



Penerapan Algoritma K-Medoids untuk Pengelompokan Genre Game Berdasarkan Pola Penjualan

Zamzamil Amin^{1*}, Satria Eka Pangestu², Muhammad Syafiq Alfaruq³, Lusiana Efrizon⁴, Rahmadenni⁵

¹⁻⁵Program Studi Teknik Informatika, Universitas Sains dan Teknologi Indonesia, Indonesia

Alamat: Jl. Purwodadi Indah No.KM. 10, Sidomulyo Bar., Kec. Tampan, Kota Pekanbaru, Riau 28294

Korespondensi penulis: zamzamilamin@gmail.com*

Abstract: *The gaming industry continues to grow rapidly, with various genres having different sales patterns. This research aims to group the best game genres based on sales patterns using the K-Means Medoid algorithm. This method was chosen because of its ability to overcome data diversity and reduce the influence of outliers compared to the conventional K-Means algorithm. The data used in this research includes game sales information from various platforms, which is then analyzed to find distribution patterns and market trends. The results of this research show that certain game genres have superior sales patterns compared to others, and there are special characteristics in the groups that are formed. These findings are expected to help game developers and publishers in making strategic decisions regarding the development and marketing of their products. In addition, this research also contributes to the development of a data-based recommendation system that can be used to understand market preferences more accurately.*

Keywords: *K-Means Medoid, Game Genre Grouping, Sales Patterns, Data Analysis, Game Industry.*

Abstrak: Industri game terus berkembang pesat, dengan berbagai genre yang memiliki pola penjualan berbeda-beda. Penelitian ini bertujuan untuk mengelompokkan genre game terbaik berdasarkan pola penjualan menggunakan algoritma K-Means Medoid. Metode ini dipilih karena kemampuannya dalam mengatasi keberagaman data serta mengurangi pengaruh outlier dibandingkan algoritma K-Means konvensional. Data yang digunakan dalam penelitian ini mencakup informasi penjualan game dari berbagai platform, yang kemudian dianalisis untuk menemukan pola distribusi dan tren pasar. Hasil dari penelitian ini menunjukkan bahwa genre game tertentu memiliki pola penjualan yang lebih unggul dibandingkan yang lain, serta terdapat karakteristik khusus dalam kelompok yang terbentuk. Temuan ini diharapkan dapat membantu pengembang dan penerbit game dalam mengambil keputusan strategis terkait pengembangan dan pemasaran produk mereka. Selain itu, penelitian ini juga berkontribusi dalam pengembangan sistem rekomendasi berbasis data yang dapat digunakan untuk memahami preferensi pasar secara lebih akurat.

Kata kunci: K-Means Medoid, Pengelompokan Genre Game, Pola Penjualan, Analisis Data, Industri Game

1. LATAR BELAKANG

Video game telah menjadi salah satu bentuk hiburan digital paling populer di dunia, menarik pemain dari berbagai kalangan usia dan latar belakang. Seiring perkembangan teknologi, industri video game telah mencatat pertumbuhan pesat, baik dari segi jumlah pemain maupun nilai ekonominya. Pada tahun-tahun terakhir, industri ini mengalami lonjakan signifikan dengan hadirnya berbagai genre dan platform yang menjawab kebutuhan hiburan modern.

Industri game telah mengalami pertumbuhan yang signifikan dalam beberapa tahun terakhir, memberikan dampak yang substansial pada perekonomian global dan nasional. Pada tahun 2024, jumlah gamer di seluruh dunia mencapai 3,32 miliar, dengan Asia sebagai penyumbang terbesar, yaitu 1,48 miliar gamer (Abril Geralin, 2024).

Dari segi pendapatan, industri game global menghasilkan sekitar USD 187,7 miliar pada tahun 2023, meningkat 2,6% dibandingkan tahun sebelumnya. Pendapatan terbesar berasal dari mobile games dengan total USD 92,6 miliar, diikuti oleh console games sebesar USD 56,1 miliar (Rahayu, 2023) .

Pertumbuhan ini tidak hanya berdampak pada peningkatan pendapatan, tetapi juga menciptakan lapangan kerja baru dan mendorong inovasi dalam teknologi serta konten kreatif. Industri game telah menjadi salah satu pilar penting dalam ekonomi digital dan ekonomi kreatif, memberikan kontribusi signifikan terhadap Produk Domestik Bruto (PDB) dan membuka peluang investasi yang menjanjikan (Rezky Pratama et al., 2023) .

Dalam konteks penelitian mengenai pengelompokan genre game berdasarkan pola penjualan menggunakan algoritma k-medoids, penting untuk memahami tren penjualan dan preferensi pasar. Algoritma k-medoids, sebagai salah satu metode clustering, dapat digunakan untuk mengelompokkan data penjualan game berdasarkan genre, platform, dan wilayah penjualan. Misalnya, sebuah penelitian membandingkan algoritma k-means dan k-medoids untuk pengelompokan data penjualan video game di Amerika Utara, yang dapat memberikan wawasan tentang preferensi pasar di wilayah tersebut

