

PENGELOMPOKAN GAME BERDASARKAN DATA TOP SELLER PADA WEBSITE "STEAM" MENGGUNAKAN METODE K-MEDOIDS

Manggala Bhakti W, Joseph Dedy Irawan, Deddy Rudhistiar

Teknik Informatika, Institut Teknologi Nasional Malang

Jalan Raya Karanglo km 2 Malang, Indonesia

8himsel8@gmail.com

ABSTRAK

Bisnis *Game* di Indonesia telah berkembang pesat dengan munculnya banyak industri yang berusaha masuk ke dalamnya. Salah satu platform terbesar yang digunakan untuk penjualan *Game* saat ini adalah "Steam". Namun, dengan banyaknya transaksi dan data yang dihasilkan. Oleh karena itu, diperlukan sebuah solusi menggunakan teknik data *scraping* untuk mengumpulkan data dan metode *Cluster K-Medoids* untuk menganalisis pola transaksi. Teknik data *scraping* akan digunakan untuk mengumpulkan data dari situs web *Game* market seperti Steam. Data yang terkumpul dikategorikan berdasarkan harga, lokasi, dan jumlah pembelian. Setelah data terkumpul, metode *K-Medoids* akan digunakan untuk mengelompokkan data penjualan *Game* sesuai yang telah ditentukan dengan jumlah *Cluster* direkomendasikan didasarkan pada *Elbow Plot*. Dengan menerapkan sistem *K-Medoids* ini waktu yang diperlukan untuk mengumpulkan semua data dan menganalisisnya dapat dilakukan secara efisien dengan akurasi data *Silhouette Score* yang dihasilkan 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.

Kata kunci : Data, Transaksi, Review, Scraping Data, Elbow Plot, Metode K-Medoids

1. PENDAHULUAN

Perkembangan internet saat ini sangat berefek pada bagaimana bisnis berjalan dan salah satu sektor yang terpengaruh oleh perkembangan ini salah satunya adalah *Game*. *Game* telah menjadi bagian dari budaya modern, mempengaruhi berbagai aspek kehidupan kita mulai dari hiburan hingga ekonomi. Dengan pemahaman yang lebih dalam tentang peran dan pengaruh *Game*, kita dapat menghargai kompleksitas dan potensi yang dimilikinya dalam membentuk dunia yang kita tinggali.

Marketplace adalah sebuah platform digital dimana penjual dan pembeli dapat bertemu dan melakukan jual-beli pada berbagai produk dan layanan yang ada. Biasanya, *Marketplace* ini akan menyediakan berbagai fitur untuk memfasilitasi interaksi antara penjual dan pembeli, termasuk sistem pembayaran, sistem pengiriman barang, sistem penilaian, dan ulasan pelanggan dalam konteks ini adalah produk *Game*.

Studi ini menyelidiki hubungan antara karakteristik pasar *Game* dan perilaku pengguna dengan fokus pada platform Steam. Penelitian ini menggunakan data transaksi dan perilaku pengguna dari Steam untuk menganalisis bagaimana fitur-fitur pasar, seperti ulasan pengguna, diskon, dan rekomendasi *Game*, memengaruhi keputusan pembelian pengguna.[1]

Data *scraping*, juga dikenal sebagai web *scraping*, adalah teknik yang digunakan untuk mengekstrak informasi dari berbagai situs web secara otomatis. Teknik ini telah menjadi alat yang sangat penting dalam analisis data, penelitian pasar, dan pengambilan keputusan bisnis di berbagai industri. Dalam konteks ini, tinjauan pustaka berikut

menyajikan pemahaman yang lebih dalam tentang konsep, teknik, aplikasi, dan tantangan yang terkait dengan data *scraping*.

Clustering Medoids, juga dikenal sebagai *K-Medoids*, adalah salah satu teknik kluster yang paling sering digunakan dalam menganalisis data. Teknik ini membagi dataset menjadi kelompok-kelompok yang homogen berdasarkan kemiripan antarobservasi. Dalam literatur, teknik ini telah mendapatkan perhatian besar karena kemudahan implementasinya dan kegunaannya dalam berbagai aplikasi.

Dalam bukunya yang terkenal, "Introduction to Data Mining" menjelaskan langkah-langkah detail dalam penerapan *Clustering Medoids*. Mereka membahas tentang pemilihan jumlah kelompok yang tepat (K), pengukuran jarak antar titik observasi, strategi inisialisasi centroid, dan kriteria penghentian iterasi. Buku ini memberikan pemahaman yang mendalam tentang konsep dan aplikasi praktis *Clustering Medoids* dalam analisis data.[3]

Secara keseluruhan, penggunaan data *scraping* berpotensi terhadap efisiensi dalam pengumpulan data dan data tersebut akan dikelola dengan metode *Clustering Medoids* yang nantinya hasil data tersebut digunakan untuk insight potensi strategi penjualan *Game* kedepan.

