

Penerapan Algoritma K-Medoids Dalam Klasterisasi Penjualan Laptop

Application of K-Medoids Algorithm in Laptop Sales Clustering

Abizar Ar Rifa'i¹, Muhamad Fatchan², Nanang Tedi Kurniadi³

^{1,2,3}Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa

¹abizararrifai@mhs.pelitabangsa.ac.id*, ²fatchan@pelitabangsa.ac.id, ³nanang@pelitabangsa.ac.id

Abstract

Technological developments have made the use of laptops a basic necessity that must be owned to assist in completing a job as a substitute for a PC (Personal Computer). Along with technological advances, many laptop brands have sprung up, from each brand to launch laptops with various advantages. This has resulted in more and more emerging various types of laptop brands that compete with each other to be able to meet the needs of today's consumers. Therefore there must be a system that can provide advice or laptop recommendations in searching for references. This study aims to classify laptop sales data using the k-medoids clustering algorithm data mining method. laptop sales data are grouped based on the similarity of the data so that data with the same characteristics will be in one cluster. Based on the calculations that have been carried out by researchers on 1000 sample data, it can be categorized into 3 clusters. cluster 1 is a category of low laptop sales, which is 137 data, then cluster 2 is a category of high laptop sales, which is 669 data, and cluster 3 is a category of medium laptop sales, which is 194 data from 1000 categories of laptop sales. It can be concluded that the grouping of laptop sales data in cluster 2 is the most widely sold because the specifications and prices of laptops are more affordable than cluster 1 and cluster 3. And it has been tested using the Rapid Miner application with the same results as manual calculations using Microsoft Excel.

Keywords: K-Medoids, Clustering, Laptop Sales

Abstrak

Perkembangan teknologi menjadikan penggunaan laptop sudah menjadi kebutuhan mendasar yang harus dimiliki untuk membantu dalam menyelesaikan suatu pekerjaan sebagai pengganti PC (Personal Computer). Seiring dengan kemajuan teknologi banyak merek laptop bermunculan, dari setiap merek meluncurkan laptop dengan berbagai keunggulannya. Hal tersebut mengakibatkan semakin banyak bermunculan berbagai jenis merek laptop yang saling bersaing untuk dapat memenuhi kebutuhan konsumen saat ini. Maka dari itu harus ada sebuah sistem yang dapat memberikan saran ataupun rekomendasi laptop dalam pencarian referensi. Penelitian ini bertujuan untuk mengelompokkan data penjualan laptop menggunakan metode data mining algoritma k-medoids clustering. data penjualan laptop dikelompokkan berdasarkan kemiripan data tersebut sehingga data dengan karakteristik yang sama akan berada dalam satu cluster. Berdasarkan perhitungan yang telah dilakukan peneliti pada 1000 sample data dapat dikategorikan dalam 3 cluster. cluster 1 merupakan kategori penjualan laptop rendah yaitu 137 data, kemudian cluster 2 merupakan kategori penjualan laptop tinggi yaitu 669 data dan cluster 3 merupakan kategori penjualan laptop sedang yaitu 194 data dari 1000 kategori penjualan laptop. Dapat disimpulkan bahwa pengelompokkan data penjualan laptop pada cluster 2 merupakan paling banyak terjual karena spesifikasi dan harga laptop yang lebih terjangkau dibandingkan cluster 1 dan cluster 3. Dan telah diuji menggunakan aplikasi Rapid Miner yang hasilnya sama dengan perhitungan manual menggunakan Microsoft Excel.

Kata kunci: K-Medoids, Klasterisasi, Penjualan Laptop

Pendahuluan

Kebutuhan teknologi saat ini sangat meningkat seiring dengan berkembangnya teknologi. Perkembangan teknologi menjadikan pengguna laptop sudah menjadi kebutuhan mendasar yang harus dimiliki untuk membantu dalam menyelesaikan suatu pekerjaan sebagai pengganti PC (Personal Computer). Laptop

merupakan perangkat komputer portable yang cocok untuk digunakan saat berpergian atau untuk penggunaan mobile dan telah menjadi kebutuhan dalam membantu pekerjaan sehari-hari. Komponen yang terdapat didalam laptop ukurannya diperkecil, dibuat lebih ringan dan hemat daya[1].

Seiring dengan kemajuan teknologi banyak merek-merek laptop bermunculan, dari setiap merek meluncurkan laptop dengan berbagai keunggulannya. Dari berbagai jenis merek laptop, spesifikasi dan fungsinya sering kali konsumen tidak bisa memilih antara merek satu dengan yang lainnya bahkan sering juga membeli laptop yang tidak sesuai dengan kebutuhannya[2].

Sekarang ini laptop sangat diperlukan oleh semua kalangan seperti anak sekolah, mahasiswa dan para pekerja kantoran karena kelebihanannya dibandingkan dengan PC (Personal Computer)[3]. Hal tersebut mengakibatkan semakin banyak bermunculan berbagai jenis merek laptop yang saling bersaing untuk dapat memenuhi kebutuhan konsumen saat ini[4]. Dengan semakin berkembangnya zaman, laptop sekarang ini mempunyai beragam merek dan spesifikasi yang terkadang membuat orang-orang mengalami kesusahan dan kesulitan dalam mencari, memilih ataupun membeli laptop yang tepat sesuai dengan kebutuhannya[5]. Dengan masalah tersebut tentunya orang-orang akan terbantu jika ada yang dapat merekomendasikan laptop yang sesuai bagi kebutuhannya. Maka dari itu harus ada sebuah sistem rekomendasi yang dapat memberikan saran ataupun rekomendasi laptop berdasarkan ketertarikan dan kebutuhan dalam pencarian referensi[6].

