

Pemanfaatan Data Mining untuk Customer Intelligence Dalam Segmentasi Penjualan Video Game Menggunakan Algoritma K-Means

Muhammad Erlangga^{1*}, Krisna Arya Vashesa², Aditya Ibanez Saputra³

^{1,2,3}Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Pamulang, Tangerang Selatan, Indonesia

^{1*}erllangga20@gmail.com, ²krisnaarya124@gmail.com, ³ibanezaditya06@gmail.com
(Email Correspondence: erllangga20@gmail.com)

Received: 4 April 2026 | Revised: 8 April 2026 | Accepted: 30 April 2026

Abstrak

Penelitian ini bertujuan untuk menerapkan teknik data mining dalam mendukung customer intelligence melalui segmentasi penjualan video game menggunakan algoritma K-Means. Dataset yang digunakan mencakup data penjualan video game dari berbagai wilayah, yaitu Amerika Utara, Eropa, Jepang, dan wilayah lainnya, yang dianalisis menggunakan pendekatan CRISP-DM yang meliputi tahapan business understanding, data understanding, data preparation, modeling, dan evaluation. Proses data preparation dilakukan dengan menghapus missing value dan data duplikat untuk meningkatkan kualitas data, kemudian dilakukan pemodelan menggunakan algoritma K-Means dengan penentuan jumlah cluster optimal menggunakan metode Elbow. Hasil penelitian menunjukkan bahwa data dapat dikelompokkan menjadi tiga cluster utama yang merepresentasikan tingkat penjualan rendah, menengah, dan tinggi, dimana sebagian besar data berada pada cluster dengan penjualan rendah, sementara hanya sebagian kecil yang berada pada cluster dengan penjualan tinggi, bahkan ditemukan satu data dengan penjualan yang sangat tinggi yang membentuk cluster tersendiri. Temuan ini menunjukkan bahwa distribusi penjualan video game bersifat tidak merata dan cenderung didominasi oleh sedikit produk yang sangat sukses, sehingga hasil analisis ini dapat memberikan insight yang berguna bagi perusahaan dalam merancang strategi bisnis yang lebih efektif berbasis customer intelligence.

Kata kunci: *Data Mining, Customer Intelligence, K-Means, Clustering*

Abstract

This study aims to apply data mining techniques to support customer intelligence through the segmentation of video game sales using the K-Means algorithm. The dataset used consists of global video game sales data obtained from Kaggle, covering various attributes such as game name, platform, release year, genre, and regional sales including North America, Europe, Japan, and other regions, with a time span from the early release period up to 2017. The research adopts the CRISP-DM framework, which includes business understanding, data understanding, data preparation, modeling, and evaluation. Data preparation involves cleaning processes such as removing missing values and duplicate records to ensure data quality. The K-Means algorithm is then implemented, with the optimal number of clusters determined using the Elbow method. The results show that the data can be grouped into three main clusters representing low, medium, and high sales levels, where the majority of video games fall into the low-sales cluster, while only a small number achieve high sales, including one extreme case with significantly higher sales than others. These findings indicate that the distribution of video game sales is highly imbalanced and dominated by a few successful products. Therefore, this study demonstrates that data mining can provide valuable insights to support business strategies through a customer intelligence approach.

Keywords: *Data Mining, Customer Intelligence, K-Means, Clustering*

1. PENDAHULUAN

Perkembangan teknologi informasi dan transformasi digital telah mendorong pertumbuhan berbagai sektor industri, termasuk industri video game yang saat ini menjadi salah satu industri hiburan digital terbesar di dunia. Video game tidak lagi dipandang hanya sebagai sarana hiburan, tetapi telah berkembang menjadi produk digital yang memiliki nilai ekonomi tinggi dan mampu menciptakan ekosistem bisnis yang kompleks. Pertumbuhan pengguna internet, perkembangan perangkat keras komputer dan konsol, serta meningkatnya penggunaan perangkat mobile telah memperluas jangkauan pasar video game secara global.

Kondisi tersebut menyebabkan volume transaksi dan data penjualan video game terus meningkat dari tahun ke tahun sehingga menghasilkan kumpulan data yang sangat besar dan beragam. Dalam lingkungan bisnis yang kompetitif, data penjualan tersebut memiliki nilai strategis karena dapat dimanfaatkan sebagai dasar dalam proses pengambilan keputusan yang lebih akurat dan berbasis data.

Seiring dengan meningkatnya volume data yang dihasilkan oleh aktivitas bisnis, perusahaan menghadapi tantangan dalam mengolah dan mengekstraksi informasi yang bernilai dari data tersebut. Data yang hanya disimpan tanpa dianalisis tidak akan memberikan manfaat yang optimal bagi organisasi. Oleh karena itu, diperlukan suatu pendekatan yang mampu mengidentifikasi pola, hubungan, serta pengetahuan tersembunyi yang terdapat dalam kumpulan data berukuran besar. Dalam konteks bisnis modern, pemanfaatan data mining sering dikaitkan dengan customer intelligence yang berfokus pada pengumpulan dan analisis data pelanggan guna memahami perilaku konsumen secara lebih mendalam [1]. Melalui customer intelligence, perusahaan dapat mengidentifikasi preferensi pelanggan, melakukan segmentasi pasar, menyusun strategi pemasaran yang lebih tepat sasaran, serta meningkatkan kualitas produk dan layanan yang ditawarkan.