2. TINJAUAN PUSTAKA

2.1. Game

Game adalah bentuk hiburan yang melibatkan interaksi antara pemain dan aturan yang ditetapkan dalam lingkungan yang biasanya disebut sebagai "permainan". Dalam *Game*, pemain memiliki tujuan tertentu yang harus dicapai atau tantangan yang harus diatasi. Sementara itu, aturan dalam *Game*

menentukan batasan dan cara pemain dapat berinteraksi dengan lingkungan serta bagaimana lingkungan tersebut merespons tindakan pemain.[4]

2.2. Platform Distribusi Digital

Platform distribusi digital adalah sebuah sistem atau layanan yang memungkinkan para pengguna untuk membeli, mengunduh, dan mengakses konten digital seperti permainan video, aplikasi, musik, film, buku elektronik, dan lainnya secara daring (online). Platform ini memberikan kemudahan bagi pengguna untuk mengakses konten digital tanpa perlu membeli fisik atau pergi ke toko konvensional.

Platform distribusi digital ini tidak hanya mempermudah pengguna untuk mengakses konten, tetapi juga memberikan kesempatan bagi pembuat konten untuk menyebarkan karyanya secara global tanpa memerlukan infrastruktur fisik yang kompleks. [4]

2.3. Steam

Steam adalah platform distribusi digital yang dikembangkan oleh Valve Corporation, merupakan platform yang didesain khusus untuk mendistribusikan permainan video, konten digital, serta menyediakan layanan terkait *Game* seperti komunikasi antar pemain, manajemen teman, pencapaian, dan lainnya. Steam juga berfungsi sebagai toko daring di mana pengguna dapat membeli, mengunduh, dan memainkan berbagai judul *Game* yang tersedia. Platform ini telah menjadi salah satu distribusi digital paling populer di dunia, digunakan oleh jutaan pengguna global untuk mendapatkan akses ke permainan dan konten digital.[4]

2.4. Python

Python artinya bahasa pemrograman tingkat tinggi yang dikembangkan di awal tahun 1990-an sang Guido van Rossum. dia dirancang menggunakan fokus pada keterbacaan kode dan kesederhanaan sintaks. Python mendukung berbagai paradigma pemrograman, termasuk pemrograman berorientasi objek, fungsional, dan berbasis kejadian.

Python bersifat open *Source*, yg berarti kode Sumbernya bisa diakses serta dimodifikasi oleh siapa pun. Hal ini sudah mendorong pertumbuhan serta adopsi Python di semua dunia. Hal ini memungkinkan pengembang untuk menentukan gaya pemrograman yg sesuai menggunakan kebutuhan proyek mereka. [5]

2.5. Data Scraping

Data *scraping* adalah proses ekstraksi informasi dari sebuah situs web. Teknik ini umumnya dilakukan secara otomatis menggunakan perangkat lunak khusus yang disebut sebagai web *scraper*. Web *scraper* dapat mengambil data dari berbagai jenis situs web, termasuk situs e-commerce, media sosial, situs berita, forum, dan lain sebagainya.

Proses data *scraping* melibatkan pengambilan informasi dari halaman web dengan menggunakan

teknik parsing HTML dan menyimpannya dalam format yang dapat digunakan lebih lanjut, seperti dalam basis data, spreadsheet, atau format lainnya. Data yang diambil dapat berupa teks, gambar, hyperlink, atau elemen lain yang ada di halaman web.[2]

2.6. Data Mining

Data Mining adalah proses untuk mencari dan mendapatkan suatu pola samar atau pengetahuan berharga dari sejumlah besar data yang tersimpan dalam berbagai bentuk, seperti basis data, data warehousing, dan koleksi informasi lainnya. Dasar teori tentang *Data Mining* melibatkan prinsip-prinsip dan teknik-teknik yang biasanya digunakan dalam hal mengekstraksi data yang berguna untuk membentuk suatu informasi.

Data Mining menjadi alat yang sangat penting dalam analisis data di berbagai sektor, termasuk bisnis, keuangan, ilmu pengetahuan, kesehatan, pemasaran, dan lain-lain. Hal ini memungkinkan organisasi dan individu untuk memanfaatkan potensi besar dari data yang mereka miliki untuk mendapatkan wawasan yang berharga dan mendukung pengambilan keputusan yang lebih baik. [3]

2.7. 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. [9]

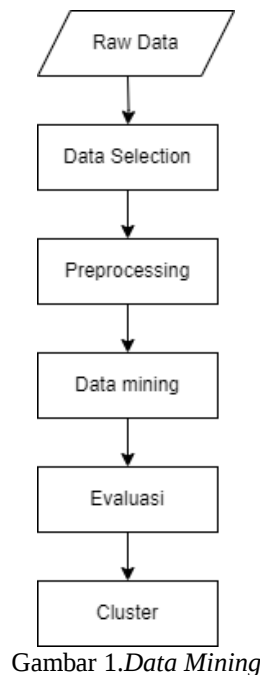
2.8. Metode Clustering K-Medoids

K-Medoids merupakan salah satu teknik dalam analisis kluster yang digunakan untuk mengelompokkan data ke dalam kelompok-kelompok berdasarkan kesamaan di antara titik-titik data. Metode ini memiliki kesamaan dengan metode k-means, namun berbeda dalam cara menentukan pusat kluster.

