bulat (integer). Kumpulan data ini dapat dimanfaatkan untuk menganalisis popularitas permainan berdasarkan penilaian dan jumlah unduhan, serta mengembangkan model prediksi dalam lingkup penambangan data.

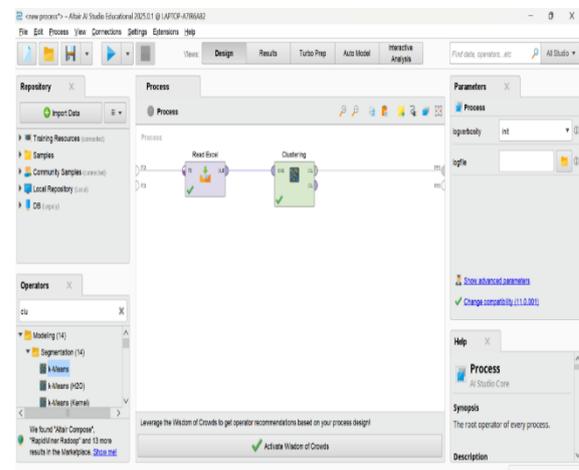


game_name	rating	estimated_downloads
Counter-Strike 2	3.200	306170000
PUBG: BATTLEGROUNDS	3.100	162350000
ELDEN RING NIGHTREIGN	3.960	940000
The Last of Us™ Part I	4.100	2000000
Red Dead Redemption 2	4.320	21610000
Dota 2	3.710	234950000
EA SPORTS FC™ 25	3.200	2320000
Project Zomboid	3.660	21370000
Game of Thrones: Kingsroad	3.400	361500
Rust	3.200	36400000
F1® 25	3.800	30800
Dead by Daylight	3.420	301000

Gambar 1 *Dataset*

2. Implementasi

Setelah menginput dan melakukan pemeriksaan data, untuk melakukan pemodelan terhadap *Dataset* Game terpopuler. Algoritma yang di gunakan adalah *K-Means Clustering*



Gambar 2 Implementasi

Dimulai dengan memasukkan data ke dalam RapidMiner, kemudian data tersebut diolah menggunakan metode k-means dengan parameter $k = 2$ dan jenis ukuran = ukuran campuran. Hal ini dilakukan karena dalam analisis Game Terpopuler, kita ingin mengidentifikasi $C0 =$ Game yang Kurang Populer dan $C1 =$ Game yang Paling Populer. Selanjutnya, data dapat diproses menggunakan *RapidMiner* dan menghasilkan dua model kluster, yaitu kluster 0 berisi 2.337 item dan kluster 1 berisi 3 item. Penomoran dimulai dari 0 karena dalam bahasa

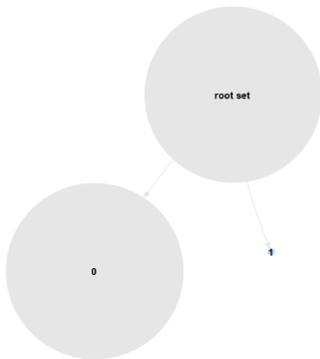
pemrograman, angka 0 merupakan angka pertama dalam urutan penomoran.

Cluster 0	2377 Items
Cluster 1	3 Items
Total	2380 Items

Tabel 1 Tabel Cluster

3. Tree

Hasil yang berbentuk pohon dalam pilihan yang disediakan oleh *RapidMiners* bertujuan untuk memberikan gambaran mengenai perbandingan antara kluster dengan seluruh data. Melalui pilihan ini, kita dapat melihat bagian dari data induk atau *root set*, yang terurai menjadi beberapa kluster.



Gambar 3 Tree

4. Table Centroid

Dari 2 (dua) cluster yang ditentukan, Game kemudian dikelompokkan sesuai hasil perhitungan k means, Menunjukkan bahwa cluster 0 (Game yang kurang populer) memiliki rating 3.228 dan estimasi unduhan 2230242.176 sedangkan cluster 1 (Game Paling Populer) memiliki rating 3.337 dan estimasi unduhan 234490000.