Pada penelitian (Aripin et al., 2023) Penelitian terdahulu yang telah dilakukan oleh Sevi Dian Nirwana, dkk pada tahun 2022 dimana pada hasil penelitian didapatkan bahwa dengan menggunakan algoritma K-Medoids dapat dipergunakan untuk membentuk 3 cluster dan kinerjanya lebih baik dibandingkan dengan algoritma clustering lainnya . Penelitian lainnya juga telah dilakukan di tahun yang sama oleh Fahikra Faisal, dkk dengan hasil penelitian yang didapatkan yaitu Algoritma K- Medoids dapat untuk membentuk clusterisasi dengan kinerja lebih baik dibandingkan dengan algoritma lainnya

Pada penelitian (Hastari et al., 2023) penggunaan algoritma K-Medoids dinilai lebih unggul untuk pengelompokkan data ketika terdapat outlier jika dibandingkan dengan penggunaan algoritma K-Means, selain itu algoritma K-Medoids bekerja lebih baik dalam skalabilitas untuk kumpulan data dengan jumlah besar dan dinilai lebih efisien dibanding algoritma K-Means.

Pada penelitian (Bhakti et al., 2024a) terkait penerapan K-Medoid untuk pengelompokkan game berdasarkan data top seller di simpulkan bahwa Silhouette Score yang dihasilkan dari data Game harga dan Review mendapat nilai 0,19 apabila Cluster tersebut dipisah menjadi Cluster data Game harga dan Game Review maka didapatkan nilai Silhouette Score untuk harga mendapat nilai 0,35 sedangkan Review mendapat nilai 0.36 yang dapat

disimpulkan terjadi kenaikan sekitar 0,15. Disebabkan karena perbedaan signifikan pada Cluster Harga & Review.

pada penelitiannya (Rezky Pratama et al., 2023) mencoba melakukan Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Penjualan Video Game di Amerika Utara Aditya menyimpulkan bahwa Pengelompokan data yang ditunjukkan dari perbandingan antara algoritma clustering, K-Medoids dan K-Means tidak terlalu signifikan. Rata-rata waktu yang diperlukan yakni pada K-Means hanya 1 detik sedangkan pada K-Medoids memerlukan 2 menit 21 detik dengan artian semakin tinggi iterasi dari grup/kelompok yang ditentukan, maka pemrosesan data akan semakin lambat.

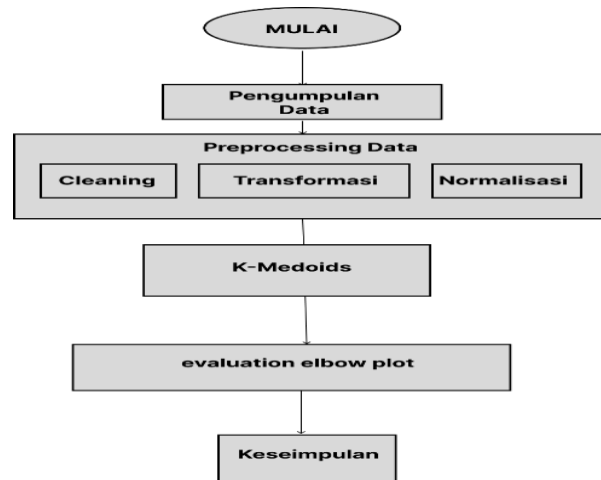
Peneliti menjadikan subsektor games

Pada penelitian (Muhammad et al., 2021) peneliti menyebutkan bahwa sebagai objek penelitian yang dapat memberikan dampak positif bagi pertumbuhan ekonomi masyarakat Kota Malang dan dapat menunjang kegiatan industri kreatif sebagai penggerak Kota Malang sebagai kota kreatif. Wilayah penelitian ini berada di Kota Malang, Provinsi Jawa Timur. Keberadaan industri kreatif permainan digital di Kota Malang memiliki berbagai permasalahan antara lain, kurangnya kolaborasi antar stakeholder, rendahnya daya saing produk dengan kota-kota lain, dan kurangnya SDM dalam pengerjaannya. Untuk itu diperlukan arahan pengembangan industri kreatif sub sektor permainan digital sebagai penggerak Kota Malang menuju kota kreatif

Penelitian ini bertujuan penelitian ini berfokus pada analisis tren pasar dengan mengelompokkan genre game berdasarkan pola penjualan. Dengan menggunakan algoritma k-medoids, penelitian ini akan mengidentifikasi bagaimana berbagai genre game memiliki pola penjualan yang berbeda di berbagai platform dan wilayah.. Selain itu, metode K-Means Medoid dipilih karena kemampuannya dalam menangani data dengan keberagaman yang tinggi serta mengurangi pengaruh outlier dibandingkan K-Means biasa . Berdasarkan permasalahan dan penelitian terdahulu yang dijadikan acuan maka penelitian ini bertujuan untuk melihat pengelompokan metode K-medoids clustering dalam mengelompokkan genre game berdasarkan pada pola penjualan game.

2. METODE PENELITIAN

Diagram alur penelitian di sajikan pada gambar 1.