Dalam *K-Medoids*, titik yang dipilih sebagai pusat kluster disebut sebagai "*Medoids*". *Medoids* adalah titik data aktual dalam himpunan data yang paling mewakili kluster secara keseluruhan. Artinya, *Medoids* adalah anggota dari kluster itu sendiri, bukan hanya titik pusat berdasarkan mean (seperti dalam k-means). [10]

3. METODE PENELITIAN

3.1. Flowchart Data Mining



Gambar 1. Data Mining

3.2. Pengumpulan Data Mentah

Pengumpulan data yang didapat dapat dilakukan menggunakan Teknik *Scraping* pada halaman web yang akan dituju untuk mengambil data yang diinginkan secara otomatis. Data yang digunakan pada penelitian adalah Harga, Diskon, dan Review pada halaman *Top Seller* yang diperoleh dari Website resmi *Game Marketplace Steam* yang didapat pada situs web <https://store.steampowered.com/>

3.3. Data Preprocessing

Preprocessing data adalah salah satu langkah dalam proses *Data Mining*, dimana hasil akhir analisa dapat optimal jika data memiliki kualitas yang baik. Proses preprocessing dilakukan untuk menghindari masalah yang datang akibat ketidak-konsistenan format data menyebabkan memperlambat waktu pemrosesan data dan mengurangi tingkat akurasi hasil analisis. Teknik preprocessing biasanya digunakan dalam penelitian ini mencakup membersihkan data untuk mengatasi nilai yang hilang (missing value), mentransformasi data agar memiliki format yang konsisten selama proses *Data Mining*, serta mengintegrasikan beberapa dataset menjadi satu dataset yang lengkap. Dengan melakukan langkah-langkah ini secara efektif, data siap untuk diproses lebih lanjut dalam analisis *Data Mining*, yang akan menghasilkan hasil yang lebih akurat dan bermakna.

3.4. Implemetasi Data Mining

Pada tahap ini, proses *Data Mining* dilakukan dengan menerapkan algoritma *K-Medoids* pada data yang telah melalui proses preprocessing. Pemilihan algoritma yang tepat dalam *Data Mining* sangat krusial

karena akan memengaruhi hasil yang berdampak pada keputusan bisnis.

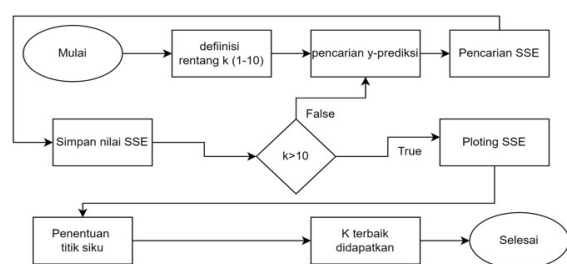
3.5. Penentuan Algoritma Terbaik

Penentuan Jumlah *Cluster* terbaik ditentukan menggunakan metode *Elbow Plot* yang dimunculkan sebelum *Cluster* ditampilkan agar hasil *Clustering* yang dicapai optimal. Penelitian ini mengukur kinerja algoritma *Clustering K-Medoids* menggunakan lima *Cluster* berdasarkan peringkat *Top Seller* pada website tersebut. Nilai *k* secara bertahap dimulai dari *Game* dengan peringkat *Top Seller* 1, 2, 3, 4, dan 5 agar memunculkan nilai *Sillhouette Score* untuk setiap jumlah *Cluster*, dan terakhir hasilnya dibandingkan untuk menentukan konfigurasi *Clustering* yang optimal.

3.6. Hasil Clustering

Pada tahap akhir, hasil dari proses ini diperoleh melalui penerapan algoritma *K-Medoids*, yaitu pengelompokan penjualan *Game* berdasarkan rentang harga, diskon, dan ulasan. Data awal dikumpulkan dengan menggunakan teknik *scraping* pada halaman *Top Seller Steam*. Dalam tahap preprocessing data, dilakukan pengecekan untuk memastikan kebersihan data, serta dilakukan tahap transformasi dan integrasi tiap datanya agar data siap untuk diproses lebih lanjut. Proses pemrosesan data menggunakan library Python, dimulai dengan mengimpor data ke dalam database web dalam format Microsoft Excel CSV, kemudian menerapkan berbagai operasi untuk pengolahan data. Dengan proses yang dirancang dengan baik, proses ini dapat dilaksanakan secara efektif untuk menghasilkan output yang dapat dianalisis dan dievaluasi. Anda dapat menarik kesimpulan serta menganalisis hasil dari proses *Data Mining* tersebut.

