



ANALISIS INTERPRETASI KARAKTERISTIK PRODUK MENGUNAKAN METODE CLUSTERING K-MEANS DALAM PENJUALAN BARANG PADA AQUSPACE

Muhammad Renaldi¹, Suzuki Syofian²

^{1,2}Teknologi Informasi, Universitas Darma Persada

muhammadrenaldi588@gmail.com

Abstract

CSS Aquatic is a store that specializes in selling aquascape materials. However, not all types of products sold have high sales values; some products have medium and low sales levels. The sales data at CSS Aquatic is not well-organized, so it only serves as an archive for the store and cannot be used to develop marketing strategies. Therefore, it is necessary to apply data mining using the CRISP-DM (Cross-Industry Standard Process Model for Data Mining) method with the K-Means algorithm at CSS Aquatic. The K-Means algorithm can be applied to determine which products have high, medium, and low sales values. The application of the K-Means method at CSS Aquatic involves grouping the sales data of the products, then selecting three clusters randomly as initial centroids. Once the data in each cluster stabilizes, the final results can be interpreted. This modeling results in an application program with a reporting system that displays clustering results for each type of product based on the quantity sold, which can be interpreted to address the issue.

Keywords: Clustering, CRISP-DM, K-Means, Data Mining, Data Interpretation

ABSTRAK

CSS Aquatic merupakan sebuah toko yang bergerak dalam bidang penjualan material aquascape, namun dari berbagai jenis produk yang dijual tidak semuanya memiliki nilai penjualan yang tinggi, ada produk yang memiliki daya jual sedang dan rendah. Data penjualan pada CSS Aquatic tidak tersusun dengan baik, sehingga data tersebut hanya berfungsi sebagai arsip bagi toko dan tidak dapat dimanfaatkan untuk pengembangan strategi pemasaran. Oleh karena itu perlu diterapkan data mining menggunakan metode CRISP-DM (*Cross-Industry Standard Process Model for Data Mining*) dengan algoritma *K-Means* pada CSS Aquatic. Algoritma *K-Means* dapat diterapkan pada CSS Aquatic untuk menentukan penjualan produk mana yang memiliki nilai penjualan tinggi, sedang, dan rendah. Penerapan metode *K-Means* pada CSS Aquatic yaitu dengan cara mengelompokkan data penjualan produk. Kemudian memilih 3 cluster secara acak sebagai centroid awal. Setelah data pada setiap cluster tidak berubah, maka dapat diinterpretasikan hasil akhirnya. Hasil pemodelan menghasilkan program aplikasi dengan sistem laporan yang menampilkan hasil clustering pada setiap jenis produk barang berdasarkan penjualan quantity barang untuk diinterpretasikan pada masalah tersebut.

Kata kunci: Clustering, CRISP-DM, K-Means, Data Mining, Interpretasi Data

I. PENDAHULUAN

Bisnis dalam sektor perdagangan adalah kegiatan usaha dengan membeli produk atau barang pada suatu perusahaan atau pihak tertentu dan dijual kembali kepada masyarakat dengan tujuan untuk menghasilkan laba secara optimal agar dapat mengembangkan, memajukan dan mempertahankan kelangsungan hidup usahanya ketingkat yang lebih tinggi. Dalam persaingan dunia bisnis pada saat ini, kita dituntut untuk selalu mengembangkan bisnis agar bertahan dalam persaingan, khususnya dalam persaingan penjualan menuntut para pengusaha untuk menemukan suatu pola yang dapat meningkatkan penjualan dan pemasaran di perusahaan, salah satunya adalah dengan pemanfaatan data penjualan (Siregar, 2018).

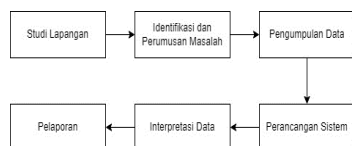
II. METODE PENELITIAN

Penelitian ini menggunakan metode CRISP-DM dan algoritma K-Means Clustering.

Tahapan yang dilakukan adalah sebagai berikut:

1. *Business Understanding*: Mengidentifikasi tujuan bisnis dan memahami masalah yang akan dipecahkan.
2. *Data Understanding*: Mengumpulkan dan memeriksa data penjualan dari CSS Aquatic.

Berikut ini gambar 1 merupakan tahapan daripada penelitian.