Gambar 1. Diagram alur penelitian

Data yang berasal dari website kaggle, data tersebut berisikan data penjualan penjualan video game dengan tahun liris game tersebut dari tahun 2000-2024. Data tersebut di lakukan preprocessing data yang meliputi :

- Cleaning data yang bertujuan untuk menghapus data yang tidak relevan atau memiliki nilai yang hilang.
- Transformasi data yang bertujuan untuk mengubah data non-numerik menjadi numerik.
- Normalisasi data dengan menggunakan metode min-max normalization yang bertujuan untuk mengubah data ke dalam skala yang sama agar dapat di bandingkan secara adil.
- Data yang sudah di proses kemudian di kelompok kan berdasarkan pola penjualan menggunakan algoritma K-Medoids.
- evaluation menggunakan elbow plot untuk menentukan jumlah cluster yang optimal.
- Analisis data yang terbagi menjadi 3 cluster utama yaitu :
 - a) Cluster 1 : game dengan penjualan tinggi di amerika serikat dan uni eropa.
 - b) Cluster 2 : game dengan penjualan tinggi secara global tanpa mendominasi di satu wilayah tertentu.
 - c) Cluster 3 : game dengan penjualan tinggi di jepang.

Kesimpulan dari penelitian ini dapat memberikan langkah strategis untuk pamarasan dan pengembangan game yang lebih tepat sasaran.

Pengumpulan data

Pada tahap pertama di mulai dengan mengumpul kan data, dalam penelitian ini data yang di gunakan adalah data penjualan video game yang di ambil dari <https://www.kaggle.com/datasets/gregorut/videogamesales> dengan data penjualan yang berada pada amerika serikat, uni eropa, jepang, global, dan lain-lain. Data set ini di pilih karna kelengkapan, kredibilitas, dan relevansinya dalam menganalisis pola penjualan video game berdasarkan genre dengan struktur yang cocok untuk metode algoritma K-Medoids.

Preprocessing Data

Data preprocessing adalah serangkaian langkah yang dilakukan untuk mempersiapkan data mentah menjadi format yang lebih baik untuk dianalisis atau digunakan oleh model pembelajaran mesin. Tujuan utama dari data preprocessing adalah meningkatkan kualitas data, mengatasi masalah yang mungkin ada, dan memastikan bahwa data siap digunakan dalam proses analisis atau pelatihan model (Ramadhani et al., 2024a).

Cleaning Data

Cleaning data di lakukan untuk menghilangkan nilai yang hilang, duplikasi, atau data yang tidak relevan agar analisis yang di lakukan lebih akurat. Dalam penelitian ini, proses cleaning data mencakup penghapusan kolom yang tidak digunakan seperti *Rank* dan *Publisher* karna tidak berkontribusi langsung dalam analisis pola penjualan berdasarkan genre game. Data yang hilang di atasi dengan menggunakan estimasi berbasis rata-rata atau median untuk menjaga konsistensi data set, misal nya mengganti nilai kosong pada penjualan di wilayah tertentu dengan rata-rata penjualan dari genre yang sama. Selain itu data yang tidak valid seperti game yang memiliki nilai penjualan 0 pada semua wilayah di hapus karna tidak memberikan informasi yang berguna untuk analisis clustering. Pada (Ramadhani et al., 2024a) Menyatakan bahwa cleaning data sangat penting dalam meningkatkan akurasi hasil clustering dengan menghilangkan noise atau data yang tidak valid.

Transformasi Data

Transformasi data di lakukan untuk mengubah format data agar lebih mudah di analisis, terutama dalam mengonversi variabel non-numerik menjadi numerik. Dalam penelitian ini, kolom genre di ubah menjadi kode angka, seperti *action = 0*, *adventure = 1*, *sports = 10*, dan seterusnya. Selain itu, kategori yang terlalu spesifik digabungkan dengan genre yang serupa jika jumlahnya terlalu sedikit, untuk menghindari ketidakseimbangan data. Misalnya, genre

Shooter dikonversi menjadi angka 9, sedangkan *Racing* menjadi angka 6, sehingga algoritma K-Medoids dapat lebih mudah dalam mengelompokkan data berdasarkan jarak antar titik. (Dwiatmoko, 2021) menyatakan transformasi data sangat penting dalam proses clustering karena mengonversi data non-numerik menjadi numerik mempermudah analisis dan meningkatkan efisiensi algoritma seperti K-Means dan K-Medoids.

Normalisasi Data

Normalisasi data memastikan semua fitur memiliki skala yang seragam agar tidak ada variabel yang mendominasi proses clustering. Penelitian ini menggunakan fitur min-max normalization yang mengubah nilai setiap fitur ke rentang 0 hingga 1 dengan menyesuaikan nilai minimum dan maksimum. Misalnya, pada penjualan tertinggi adalah 41.49 juta unit dan terendah adalah 0.01 juta unit, nilai-nilai ini di normalisasikan sehingga yang tertinggi mendekati 1 dan yang terendah mendekati 0. Dengan normalisasi ini, variabel memiliki bobot seimbang dalam clustering dalam meningkatkan akurasi analisis. (Ahmad Harmain et al., 2022) menjelaskan bahwa normalisasi data sangat penting dalam meningkatkan efisiensi dalam algoritma yang berbasis clustering, karena membantu menghindari bias akibat perbedaan skala dan memastikan hasil yang lebih akurat dalam mengelompokkan data.

Clustering

Clustering adalah metode segmentasi data yang sangat berguna dalam prediksi dan analisis masalah bisnis tertentu, seperti segmentasi pasar, pemasaran, dan pemetaan zonasi wilayah (Juliandri Saputra et al., 2024). Dalam penelitian ini. Proses clustering dilakukan menggunakan algoritma K-Medoids untuk mengelompokkan genre game berdasarkan pola penjualan di berbagai wilayah. Algoritma K-Medoids dimulai dengan pemilihan medoid awal secara acak, kemudian setiap game ditetapkan sebagai medoid terdekat berdasarkan jarak euclidean. Proses ini terus berulang hingga mencapai konvergensi. Karakteristik data seperti variasi penjualan antar wilayah dan potensi outlier dapat dipertimbangkan untuk memastikan bahwa pengelompokan mencerminkan pola penjualan yang sebenarnya.