3.7. Elbow Plot



Gambar 2. Elbow Plot

Metode *Elbow Plot* berguna untuk menentukan total *Cluster* yang efisien serta optimal dengan melihat perbandingan hasil perubahan jumlah *Cluster* yang membentuk bentuk menyerupai siku-siku pada suatu titik yang ditentukan [4].

Berikut adalah langkah-langkah algoritma siku untuk menentukan nilai *k* pada *K-medoid*:

- Inialisasi nilai *k* awal;
- Meningkatkan nilai *k*;

- Hitung jumlah kesalahan kuadrat untuk setiap nilai k ;
- Analisislah hasil penjumlahan kesalahan kuadrat untuk nilai k yang tereduksi secara signifikan;
- Cari dan tentukan nilai k yang merupakan sudut.

Dalam metode Elbow, untuk menentukan nilai *Cluster* optimal dipilih berdasarkan nilai *Sum of Square Error (SSE)*, nilai ini akan mengukur selisih antara data yang didapat dengan model estimasi yang dijalankan proses sebelumnya. *SSE* berguna sebagai pertimbangan dalam menentukan jumlah *Cluster* yang optimal, diambil dari titik di mana terjadi penurunan *SSE* yang signifikan dan membentuk siku. (3):

$$SSE = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - C_k\|^2 \dots (1)$$

Dimana:

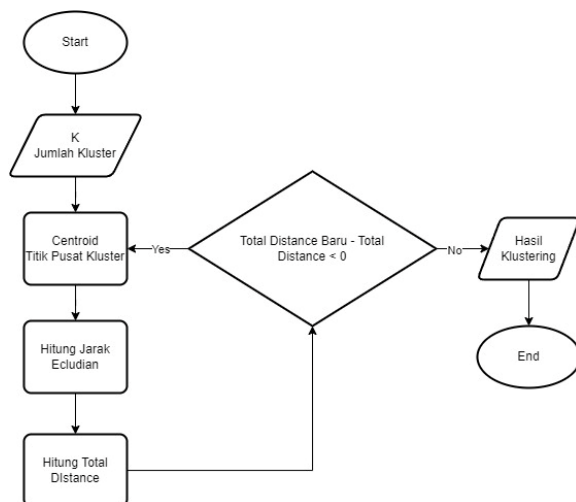
K = jumlah *Cluster*

x_i = data ke - i

C_k = centroid *Cluster*

3.8. K-Medoids

Metode K-Medoid digunakan untuk menentukan pengelompokannya data *Game* harga dan *Review* ke dalam sejumlah *Cluster* yang akan ditentukan dengan metode *Elbow Plot* [11].



Gambar 3. K-Medoids

Proses Metode K-Medoids :

- Mulai dengan menginisialisasi k pusat *Cluster* (jumlah *Cluster*)
- Tentukan nilai titik setiap data (objek) ke *Cluster* terdekat dengan menggunakan rumus Jarak Euclidean:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2};$$

$$i = 1, 2, 3, \dots, n \dots \dots \dots (2)$$

- Pilih secara acak salah satu objek dari setiap *Cluster* sebagai titik pusat medoid kandidat baru.
- Hitung jarak setiap objek dalam masing-masing *Cluster* ke medoid kandidat baru.

- Hitung total deviasi (S) dengan membandingkan total jarak baru dengan total jarak lama. Jika $S < 0$, tukar objek dengan data *Cluster* untuk membentuk kumpulan baru sebanyak k objek sebagai medoid.
- Ulangi langkah 3 sampai 5 sampai tidak ada perubahan pada medoid, sehingga terbentuklah *Cluster* beserta anggota *Cluster* masing-masing.

3.9. Silhouette Score

Berikut adalah rumus perhitungan Euclidean Distance pada *Cluster* harga, yaitu :

$$d = \sqrt{(a_1 - a_2)^2 + (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \dots (3)$$

Dimana:

a_1 = nilai titik *rank* kulster 1

a_2 = nilai titik *rank* kulster 2

x_1 = nilai titik price kluster 1

x_2 = nilai titik price kluster 2

y_1 = nilai titik disc price

y_2 = nilai titik disc price kluster 2

z_1 = nilai titik persen disc kluster 1

z_2 = nilai titik persen disc kluster 2

Untuk perhitungan pada kluster harga sesuaikan jumlah dan data dengan variable *Game Review*. lalu setelah hitung jarak antar titik hitung jarak rata-rata dari titik i ke titik lain dalam kluster yang sama.

$$a(i) = \frac{1}{|C_i| - 1} \sum_{j \in C_i, j \neq i} d(i, j) \dots (4)$$

Dimana:

C_i = titik pada kluster i

$|C_i|$ = jumlah jarak titik pada kluster i

$d(i, j)$ = adalah jarak antara titik i dan j

$j \neq i$ = memastikan tidak menghitung kluster dirinya sendiri

Lalu hitung *Sillhouette score* dengan menggunakan rata-rata jarak yang telah dihitung sebelumnya yaitu