Gambar 1 Tahapan Penelitian

2.1. Implementasi CRISP-DM

Penelitian ini menggunakan teknik data mining clustering dengan algoritma K-means, dan menggunakan metode CRISP-DM dengan enam tahapan yaitu business understanding, data understanding, data preparation, modeling, evaluation, dan deployment.

1. *Business Understanding* pada penelitian yang ditunjukkan pada Tabel 1:

Tabel 1 Dataset Awal

No.	Tanggal	Status	Barang
1	2/1/2022	Retail	Batu Pasir
2	2/1/2022	Retail	Pasir Silika - Halus
3	2/1/2022	Retail	Kayu Rasamala - XXL
4	2/1/2022	Retail	Kayu Rasamala - L
5	2/1/2022	Retail	Batu Turly
6	2/1/2022	Retail	Kayu Rasamala - S
7	2/1/2022	Retail	Kayu Rentek - XXL
8	2/1/2022	Retail	Moss Weeping
9	2/1/2022	Retail	Batu Serventine
...
22	31/1/2022	Retail	Pasir Malang
6	2	Retail	Hitam - Halus

2. *Data Preparation*

Proses ini melakukan persiapan data dengan menyesuaikan dataset agar dapat sesuai dengan kebutuhan yang akan digunakan saat tahap pemodelan.

a. *Data Selection*

Pada tahap data selection akan dilakukan penyeleksian data. Tahap selection berupa proses pemilihan data dengan memilih data apa saja yang digunakan dan yang tidak diperlukan pada dataset awal untuk kemudian dijadikan dataset yang dibutuhkan untuk pemodelan. Dataset penjualan yang didapatkan memiliki 226 data. Kemudian data ini akan diseleksi menjadi 37 data, yaitu data penjualan batu, pasir, dan kayu. Data hasil seleksi dapat dilihat pada Tabel 2 berikut:

Tabel 2 *Data Selection*

No.	JENIS BARANG	01-01-2022	02-01-2022	31-01-2022
1	<i>Batu Pasir</i>	(Libur)	27	34
2	<i>Batu Kriket</i>	(Libur)	13	20
3	<i>Batu Erangga</i>	(Libur)	10	40

4	Batu Serventine	(Libur)	0	8
5	Batu Fosil Kayu	(Libur)	7	5
6	Batu Turly	(Libur)	10	10
7	Batu Lava	(Libur)	15	25
8	Batu Besi	(Libur)	0	4
9	Batu Apung	(Libur)	3	2
...
37	Kayu Santigi - XXL	(Libur)	8	15

b. *Data Preprocessing*

Data yang didapatkan setelah melewati proses data selection memiliki sebanyak 37 data dengan total 1.147 record, kemudian data ini akan direduksi menjadi dataset batu, kayu, dan pasir dari total terdapat 32 field kemudian menjadi hanya 5 field saja yaitu atribut jenis barang, penjualan minggu ke-1, minggu ke-2, minggu ke-3, dan minggu ke-4. Data hasil preprocessing dapat dilihat pada Tabel 3 berikut:

Tabel 3 Data Hasil Preprocessing

JENIS BARANG	MINGGU-1	MINGGU 2	MINGGU 3	MINGGU 4
Batu Pasir	178	82	662	219
Batu Kriket	73	56	371	117
Batu Erangga	68	74	350	40
Batu Serventine	31	27	118	34
Batu Fosil Kayu	44	38	249	33
Batu Turly	83	72	294	54
Batu Lava	97	88	569	102
Batu Besi	28	14	8	18
Batu Apung	22	26	121	19
Batu Zeolit	46	17	133	25
Batu Karang Jahe	32	21	111	34

c. *Data Transformation*

Pada tahap Transformasi Data akan dilakukan perubahan data menjadi data yang dapat diolah menggunakan algoritma yang akan dipakai, apakah dalam bentuk numerik, klasifikasi dan lain-lain. Pada dataset yang digunakan ini sudah menggunakan algoritma yang sesuai dengan model, sehingga dapat lanjut ke tahap berikutnya.

3. *Modeling*

Data yang digunakan untuk pemodelan dengan algoritma adalah data penjualan produk batu pada CSS Aquatic dibulan agustus 2022, Berikut adalah data penjualannya pada Tabel 4:

Tabel 4 Data Penjualan Batu

JENIS BARANG	MINGGU 1	MINGGU 2	MINGGU 3	MINGGU 4
Batu Pasir	178	82	662	219
Batu Kriket	73	56	371	117
Batu Erangga	68	74	350	40
Batu Serventine	31	27	118	34
Batu Fosil Kayu	44	38	249	33
Batu Turly	83	72	294	54
Batu Lava	97	88	569	102
Batu Besi	28	14	8	18
Batu Apung	22	26	121	19
Batu Zeolit	46	17	133	25
Batu Karang Jahe	32	21	111	34

Pada tahap ini data yang didapatkan akan digunakan dalam analisis data mining. Transformasi ini dilakukan dengan cara memasukkan data-data transaksi penjualan ke dalam aplikasi data mining.