Dalam penjualan laptop sering kali ditemukan, tidak seimbang nya penjualan antar merek satu dengan merek lainya. Hal ini menyebabkan penjual sering kali menyimpan perangkat laptop tersebut cukup lama yang mengakibatkan berkurangnya nilai jual[7]. Maksudnya adalah pada saat penyimpanan yang cukup lama jika tidak di maintenance dengan baik mulai dari kardus, aksesoris akan memudar warnanya bahkan bisa menjadi rusak, hal tersebutlah yang membuat penjual merugi. Dari permasalahan tersebut perlu adanya pengelompokan jenis laptop yang akan di pasarkan yang sesuai minat dari pembeli. Sehingga penjual tidak harus menyimpan cukup lama jika perangkat tersebut tidak laku terjual.

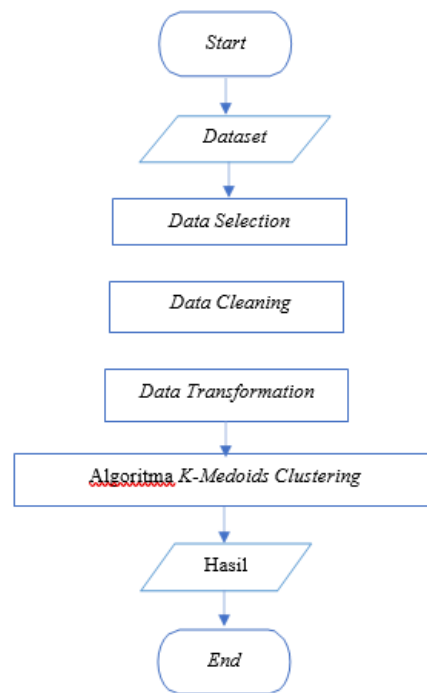
Data mining adalah operasi penggunaan dan resourcing data untuk mencari hubungan atau pola dari sekumpulan data yang berukuran besar.[8]. Data mining merupakan proses untuk mendapatkan informasi dari basis data yang besar dengan cara penggalian data dengan memanfaatkan kumpulan data[9]. Dengan kemajuan teknologi informasi sekarang sudah semakin berkembang pesat dan hampir mencakup di segala bidang. Kemajuan teknologi tersebut menghasilkan data yang besar dan banyak mulai dari bidang industri, ekonomi, ilmu dan teknologi serta berbagai bidang lainnya. Salah satu algoritma yang digunakan dalam teknik data mining yang memakai teori klustersasi adalah K-medoids[10].

Metode Penelitian

Dalam metode penelitian ini bertujuan untuk lebih memahami bagaimana mengimplementasikan data penjualan laptop dan dapat diartikan sebagai proses pengumpulan data atau analisis data yang dilakukan secara teratur dan logis untuk mencapai tujuan tertentu dan memperkaya pengetahuan itu sendiri sehingga dapat menghasilkan fakta dan wawasan yang tidak biasa, tahapan yang akan digunakan dalam melakukan klusterisasi dan penentuan atribut untuk mempermudah penelitian sehingga penelitian dapat berjalan dengan baik dan sistematis, serta memenuhi tujuan yang diinginkan, penelitian deskriptif ditujukan untuk mendeskripsikan suatu keadaan atau fenomena-fenomena apa adanya. Kategori penelitian deskriptif merupakan kategori yang digunakan untuk menggambarkan atau menganalisis suatu hasil penelitian tetapi tidak digunakan untuk membuat kesimpulan yang lebih luas. Pendekatan kuantitatif banyak dituntut menggunakan angka, mulai dari pengumpulan data, penafsiran terhadap data tersebut, serta penampilan hasilnya. Penelitian deskriptif dengan pendekatan kuantitatif dapat dikatakan sebagai kategori penelitian dengan menggambarkan sesuatu berdasarkan data yang dikumpulkan dalam bentuk angka mengenai fakta-fakta yang ada dan menentukan atribut dan variabel yang akan diujikan.

Tahapan Penelitian

Dalam melakukan analisis dan mencari pola data untuk dijadikan sebuah dataset maka untuk memudahkan penelitian ini dan dapat berjalan dengan sistematis serta memenuhi tujuan yang diinginkan maka dibuat langkah – langkah dalam tahapan penelitian yang akan dilakukan sebagai berikut :



Gambar 1 Tahapan Penelitian

Dalam tahapan penelitian data awal yang digunakan merupakan dataset penjualan laptop yang didapat dari website kaggle. Berikut link dataset nya (<https://www.kaggle.com/muhammetvarl/laptop-price>). Setelah mendapatkan dataset nya lalu dilanjutkan dengan melakukan pengolahan data seperti data selection, data cleaning dan data transformation. Lalu dilanjutkan ke proses clustering dengan menggunakan algoritma k-medoids untuk mengetahui hasil dari klusterisasi penjualan laptop.