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \dots (5)$$

Dimana:

$a(i)$ = rata-rata jarak pada kluster yang sama

$b(i)$ = rata-rata jarak pada kluster yang berbeda

Pada tahap yang terakhir adalah hitung nilai rata-rata keseluruhan nilai *sillhouette score* yang telah didapat

$$\text{Rata - rata} = \frac{1}{N} \sum_{i=1}^N s(i) \dots (6)$$

Dimana:

N = Total kluster yang digunakan

$s(i)$ = *Sillhouette score* pada titik kluster i

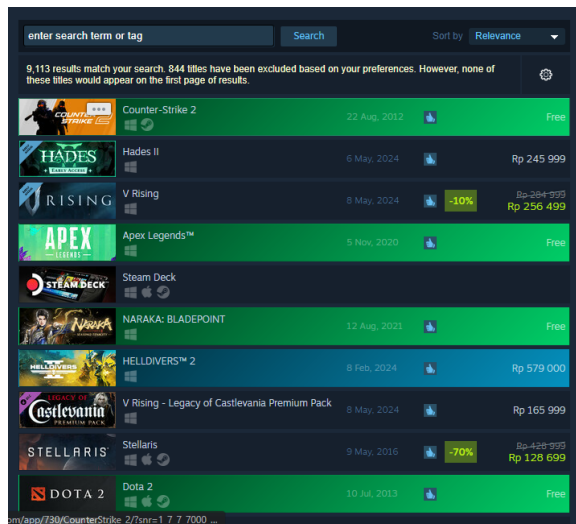
4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Proses pengumpulan data-data *Top seller* pada website <https://store.steampowered.com/> seperti pada

Gambar 4 dilakukan dengan cara *Data Scraping* ke website tujuan menargetkan data sebagai berikut yaitu nama, *Review*, dan harga menargetkan. Dikarenakan website bersifat responsive data yang didapat harus menargetkan link *jquery* agar dapat dilakukan *script* link *jquery* sebagai berikut :

https://store.steampowered.com/search/results/?query&start={x}&count=50&dynamic_data=&sort_by=_ASC&snr=1_7_7_7000_7&filter=topsellers&tags=19&infinite=1.



Gambar 4. Source Dataset

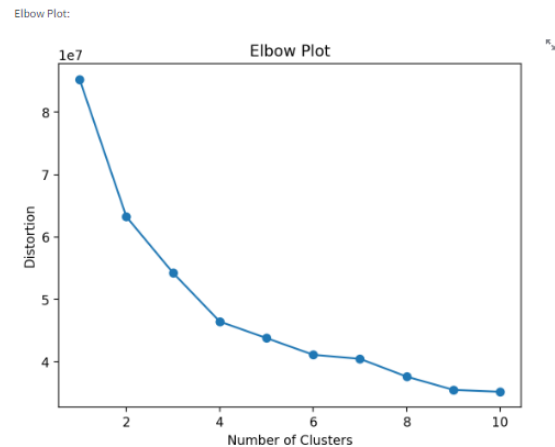
Setelah mendapat link dan dilakukan *Data scraping* data yang diperoleh dari website tujuan akan dilakukan preprocessing dan implementasi *Data Mining* pada data *script raw*. Data yang dihasilkan adalah *rank*, *title*, *harga* awal, *harga* final, *diskon*, *Review*, dan *total Review*. Pada tahapan ini hasil *script* berbentuk format *csv* file seperti pada Gambar 5.

	rank	title	price	disc_price	persen_disc	review_data	percent_review	total_review
0	1	Counter-Strike 2	0	0	0	Very Positive	87	8,049,615
1	2	Ghost of Tsushima DIRECT	879,000	879,000	0	No Data	9	0
2	3	PUBG: BATTLEGROUNDS	0	0	0	Mixed	58	2,358,311
3	4	Men of War II	338,999	338,999	0	No Data	0	0
4	5	Hades II	245,999	245,999	0	Very Positive	94	28,632
5	6	Dota 2	0	0	0	Very Positive	81	2,236,681
6	7	V Rising	284,999	256,499	10	Very Positive	88	75,301
7	8	Apex Legends	0	0	0	Mostly Positive	77	826,086
8	9	HELLDIVERS 2	579,000	579,000	0	Mostly Positive	73	634,166
9	10	ELDEN RING	599,000	599,000	0	Very Positive	92	602,587
10	11	Destiny 2	0	0	0	Very Positive	80	593,612
11	12	Grand Theft Auto V	401,000	153,600	61	Very Positive	87	1,636,136
12	13	Gray Zone Warfare	284,999	284,999	0	Mixed	67	36,177
13	14	Diablo IV	1,050,000	1,050,000	0	Mixed	69	16,223
14	15	Dead by Daylight	149,000	149,000	0	Very Positive	80	522,323
15	16	ANIMAL WELL	211,000	189,900	10	Overwhelmingly	96	4,984

Gambar 5. Hasil Scrap

4.2. Pengujian Metode

4.3. Elbow Plot



Gambar 6. Elbow Plot

Penjelasan Gambar 6 merupakan *Elbow Plot* yang digunakan untuk menentukan *K* yang akan digunakan pada klustering. Pada grafik terlihat pada Kluster 4 adalah titik belok grafik (titik elbow) yang menandakan bahwa kluster itu adalah yang dianggap paling optimal sesuai metode.