Pemilihan algoritma K-Medoids didasarkan pada keunggulannya dalam menangani outlier dibandingkan dengan K-Means. K-Means menggunakan rata-rata sebagai pusat cluster, sehingga dapat sensitif terhadap outlier yang dapat mempengaruhi hasil clustering secara signifikan. Penelitian (Kamila et al., 2019) dalam membandingkan kedua algoritma ini, menyatakan bahwa K-Medoids lebih tahan terhadap outlier karena menggunakan objek

representatif sebagai pusat cluster, sementara K-Means lebih sensitif terhadap outlier karena menggunakan rata-rata sebagai pusat cluster.

Algoritma K-Medoids

K-Medoids atau Partitioning Around Medoids (PAM) atau K-Medians merupakan algoritma clustering sebagai varian dari algoritma K-Means. Kelemahan K-Means yang sensitif terhadap outlier dianggap dapat diatasi oleh K-Medoids, dimana suatu objek dengan nilai yang besar mungkin saja secara substansi menyimpang dari distribusi data (Bahri & Midyanti, 2023).

Menurut (Simanjuntak et al., 2023) Strategi dasar dari algoritma k-medoids adalah untuk menemukan cluster k pada objek n dengan terlebih dahulu menemukan objek awal (medoid) secara acak sebagai perwakilan untuk setiap cluster. Setiap objek yang tersisa dikelompokkan dengan medoid yang paling mirip. Metode k-medoids menggunakan objek representatif sebagai titik yang menjadi acuan dan bukan rata – rata objek dari setiap cluster adalah kunci dari metode ini. Cara kerja K-Medoids adalah sebagai berikut (Tampubolon et al., 2021) :

- Tentukan jumlah pusat cluster sebanyak k (jumlah cluster)
- Tentukan medoid awal secara acak sebanyak k dari n data
- Hitung jarak objek ke tiap medoid yang telah dipilih, gunakan Euclidean Distance. Hitung Total jarak terdekat pada medoid awal (a) menggunakan Persamaan (1).

$$\sqrt{(x_i - X)^2 + \dots + (y_n - Y)^2} \quad (1)$$

- Pilih secara acak medoid baru, lalu hitung jarak objek ke tiap medoid yang telah dipilih, gunakan Euclidean Distance. Hitung Total jarak terdekat pada medoid baru (b)
- Hitung total simpangan (S) menggunakan Persamaan (2).

$$S = b - a \quad (2)$$

a : jumlah jarak terdekat antara objek ke medoid awal.

b : jumlah jarak terdekat antara objek ke medoid baru.

Jika $S < 0$, maka tukar objek dengan data untuk menghasilkan sekumpulan k baru sebagai medoid.

Ulangi 3-5, sampai tidak terjadi lagi perubahan medoids, sehingga dihasilkan sejumlah cluster beserta anggotanya.

Pseudocode algoritma K-Medoids dapat di lihat pada gambar 2.

```
# Input : Data X, Jumlah Cluster k
# Output : Cluster hasil pengelompokan

# 1. Pilih k medoid secara acak dari X
# 2. Repeat:
# 3. Untuk setiap titik data x dalam X:
# 4. Hitung jarak antara x dan setiap medoid
# 5. Tetapkan x ke cluster dengan medoid terdekat
# 6. Untuk setiap cluster:
# 7. Pilih titik dalam cluster sebagai kandidat medoid baru
# 8. Hitung total jarak dalam cluster jika menggunakan kandidat medoid baru
# 9. Jika total jarak berkurang, ganti medoid lama dengan kandidat baru
# 10. Sampai tidak ada perubahan pada medoid
```

Gambar 2. Pseudocode K-Medoids.

Kelebihan menggunakan algoritma ini adalah K-Medoids lebih tahan terhadap outlier dan lebih stabil pada dataset dengan distribusi yang tidak merata karena menggunakan titik data sebagai pusat cluster. Metode ini juga fleksibel dengan berbagai metrik jarak. Namun, algoritma ini memiliki kekurangan pada segi kecepatan pengolahan data karena pembaruan medoid lebih lambat di bandingkan dengan K-Means terutama pada dataset dengan jumlah yang besar. Selain itu, pemilihan medoid awal dapat mempengaruhi hasil clustering, dan metode ini kurang efektif untuk dataset yang berdimensi tinggi. Oleh karena itu K-Medoids lebih cocok digunakan pada dataset yang memiliki noise atau outlier dan ketika dataset memiliki interpretasi pusat cluster sebagai titik aktual yang dianggap penting.

Elbow Plot

Diagram siku (Elbow Plot) merupakan salah satu metode visual yang digunakan dalam analisis kluster untuk membantu menentukan jumlah kluster optimal dari sebuah dataset. Fungsinya adalah untuk mengenali titik di mana penambahan jumlah kluster tidak lagi memberikan peningkatan yang signifikan dalam pemahaman struktur data. Elbow Plot telah menjadi alat yang umum digunakan dalam berbagai bidang, termasuk ilmu data, statistik, dan machine learning. Dalam praktiknya, Elbow Plot membantu analisis data dalam memahami struktur yang tersembunyi dalam dataset dan memilih jumlah kluster yang optimal untuk analisis lebih lanjut (Bhakti et al., 2024b).