Attribute	cluster_0	cluster_1
rating	3.228	3.337
estimated_downloads	2230242.176	234490000

Gambar 4 Centroid TableS

D. PENUTUP

Kesimpulan

Penelitian ini berhasil mengimplementasikan algoritma *K-Means Clustering* menggunakan *RapidMiner* pada data Game terpopuler di Steam untuk mengelompokkan game berdasarkan yang paling sering dimainkan. Hasil pengelompokan menunjukkan bahwa Cluster 0 mewakili game dengan estimasi unduhan dan rating yang kurang, Cluster 1 mewakili game paling populer.

Dengan total 2380 yang berhasil dikelompokkan ke dalam dua cluster, analisis ini memberikan wawasan yang lebih mendalam tentang kelarisanebuah game. Visualisasi hasil *clustering* juga memudahkan interpretasi dengan pembagian warna yang merepresentasikan masing-masing kategori produk.

Implementasi ini membuktikan bahwa penggunaan teknologi sederhana seperti *RapidMiner* dapat membantu pemilihan game yang berkualitas. Selain itu, pendekatan ini dapat menjadi langkah awal pemain dalam memanfaatkan teknologi data mining untuk pengambilan keputusan yang lebih tepat dan berbasis data.

Saran

Agar hasil klasterisasi lebih akurat dan komprehensif, penelitian selanjutnya harus mencakup lebih banyak variabel, seperti genre game, harga, dan ulasan pengguna. Selain itu, untuk menilai performa dan keakuratan model K-Means, Anda dapat menggunakan algoritma *clustering* hierarkis seperti DBSCAN atau Hierarchical *Clustering*. Selain itu, penggunaan kumpulan data yang lebih canggih dan lebih besar akan meningkatkan validitas hasil analisis dalam menentukan tren popularitas game di platform Steam.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada dosen pembimbing yang telah memberikan bimbingan dan saran yang sangat berharga selama proses penyusunan jurnal ini. Penulis juga mengucapkan terima kasih kepada seluruh kelompok atas kerja sama dan kontribusi yang kuat dalam menyelesaikan penelitian ini. Penulis juga berterima kasih kepada pihak-pihak yang telah memberikan data dan alat bantu seperti *RapidMiner* dan *Kaggle*, yang sangat membantu dalam melaksanakan analisis data. Kami berharap temuan penelitian ini akan bermanfaat bagi pengembang game, akademisi, dan pihak lain yang tertarik dengan data mining.

E. DAFTAR PUSTAKA

- Amin, Z., Pangestu, S. E., Alfaruq, M. S., Efrizon, L., & Rahmadenni, R. (2025). Penerapan Algoritma K-Medoids untuk Pengelompokan Genre Game Berdasarkan Pola Penjualan. *JURAL Riset RUMPUN ILMU TEKNIK*, 4(1), 270–286.
- Cevin, J., Gultom, D., Orlando, N., & Herdiatmoko, H. F. (2024). Analisis Sentimen Keluhan Pengguna pada Game Supremacy 1914 Menggunakan Algoritma *K-Means Clustering*. *Jurnal Rekayasa Sistem Informasi dan Teknologi*, 2(2), 769–776.