- Menetapkan jumlah cluster. Jumlah cluster ditetapkan berdasarkan hasil transaksi penjualan yaitu C1, C2, dan C3. Oleh karena itu jumlah cluster yang ditetapkan adalah 3 cluster.
- Inisialisasi pusat cluster (centroid) secara acak berdasarkan data skor hasil survei lapangan. Nilai pusat cluster dapat dilihat pada Tabel 5 berikut:

Tabel 5 Inisialisasi Cluster

Centroid 1	87.5	59.75	375.2	102.5
Centroid 2	63	53	280	51.75
Centroid 3	33.333	21.33	121.6	26

- Hitung jarak setiap data terhadap pusat cluster. Misalnya untuk menghitung jarak distance pertama dengan pusat cluster pertama adalah:

$$d_{11} = \sqrt{(178 - 87.5)^2 + (82 - 59.75)^2 + (662 - 287.75)^2 + (219 - 92.5)^2}$$

$$= 405.89484475662$$

$$d_{12} = \sqrt{(178 - 63)^2 + (82 - 53)^2 + (662 - 280)^2 + (219 - 51.75)^2}$$

$$= 433.54649404649$$

$$d_{13} = \sqrt{(178 - 33.33333333)^2 + (82 - 21.33333333)^2 + (662 - 121.6666667)^2 + (219 - 26)^2}$$

$$= 594.82602498546$$

- Tabel 6 berikut adalah hasil perhitungan lengkap dari perhitungan clustering:

Tabel 6 Perhitungan Hasil Cluster

JENIS BARANG	CENTRO ID 1	CENTRO ID 2	CENTRO ID 3
Batu	323.2384	433.5464	594.8260
Pasir	955416	9404649	2498546
Batu	21.27498	112.4613	270.5981
Kriket	5311393	8226076	0297438

Batu	71.60394	285.5592	139.4787
Erangga	5422023	1014739	9169728
Batu	274.1069	168.1043	10.72380
Serve	5905066	2028952	5294764
Batu	152.1007	43.57249	129.0516
Fosil	0677022	7059498	6923885
Batu	95.52028	31.01713	188.4586
Turly	5803593	2362615	6036526
Batu	196.0296	297.3668	463.0129
Lava	5336908	8198251	5878193
Batu	384.2474	279.0502	114.3080
Besi	7884664	8668683	6329097
Batu	277.5691	169.5982	14.13034
Apung	3553203	3849321	5596151
Batu	261.2300	154.6271	17.56891
Zeolit	2315967	7257972	1937473
Batu	281.2509	175.6731	13.40397
Karang Jahe	9999822	6955073	9508589

- Tabel 7 berikut adalah hasil klasifikasi setiap data berdasarkan kedekatannya dengan titik pusat data atau centroid, diambil berdasarkan tingkat penjualan rendah (C3), sedang (C2), dan tinggi (C1):

Tabel 7 Hasil Klasifikasi Data

JENIS BARANG	CENTRO ID 1	CENTRO ID 2	CENTRO ID 3	CLUSTER
Batu	323.238	433.5	594.826	C1
Pasir	49554	4649	02498	
	16	4046	546	
		49		
Batu	21.2749	112.4	270.598	C1
Kriket	85311	6138	10297	
	393	2260	438	
		76		
Batu	71.6039	285.5	139.478	C1
Erangga	45422	5921	79169	
	023	0147	728	
		39		
Batu	274.106	168.1	10.7238	C3
Serve	95905	0432	05294	
entine	066	0289	764	
		52		
Batu	152.100	43.57	129.051	C2
Fosil	70677	2497	66923	
Kayu	022	0594	885	
		98		

Batu Turly	95.5202 85803 593	31.01 7132 3626 15	188.458 66036 526	C2
Batu Lava	196.029 65336 908	297.3 6688 1982 51	463.012 95878 193	C1
Batu Besi	384.247 47884 664	279.0 5028 6686 83	114.308 06329 097	C3
Batu Apung	277.569 13553 203	169.5 9823 8493 21	14.1303 45596 151	C3
Batu Zeolit	261.230 02315 967	154.6 2717 2579 72	17.5689 11937 473	C3
Batu Karang Jahe	281.250 99999 822	175.6 7316 9550 73	13.4039 79508 589	C3

- Memperbaharui nilai centroid dengan menghitung ulang nilai centroid dari cluster yang baru terbentuk. Proses ini dilakukan dengan menghitung nilai mean dari setiap data points di dalam cluster tersebut.
- Ulangi langkah 3 hingga 4 sampai anggota tiap cluster tidak ada yang berubah.