Min-Max Normalization

Metode Min-Max Normalization adalah cara untuk mengubah nilai suatu variabel ke dalam rentang tertentu, antara 0 dan 1, dengan mengurangi nilai minimum dari setiap observasi dan membaginya dengan selisih antara nilai maksimum dan minimum variabel (Ramadhani et al., 2024b).

$$XSCALED = x - min / max - min \quad (1).$$

3. HASIL DAN PEMBAHASAN

Pada bagian ini akan menjelaskan tahapan-tahapan pada penelitian, di mulai dengan pengambilan data, preproseccing data, implementasi algoritma K-Medoids, dan penentuan jumlah cluster menggunakan elbow plot.

Pengambilan data

Data di ambil dari website www.kaggle.com dengan jumlah data sebanyak 16.600 data yang terdiri dari 11 variabel di antara nya : rank , name, platfrom, years, genre, Publisher, NA Sales, EU Sales, JP Sales, dan global Sales.

Preprocessing data

Preprocessing data di lakukan untuk membersihkan, mengubah, dan menormalisasikan data sebelum di lakukan analisis agar menghasilkan hasil yang relevan dan akurat.

Cleaning data

Data yang digunakan merupakan data penjualan video game yang penjualannya tersebar mulai dari amerika serikat, uni eropa, jepang, global, dan lain lain. Pada cleaning data variabel yang digunakan di antara nya adalah : name, genre, NA sales, EU sales, JP sales, global sales, dan other sales. Data ini telah di lakukan proses cleaning data sebagaimana terdapat pada tabel 1.

Tabel 1. Hasil cleaning data

Name	Genre	NA_Sale s	EU_Sale s	JP_Sale s	Other_Sale s	Global_Sal es
Wii Sports	Sports	41.49	29.02	3.77	8.46	82.74
Super Mario Bros.	Platfor m	29.08	3.58	6.81	0.77	40.24
Mario Kart Wii	Racing	15.85	12.88	3.79	3.31	35.82
Wii Sports Resort	Sports	15.75	11.01	3.28	2.96	33
Pokemon Red/Pokemon Blue	Role- Playing	11.27	8.89	10.22	1	31.37
Tetris	Puzzle	23.2	2.26	4.22	0.58	30.26
New Super Mario Bros.	Platfor m	11.38	9.23	6.5	2.9	30.01
Wii Play	Misc	14.03	9.2	2.93	2.85	29.02

New Super Mario Bros. Wii	Platform	14.59	7.06	4.7	2.26	28.62
Duck Hunt	Shooter	26.93	0.63	0.28	0.47	28.31
...
Spirits & Spells	GBA	0.01	0	0	0	0.01

Pada tabel 1 menunjukkan hasil dari proses cleaning data yang dilakukan untuk memastikan hanya data yang relevan yang di gunakan dalam analisis. Dalam proses ini, game yang memiliki data tidak lengkap atau tidak relevan di hapus, dan atribut data seperti genre dan angka penjualan di berbagai wilayah diklarifikasi. Data yang bersih ini memungkinkan algoritma K-Medoids untuk bekerja lebih efektif dalam mengelompokkan game berdasarkan pola penjualan.

Transformasi Data

Setelah data di cleaning , langkah selanjutnya adalah transformasi data. Pada data ini kolom genre yang awal nya non-numerik di ubah menjadi numerik dan setiap genre berlambangkan angka. Berikut ini adalah genre yang di transformasikan menjadi numerik.

Sports = 10

Platform = 4

Racing = 6

Role-Playing = 7

Puzzle = 5

Shooter = 9

Fighting = 2

Simulation = 8

Action = 0

Adventure = 1

Misc = 3

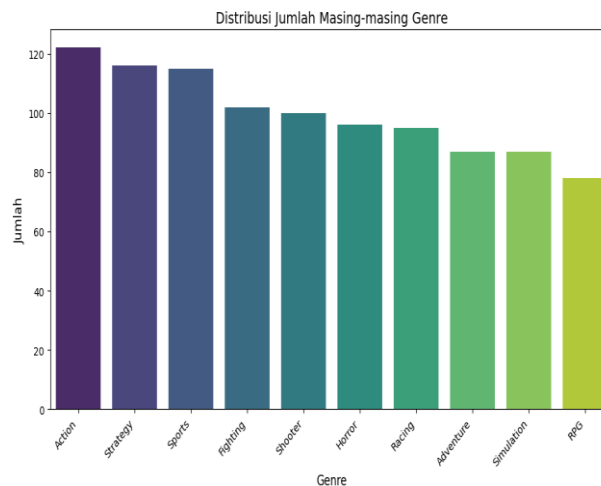
Strategy = 11

hasil transformasi data di sajikan pada tabel 2 dan gambar 3.