4.4. K-Medoids

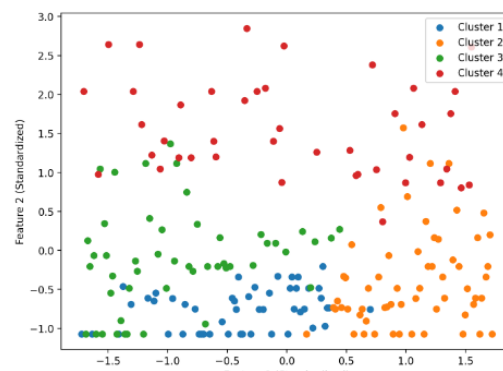
Penjelasan Gambar 7 merupakan centroid yang ditentukan untuk metode *K-Medoids*. Titik ini ditentukan secara random dengan batasan sesuai *Summary* tiap *Cluster* nya.

Iteration 4

Cluster Centers:

	rank	price	disc_price	persen_disc	percent_review	total_review
0	87	89,999	89,999	0	96	165,148
1	163	130,000	130,000	0	88	6,250
2	46	266,000	266,000	0	83	120,312
3	94	699,000	699,000	0	70	19,950

Gambar 7. Centeroid K-Medoids



Gambar 8. Grafik Jumlah Elemen Cluster

Penjelasan Gambar 8 adalah grafik area tiap *Cluster* yang ditentukan menggunakan warna. Tiap *Cluster* memiliki daerahnya masing-masing, dimana

yang mana ditentukan oleh perhitungan *K-Medoids* dan dipengaruhi oleh centroid yang telah ditentukan secara random.

$$S = \text{Total Distance Baru} - \text{Total Distance Lama} \\ = 67.38409671 - 91.36762095 = -24$$

Summary Statistics:

	rank	price	disc_price	persen_disc	review_data	percent_review	total_review	Cluster
count	50	50	50	50	50	50	50	50
mean	77.66	90,063.36	67,217.06	14.64	2.02	88.52	398,933.94	0
std	13.4375	73,759.554	59,852.4468	24.7714	1.1516	9.66	169,757.0008	0
min	1	0	0	0	1	58	141	0
25%	56.5	0	0	0	1	81.25	14,849.5	0
50%	80.5	90,999	61,499.5	0	2	92.5	131,102	0
75%	104.75	146,499	101,624	20	2	96	293,556.5	0
max	141	245,999	206,999	85	5	99	8,049,675	0

Gambar 9. Summary Cluster 1

Pada table *Summary Cluster* satu ini dapat diartikan bahwa *Cluster 1* berisi produk *Game* yang memiliki jarak harga bervariasi dengan rentang harga antara 0 hingga 200.000 rupiah yang memiliki total percent *Review* positif diantara 58% - 99% yang dapat diartikan kluster ini adalah *Game* yang memiliki reputasi yang baik dan mudah diraih oleh user Indonesia.

Summary Statistics:

	rank	price	disc_price	persen_disc	review_data	percent_review	total_review	C
count	59	59	59	59	59	59	59	59
mean	161.5985	193,616.4068	137,527.2203	14.4746	2.4407	82.5085	86,594.3051	
std	24.1335	173,648.963	120,181.1051	27.5038	1.4053	17.9005	145,018.0055	
min	110	0	0	0	0	0	0	
25%	141.5	90,999	67,499.5	0	2	79	10,255	
50%	163	130,999	108,999	0	2	88	30,045	
75%	182	277,499.5	195,649.5	15	2.5	94	87,508	
max	200	748,000	459,000	94	6	98	705,687	

Gambar 10. Summary Cluster 2

Pada table *Summary Cluster* satu ini dapat diartikan bahwa *Cluster 2* berisi produk *Game* yang memiliki jarak harga bervariasi dengan rentang harga antara 0 hingga 500.000 rupiah yang memiliki total percent *Review* positif diantara 0% - 98% yang dapat diartikan kluster ini adalah *Game* yang memiliki reputasi yang baik dan cukup mudah diraih oleh user Indonesia tetapi masih dibawah kluster 1.

