



PENERAPAN DATA MINING CLUSTERING UNTUK MENGELOMPOKKAN BERBAGAI JENIS MERK MESIN CUCI

Wulan Alinda Wahyuni¹⁾, Sudin Saepudin²⁾

^{1, 2)}Program Studi Sistem Informasi, Universitas Nusa Putra

Jl. Raya Cibolang No. 21 Cibolang Kaler, Cisaat, Sukabumi, Jawa Barat 43152
e-mail: Wulan.alinda_si18@nusaputra.ac.id¹⁾, Sudinsaepudin@nusaputra.ac.id²⁾

* Korespondensi: e-mail: Sudinsaepudin@nusaputra.ac.id

ABSTRAK

Seiring berkembangnya zaman berkembang pula teknologi-teknologi yang mempermudah manusia untuk melakukan sesuatu hal, salah satunya adalah mesin cuci. Mesin cuci saat ini menjadi kebutuhan pokok bagi ibu rumah tangga, karena dengan adanya mesin cuci mempermudah mereka untuk mencuci pakaian, tidak perlu repot dan capek dengan adanya mesin cuci mereka tinggal duduk manis menunggu atau bisa juga melakukan pekerjaan lain sambil menunggu. Namun mesin cuci juga ada beberapa tipe dengan berbeda keunggulan. Hal ini membuat para ibu rumah tangga bingung untuk memilih mesin cuci tipe apa yang mereka inginkan. mesin cuci mana yang terbaik. Pada penelitian ini penulis menggunakan metode K-Means Clustering. Maka dengan ini penulis ingin mengelompokkan berbagai jenis dan merk mesin cuci. Hal ini diharapkan agar mempermudah bagi para ibu rumah tangga untuk memilih mesin cuci yang mereka butuhkan.

Kata Kunci: *Mesin cuci, K-Means, Clustering*

ABSTRACT

Along with the development of the times, technologies that make it easier for humans to do things, one of which is a washing machine. Washing machines are now a basic need for housewives, because with a washing machine it makes it easier for them to wash clothes, they don't need to be bothered and tired of having a washing machine, they just sit back and wait or they can do other work while waiting. but there are also several types of washing machines with different advantages. This makes housewives confused to choose what type of washing machine they want. Which washing machine is the best In this study, the author uses the K-Means Clustering method. So with this the author wants to classify various types and brands of washing machines. This is expected to make it easier for housewives to choose the washing machine they need.

Keywords: *Washing Machine, K-Means, Clustering*

I. PENDAHULUAN

Dengan berkembangnya teknologi saat ini, persaingan dalam dunia teknologi semakin hari semakin meningkat, hampir semua orang tidak mau ketinggalan zaman sehingga teknologi pada saat ini merupakan kebutuhan pokok yang tidak bisa terlewatkan. Salah satu teknologi yang sering dibutuhkan oleh orang banyak adalah penggunaan mesin cuci. Mesin cuci saat ini banyak sekali jenisnya, berbeda merk dan tipe maka berbeda pula keunggulan mesin cuci tersebut. Hal ini mengharuskan setiap orang yang membelinya harus selektif memilih mesin cuci yang terbaik untuk kebutuhannya.

Dengan itu dibutuhkan cara khusus untuk dapat mempermudah seseorang untuk memilih mesin cuci yang terbaik, dengan itu penulis akan melakukan pengelompokan merk mesin cuci menggunakan data mining K-Means Clustering dengan kriteria : Merk dan tipe, Watt, dan kapasitas pakaian dengan data yang diambil adalah bersumber dari internet. Diharapkan pengelompokan ini dapat mempermudah orang banyak untuk dapat memilih mesin cuci terbaik untuk kebutuhannya.

II. TINJAUAN PUSTAKA

A. Clustering

Clustering merupakan metode penganalisaan data yang sering dimasukan sebagai salah satu metode data mining yang tujuannya adalah untuk mengelompokkan data dengan karakteristik yang sama.