Tabel 2. Hasil transformasi data

Rank	Genre_game	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
1	0	877.83	525	159.95	187.38	1751.18
2	10	683.35	376.85	135.37	134.97	1330.93
3	8	582.6	313.27	38.28	102.69	1037.37
4	7	327.28	188.06	352.31	59.61	927.37
5	4	447.05	201.63	130.77	51.59	831.37

Pada tabel 2 menunjukkan hasil transformasi data, di mana kategori genre yang semula berbentuk teks telah di konversi menjadi representasi numerik untuk memudahkan proses analisis dan clustering. Setiap genre diberikan kode angka tertentu untuk memastikan bahwa data dapat di proses oleh algoritma yang di gunakan. Untuk distribusi jumlah masing-masing genre dapat di lihat pada gambar 3.

**Gambar 3.** Visualisasi hasil distribusi jumlah masing-masing genre.

Dari grafik ini dapat di simpulkan bahwa genre game *action* memiliki jumlah game terbanyak, di susul oleh *strategy dan sports*, sedangkan *RPG* memiliki jumlah game paling sedikit. Dengan adanya grafik ini dapat digunakan untuk memahami distribusi genre dalam dataset, yang dapat membantu dalam analisis tren industri game atau strategi untuk pemasaran berdasarkan popularitas genre.

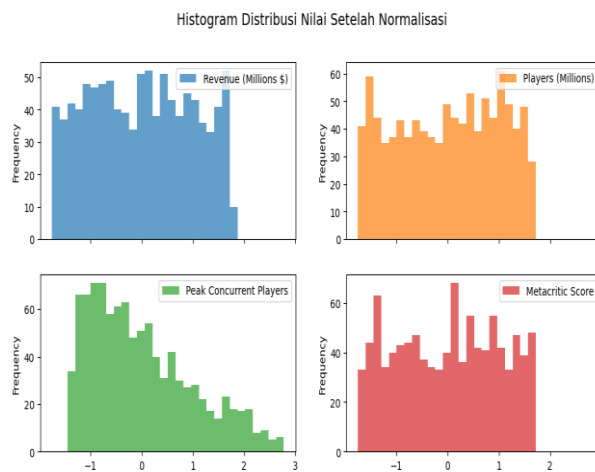
Normalisasi Data

Setelah dilakukan transformasi data, selanjutnya adalah normalisasi data dengan menggunakan metode Min-Max normalization yang memiliki range nilai dari 0-1. Hasil dari normalisasi data terdapat pada tabel 3.

Tabel 3. Hasil normalisasi data

Rank	Genre_game	NA_Sales (Normalized)	EU_Sales (Normalized)	JP_Sales (Normalized)	Other_Sales (Normalized)	Global_Sales (Normalized)
1	0	1.000	1.000	0.394	1.000	1.000
2	10	0.779	0.718	0.338	0.72	0.76
3	8	0.664	0.597	0.059	0.547	0.592
4	7	0.373	0.358	1.000	0.318	0.53
5	4	0.509	0.384	0.319	0.275	0.474

Pada tabel 3 ini hasil normalisasi data memastikan bahwa tidak ada variabel yang mendominasi proses clustering karena perbedaan skala yang signifikan. Misalnya, nilai penjualan tertinggi pada NA Sales di ubah menjadi 1, sementara nilai terendah mendekati 0, hal ini memungkinkan algoritma untuk lebih mudah membandingkan data antar game tanpa bias dari perbedaan skala. Untuk distribusi nilai setelah normalisasi data dapat di lihat pada gambar 4.



Gambar 4. Visualisasi hasil distribusi nilai setelah normalisasi

Pada histogram ini distribusi nilai setelah normalisasi dengan metode min-max berhasil menyelaraskan skala variabel sehingga tidak ada fitur yang mendominasi proses clustering sehingga analisis data menjadi lebih akurat dan representatif terhadap pola penjualan di berbagai wilayah.

Analisis Hasil Preprocessing

Hasil preprocessing data yang di lakukan memberikan dampak signifikan terhadap kualitas clustering dengan meningkat kan akurasi dalam mengenali pola penjualan berdasarkan genre. Efisiensi algoritma K-Medoids juga meningkat karna data yang sudah di proses lebih terorganisir, mempercepat iterasi tanpa kehilangan informasi penting. Dengan adanya preprocessing yang tepat, hasil clustering menjadi lebih representatif terhadap tren penjualan game di berbagai wilayah, memberikan wawasan yang lebih akurat bagi pengembang dan penerbit game dalam menyusun strategi pemasaran dan pengembangan produk.

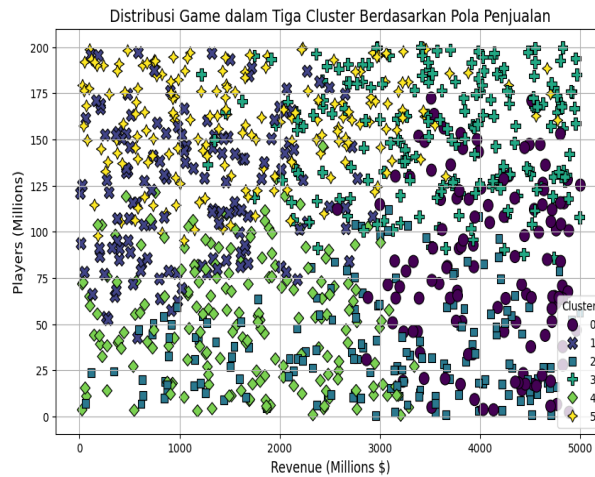
K-Medoids

Mengelompokkan genre game yang dimana pengelompokan ini berdasarkan kemiripan dalam penjualan di berbagai wilayah yang ada. Algoritma yang di gunakan adalah K-Medoids dengan nilai $K = 3$, Dengan K-Medoids terbentuk 3 cluster yang menghasilkan game dengan penjualan tinggi di wilayah tertentu. Hasil clustering K-Medoids di sajikan pada tabel 4.