Summary Statistics:

	rank	price	disc_price	persen_disc	review_data	percent_review	total_review	Clu
count	49	49	49	49	49	49	49	
mean	48.2857	274,448.551	199,148.5714	17.449	2.7551	75.2653	253,242.898	
std	35.0785	176,076.861	131,004.0011	29.5868	1.665	24.3283	446,363.7986	
min	3	0	0	0	0	0	0	
25%	18	169,000	103,000	0	2	75	21,855	
50%	42	266,000	229,000	0	2	82	71,858	
75%	70	362,000	284,999	25	4	88	257,605	
max	126	690,000	599,000	85	9	98	2,358,311	

Gambar 11. Summary Cluster 3

Pada table *Summary Cluster* satu ini dapat diartikan bahwa *Cluster 3* berisi produk *Game* yang

memiliki jarak harga bervariasi dengan rentang harga antara 0 hingga 600.000 rupiah yang memiliki total percent *Review* positif diantara 0% - 95% yang dapat diartikan kluster ini adalah *Game* yang memiliki reputasi yang cukup baik tetapi memiliki harga yang cukup sulit diraih oleh user Indonesia dan dibawah kluster 2.

Summary Statistics:

	rank	price	disc_price	persen_disc	review_data	percent_review	total_review	C
count	42	42	42	42	42	42	42	
mean	102.9048	744,552.1667	674,165.2619	8.0714	3.0476	64.9762	194,017.6429	
std	58.3312	173,951.1341	203,100.0241	20.0078	1.9749	29.6257	161,310.8467	
min	2	408,000	107,360	0	0	0	0	
25%	50.25	599,000	549,250	0	2	56.5	5,473.75	
50%	97.5	699,499.5	653,999.5	0	2	72	34,789.5	
75%	156.75	879,000	799,000	0	5	87.75	125,834.75	
max	190	1,106,800	1,050,000	89	9	97	652,023	

Gambar 12. Summary Cluster 4

Pada table *summa4y Cluster* satu ini dapat diartikan bahwa *Cluster 3* berisi produk *Game* yang memiliki jarak harga bervariasi dengan rentang harga antara 0 hingga 1.100.000 rupiah yang memiliki total percent *Review* positif diantara 0% - 97% yang dapat diartikan kluster ini adalah *Game* yang memiliki reputasi yang cukup baik namun memiliki harga yang sulit diraih oleh user Indonesia.

4.5. Silhouette Score

Untuk mengetahui keakuratan penggunaan Metode *K-Medoids* untuk menentukan Kluster digunakan perbandingan menggunakan Metode *Silhouette Score* yang diterapkan ke 3 kelompok data yaitu *Cluster Harga* dan *Review*, *Harga*, dan *Review Silhouette score* *Harga* dan *Review* dengan nilai *Elbow Plot* sebesar 4 Kluster adalah 0.196.

Silhouette Score: 0.19606090680798147

Gambar 13. Silhouette Score Harga dan Review

Silhouette score pada *K-Medoids* yang hanya menargetkan Data *Harga* dengan nilai *Elbow Plot* sebesar Kluster 6 adalah 0.353.

Silhouette Score: 0.3534429721983339

Gambar 14. Silhouette Score Harga

Silhouette score pada *K-Medoids* yang hanya menargetkan Data *Review* dengan nilai *Elbow Plot* sebesar Kluster 3 adalah 0.363

Silhouette Score: 0.36305403469736147

Gambar 15. Silhouette Score Review

Silhouette score Rentang nilai *Silhouette Score* adalah antara -1 hingga 1. Nilai *Silhouette Score* digunakan demi kelancaran mengevaluasi seberapa baik setiap objek data telah dikelompokkan pada

klaster yang sesuai dengan objek-objek serupa. Nilai Mendekati nilai 1 menunjukkan bahwa objek data berada di dalam klaster yang sesuai dengan objek-objek serupa dan cukup terpisah dari objek dalam klaster lain. Semakin mendekati 1 semakin baik pengelompokan klaster. Dan apabila Nilai Mendekati -1 menunjukkan bahwa objek data mungkin telah salah dikelompokkan ke dalam klaster yang salah. Semakin mendekati -1 semakin tidak sesuai objek data tersebut dengan klaster tempat dia ditempatkan.

Dilihat Dari 3 *sillhouette score* diatas dapat dinyatakan data dan metode menghasilkan sebuah Kluster yang cukup baik dimana nilai yang didapat diatas 0. Penyebab nilai *Sillhouette score* Data dan Review memiliki nilai yang cukup rendah disebabkan ada 2 kelompok data yang memiliki property yang berbeda dimana Harga semakin rendah semakin bagus sedangkan untuk Review semakin tinggi semakin baik hal yang bertolak belakang ini lah yang menyebabkan jatuhnya nilai tersebut dan dapat dilihat apabila 2 data tersebut dipisah dapat menghasilkan kenaikan *Sillhouette Score* hingga 1.7.