Salah satu sumber data yang dapat dimanfaatkan untuk mendukung customer intelligence adalah data penjualan video game. Data penjualan tidak hanya menunjukkan tingkat keberhasilan suatu produk, tetapi juga dapat menggambarkan karakteristik pasar pada berbagai wilayah geografis. Perbedaan budaya, preferensi pemain, kondisi ekonomi, dan perkembangan teknologi menyebabkan pola penjualan video game pada setiap wilayah memiliki karakteristik yang berbeda. Meskipun demikian, pemanfaatan data penjualan video game untuk mengidentifikasi pola segmentasi pasar masih belum dilakukan secara optimal. Sebagian besar data penjualan hanya digunakan sebagai laporan historis tanpa dianalisis lebih lanjut untuk menemukan karakteristik kelompok pasar yang terbentuk [2]. Akibatnya, perusahaan belum memperoleh gambaran yang jelas mengenai distribusi penjualan pada berbagai wilayah serta perbedaan karakteristik pasar yang ada. Kondisi tersebut menunjukkan adanya kebutuhan untuk melakukan analisis yang mampu mengelompokkan data penjualan berdasarkan kemiripan karakteristik sehingga informasi yang terkandung dalam data dapat dimanfaatkan secara lebih maksimal untuk mendukung pengambilan keputusan bisnis [3]. Untuk memperoleh informasi yang bermakna dari data penjualan video game, diperlukan metode analisis yang mampu mengelompokkan data berdasarkan karakteristik tertentu. Salah satu teknik data mining yang banyak digunakan adalah clustering, yaitu metode unsupervised learning yang bertujuan mengelompokkan data berdasarkan tingkat kemiripan karakteristik tanpa memerlukan label kelas. Di antara berbagai algoritma clustering yang tersedia, K-Means merupakan salah satu algoritma yang paling populer karena memiliki proses komputasi yang relatif cepat, mudah diimplementasikan, serta mampu menangani data dalam jumlah besar dengan baik [4]. Algoritma ini telah banyak digunakan dalam segmentasi pelanggan, analisis pasar, dan pengelompokan data penjualan untuk menghasilkan informasi yang mendukung pengambilan keputusan bisnis. Dalam penelitian ini, jumlah cluster optimal ditentukan menggunakan metode Elbow yang memanfaatkan nilai inerti untuk menemukan titik siku (elbow point), sedangkan kualitas hasil clustering dievaluasi menggunakan Silhouette Score guna memastikan bahwa cluster yang terbentuk memiliki tingkat pemisahan yang baik.

Berdasarkan permasalahan tersebut, penelitian ini menerapkan algoritma K-Means untuk melakukan segmentasi data penjualan video game menggunakan dataset yang diperoleh dari Kaggle. Dataset tersebut berisi informasi penjualan video game pada beberapa wilayah utama, yaitu Amerika Utara, Eropa, Jepang, dan wilayah lainnya. Penelitian dilaksanakan menggunakan framework CRISP-DM (Cross Industry Standard Process for Data Mining) agar proses analisis data dapat dilakukan secara sistematis dan terstruktur [5]. Melalui penelitian ini diharapkan dapat diperoleh segmentasi penjualan video game yang mampu menggambarkan pola distribusi pasar berdasarkan wilayah penjualan. Hasil segmentasi tersebut diharapkan dapat memberikan insight yang bermanfaat bagi perusahaan dalam mendukung implementasi customer intelligence, khususnya dalam penyusunan strategi pemasaran, pengembangan produk, dan pengambilan keputusan bisnis berbasis data sehingga perusahaan dapat meningkatkan daya saing di tengah persaingan industri video game yang semakin kompetitif.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining untuk melakukan segmentasi data penjualan video game berdasarkan karakteristik penjualannya pada berbagai wilayah. Dataset yang digunakan diperoleh dari platform Kaggle dan berisi informasi penjualan video game pada beberapa wilayah utama, yaitu Amerika Utara (NA_Sales), Eropa (EU_Sales), Jepang (JP_Sales), dan wilayah lainnya (Other_Sales). Teknik pengambilan sampel yang digunakan adalah total sampling, yaitu seluruh data yang memenuhi kriteria penelitian digunakan sebagai sampel analisis. Setelah dilakukan proses pembersihan data, diperoleh sebanyak 16.291 data yang digunakan dalam penelitian. Proses analisis

dilakukan menggunakan algoritma K-Means Clustering dengan framework CRISP-DM (Cross Industry Standard Process for Data Mining) sebagai pedoman pelaksanaan penelitian. Framework ini dipilih karena menyediakan tahapan yang sistematis dalam proses pengolahan data, mulai dari identifikasi permasalahan hingga interpretasi hasil analisis [6]. Adapun tahapan penelitian yang digunakan terdiri atas Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment.