Tabel 4. Hasil clustering K-medoid

Cluster	Genre yang Dominan	Pola Penjualan
1	Sports, Racing, Platform	Tinggi di Amerika Serikat dan Uni Eropa
2	Action, Shooter, Fighting	Tinggi secara global tanpa dominasi di satu wilayah tertentu
3	RPG, Adventure, Strategy	Tinggi di Jepang

Berdasarkan tabel 4 dapat diketahui bahwa cluster 1 adalah game dengan genre *Sports*, *Racing*, dan *Platform* yang sangat populer pada pasar Amerika serikat dan uni eropa, pada cluster 2 adalah game dengan genre game *Action*, *Shooter*, dan *Fighting* dengan kesuksesan besar penjualan berada pada wilayah amerika utara, dan untuk cluster 3 adalah game dengan genre *RPG*, *Adventure*, dan *Strategy* yang memiliki pasar penjualan yang sangat kuat di wilayah jepang dan sekitar nya. Visualisasi distribusi game dalam tiga cluster berdasarkan pola penjualan dapat di lihat pada gambar 5.



Gambar 5. Visualisasi distribusi game dalam tiga cluster.

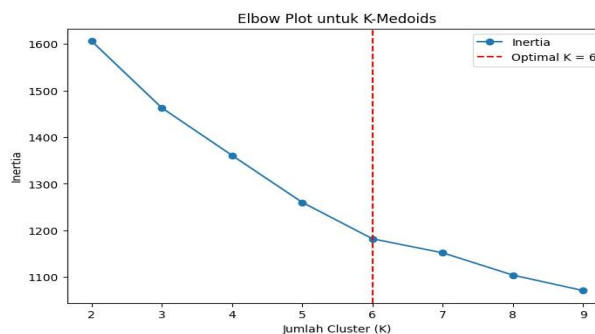
Pada visualisasi ini menunjukkan perbedaan pada preferensi pasar di suatu wilayah yang sebagai berikut :

- Cluster 1 (Sports, Racing, Platform) laris di Amerika Serikat dan Uni Eropa, mencerminkan minat pada game kompetitif dan kasual.
- Cluster 2 (Action, Shooter, Fighting) memiliki penjualan merata secara global, menandakan daya tarik universal.
- Cluster 3 (RPG, Adventure, Strategy) dominan di Jepang, menunjukkan preferensi terhadap game dengan alur cerita mendalam dan strategi kompleks.

Dengan distribusi ini, pengembang dan penerbit game dapat lebih memahami preferensi pasar di setiap wilayah, sehingga dapat menyusun strategi pemasaran dan pengembangan produk yang lebih tepat sasaran.

Evaluation Elbow Plot

Elbow plot digunakan untuk mengukur kualitas dari hasil dari clustering yang ada. Elbow plot digunakan sesuai dengan jumlah K setelah hasil dari perhitungan K-Medoids di peroleh. Visualisasi elbow plot dapat di lihat pada gambar 2.



Gambar 4. Visualisasi elbow plot

Dalam penelitian ini, hasil Elbow Plot menunjukkan bahwa $k = 6$ adalah jumlah jumlah cluster yang optimal. Pemilihan k yang lebih kecil seperti $k = 3$ dan $k = 4$ dapat menyebabkan dataset tidak cukup terproses dan tersegmentasi sehingga dapat menyebabkan pola penjualan yang lebih spesifik tidak terlihat dengan jelas. Sebaliknya, dengan $k = 7$ atau lebih dapat menyebabkan overfitting, di mana model terlalu spesifik terhadap data latih dan kehilangan generalisasi terhadap data baru. Pada hasil Elbow Plot ini dapat disimpulkan bahwa jumlah cluster yang optimal dan berperan krusial untuk mendapatkan insight yang tepat dari dataset yang dimiliki adalah $k = 6$. Dengan pemilihan $k = 6$ memungkinkan keseimbangan antara akurasi dan kompleksitas, serta memberikan manfaat langsung dalam pemahaman pada pola penjualan dan pengambilan keputusan bisnis berbasis data.

4. KESIMPULAN

Dalam penelitian yang dilakukan dapat disimpulkan bahwa setiap genre memiliki pola penjualan yang berbeda di berbagai wilayah dan menunjukkan pentingnya strategi dalam pemasaran produk game yang akan di pasarkan. Implikasi praktis dari penelitian ini adalah bahwa pengembang dan penerbit game dapat menggunakan wawasan ini untuk mengoptimalkan strategi pemasaran, menentukan wilayah prioritas distribusi, serta menyesuaikan fitur game dengan preferensi pasar yang berbeda. Untuk meningkatkan akurasi dan relevansi penelitian, disarankan untuk menggunakan dataset dan mencakup periode waktu yang lebih luas guna menangkap tren industri yang lebih jelas, penggunaan teknik machine learning lain seperti Random Forest atau Neural Networks juga dapat dipertimbangkan untuk membandingkan efektivitas model clustering yang digunakan dalam penelitian ini.