Pengumpulan Data

Pada saat pengumpulan data, cara yang dapat digunakan oleh peneliti untuk mengumpulkan data penjualan laptop yang memiliki atribut atau variabel dan nilai dari suatu objek, kegiatan yang mempunyai variasi tertentu yang ditetapkan oleh peneliti sehingga dapat dipelajari kemudian ditarik kesimpulannya, teknik pengumpulan data yang digunakan dalam penelitian ini diambil dari kaggle.com, dalam melakukan analisis ini kita dapat mencari pola agar memudahkan penelitian ini dan dapat berjalan dengan sistematis dan memenuhi tujuan yang diinginkan maka dibuat langkah – langkah dalam proses pengolahan data dan himpunan data dalam penelitian yang akan dilakukan dapat dijelaskan sebagai berikut :

Tabel 1 Dataset Penjualan Laptop

Dataset	Data 1	Data 2	Data 3	Dn
Company	Apple	HP	Acer	...
Product	Macbook	250 G6	Aspire 3	...
Type Name	Ultrabook	Notebook	Notebook	...
Inches	13.03	15.06	15.06	...
Screen Resolution	IPS panel Retina Display 2560x1600	Full HD 1920x1080	1366x768	...
CPU	Intel Core	Intel Core i5 7200U 2.5GHz	AMD A9-Series 9420 3GHz	...
RAM	8GB	8GB	4GB	...
Memory	128 GB SSD	256GB	500GB HDD	...
GPU	Intel Iris Plus Graphics 640	Intel HD Graphics 620	AMD Radeon R5	...
Operasi System	MacOS	No OS	Windows 10	...
Weight	1.37kg	1.86kg	2.1kg	...
Price Euros	1339.69	575	400	...

Pengolahan Data

Pengolahan data pada penelitian ini menggunakan teknik klusterisasi data untuk mengetahui pengelompokan data dengan menggunakan Algoritma K-Medoids. Data dapat diproses dan dijadikan dataset dalam penelitian ini, data tersebut akan di sesuaikan kebutuhan dalam penelitian ini agar dapat model yang diperoleh, memiliki kemampuan generalisasi yang baik dalam melakukan klusterisasi data, data yang dilatih untuk membuat kelas, C1, C2 dan C3 atau menjalankan fungsi dari sebuah algoritma sesuai dengan tujuannya dalam penelitian ini sehingga dapat melihat keakuratan atau performa dari suatu data[11]. Data Selection merupakan proses pemilihan data dari sekumpulan data operasional yang ada sebelum masuk ke tahap mining data maupun informasi.

Tabel 2 Data Selection

Dataset	Data 1	Data 2	Data 3	Dn
Company	Apple	HP	Acer	...
Product	Macbook	250 G6	Aspire 3	...
Type Name	Ultrabook	Notebook	Notebook	...
Inches	13.03	15.06	15.06	...
Screen Resolution	IPS panel Retina Display 2560x1600	Full HD 1920x1080	1366x768	...
CPU	Intel Core	Intel Core i5 7200U 2.5GHz	AMD A9-Series 9420 3GHz	...
RAM	8GB	8GB	4GB	...
Memory	128GB SSD	256GB	500GB HDD	...
GPU	Intel Iris Plus Graphics 640	Intel HD Graphics 620	AMD Radeon R5	...
Operasi System	MacOS	No OS	Windows 10	...
Weight	1.37kg	1.86kg	2.1kg	...
Price Euros	1339.69	575	400	...

Data Cleaning pada tahap data cleaning merupakan proses pembersihan dari data yang akan dipakai, untuk penghapusan data dengan membuang missing value, duplikasi data, dan memeriksa inkonsistensi dari sebuah data dan memperbaiki kesalahan pada data[12]. Proses pembersihan data dilakukan secara manual dengan bantuan software spreadsheet atau microsoft excel.

Tabel 3 Data Cleaning

Company	Product	RAM	Memory	Price Euro
Apple	MacBook Pro	8GB	128GB SSD	1339.69
Apple	MacBook Air	8GB	128GB Flash Storage	898.94
HP	250 G6	8GB	256GB SSD	575
Apple	MacBook Pro	16GB	512GB SSD	2537.45
Asus	E402WA-GA010T (E2-6110/2GB/32GB/W10)	2GB	32GB Flash Storage	1803.6
Acer	Aspire 3	4GB	500GB HDD	400
Lenovo	V310-15ISK (i3-6006U/4GB/1TB/FHD/W10)	4GB	1TB HDD	499
HP	15-BS101nv (i7-8550U/8GB/256GB/FHD/W10)	8GB	256GB SSD	659
Asus	ZenBook UX430UN	16GB	512GB SSD	1496
Dell	Latitude	8GB	256GB	0

Data Transformation merupakan proses mengubah format data awal menjadi sebuah format data standar untuk proses pembacaan data dengan algoritma pada program maupun tool yang digunakan[13]. Berikut adalah hasil pengolahan data awal setelah melawati tahapan diatas untuk dijadikan dataset pada tahap selanjutnya, ditunjukkan pada tabel berikut ini :

Tabel 4 Data Transformation (Dataset)

Company	Product	RAM	Memory	Price Euro
Apple	MacBook Pro	8GB	128GB SSD	1339.69
Apple	MacBook Air	8GB	128GB Flash Storage	898.94
Hp	250 G6	8GB	256GB SSD	575
Apple	MacBook Pro	16GB	512GB SSD	2537.45