2.1 Business Understanding

Penelitian diawali dengan mengidentifikasi permasalahan yang melatarbelakangi kebutuhan analisis data penjualan video game. Meskipun industri video game menghasilkan data penjualan dalam jumlah besar, informasi yang terkandung di dalamnya belum selalu dimanfaatkan secara optimal untuk mendukung pengambilan keputusan bisnis [7]. Kondisi tersebut mendorong perlunya proses segmentasi data guna memperoleh pemahaman yang lebih baik mengenai karakteristik pasar pada berbagai wilayah penjualan. Oleh karena itu, penelitian ini difokuskan pada pengelompokan data penjualan video game untuk menghasilkan informasi yang dapat mendukung implementasi customer intelligence dan penyusunan strategi bisnis berbasis data.

2.2 Data Understanding

Pemahaman terhadap karakteristik data menjadi langkah penting sebelum proses analisis dilakukan. Dataset yang digunakan diperoleh dari platform Kaggle dan berisi informasi mengenai penjualan video game pada berbagai wilayah pasar global. Pada tahap ini dilakukan eksplorasi data untuk mengetahui struktur dataset, jenis atribut yang tersedia, serta kondisi kualitas data yang akan digunakan dalam penelitian. Atribut NA_Sales, EU_Sales, JP_Sales, dan Other_Sales dipilih karena mampu merepresentasikan performa penjualan video game pada masing-masing wilayah sehingga relevan dengan tujuan segmentasi yang ingin dicapai.

2.3 Data Preparation

Sebelum memasuki proses pemodelan, data terlebih dahulu dipersiapkan agar memenuhi standar kualitas yang dibutuhkan dalam analisis clustering. Proses yang dilakukan meliputi identifikasi dan penghapusan data yang memiliki nilai kosong (missing value), pemeriksaan kemungkinan adanya data duplikat, serta pemilihan atribut yang sesuai dengan tujuan penelitian. Selain itu, dilakukan transformasi data untuk memastikan data berada dalam kondisi yang siap digunakan pada proses pengelompokan. Tahap ini memiliki peran penting karena kualitas data yang baik akan berpengaruh langsung terhadap akurasi dan validitas hasil clustering yang dihasilkan [8], [9].

2.4 Modeling

Proses pemodelan dalam penelitian ini dilakukan menggunakan algoritma K-Means Clustering untuk mengelompokkan data penjualan video game berdasarkan tingkat kemiripan karakteristik penjualan pada setiap wilayah pemasaran. Algoritma K-Means dipilih karena memiliki kemampuan yang baik dalam mengidentifikasi pola pengelompokan pada data tanpa label (unsupervised learning) serta mampu mengolah data dalam jumlah besar dengan tingkat efisiensi yang tinggi [10]. Mekanisme kerja algoritma ini dimulai dengan menentukan sejumlah pusat cluster (centroid) yang kemudian digunakan sebagai acuan dalam mengelompokkan data. Setiap data akan ditempatkan pada cluster yang memiliki jarak terdekat terhadap centroid, kemudian posisi centroid diperbarui secara iteratif hingga diperoleh kondisi yang stabil dan tidak terjadi perubahan keanggotaan cluster secara signifikan. Melalui proses tersebut, data yang memiliki karakteristik penjualan serupa akan terkumpul dalam kelompok yang sama sehingga pola distribusi penjualan dapat diamati dengan lebih jelas.

Sebelum proses clustering dilakukan, terlebih dahulu ditentukan jumlah cluster yang optimal menggunakan metode Elbow. Metode ini diterapkan dengan menganalisis nilai inerti atau Sum of Squared Error (SSE) pada beberapa variasi jumlah cluster yang diuji [11]. Nilai inerti menggambarkan total jarak antara setiap data dengan centroid pada cluster tempat data tersebut berada, sehingga semakin kecil nilai yang dihasilkan menunjukkan semakin baik tingkat homogenitas dalam cluster. Penentuan jumlah cluster optimal dilakukan dengan mengamati titik siku (elbow point) pada grafik Elbow, yaitu titik ketika penurunan nilai inerti mulai melambat secara signifikan. Jumlah cluster yang diperoleh dari metode tersebut kemudian digunakan sebagai parameter utama dalam proses K-Means sehingga hasil pengelompokan yang dihasilkan mampu

merepresentasikan struktur data secara lebih efektif dan memudahkan proses interpretasi terhadap pola penjualan video game [12].