DAFTAR REFERENSI

- Abril Geralin. (2024). *Industri Gaming, Perkembangan Pesat dan Dampaknya di 2024*. LEET MEDIA.
- Ahmad Harmain, Paiman, P., Kurniawan, H., Kusriani, K., & Dina Maulina. (2022). Normalisasi Data Untuk Efisiensi K-Means Pada Pengelompokan Wilayah Berpotensi Kebakaran Hutan Dan Lahan Berdasarkan Sebaran Titik Panas. *TEKNIMEDIA: Teknologi Informasi Dan Multimedia*, 2(2), 83–89. <https://doi.org/10.46764/teknimedia.v2i2.49>
- Aripin, S., Gulo, T., & Angin, G. P. N. S. P. (2023). Penerapan Metode K-Medoids Clustering Pada Penanganan Kasus Demam Berdarah. *BEES: Bulletin of Electrical and Electronics Engineering*, 3(3), 139–146. <https://doi.org/10.47065/bees.v3i3.3173>

- Bahri, S., & Midyanti, D. M. (2023). Penerapan Metode K-Medoids untuk Pengelompokan Mahasiswa Berpotensi Drop Out. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(1), 165–172. <https://doi.org/10.25126/jtiik.20231016643>
- Bhakti, M., Irawan, J. D., & Rudhistiar, D. (2024a). PENGELOMPOKAN GAME BERDASARKAN DATA TOP SELLER PADA WEBSITE " STEAM " MENGGUNAKAN METODE K-MEDOIDS. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(5), 8242–8249.
- Bhakti, M., Irawan, J. D., & Rudhistiar, D. (2024b). PENGELOMPOKAN GAME BERDASARKAN DATA TOP SELLER PADA WEBSITE " STEAM " MENGGUNAKAN METODE K-MEDOIDS. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(5), 8242–8249.
- Dwiatmoko, F. (2021). Preprocessing Tranformasi Data Menggunakan K-Means Clustering. *Explore*, 11(2), 141. <https://doi.org/10.35200/explore.v11i2.544>
- Hastari, D., Nurunnisa, F., Winanda, S., & Dwi Aprillia, D. (2023). Penerapan Algoritma K-Means dan K-Medoids untuk Mengelompokkan Data Negara Berdasarkan Faktor Sosial-Ekonomi dan Kesehatan. *SENTIMAS: Seminar Nasional Penelitian Dan Pengabdian Masyarakat*, 274–281.
- Juliandri Saputra, Muhammad Iqbal Al Aksha, Lily Maryani, Gilang, & Rahmaddeni. (2024). Analisis Perbandingan Efektivitas Metode Fuzzy C-Means dan K-Means dalam Mengelompokkan Buku Berdasarkan Frekuensi Peminjaman di Perpustakaan SMKN 1 Mandau. *Explore*, 14(2), 87–92. <https://doi.org/10.35200/ex.v14i2.121>
- Kamila, I., Khairunnisa, U., & Mustakim, M. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 5(1), 119. <https://doi.org/10.24014/rmsi.v5i1.7381>
- Muhammad, O., Waloejo, B. S., & ... (2021). Strategi Pengembangan Industri Kreatif Subsektor Games Di Kota Malang. *Planning for Urban Region ...*, 10, 69–76.
- Rahayu, A. (2023). *Perkembangan Industri Game Dunia, 48% Pemainnya Berasal dari Asia-Pasifik*. Good Stat.
- Ramadhani, J., Safitra Anugraha, Y., Fauzan, A., & Efrizoni, L. (2024a). Perbandingan Algoritma K-Means Clustering Dan K-Medoids Dalam Mengelompokkan Tingkat Kemiskinan Di Provinsi Riau. *Jurnal Jaringan Sistem Informasi Robotik (Jsir)*, 8(1), 114–125.
- Ramadhani, J., Safitra Anugraha, Y., Fauzan, A., & Efrizoni, L. (2024b). Perbandingan Algoritma K-Means Clustering Dan K-Medoids Dalam Mengelompokkan Tingkat Kemiskinan Di Provinsi Riau. *Jurnal Jaringan Sistem Informasi Robotik (Jsir)*, 8(1), 114–125.
- Rezky Pratama, A., Maulana, B., Didho Rianda, R., & El Hasyim, S. (2023). Comparison of K-Means and K-Medoids Algorithms for Grouping Video Game Sales Data in North America. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, 3(2), 111–118. <https://doi.org/10.57152/ijirse.v3i2.898>

- Simanjuntak, D. S. M., Gunawan, I., Sumarno, S., Poningsih, P., & Sari, I. P. (2023). Penerapan Algoritma K-Medoids Untuk Pengelompokan Pengangguran Umur 25 tahun Keatas Di Sumatera Utara. *Jurnal Krisnadana*, 2(2). <https://doi.org/10.58982/krisnadana.v2i2.264>
- Tampubolon, H. D., Suhada, S., Safii, M., Solikhun, S., & Suhendro, D. (2021). Penerapan Algoritma K-Means dan K-Medoids Clustering untuk Mengelompokkan Tindak Kriminalitas Berdasarkan Provinsi. *Jurnal Ilmu Komputer Dan Teknologi*, 2(2), 6–12. <https://doi.org/10.35960/ikomti.v2i2.703>