Apple	MacBook Pro	8GB	256GB SSD	1803.6
Acer	Aspire 3	4GB	500GB HDD	400
Apple	MacBook Pro	16GB	256GB Flash Storage	2139.97
Apple	MacBook Air	8GB	256GB SSD	1158.7
Asus	ZenBook UX430UN	16GB	512GB SSD	1496
Acer	Swift 3	8GB	256GB SSD	456
Asus	Rog Strix	8GB	256GB SSD	1229
Dell	Inspiron 3567	4GB	256GB	639
Lenovo	Yoga Book	4GB	64GB	319
HP	ProBook 430	8GB	512GB	1103
Asus	VivoBook Max	4GB	256GB	586,19

Pemodelan

Pada tahap ini dilakukan pemodelan data, metode yang dipakai adalah Algoritma K-Medoids atau sering disebut juga dengan algoritma PAM (Partitioning Around Medoid) dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw, dimana merupakan algoritma yang mirip dengan k-means karena kedua algoritma ini partitional yang memecah dataset menjadi kelompok – kelompok[14]. Perbedaan dari algoritma k-means dengan algoritma K- Medoids terletak pada penentuan pusat cluster, dimana algoritma k-means menggunakan nilai rata – rata (means) dari setiap cluster sebagai pusat cluster dan algoritma K-Medoids menggunakan objek data sebagai perwakilan (medoids) sebagai pusat cluster.[15]. Algoritma K- Medoids digunakan untuk mengatasi kelemahan dari algoritma k-means yang sangat sensitif terhadap penciran (outlier) karena objek – objek ini sangat jauh letaknya atau karakteristiknya dari mayoritas data lainnya, sehingga jika dimasukkan ke suatu cluster data semacam ini bisa mendistorsi nilai rata-rata (mean) cluster tersebut[16].

Evaluasi Pengujian

Evaluasi dapat dilakukan dengan cara mengamati dan menganalisis hasil dari algoritma yang digunakan untuk memastikan bahwa hasil pengujian benar - benar sesuai dengan pembahasan. Melakukan pengecekan terhadap setiap nilai atribut dan model yang sudah dibangun. Kemudian melakukan evaluasi dengan cara mengamati dan menganalisis hasil dari algoritma yang digunakan untuk memastikan bahwa hasil pengujian menghasilkan nilai C1, C2 dan C3 benar dan sesuai hasil pembahasan, pengujian dilakukan untuk mengukur keakuratan hasil dari tiap model yang diusulkan[17]. K-Medoids algoritma yang mirip dengan k-means karena kedua algoritma ini partitional yang memecah dataset menjadi kelompok – kelompok[18]. Perbedaan dari algoritma k-means dengan algoritma K- Medoids terletak pada penentuan pusat cluster, dimana algoritma k-means menggunakan nilai rata – rata (means) dari setiap cluster sebagai pusat cluster dan algoritma K- Medoids menggunakan objek data sebagai perwakilan (medoids) sebagai pusat cluster.

Hasil dan Pembahasan

Analisis Penjualan Laptop

Penelitian ini menggunakan algoritma K-Medoids untuk menganalisis pengelompokan dalam penjualan laptop yang sesuai dengan kebutuhan pelanggan. Sumber data sebagai objek pada penelitian ini adalah data yang diambil dari situs kaggle.com. Data yang digunakan dalam penelitian ini terdiri dari atribut atau variabel seperti Product, Ram, Memory, dan Price euros.

Tabel 5 Dataset Penjualan Laptop

Company	Product	RAM	Memory	Price Euro
Apple	MacBook Pro	8	128	1339.69
Apple	MacBook Air	8	128	898.94
HP	250 G6	8	256	575
Apple	MacBook Pro	16	512	2537.45
Apple	MacBook Pro	8	256	1803.6
Acer	Aspire 3	4	500	400
Apple	MacBook Pro	16	256	2139.97
Apple	MacBook Air	8	256	1158.7
Asus	ZenBook UX430UN	16	512	1496
Acer	Swift 3	8	256	770

Apple	MacBook Pro	16	256	2439,97
Dell	Inspiron 3567	4	256	498,9
HP	250 G6	4	500	393,9
Lenovo	IdeaPad 320	8	512	499
Dell	XPS 13	8	128	979

Implementasi Algoritma K-Medoids

Pada perhitungan algoritma K-Medoids ini menggunakan dataset yang akan diolah yaitu sebanyak 1000 yang akan dikelompokkan kedalam tiga cluster yaitu low, medium, dan high. Dengan pemodelan yang sudah ditetapkan sebelumnya, maka berikut adalah contoh untuk perhitungan Euclidean Distance pada record ke 1 dan 1000 dari proses iterasi pertama.

Tabel 6 Inisialisasi Pusat Cluster pada Medoid

Titik Pusat Awal	Ram	Memory	Prince_Euros
Cluster ke-1	16	512	2537
Cluster ke-2	8	256	1803
Cluster ke-3	4	500	400

Nilai titik cluster tersebut digunakan sebagai perhitungan iterasi ke pertama, maka dibawah ini adalah contoh untuk perhitungan Euclidean Distance terhadap data ke-1 dan ke-1000 pada iterasi pertama :