Tabel 8 berikut adalah hasil dari perulangan kedua yang telah diperbarui dengan cara menghitung ulang jarak centroid dengan cluster pada tabel 7.

Tabel 8 Perhitungan Ulang Centroid ke-2

Centroid 1	104	75	488	119.5
Centroid 2	63.5	55	271.5	43.5
Centroid 3	31.8	21	98.2	26

Tabel 9 berikut adalah hasil klasifikasi data pada perhitungan ke-2 dan terdapat perbedaan anggota cluster dengan perhitungan pertama, maka akan dilakukan perhitungan selanjutnya.

Tabel 9 Hasil Klasifikasi Perhitungan Ke-2

JENIS BARANG	CENTR OID 1	CENT ROID 2	CENTR OID 3	CL USTER
Batu Pasir	213.77 85069	443.9 92961 7	616.61 56664	C1
Batu Kriket	122.54 48897	124.0 71551 9	292.61 45588	C1
Batu Eranga	163.28 27303	80.96 75861 1	260.22 81307	C2
Batu Serventine	389.67 06943	159.6 64492	22.196 3961	C3
Batu Fosil Kayu	263.76 55209	35.85 73562 9	152.40 56429	C2
Batu Turly	205.85 49246	35.85 73562 9	210.58 0341	C2
Batu Lava	84.173 92708	306.8 22016 8	485.97 18922	C1
Batu Besi	500.19 92103	270.2 29069 5	90.903 68529	C3
Batu Apung	392.31 90666	160.6 66580 2	26.265 56681	C3
Batu Zeolit	376.40 83554	145.8 58664 5	37.811 10948	C3
Batu Karang Jahe	396.91 21439	167.3 28270 2	15.095 69475	C3

Tabel 10 berikut adalah nilai centroid pada perhitungan data ke-3 dengan menghitung ulang jarak centroid dari cluster yang baru terbentuk pada tabel 9.

Tabel 10 Perhitungan Ulang Centroid Ke-3

Centroid 1	116	75.33 33333 3	534	146
Centroid 2	65	61.33 33333 3	297.6 6666 67	42.3 333 333 3
Centroid 3	31.8	21	98.2	26

Tabel 11 berikut adalah hasil clustering data pada perhitungan ke-3 dan masih terdapat

perbedaan anggota clusternya dengan perulangan ke-2, maka perulangan masih dilanjutkan karna masih ada perbedaan.

Tabel 11 Hasil Klasifikasi Perhitungan Ke-3

JENIS BARA	CENTR OID 1	CENT ROID 2	CENTR OID 3	CL UST ER
Batu Pasir	160.00 45138	420.8 87158 3	616.61 56664	C1
Batu Kriket	172.14 17375	105.0 96780 8	292.61 45588	C2
Batu Eranga	217.71 0307	53.97 83907 4	260.22 81307	C2
Batu Serventine	441.77 04281	186.2 37303 8	22.196 3961	C3
Batu Fosil Kayu	317.13 05374	58.66 00375	152.40 56429	C2
Batu Turly	259.16 03965	24.23 49609 7	210.58 0341	C2
Batu Lava	60.683 14794	280.9 21697 3	485.97 18922	C1
Batu Besi	551.87 47845	296.8 30479 1	90.903 68529	C3
Batu Apung	444.93 57007	186.6 89581 9	26.265 56681	C3
Batu Zeolit	428.65 46136	172.4 58690 7	37.811 10948	C3
Batu Karang Jahe	448.86 64736	193.9 83676 3	15.095 69475	C3

Tabel 12 berikut adalah nilai centroid pada perhitungan data ke-4 dengan menghitung ulang jarak centroid dari cluster yang baru terbentuk pada table 11

Tabel 12 Perhitungan Ulang Centroid Ke-4

Centroid 1	137.5	85	615.5	160.5
Centroid 2	67	60	316	61
Centroid 3	31.8	21	98.2	26