2.5 Evaluation

Evaluasi dilakukan untuk mengukur kualitas hasil clustering yang diperoleh dari proses pemodelan serta memastikan bahwa kelompok yang terbentuk mampu merepresentasikan karakteristik data secara optimal. Pada tahap ini, setiap cluster dianalisis berdasarkan tingkat kemiripan data penjualan di dalam kelompok yang sama dan perbedaannya dengan kelompok lain [13]. Proses evaluasi bertujuan untuk mengidentifikasi sejauh mana hasil pengelompokan mampu membentuk cluster yang memiliki karakteristik yang homogen di dalam kelompok serta heterogen antar kelompok. Selain itu, evaluasi juga dilakukan untuk memahami pola distribusi data yang terbentuk sehingga karakteristik masing-masing cluster dapat diinterpretasikan dengan lebih jelas [14]. Hasil evaluasi menjadi dasar dalam menilai efektivitas model clustering yang digunakan serta mendukung proses analisis dan pengambilan keputusan berdasarkan pola penjualan video game yang berhasil diidentifikasi.

2.6 Deployment

Tahap akhir penelitian difokuskan pada interpretasi hasil clustering untuk menghasilkan informasi yang bernilai bagi kebutuhan bisnis. Setiap cluster yang terbentuk dianalisis berdasarkan karakteristik penjualannya sehingga dapat diketahui pola distribusi pasar pada berbagai wilayah. Informasi tersebut selanjutnya digunakan untuk menghasilkan insight yang mendukung pengambilan keputusan, khususnya dalam penyusunan strategi pemasaran, penentuan target pasar, dan pengembangan produk yang lebih sesuai dengan kebutuhan konsumen [15]. Dengan demikian, hasil penelitian tidak hanya menghasilkan pengelompokan data, tetapi juga memberikan manfaat praktis dalam mendukung implementasi customer intelligence..

3. HASIL DAN PEMBAHASAN

3.1 Hasil Data Preparation

Tahap data preparation dilakukan untuk memastikan bahwa data yang digunakan dalam proses clustering memiliki kualitas yang baik dan sesuai untuk dianalisis. Proses ini diawali dengan pemeriksaan terhadap kelengkapan data untuk mengidentifikasi keberadaan missing value pada setiap atribut. Hasil pemeriksaan menunjukkan bahwa terdapat sejumlah data yang memiliki nilai kosong sehingga berpotensi memengaruhi proses pengelompokan dan menghasilkan interpretasi yang kurang akurat. Oleh karena itu, data yang tidak memenuhi kriteria kelengkapan dihapus dari dataset agar kualitas data yang digunakan tetap terjaga. Setelah proses pembersihan dilakukan, jumlah data berkurang dari 16.598 data menjadi 16.291 data. Selain itu, dilakukan pemeriksaan terhadap kemungkinan adanya data duplikat untuk menghindari terjadinya redundansi yang dapat memengaruhi hasil analisis. Hasil pemeriksaan menunjukkan bahwa tidak ditemukan data duplikat sehingga seluruh data yang tersisa dapat langsung digunakan pada tahap berikutnya. Dataset yang telah melalui proses pembersihan ini dinilai lebih representatif dan siap digunakan dalam proses pemodelan menggunakan algoritma K-Means.

Data Awal:

	Rank	Name	Platform	Year	Genre	Publisher
0	1	Wii Sports	Wii	2006.0	Sports	Nintendo
1	2	Super Mario Bros.	NES	1985.0	Platform	Nintendo
2	3	Mario Kart Wii	Wii	2008.0	Racing	Nintendo
3	4	Wii Sports Resort	Wii	2009.0	Sports	Nintendo
4	5	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo

	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	41.49	29.02	3.77	8.46	82.74
1	29.08	3.58	6.81	0.77	40.24
2	15.85	12.88	3.79	3.31	35.82
3	15.75	11.01	3.28	2.96	33.00
4	11.27	8.89	10.22	1.00	31.37

Gambar 1. Data Sebelum Cleaning

Data Setelah Cleaning:

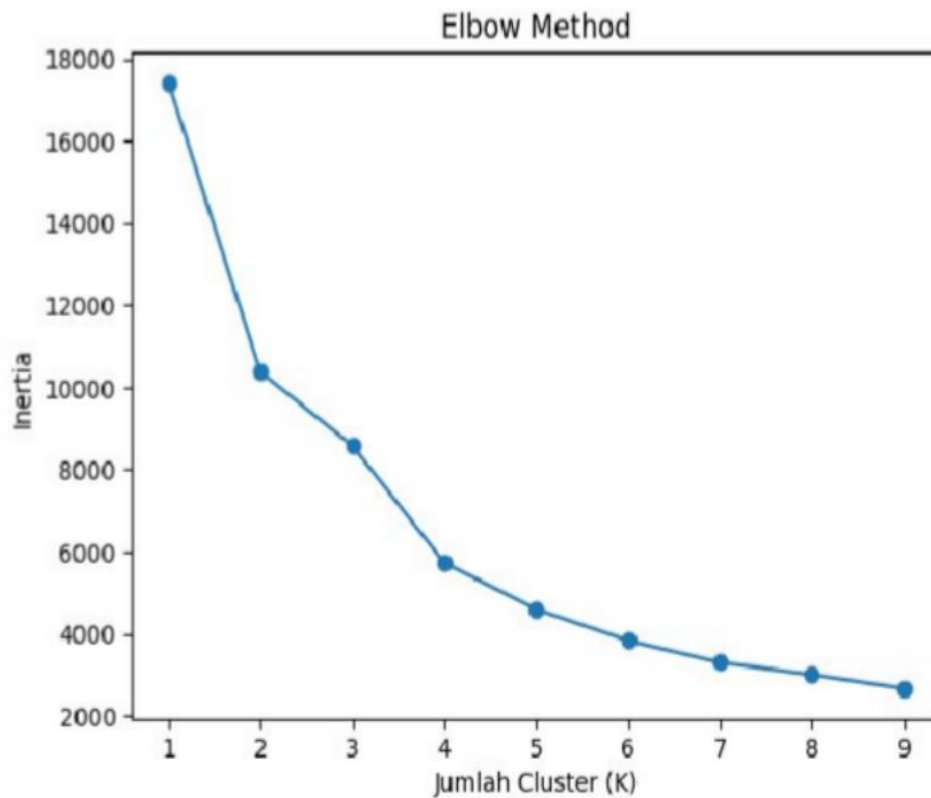
	Rank	Name	Platform	Year	Genre	Publisher
0	1	Wii Sports	Wii	2006.0	Sports	Nintendo
1	2	Super Mario Bros.	NES	1985.0	Platform	Nintendo
2	3	Mario Kart Wii	Wii	2008.0	Racing	Nintendo
3	4	Wii Sports Resort	Wii	2009.0	Sports	Nintendo
4	5	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo

	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	41.49	29.02	3.77	8.46	82.74
1	29.08	3.58	6.81	0.77	40.24
2	15.85	12.88	3.79	3.31	35.82
3	15.75	11.01	3.28	2.96	33.00
4	11.27	8.89	10.22	1.00	31.37

Gambar 2. Data Setelah Cleaning

3.2 Penentuan Jumlah Cluster Menggunakan Metode Elbow

Penentuan jumlah cluster merupakan salah satu tahapan penting dalam penerapan algoritma K-Means karena jumlah cluster yang dipilih akan memengaruhi kualitas segmentasi yang dihasilkan. Pada penelitian ini, jumlah cluster optimal ditentukan menggunakan metode Elbow dengan menganalisis perubahan nilai inertia pada beberapa variasi jumlah cluster. Nilai inertia digunakan untuk mengukur tingkat kedekatan data terhadap centroid dalam cluster yang sama, sehingga nilai yang lebih rendah menunjukkan tingkat homogenitas yang lebih baik.



Gambar 3. Grafik Elbow Method

Berdasarkan grafik Elbow yang ditunjukkan pada Gambar 3, penurunan nilai inertia terlihat cukup signifikan pada rentang $K=1$ hingga $K=3$. Setelah titik tersebut, laju penurunan mulai melambat dan tidak menunjukkan perubahan yang berarti. Pola tersebut membentuk titik siku (elbow point) pada $K=3$ yang mengindikasikan bahwa tiga cluster merupakan jumlah yang paling optimal untuk merepresentasikan struktur data. Dengan demikian, penelitian ini menggunakan tiga cluster sebagai dasar dalam proses pengelompokan data penjualan video game.

3.3 Evaluasi Clustering Menggunakan Silhouette Score

Kualitas hasil clustering dievaluasi menggunakan Silhouette Score untuk mengetahui tingkat kohesi dan separasi dari cluster yang terbentuk. Metode ini mengukur seberapa dekat suatu data dengan anggota dalam cluster yang sama dibandingkan dengan anggota pada cluster lainnya. Hasil pengujian menunjukkan bahwa model K-Means dengan tiga cluster menghasilkan nilai Silhouette Score sebesar 0,915. Nilai tersebut berada sangat dekat dengan angka 1, yang menunjukkan bahwa cluster yang terbentuk memiliki tingkat pemisahan yang baik serta karakteristik yang relatif homogen di dalam masing-masing kelompok. Dengan demikian, hasil evaluasi menunjukkan bahwa model clustering yang digunakan mampu merepresentasikan pola data penjualan video game secara optimal.

3.4 Hasil Clustering

Berdasarkan proses clustering menggunakan algoritma K-Means, data penjualan video game berhasil dikelompokkan ke dalam tiga cluster yang memiliki karakteristik berbeda. Pengelompokan dilakukan berdasarkan atribut penjualan pada wilayah Amerika Utara, Eropa, Jepang, dan wilayah lainnya. Hasil segmentasi menunjukkan adanya variasi tingkat performa penjualan yang cukup jelas di antara kelompok yang terbentuk, sehingga memungkinkan dilakukan identifikasi pola pasar yang lebih terstruktur. Cluster 0 merupakan kelompok dengan jumlah anggota terbesar dan didominasi oleh video game yang memiliki tingkat penjualan relatif rendah pada seluruh wilayah. Cluster 1 berisi video game dengan tingkat penjualan yang lebih tinggi dan menunjukkan performa pasar yang lebih baik dibandingkan Cluster 0. Sementara itu, Cluster 2 hanya terdiri dari satu data yang memiliki nilai penjualan jauh lebih tinggi dibandingkan seluruh data lainnya. Keberadaan cluster tersebut menunjukkan adanya data dengan