Data ke-1

$$D1(c1) = \sqrt{(8 - 16)^2 + (128 - 512)^2 + (1339.69 - 2537)^2} = 1257$$

$$D1(c2) = \sqrt{(8 - 8)^2 + (128 - 256)^2 + (1339.69 - 1803)^2} = 481$$

$$D1(c3) = \sqrt{(8 - 4)^2 + (128 - 500)^2 + (1339.69 - 400)^2} = 1010$$

Data ke-1000

$$D1000(c1) = \sqrt{(8 - 16)^2 + (512 - 512)^2 + (2229 - 2537)^2} = 308$$

$$D1000(c2) = \sqrt{(8 - 8)^2 + (512 - 256)^2 + (2229 - 1803)^2} = 496$$

$$D1000(c3) = \sqrt{(8 - 4)^2 + (512 - 500)^2 + (2229 - 400)^2} = 1829$$

Berikut tabel dibawah ini dapat dilihat untuk hasil dari perhitungan Euclidean Distance pada tahap iterasi pertama :

Tabel 7 Euclidean Distance Iterasi 1

C1	C2	C3	Jarak Terdekat
1257	481	1010	481
1682	913	622	622
1979	1228	300	300
0	777	2137	0
777	0	1424	0
2137	1424	0	0
472	336	1757	336
1402	644	796	644
1042	401	1095	401
1758	1033	433	433
.....
691	91	1514	91
1595	824	696	696
1459	686	818	686
943	174	1252	174
308	496	1829	308
Jumlah Kedekatan			418625

Jarak hasil perhitungan akan dilakukan perbandingan antara 3 cluster dari data terdekat dengan pusat cluster. Secara umum berikut adalah matriks kelompok cluster yang terjadi pada iterasi pertama.

Tabel 8 Matriks Kelompok Data Iterasi ke-1

C1	C2	C3
	T	
		T
		T
T		
	T	
		T
	T	
	T	
	T	
		T
...
	T	
		T
	T	
	T	
T		
91	406	503

Setelah didapatkan hasil jarak dari setiap objek (cost) pada iterasi pertama maka langkah berikutnya perlu ditentukan kembali titik cluster baru dengan menginisialisasikan pusat cluster pada kandidat medoid baru (non medoid) untuk iterasi kedua.

Tabel 9 Inisialisasi Pusat Cluster pada Medoid Baru

Titik Pusat Awal	Ram	Memory	Prince_Euros
Cluster baru ke-1	8	128	1339
Cluster baru ke-2	8	128	898
Cluster baru ke-3	8	265	575

Nilai titik cluster baru tersebut digunakan kembali untuk perhitungan iterasi kedua, dengan langkah yang sama seperti sebelumnya, maka dibawah ini adalah contoh untuk perhitungan Euclidean Distance terhadap data ke-1 dan ke-1000 pada proses iterasi kedua :

Data ke-1

$$D1(c1) = \sqrt{(8 - 8)^2 + (128 - 128)^2 + (1339.69 - 1339.69)^2} = 0$$

$$D1(c2) = \sqrt{(8 - 8)^2 + (128 - 128)^2 + (1339.69 - 898)^2} = 440$$

$$D1(c3) = \sqrt{(8 - 8)^2 + (128 - 256)^2 + (1339.69 - 575)^2} = 775$$

Data ke-1000

$$D1000(c1) = \sqrt{(8 - 8)^2 + (512 - 128)^2 + (2229 - 1339.69)^2} = 968$$

$$D1000(c2) = \sqrt{(8 - 8)^2 + (512 - 128)^2 + (2229 - 898)^2} = 1384$$

$$D1000(c3) = \sqrt{(8 - 8)^2 + (512 - 256)^2 + (2229 - 575)^2} = 1673$$

Berikut tabel dibawah ini dapat dilihat untuk hasil dari perhitungan Euclidean Distance pada tahap iterasi kedua :

Tabel 10 Euclidean Distance Iterasi 2

C1	C2	C3	Jarak Terdekat
0	440	775	0
440	0	348	0
775	348	0	0
1257	1682	1979	1257
481	913	1228	481
1010	622	300	300
810	1247	1564	810
221	289	583	221

414	709	954	414
583	181	195	181
.....
569	1004	1320	569
350	90	433	90
210	230	568	210
316	741	1054	316
968	1384	1673	968
Jumlah Kedekatan			372726

Jarak hasil perhitungan akan dilakukan perbandingan diantara 3 cluster dari data terdekat dengan pusat cluster. Secara umum berikut adalah matriks kelompok cluster yang terjadi pada iterasi kedua.

Tabel 11 Matriks Kelompok Data Iterasi ke-2

C1	C2	C3
T		
	T	
		T
T		
T		
		T
T		
T		
T		
	T	
.....
T		
	T	
T		
T		
T		
463	236	301

Setelah didapatkan nilai jarak antara iterasi ke-1 dan iterasi ke-2, maka langkah selanjutnya menghitung total simpangan (S) dengan mencari selisih dari nilai total kedekatan baru – nilai total kedekatan lama. Dengan ketentuan jika $S > 0$ maka proses cluster dihentikan.

$$\begin{aligned}
 S &= \text{Total kedekatan baru} - \text{Total kedekatan lama} \\
 &= 372726 - 418625 \\
 &= -45898
 \end{aligned}$$

Karena nilai $S < 0$ maka tukar nilai objek dengan menentukan medoid baru dan proses clustering dilanjutkan ke iterasi ketiga. Setelah didapatkan hasil jarak dari setiap objek (cost) pada iterasi kedua maka langkah berikutnya perlu ditentukan kembali titik cluster baru dengan menginisialisasikan pusat cluster pada kandidat medoid baru (non medoid) untuk iterasi ketiga.