- Devila, L. E., Cholil, S. R., Athallah, R. D., & Irawan, A. A. (2022). Implementasi Algoritma K-Means untuk Menganalisa Pemain Video Game Mobile Legend untuk Mengetahui Tipe Hero dan Role yang Sering Digunakan pada Setiap Kalangan. *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, 6(3), 261–268.
- Dwipayana, L. (2021). *Clustering dengan Metode K-Means untuk Menentukan Matchmaking pada Game Online Mobile Legends* (Skripsi, Politeknik Negeri Jember).
- Dwynne, Z. C., Aini, D. N., Pertiwi, T. A., Suryani, S., & Pramana, D. (2023, August). Cluster Tingkat Kecanduan Game Online pada Mahasiswa Fakultas Sains dan Teknologi dan Korelasinya terhadap Minat Belajar. In *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat* (pp. 126–132).
- Febriansyah, A. R., & Arifin, T. (2023). Penerapan Algoritma K-Means Clustering pada Agent Video Game Valorant. *eProsiding Teknik Informatika (PROTEKTIF)*, 4(1), 106–214.
- Hermawan, M. A. Analisis Klasterisasi Pemain Mobile Legends Menggunakan Algoritma K-Means terhadap Penyesuaian Sistem Rekomendasi Hero.
- Jamil, H. M., & Sulianta, F. Analisis Pengaruh Harga dan Ulasan terhadap Penjualan Video Game Menggunakan Algoritma K-Means Clustering.
- Maulida, L. (2025). Penggunaan Algoritma K-Means dalam Menentukan Karakter Game Zenless Zone Zero. *REMIK: Riset dan E-Jurnal Manajemen Informatika Komputer*, 9(1), 403–414.
- Muharizki, M. I., & Arianto, D. B. (2023). Analisis Clustering dengan Metode K-Means terhadap Statistik Permainan Pro-Player Valorant pada Kompetisi Valorant Champions 2022. *Serunai: Jurnal Ilmiah Ilmu Pendidikan*, 9(1), 40–47.
- Mustofa, M. (2019). Penerapan Algoritma K-Means Clustering pada Karakter Permainan Multiplayer Online Battle Arena. *Jurnal Khatulistiwa Informatika*, 6(2), 246–254.
- Orlando, N., Prasetya, B. N., & Hermawan, L. (2025, April). Analisis Sentimen Keluhan Pengguna pada Game Mobile Legends Menggunakan Metode K-Means Clustering. In *MDP Student Conference* (Vol. 4, No. 1, pp. 292–298).
- Prastyo, Y. (2017). Pembagian Tingkat Kecanduan Game Online Menggunakan K-Means Clustering serta Korelasinya terhadap Prestasi Akademik. *Elinvo (Electronics, Informatics, and Vocational Education)*, 2(2), 138–148.
- Pratama, A. R., Maulana, B., Rianda, R. D., & El Hasyim, S. (2023). Comparison of K-Means and K-Medoids Algorithms for Grouping Video Game Sales Data in North America: Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Penjualan Video Game di Amerika Utara. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, 3(2), 111–118.
- Qasthari, M. W., & Kurniawan, R. (2024). Penggunaan Algoritma K-Means Clustering untuk Mengelompokkan Pemain Berdasarkan Gaya Bermain pada Battle Royale Call of Duty Mobile. *Future Academia: The Journal of Multidisciplinary Research on Scientific and Advanced*, 2(3), 280–292.

RapidMiner Documentation. (2023). *RapidMiner official documentation*. Retrieved from <https://docs.rapidminer.com>

Reynaldo, Y., Triayudi, A., & Ningsih, S. (2022). Analisis Faktor yang Mempengaruhi Gamers PC dan Konsol Beralih ke Game Mobile Menggunakan Metode *K-Means Clustering*. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, 6(1), 42–48.

Saputra, P. A. H. (2024). *Penerapan Algoritma Clustering K-Means pada Pengelompokan Hero di Game Mobile Legends* (Doctoral dissertation, Universitas Nahdlatul Ulama Sunan Giri).

Shahibah, S., Triyasri, N., Inanti, A. A., Rachel, J., & Pratama, N. D. (2025). Analisis Klastering dari Data Behavior Online Gaming Menggunakan Algoritma K-Means. *JoDMApps (Journal of Data Science Methods and Applications)*, 1(1), 18–24.

Waode, Y. S. Kajian Algoritma Optimasi K-Means untuk Meningkatkan Kualitas Cluster pada Data Sparse.