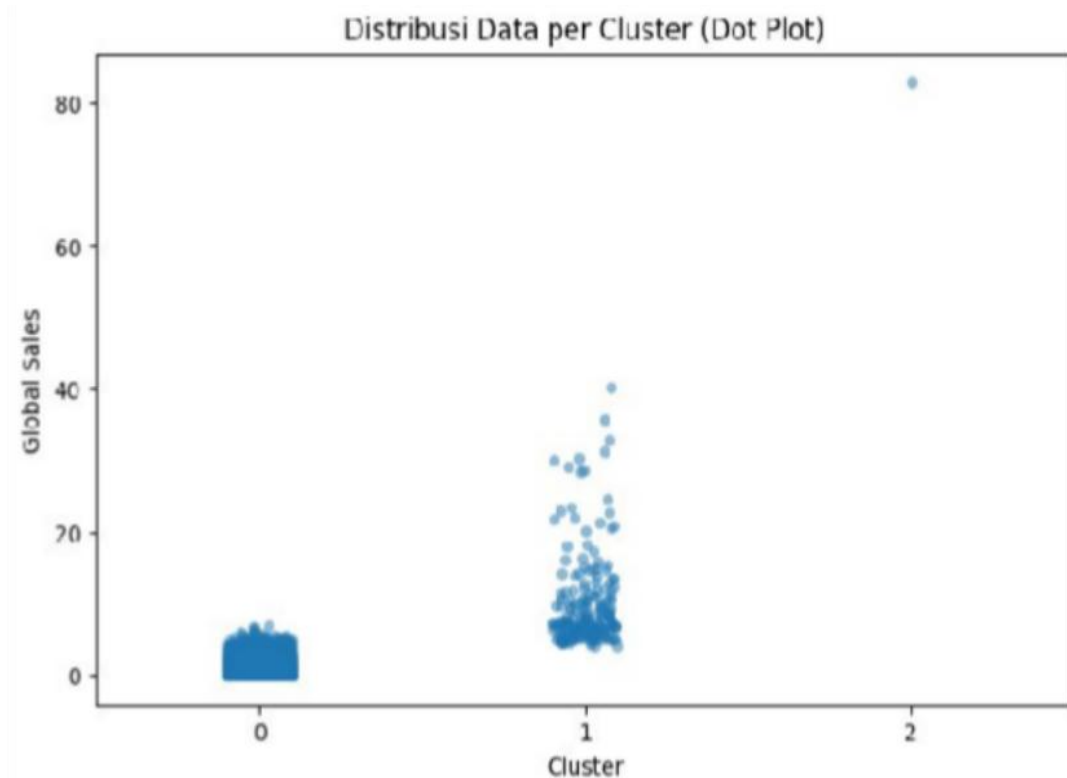
karakteristik ekstrem (outlier) yang memiliki performa penjualan luar biasa dibandingkan mayoritas video game dalam dataset.

	Name	Cluster
0	Wii Sports	2
1	Super Mario Bros.	1
2	Mario Kart Wii	1
3	Wii Sports Resort	1
4	Pokemon Red/Pokemon Blue	1
5	Tetris	1
6	New Super Mario Bros.	1
7	Wii Play	1
8	New Super Mario Bros. Wii	1
9	Duck Hunt	1
10	Nintendogs	1
11	Mario Kart DS	1
12	Pokemon Gold/Pokemon Silver	1
13	Wii Fit	1
14	Wii Fit Plus	1
15	Kinect Adventures!	1
16	Grand Theft Auto V	1
17	Grand Theft Auto: San Andreas	1
18	Super Mario World	1
19	Brain Age: Train Your Brain in Minutes a Day	1
20	Pokemon Diamond/Pokemon Pearl	1

Gambar 4. Hasil Clustering Data

3.5 Distribusi Cluster

Analisis distribusi cluster dilakukan untuk mengetahui proporsi jumlah data pada setiap kelompok yang terbentuk. Hasil visualisasi menunjukkan bahwa sebagian besar data terkonsentrasi pada Cluster 0, sedangkan Cluster 1 hanya mencakup sebagian kecil data. Kondisi ini menunjukkan bahwa mayoritas video game dalam dataset memiliki tingkat penjualan yang relatif rendah dibandingkan kelompok lainnya. Di sisi lain, Cluster 2 hanya terdiri dari satu data yang memiliki karakteristik sangat berbeda dari data lainnya. Perbedaan jumlah anggota yang cukup besar antar cluster menunjukkan bahwa distribusi penjualan video game bersifat tidak merata. Untuk mempermudah interpretasi, visualisasi distribusi ditampilkan menggunakan skala logaritmik sehingga seluruh cluster dapat diamati secara lebih jelas. Temuan ini mengindikasikan bahwa industri video game cenderung didominasi oleh sejumlah kecil produk yang berhasil mencapai tingkat penjualan yang sangat tinggi.