Tabel 12 Inisialisasi Pusat Cluster pada Medoid Baru

Titik Pusat Awal	Ram	Memory	Prince_Euros
Cluster baru ke-1	8	128	1129
Cluster baru ke-2	8	256	1629
Cluster baru ke-3	8	512	2229

Data ke - 1

$$D1(c1) = \sqrt{(8 - 8)^2 + (128 - 128)^2 + (1339.69 - 1129)^2} = 210$$

$$D1(c2) = \sqrt{(8 - 8)^2 + (128 - 256)^2 + (1339.69 - 1629)^2} = 316$$

$$D1(c3) = \sqrt{(8 - 8)^2 + (128 - 512)^2 + (1339.69 - 2229)^2} = 968$$

Data ke-1000

$$D1000(c1) = \sqrt{(8 - 8)^2 + (512 - 128)^2 + (2229 - 1129)^2} = 1165$$

$$D1000(c2) = \sqrt{(8 - 8)^2 + (512 - 256)^2 + (2229 - 1629)^2} = 652$$

$$D1000(c3) = \sqrt{(8 - 8)^2 + (512 - 512)^2 + (2229 - 2229)^2} = 0$$

Berikut tabel dibawah ini dapat dilihat untuk hasil dari perhitungan Euclidean Distance pada tahap iterasi ketiga :

Tabel 13 Euclidean Distance Iterasi 3

C1	C2	C3	Jarak Terdekat
210	316	968	210
230	741	1384	230
568	1054	1673	568
1459	943	308	308
686	174	496	174
818	1252	1829	818
1019	551	271	271
131	470	1100	131
530	289	734	289
381	859	1481	381
.....
776	266	420	266
140	652	1298	140
0	516	1165	0
516	0	652	0
1165	652	0	0
Jumlah Kedekatan			387656

Selanjutnya jarak dari hasil perhitungan akan dilakukan perbandingan antara 3 cluster data terdekat dengan pusat cluster dengan mengambil nilai terkecil, jarak ini menunjukkan bahwa data tersebut berada dalam 1 kelompok dengan pusat cluster terdekat. Berikut adalah Matriks kelompok cluster yang terjadi pada iterasi ke-3, dilambangkan dengan simbol "T".

Tabel 14 Matriks kelompok data Iterasi ke-3

C1	C2	C3
T		
T		
T		
		T
	T	
T		T
T		
	T	
T		
.....
	T	
T		
T		
	T	
		T
669	194	137

Setelah didapatkan nilai jarak antara iterasi ke-2 dan iterasi ke-3, maka langkah selanjutnya menghitung total simpangan (S) dengan mencari selisih dari nilai total kedekatan baru – nilai total kedekatan lama. Dengan ketentuan jika $S > 0$ maka proses cluster dihentikan dan jika nilai $S < 0$ maka tukar nilai objek dengan menentukan medoid baru.

$$\begin{aligned}
 S &= \text{Total kedekatan baru} - \text{Total kedekatan lama} \\
 &= 387656 - 372726 \\
 &= 14929
 \end{aligned}$$

Dari perhitungan diatas, dengan ketentuan jika $S > 0$ maka proses cluster dihentikan, maka perhitungan iterasi berhenti di iterasi ke-3 dengan titik pusat cluster pada tabel dibawah ini :

Tabel 15 Titik Pusat Cluster pada Iterasi ke-3

Titik Pusat Awal	Ram	Memory	Prince_Euros
Cluster ke-1	8	128	1129
Cluster ke-2	8	256	1629
Cluster ke-3	8	512	2229

Pembentukan Kelompok (Cluster)

Dari proses pra-processing data didapatkan 1000 data yang akan diproses menggunakan algoritma K-Medoids. Melalui beberapa tahapan seperti pada bagian pemodelan data, didapatkan hasil bahwa proses clustering dengan algoritma K-Medoids berhenti pada iterasi ke-3, karena total simpangan (S) dengan mencari selisih dari (nilai total kedekatan baru – nilai total kedekatan lama) sudah memenuhi ketentuan $S > 0$. Berikut adalah bentuk cluster yang didapat :

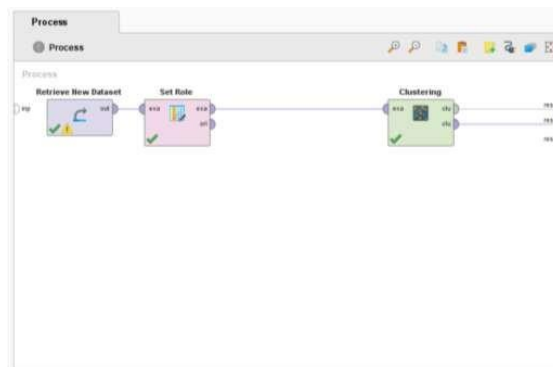
1. Cluster pertama memiliki pusat cluster (8, 128, 1129) sehingga dapat diartikan pada cluster ini adalah kelompok yang memiliki penjualan laptop dengan tingkat tinggi atau berkategori high. Ada 669 jumlah spesifikasi laptop dari katagori penjualan laptop yang masuk dalam cluster pertama.
2. Cluster kedua memiliki pusat cluster (8, 256, 1629) sehingga dapat diartikan pada cluster ini adalah kelompok yang memiliki penjualan laptop dengan tingkat sedang atau berkategori medium. Ada 194 jumlah spesifikasi laptop dari katagori penjualan laptop yang masuk dalam cluster kedua.
3. Cluster ketiga memiliki pusat cluster (8, 512, 2229) sehingga dapat diartikan pada cluster ini adalah kelompok yang memiliki penjualan laptop dengan tingkat rendah atau berkategori low. Ada 137 jumlah spesifikasi laptop dari katagori penjualan laptop yang masuk dalam cluster ketiga.