3.1 Interpretasi Cluster

Interpretasi cluster merupakan proses terakhir dari pengklasteran yang bertujuan untuk memberi ciri spesifik atau menggambarkan isi kluster yang terbentuk. Tabel 12 berikut adalah data dari hasil pemodelan:

Berdasarkan data dari masing masing cluster pada tabel 12, maka dapat diinterpretasikan bahwa:

1. Produk dengan Tema Populer *Waterfall* dan bobot produk yang tidak berat memiliki daya penjualan paling tinggi dan semua anggotanya berada pada cluster 1, ini menginterpretasikan bahwa tema *Waterfall* memiliki peminatan paling tinggi bagi customer dan memiliki market yang lebih luas, sehingga produk-produk yang berkaitan dengan bahan baku pembuatan tema tersebut berdampak berupa meningkatnya penjualan pada produk tersebut.

Berdasarkan data interpretasi ini dapat dijadikan acuan untuk menggali potensi segmentasi ini dalam hal penjualan dan peningkatan popularitas toko. Yaitu dengan cara menjadikan produk bahan-bahan *Waterfall* sebagai produk andalan dengan memberikan kualitas yang baik dan harga yang terbaik dibanding pesaing bisnis, lalu mengiklankan produk dengan popularitas paling tinggi ini melalui iklan online pada e-commerce agar memudahkan calon pelanggan untuk mengenali keunggulan produk pada CSS Aquatic, dengan ini produk dengan tema populer selain *Waterfall* juga memungkinkan mendapat manfaatnya seiring dengan dengan meningkatnya popularitas toko.

2. Produk dengan Cluster ke-2 adalah produk dengan penjualan sedang dan diisikan anggota kelompok dengan Tema Populer Natural sebanyak 75% anggota, dan 25% dengan Tema Populer Tebing. Pada kelompok ini semua anggotanya memiliki kesamaan pada data warna, yaitu warna cerah dan bobot yang berat. Berdasarkan data

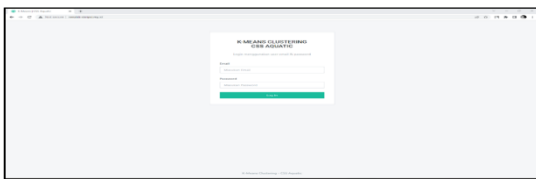
tersebut dapat diinterpretasikan bahwa walaupun produk pada cluster ke-2 memiliki bobot yang berat, namun produk pada cluster ini tetap memiliki peminat yang cukup banyak karna memiliki warna yang cerah sebagai daya tarik. Oleh karna itu berdasarkan data hasil pemodelan, disarankan bagi CSS Aquatic untuk menstabilkan produk produk bahan lain dari tema *Natural* agar tidak terjadi penurunan penjualan pada segmentasi ini.

III. HASIL PEMBAHASAN

3.RANCANGAN DAN HASIL

a. Pembuatan *User Interface*

Pada *User Interface* yang dibuat ini digunakan sebagai akses user ke aplikasi, berikut adalah gambar 2 yang menampilkan *User Interface Form Login*, pada form login ini admin akan diminta untuk menginputkan username dan password.



Gambar 2 *Form Login*

b. Halaman Dashboard

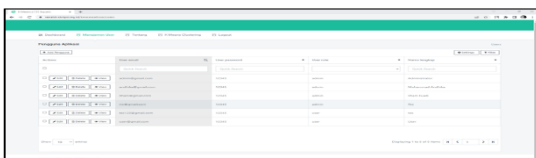
Pada gambar 3 berikut ini adalah tampilan awal pada saat admin login ke aplikasi dan ditampilkan halaman dashboard.



Gambar 3 *Halaman Dashboard*

c. Halaman Manajemen User

Pada gambar 4 berikut adalah tampilan menu manajemen user, dimana admin dapat melihat dan mengontrol semua data user yang terdaftar pada program.



Gambar 4 *Halaman Manajemen User*

d. Halaman Tentang K-Means

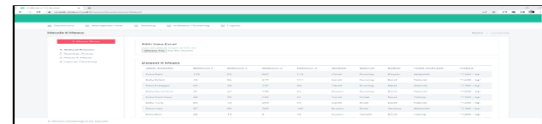
Gambar 5 berikut adalah Halaman Tentang K-Means, halaman ini berisi informasi-informasi tentang K-Means dan bagaimana proses K-Means dilakukan, hingga contoh penerapannya.