Gambar 5. Distribusi Cluster

3.6 Analisis Karakteristik Cluster

Analisis karakteristik cluster dilakukan dengan membandingkan nilai rata-rata penjualan pada setiap wilayah untuk memahami perbedaan pola yang terbentuk. Hasil analisis menunjukkan bahwa Cluster 0 memiliki rata-rata penjualan terendah pada seluruh atribut yang digunakan. Kondisi ini mengindikasikan bahwa sebagian besar video game dalam dataset termasuk ke dalam kategori produk dengan performa pasar yang relatif rendah. Temuan tersebut menunjukkan bahwa tidak semua video game yang dirilis mampu memperoleh penerimaan pasar yang luas, sehingga tingkat persaingan dalam industri video game dapat dikategorikan cukup tinggi.

Karakteristik yang berbeda terlihat pada Cluster 1 yang memiliki nilai rata-rata penjualan lebih tinggi dibandingkan Cluster 0. Kelompok ini merepresentasikan video game yang berhasil memperoleh respons pasar yang lebih baik dan memiliki tingkat penjualan yang relatif stabil. Meskipun jumlahnya lebih sedikit, keberadaan Cluster 1 menunjukkan adanya produk-produk yang mampu mencapai performa komersial yang baik dan memiliki daya tarik yang lebih tinggi di pasar dibandingkan mayoritas game lainnya.

Perbedaan paling mencolok ditemukan pada Cluster 2 yang hanya terdiri dari satu data, yaitu Wii Sports. Game tersebut memiliki nilai penjualan yang jauh melampaui seluruh data lain dalam dataset sehingga membentuk cluster tersendiri. Fenomena ini menunjukkan adanya distribusi penjualan yang tidak merata, dimana hanya sedikit produk yang mampu mencapai tingkat kesuksesan yang sangat tinggi. Keberadaan Wii Sports sebagai outlier mengindikasikan bahwa pasar video game cenderung bersifat hit-driven market, yaitu pasar yang didominasi oleh sejumlah kecil produk dengan performa penjualan yang sangat tinggi dibandingkan produk lainnya.

	NA_Sales	EU_Sales	JP_Sales	Other_Sales
Cluster				
0	0.207079	0.114291	0.064805	0.038029
1	5.088610	2.872941	1.267059	0.898770
2	41.490000	29.020000	3.770000	8.460000

Gambar 6. Karakteristik Cluster

3.7 Insight Bisnis

Hasil segmentasi menunjukkan bahwa sebagian besar video game berada pada kelompok dengan tingkat penjualan rendah. Kondisi tersebut mengindikasikan bahwa persaingan dalam industri video game sangat kompetitif dan tidak semua produk mampu memperoleh keberhasilan komersial yang signifikan. Oleh karena itu, perusahaan perlu memahami faktor-faktor yang memengaruhi keberhasilan suatu produk agar dapat meningkatkan peluang keberhasilan pada peluncuran produk berikutnya.

Selain itu, keberadaan Cluster 1 menunjukkan bahwa terdapat kelompok video game yang mampu mencapai performa penjualan yang relatif baik meskipun jumlahnya tidak dominan. Kelompok ini dapat dijadikan sebagai referensi dalam mengidentifikasi karakteristik produk yang memiliki potensi pasar lebih tinggi. Informasi tersebut dapat dimanfaatkan oleh perusahaan dalam merancang strategi pengembangan produk maupun menentukan segmentasi pasar yang lebih tepat.

Temuan mengenai Wii Sports sebagai satu-satunya anggota pada Cluster 2 menunjukkan bahwa pasar video game cenderung terpusat pada produk-produk unggulan yang memiliki daya tarik luar biasa. Kondisi ini mengindikasikan bahwa inovasi produk, kualitas pengalaman pengguna, strategi pemasaran, dan dukungan platform menjadi faktor penting yang dapat mendorong peningkatan penjualan. Dengan memanfaatkan hasil segmentasi yang diperoleh, perusahaan dapat menyusun strategi bisnis yang lebih terarah dan berbasis data dalam menghadapi dinamika pasar video game yang terus berkembang.

4. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma K-Means Clustering untuk melakukan segmentasi data penjualan video game berdasarkan karakteristik penjualan pada berbagai wilayah. Hasil penentuan jumlah cluster menggunakan metode Elbow menunjukkan bahwa tiga cluster merupakan jumlah yang paling optimal untuk digunakan dalam proses pengelompokan. Evaluasi menggunakan Silhouette Score menghasilkan nilai sebesar 0,915 yang menunjukkan bahwa model memiliki kualitas clustering yang sangat baik. Hasil segmentasi menunjukkan bahwa mayoritas video game berada pada kategori penjualan rendah, sedangkan hanya sebagian kecil yang mampu mencapai tingkat penjualan tinggi. Temuan tersebut menunjukkan bahwa distribusi penjualan video game bersifat tidak merata dan cenderung didominasi oleh sejumlah kecil produk yang sangat sukses. Selain menghasilkan segmentasi data, penelitian ini juga memberikan insight yang dapat dimanfaatkan untuk mendukung strategi pemasaran dan pengambilan keputusan bisnis berbasis data. Penelitian selanjutnya dapat mengembangkan analisis dengan menambahkan atribut lain yang berpotensi memengaruhi tingkat penjualan video game, seperti genre, publisher, rating pengguna, maupun ulasan pelanggan. Selain itu, penggunaan metode clustering lain seperti Hierarchical Clustering atau DBSCAN dapat dilakukan untuk membandingkan kualitas hasil segmentasi yang diperoleh. Pengembangan penelitian ke arah analisis prediktif menggunakan metode klasifikasi atau

regresi juga dapat menjadi alternatif untuk mengidentifikasi faktor-faktor yang berpengaruh terhadap keberhasilan penjualan video game pada pasar global.