Pembahasan

Proses Pengujian Data (Rapid Miner)

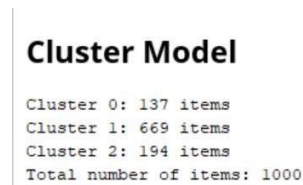
Pada proses ini metode klasterisasi dengan algoritma K-Medoids diterapkan untuk membentuk kelompok cluster dengan keakurasian yang tepat. Dalam penelitian ini menggunakan pengujian perhitungan tools Rapid Miner, hasil pengujian yang didapat dengan menggunakan tools Rapid Miner adalah dengan tahapan langkah – langkah sebagai berikut :

1. Melakukan import data yang diperlukan untuk proses pada tools Rapid Miner. Pada aplikasi Rapid Miner pilih dan klik Import Data, kemudian pilih data yang akan dipakai serta kemudian menentukan atribut label yang akan digunakan.
2. Klik menu Design, pada tampilan proses, tambahkan dataset pada folder ke layar tampilan proses. Pada menu Names & Roles cari fungsi Set Role yang nantinya akan dipakai untuk mengatur role atribut, kemudian drag ke layar tampilan proses.
3. Selanjutnya pada menu Modeling. Dalam submenu Segmentation, pilih fungsi K-Medoids, untuk menerapkan algoritma K-Medoids terhadap proses klasterisasi yang akan dilakukan.
4. Koneksikan semua perintah tersebut sehingga pada layar tampilan proses terlihat alur sebagai berikut:



Gambar 2 Proses Rapid Miner Klasterisasi K-Medoids

5. Setelah dilakukan Running Proses pada tools Rapid Miner, didapatkan hasil pengelompokan cluster terhadap 1000 record data yang diproses dan dapat dilihat pada gambar berikut :



Gambar 3 Cluster model dari 1000 record data

6. Dapat dilihat melalui hasil cluster model yang terbentuk terhadap 3 cluster yang sudah ditentukan sebelumnya, tiap masing-masing cluster memiliki anggota cluster (wilayah) yang sesuai dengan kelompoknya.
7. Adapun bentuk visualisasi menggunakan scatter plot untuk masing-masing cluster juga dapat dilihat pada gambar berikut ini :



Gambar 4 Grafik Scatter Plot dari Cluster yang terbentuk

Analisis Hasil

Setelah melakukan tahapan dalam mencari cluster penjualan laptop melalui metode cluster, pemanfaatan algoritma K-Medoids yang digunakan menghasilkan suatu pengelompokan cluster terhadap hasil bahwa proses clustering dengan algoritma K-Medoids berhenti pada iterasi ke-3, karena total simpangan (S) dengan mencari selisih dari (nilai total kedekatan baru – nilai total kedekatan lama) sudah memenuhi ketentuan $S > 0$. Berikut adalah bentuk cluster yang didapat yaitu :

1. Cluster pertama memiliki pusat cluster (8, 128, 1129) pada cluster ini memiliki jumlah spesifikasi laptop dari katagori penjualan yang paling banyak diantara 2 cluster lainnya yaitu ada 669 jumlah spesifikasi laptop dari katagori penjualan laptop.
2. Cluster kedua memiliki pusat cluster (8, 512, 2229) dan dari hasil perhitungan data yang digunakan ada 194 jumlah spesifikasi laptop dari katagori penjualan laptop yang masuk dalam cluster kedua.
3. Cluster ketiga memiliki pusat cluster (8, 256, 1629) dan terdiri dari 137 jumlah spesifikasi laptop dari katagori penjualan laptop yang masuk dalam cluster ketiga.

Pada pengujian dengan tools Rapid Miner juga didapatkan hasil yang serupa. Dalam Rapid Miner ditentukan sebanyak 3 cluster untuk pembentukan kelompok cluster produk yang ada dari dataset sebanyak 1000 data yang digunakan. Untuk Rapid Miner sendiri penamaan cluster dimulai dari Cluster_0, Cluster_1 dan Cluster_2. Anggota dari masing-masing cluster juga memiliki kemiripan dengan perhitungan manual yang dilakukan. Hanya saja dalam proses menggunakan tools Rapid Miner tidak ditentukan nilai cluster awal sebagaimana yang dilakukan dalam proses perhitungan manual. Namun untuk hasil yang didapatkan tidak jauh berbeda yaitu sebagai berikut :

1. Cluster_0 terdiri dari 137 anggota kelompok cluster penjualan laptop yang dimana ini sesuai dengan kelompok cluster low pada tahap perhitungan manual.
2. Cluster_1 terdiri dari 669 anggota kelompok cluster penjualan laptop yang dimana ini sesuai dengan kelompok cluster high pada tahap perhitungan manual.
3. Cluster_2 terdiri dari 194 anggota kelompok cluster penjualan laptop yang dimana ini sesuai dengan kelompok cluster medium pada tahap perhitungan manual.