4.6. Perbandingan Aplikasi Website dengan Rapid Miner

Total Cluster Metode K-Medoids

Pivot Table:

Cluster	0
0	36
1	32
2	64
3	68

Cluster Model

Cluster 0: 105 items
Cluster 1: 14 items
Cluster 2: 71 items
Cluster 3: 10 items
Total number of items: 200

Gambar 16. Perbandingan Cluster Harga & Review

Perbandingan pada Cluster harga dan Review yang memiliki Cluster berjumlah 4 memiliki perbedaan signifikan pada Cluster 1,2,4. Untuk Cluster 3 memiliki selisih yang tidak terlalu signifikan. Perbedaan ini disebabkan ada 2 data yaitu Harga dan Review yang menjadi satu.

Pivot Table:

Cluster	0
0	35
1	48
2	28
3	32
4	28
5	29

Cluster Model

Cluster 0: 39 items
Cluster 1: 21 items
Cluster 2: 19 items
Cluster 3: 70 items
Cluster 4: 22 items
Cluster 5: 29 items
Total number of items: 200

Gambar 17. Perbandingan Cluster Harga

Bisa dilihat dari Cluster harga Cluster yang memiliki Cluster berjumlah 6. Perbandingan semua Cluster kecuali Cluster ke 4 tidak memiliki selisih yang begitu besar dikarenakan hanya mengClusterkan data harga yang memiliki sifat semakin rendah semakin bagus.

Total Cluster Method K-Medoids

Pivot Table:

Cluster	0
0	69
1	54
2	77

Cluster Model

Cluster 0: 61 items
Cluster 1: 113 items
Cluster 2: 26 items
Total number of items: 200

Gambar 18. Perbandingan Cluster Review

Tetapi untuk Cluster dari data Review Cluster yang memiliki Cluster berjumlah 3 memiliki suatu perbedaan pada Cluster 2 dan 3. Walaupun pada Cluster memiliki selisih yang tidak begitu signifikan tetapi perbedaan ini dapat memberikan efek apabila data harga dan Review dijadikan satu dan dilakukan Cluster. Efeknya dapat dilihat dari perbandingan pada Cluster Harga dan Review yang memiliki selisih yang cukup signifikan.

4. KESIMPULAN DAN SARAN

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 serta Review sedangkan Cluster untuk Data Harga memiliki perbedaan yang tidak terlalu signifikan. Perbedaan Harga & Review disebabkan karena efek dari Perbedaan Hasil Cluster dari data Review yang berimbas apabila kedua data tersebut dijadikan satu. Pada penelitian ini dapat ditingkatkan dengan meningkatkan kualitas cara menentukan variable ke dalam peringkat yang berbeda, atau berbagai menggunakan teknik *Data Mining* yang digabung dengan sistem mengelolah data warehousing serta intelijen bisnis.

DAFTAR PUSTAKA

- [1] Ramachandran, M., Rajendran, R., & Surya, S. (2020). A Comprehensive Review on Web Scraping Techniques. *Journal of Physics: Conference Series*, 1511(1), 012039.
- [2] Galit Shmueli, Peter C. Bruce, Peter Gedeck (Edisi ke-1, 2019) "Data Mining for Business Analytics: Concepts, Techniques, and Applications in Python"
- [3] Emily Brown (2022) "Exploring the Economic Impact of Steam Sales on Indie Game Developers"
- [4] Alok Singh (2019), Web Scraping with Python Using Scrapy"
- [5] Fahmi Firzada (2021) "Klasterisasi Tingkat Masa Studi Tepat Waktu Mahasiswa Menggunakan Algoritma K-Medoids" jsisfotek (Jurnal Sistem Informasi dan Teknologi) e-ISSN: 2686-3154
- [6] Samudi (2020) The K-Medoids Clustering Method for Learning Applications during the COVID-19 Pandemic Sinkron : Jurnal dan

- Penelitian Teknik Informatika e-ISSN : 2541-2019
- [7] Cici Astria (2022) "Pengembangan Metode Datamining K-Medoid Pada Kasus Distribusi Listrik di Indonesia" ISBN: 978-602-52720-7-3
- [8] Limbawan Sitorus (2022) "PENGELOMPOKAN PRODUKSI TANAMAN PERKEBUNAN KELAPA BERDASARKAN PROVINSI MENGGUNAKAN ALGORITMA K-MEDOID CLUSTERING" skripsi tunasbangsa
- [9] Fahmi Firzada (2021) "Klasterisasi Tingkat Masa Studi Tepat Waktu Mahasiswa Menggunakan Algoritma *K-Medoids*" jsisfotek (*Jurnal Sistem Informasi dan Teknologi*) e-ISSN: 2686-3154
- [10] Farahdinna, F., Nurdiansyah, I., Suryani, A., & Wibowo, A. (2019). Perbandingan Algoritma K-Means dan *K-Medoids* dalam Klasterisasi Produk Asuransi Perusahaan Nasional. *Jurnal Ilmiah Fifo*, 11(2). DOI: <http://dx.doi.org/10.22441/fifo.2019.v11i2.010>