DAFTAR PUSTAKA

- [1] B. Barczak and K. Kwiatkowska, "Artificial Intelligence and the Development of Business Models in the Gaming Industry - Case Study," *SSRN*, pp. 01–23, 2024.
- [2] A. C. R. Hash, D. I. C. Baskoro, I. P. G. Weda, M. Fakhri, M. Fauzan, and M. Fansyuri, "Data Mining Menggunakan Algoritma K-Means Untuk Menentukan Game Terpopuler Pada Platform Steam Dengan RapidMiner," *Journal of Information Technology and Informatics Engineering*, vol. 1, no. 1, pp. 51–56, 2025.
- [3] M. A. Bimo and H. Maupa, "Analisis Pemasaran Digital Terhadap Penjualan Video Game Online Pada Perusahaan Toko Game di Jakarta Barat," *Jurnal Manajemen Bisnis dan Kewirausahaan*, vol. 8, no. 5, pp. 1174–1186, 2024.
- [4] N. N. F. Adzani, W. Witanti, and F. R. Umbara, "Klasifikasi Tingkat Penjualan Video Game Dengan Menggunakan Metode K – Nearest Neighbors," *INFOTECH journal*, vol. 9, no. 2, pp. 618–625, 2023, doi: 10.31949/infotech.v9i2.7371.
- [5] D. Ruswanti, D. Susilo, and Riani, "Implementasi CRISP-DM pada Data Mining untuk Melakukan Prediksi Pendapatan dengan Algoritma C.45," *Go Infotech: Jurnal Ilmiah STMIK AUB*, vol. 30, no. 1, pp. 111–121, 2024, doi: 10.36309/goi.v30i1.266.
- [6] A. Adhariani, T. Harmini, and A. Musthafa, "Implementasi Algoritma K-Means Untuk Clustering Kasus Kriminal Di Wilayah Kota Depok," *Prosiding Seminar Nasional Sains dan Teknologi Seri III*, vol. 2, no. 1, pp. 408–415, 2025.
- [7] E. Ramadhan and A. Voutama, "Visualisasi Prediksi Penjualan Game Di Dunia Menggunakan Power BI," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, pp. 887–894, 2025, doi: 10.23960/jitet.v13i2.6353.
- [8] F. D. Agustiar, B. N. Sari, and I. Maulana, "Penerapan Data Mining Untuk Pengelompokan Produk Penjualan Menggunakan Algoritma K-Means (Studi Kasus : Toko Agung Makmur Jaya)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 1, pp. 59–67, 2025.
- [9] Erlinda, H. Nopriandi, and D. J. Putra, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Nusa Indah," *Startup: Journal of Digital Business*, vol. 2, no. 1, pp. 1–10, 2025, doi: 10.32485/kopertip.v4i2.118.
- [10] T. Wahyudi, N. Sa'adah, and D. Puspitasari, "Penerapan Metode K-Means Pada Data Penjualan Untuk Mendapatkan Produk Terlaris di PT. Titian Nusantara Boga," *Jurnal Sains dan Teknologi*, vol. 5, no. 1, pp. 228–236, 2023, doi: 10.36040/jati.v7i2.6562.
- [11] N. A. Maori and Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24176/simet.v14i2.9630.
- [12] P. K. Ritonga and M. S. Hasibuan, "Analisis Perbandingan Silhouette dengan Elbow pada Algoritma K-Means dan DBSCAN," *Metik Jurnal*, vol. 9, no. 1, p. 2025, 2025, doi: 10.47002/metik.v9i1.1027.
- [13] H. Alexander, Y. Umaidah, and M. Jajuli, "Implementasi Clustering Untuk Menentukan Efektivitas Nilai Siswa Sesudah Pandemi Covid-19 Menggunakan Algoritma K-Means," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 3, pp. 1493–1500, 2023, doi: 10.36040/jati.v7i3.7174.
- [14] M. Sholeh and K. Aeni, "Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 8, no. 1, pp. 56–65, 2023, doi: 10.30998/string.v8i1.16388.
- [15] D. Handayani, A. Noeman, Rasim, A. Hidayat, and H. Lubis, "Implementasi Algoritma K-Means Untuk Pengelompokan Produk Terlaris Pada Pangkalan Sudiawati Bekasi," *Jurnal Manajemen Informatika Jayakarta*, vol. 6, pp. 160–171, 2026.