Gambar 5 *Halaman Tentang K-Means*

e. Halaman Dataset Clustering K-Means

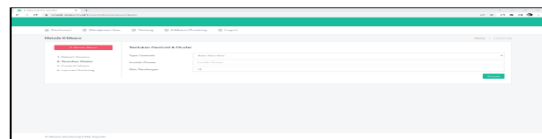
Selanjutnya pada gambar 6 berikut adalah tampilan menu input dataset pada halaman Clustering K-Means. Pada menu ini admin dapat menginput dataset untuk kemudian diproses untuk dilakukan perhitungan clustering.



Gambar 6 *Halaman Dataset Clustering K-Means*

f. Halaman Dataset Clustering K-Means

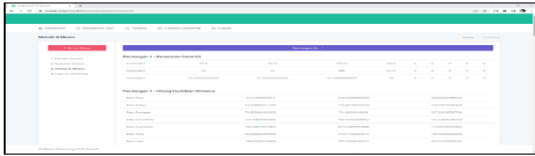
Pada gambar 7 berikut ini adalah tampilan menu Tentukan Cluster, pada halaman ini user diminta untuk menentukan jumlah cluster dan maksimal perulangan pada cluster yang diinginkan, berikut adalah tampilannya:



Gambar 7 *Menu Tentukan Cluster*

g. Menu Proses K-Means

Menu selanjutnya adalah menu Proses K-Means, pada menu Proses K-Means ini akan ditampilkan perulangan dengan metode berdasarkan jumlah yang telah diinput pada menu Tentukan Cluster sebelumnya, berikut adalah tampilan menu Proses K-Means pada gambar 8.



Gambar 8 Menu Proses K-Means

h. Menu Laporan Clustering

Pada gambar 9 berikut adalah tampilan dari menu Laporan Clustering, pada menu Clustering ini akan ditampilkan ringkasan hasil akhir dari proses clustering pada menu menu sebelumnya.



Gambar 9 Menu Laporan Clustering

I. PENGUJIAN

Pada pengujian ini dilakukan uji coba pada setiap menu yang telah dibuat, aplikasi ini diuji oleh Administrator CSS Aquatic untuk melihat fungsionalitas, kelengkapan menu dan tampilan aplikasi. Hasil pengujian tersebut memberikan keterangan kesesuaian dari sistem program yang dibuat.

Pada tabel 16 ini menjelaskan hasil uji coba dari program aplikasi yang diawali dengan login admin, kelola file dataset, kelola perulangan cluster, melihat detail proses K-Means, dan melihat laporan hasil berjalan dengan baik..

Tabel 16 Hasil Uji Coba

No	Komponen	Hasil
1	Login	Sukses
2	Kelola file dataset	Sukses
3	Kelola data user	Sukses
4	Kelola perulangan klasterisasi program	Sukses
5	Melihat detail proses K-Means	Sukses
6	Melihat laporan hasil clustering	Sukses

IV. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan bahwa:

- Perhitungan dengan algoritma kmeans clustering diperoleh cluster atau pengelompokan yang sama. Untuk cluster 1 sebagai cluster dengan peminatan paling tinggi, cluster 2 sebagai cluster dengan peminatan menengah, dan cluster 3 sebagai cluster dengan peminatan terendah.
- Aplikasi Clustering Data Mining tersebut akan membantu bagi CSS Aquatic dalam menginterpretasi data penjualan barang dan mendapati kesimpulan peminatan customer terhadap karakteristik produk dalam kelompok jenis barang yang sangat diminati, lumayan diminati, dan kurang diminati.
- Metode K-Means Clustering dapat menginterpretasi data hasil penjualan sebagai data yang dapat dimanfaatkan untuk pembangunan strategi pemasaran, dengan membuat pengelompokan jenis barang untuk mencari nilai ketertarikan customer.

DAFTAR PUSTAKA

Budiana, Nurul, D., Riki, R.A.S., & Meilia, N.I.S. (2019), 'Penetapan Instruktur Diklat Menggunakan Metode Clustering K-Means dan Topsis Pada PT PLN (Persero) Udiklat Jakarta'

Gustrianda, Reza, & Dadang, I.M. (2022), 'Penerapan Data Mining Dalam Pemilihan Produk Unggulan dengan Metode Algoritma K-Means Dan K-Medoids.' Jurnal Media Informatika Budidarma 6.1.

Handoko, Suhandio, Fauziah, F., & Endah, T.E.H. (2020), 'Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-