Kesimpulan

Berdasarkan hasil pengujian yang telah dilakukan dalam penelitian ini, maka dapat diambil suatu kesimpulan yaitu, penerapan metode algoritma k-medoids clustering yang dimana hasilnya menunjukkan sebuah wawasan baru dalam melakukan pengelompokan data pada penjualan laptop. Berdasarkan perhitungan yang telah dilakukan peneliti pada 1000 sample data dapat dikategorikan dalam 3 kelompok (cluster). cluster 1 merupakan kategori penjualan laptop dengan spesifikasi rendah (low) yaitu 137 data, kemudian cluster 2 merupakan kategori penjualan laptop dengan spesifikasi tinggi (high) yaitu 669 data dan cluster 3 merupakan kategori penjualan laptop dengan spesifikasi sedang (medium) yaitu 194 data dari 1000 kategori penjualan laptop berdasarkan spesifikasi laptop yang diuji. Dapat disimpulkan bahwa pengelompokan data penjualan laptop pada cluster 2 merupakan paling banyak terjual karena spesifikasi dan harga laptop yang lebih terjangkau dibandingkan cluster 1 dan cluster 3. Dan telah diuji menggunakan aplikasi Rapid Miner yang hasilnya sama dengan perhitungan manual menggunakan Microsoft Excel.

Daftar Rujukan

- [1] Herwinsyah, "Penerapan Fuzzy Inference System (FIS) Dengan Metode Mamdani Pada Sistem Prediksi Penjualan Laptop Implementation of Fuzzy Inference System (FIS) with the Mamdani Method in Laptop Sales Prediction System," *J. Multimed. Artif. Intell.*, vol. 3, no. 2, pp. 52–59, 2019.
- [2] L. Henando, "Algoritma Apriori Dan Fp-Growth Untuk Analisa Perbandingan Data Penjualan Leptop Berdasarkan Merk Yang Diminati Konsumen (Studi Kasus : Indocomputer Payakumbuh)," *J-Click*, vol. 6, no. 2, pp. 201–207, 2019.
- [3] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. I. R.H.Zer, and D. Hartama, "Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 166–173, 2020, doi: 10.36294/jurti.v4i1.1296.
- [4] E. Eli Lavindi and A. Rohmani, "Aplikasi Hybrid Filtering Dan Naïve Bayes Untuk Sistem Rekomendasi Pembelian Laptop Hybrid Filtering and Naïve Bayes Application for Laptop Purchase Recommendation Systems," *J. Inf. Syst.*, vol. 4, no. 1, pp. 54–64, 2019.
- [5] A. D. Andini and T. Arifin, "Implementasi Algoritma K-Medoids Untuk Klasterisasi Data Penyakit Pasien Di Rsud Kota Bandung," *J. Responsif Ris. Sains dan Inform.*, vol. 2, no. 2, pp. 128–138, 2020, doi: 10.51977/jti.v2i2.247.
- [6] A. E. Wijaya and D. Alfian, "Sistem Rekomendasi Laptop Menggunakan Collaborative Filtering Dan Content-Based Filtering," *J. Comput. Bisnis*, vol. 12, no. 1, pp. 11–27, 2018.
- [7] B. Wira, A. E. Budianto, and A. S. Wiguna, "Implementasi Metode K-Medoids Clustering untuk Mengetahui Pola Pemilihan Program Studi," *J. Terap. Sains Teknol.*, vol. 1, no. 3, pp. 54–69, 2019.
- [8] C. A. Rahardja, T. Juardi, and H. Agung, "Implementasi Algoritma K-Nearest Neighbor Pada Website Rekomendasi Laptop," *J. Buana Inform.*, vol. 10, no. 1, p. 75, 2019, doi: 10.24002/jbi.v10i1.1847.
- [9] N. L. Anggreini, "Teknik Clustering Dengan Algoritma K-Medoids Untuk Menangani Strategi Promosi Di Politeknik Tedc Bandung," *J. Teknol. Inf. dan Pendidik.*, vol. 12, no. 2, pp. 1–7, 2019, doi: 10.24036/tip.v12i2.215.
- [10] I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 119, 2019, doi: 10.24014/rmsi.v5i1.7381.
- [11] S. Aulia, "Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja)," *Djtechno J. Teknol. Inf.*, vol. 1, no. 1, pp. 1–5, 2021, doi: 10.46576/djtechno.v1i1.964.
- [12] G. A. Marcoulides, *Discovering Knowledge in Data: an Introduction to Data Mining*, vol. 100, no. 472. 2005.
- [13] P. Studi, T. Informatika, F. Teknik, D. A. N. Komputer, and U. P. Batam, "PENERAPAN DATA MINING DALAM PEMILIHAN PENERAPAN DATA MINING DALAM PEMILIHAN," 2021.
- [14] W. Gata, "Akurasi Text Mining Menggunakan Algoritma K-Nearest Neighbour pada Data Content Berita SMS," vol. 6, 2016.
- [15] KNBS, "IMPLEMENTASI ALGORITMA K-MEDOIDS DALAM MENENTUKAN TINGKAT PENYEBARAN PANDEMI COVID-19 DI INDONESIA," p. 6, 2021.
- [16] S. Saefudin and D. Fernando, "Penerapan Data Mining Rekomendasi Buku Menggunakan Algoritma Apriori," *JSiI (Jurnal Sist. Informasi)*, vol. 7, no. 1, p. 50, 2020, doi: 10.30656/jsii.v7i1.1899.
- [17] Sutha, "Bab II Landasan Teori," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2018.
- [18] B. A. B. Ii and L. Teori, "Sistem Informasi Koran Berbasis Web," *J. Edukasi*, pp. 9–45, 2014.