Baskoro 2010 [1], menyatakan bahwa ‘Clustering adalah salah satu alat bantu pada data mining yang bertujuan mengelompokkan objek-objek kedalam cluster-cluster’

Clustering merupakan salah satu teknik dalam data mining yang berfungsi untuk memetakan data pada kelompok atau cluster tertentu[2]. Setiap data dikelompokkan berdasarkan kemiripannya dengan data yang lain. Data yang memiliki karakteristik yang sama akan dimasukkan dalam cluster yang sama, dan data yang memiliki karakteristik yang berbeda akan ditempatkan dalam cluster yang lain. Banyak sekali metode yang dapat digunakan untuk metode clustering, salah satunya adalah metode K-Means.

B. K-Means

K-means merupakan salah satu metode yang sering digunakan untuk clustering. Algoritma K-Means merupakan salah satu metode clustering non hirarki yang memiliki waktu komputasi yang relatif cepat [2]. Algoritma K-Means ini dimulai dengan menentukan terlebih dahulu berapa cluster yang ingin dibuat, selanjutnya menentukan nilai awal untuk masing-masing cluster, lalu menghitung jarak dari masing-masing data dengan nilai awal yang telah ditentukan. Setelah jarak dari tiap data dihitung, data dimasukkan dalam cluster terdekat. Selanjutnya menghitung rata-rata dari tiap cluster, rata-rata tersebut nantinya akan digunakan sebagai nilai awal yang baru dalam menghitung jarak dari masing-masing data.

Pada K-Means ini, pusat cluster atau centroid dipilih pada tahap awal secara acak dari sekumpulan koleksi (populasi) data. Kemudian K-Means menguji masing-masing komponen didalam populasi data dan menandai komponen tersebut ke salah satu centroid yang telah didefinisikan sebelumnya berdasarkan jarak minimum antara komponen (data) dengan masing- masing centroid. Posisi centroid akan dihitung kembali sampai semua komponen data dikelompokkan ke setiap centroid dan terakhir akan terbentuk posisi centroid baru.

Langkah-langkah melakukan clustering dengan metode K-Means Clustering adalah sebagai berikut :

1. Menentukan nilai K untuk jumlah cluster yang ingin dibentuk
2. Memilih nilai untuk pusat cluster awal (centroid) sebanyak K
3. Menghitung jarak setiap data input terhadap masing-masing centroid menggunakan rumus jarak Euclidean hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid.
 - Jarak Euclidean yang dirumuskan sebagai berikut

$$D(i,j) = \sqrt{(X_{1j} - X_{1i})^2 + (X_{2j} - X_{2i})^2 + \dots + (X_{kj} - X_{ki})^2} \dots (1)$$

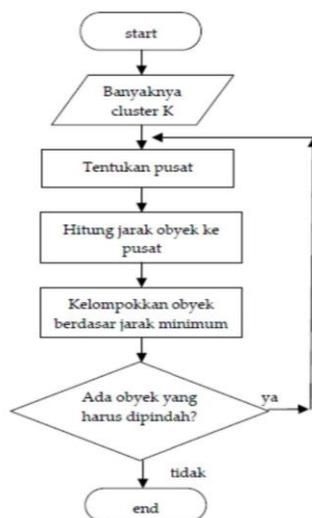
$D(i,j)$ = Jarak data ke i ke pusat cluster j

X_{ki} = Data ke i pada atribut data ke K

X_{kj} = Data ke j pada atribut data ke K

III. METODE PENELITIAN

Tahapan-tahapan yang dilakukan dalam penelitian adalah seperti pada gambar dibawah



Gambar.1. Flowchart Clustering K-Means



Gambar.1. diatas menunjukkan langkah-langkah menentukan clustering dengan metode K-Means menurut [3] yang terdiri dari beberapa tahapan sebagai berikut:

Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke cluster tertentu ditentukan jarak antara data ke tiap pusat cluster. Jarak paling dekat antara satu data dengan satu cluster tertentu akan menentukan suatu data masuk dalam cluster mana. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan jarak Euclidean yang dirumuskan sebagai berikut:

1. Tentukan jumlah cluster
2. Menentukan titik awal cluster secara random
3. Hitung jarak ke masing-masing centroid
4. Kelompokkan objek berdasarkan jarak minimum
5. Kembali ke Step 2, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai threshold yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan di atas nilai threshold yang ditentukan.

$$D = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2} \dots\dots(2)$$

Keterangan :

- D = jarak
- X₁ = koordinat latitude 1
- X₂ = koordinat latitude 2
- Y₁ = koordinat longitude 1
- Y₂ = koordinat longitude 2

Koordinat latitude dan longitude dapat dijadikan variabel guna melakukan perhitungan jarak antara dua buah titik lokasi apabila dibentangkan garis lurus diantara keduanya (Yulianto, Ramadiani, & Kridalaksana, 2018).

IV. HASIL DAN PEMBAHASAN

A. Menentukan Jumlah Cluster

Tabel.1. Data yang dianalisis yaitu data dari 25 tipe mesin cuci

N	Nama	Parameter	
		KAPASITAS	WATT
1	SANKEN TW-7750	7	150
2	DENPOO DW 8907 B	8	370
3	TOSHIBA VH - H95MN	9	180
4	DENPOO DWF-112	7	400
5	SANKEN TW-1127656	9	250
6	SANKEN TWIN TUB- TW-882N	7	230
7	SHARP EST 95CR	9	255
8	TOSHIBA HV-J140MN	9	220
9	DENPOO DW 898	8	370
10	SAMSUNG TOP LOADING WA 70H420056	7	350
11	SANYO AQUA 880XT	8	120
12	SHARP T65	7	330
13	DENPOO DW 976	8	400
14	AQUA AQW-77D	7	380
15	SHARP ES-T75NT-PK	7	330
16	SHARP ES T97CM-BL	9	225
17	SHARP MW	9	130
18	PANASONIC NA-WG-FCI	9	250
19	DENPOO DW 901	8	370



20	SANKEN TW-8660-EPK	8	250
21	POLYTRON PAW 90617	9	330
22	SANKEN TW-1127656	9	250
23	SHARP T65	7	140
24	L6 P9050R	9	350
25	AQUA QW-881XT	8	350

B. Menentukan titik awal cluster secara random

Berdasarkan data dari tabel diatas maka kita akan menentukan banyaknya cluster yang dibentuk tiga ($k=3$). Banyaknya cluster harus lebih kecil dari pada banyaknya data ($k < n$).

Tabel.2. Tabel Data

	Jumlah Cluster	3	
	Pusat Cluster	(random)	
C 1	Sanken TW-7750	7	150
C 2	Sanken Tw N 8660	8	250
C 3	LG PG05012	9	350

Inisialisasi centroid dataset pada tabel dataset diatas adalah $C1 = \{7, 150\}$, $C2 = \{8, 250\}$ dan $C3 = \{9, 350\}$. Inisialisasi centroid dapat ditentukan secara manual maupun random.

Untuk pengulangan berikutnya (pengulangan ke-1 sampai selesai), centroid baru dihitung dengan **menghitung nilai rata-rata data pada setiap cluster**. Jika centroid baru berbeda dengan centroid sebelumnya, maka proses dilanjutkan ke langkah berikutnya. Namun jika centroid yang baru dihitung sama dengan centroid sebelumnya, maka proses clustering selesai.

C. Hitung jarak ke masing-masing centroid dan mengelompokkan objek berdasarkan jarak minimum

Setelah didapat titik pusat awal cluster, kemudian dilakukan perhitungan jarak Euclidean, dan mengelompokkan berdasarkan jarak terkecil selanjutnya akan di dapat nilai centroid baru untuk acuan perhitungan berikutnya sampai nilai centroid sebelum dan sesudah bernilai sama.

Berikut ini beberapa hasil perhitungan jarak Euclidean Hasil perhitungan persamaan Euclidean Distance:

Pada titik pusat Cluster 1 (7,150)

Jarak data dengan Centroid C1 adalah

$C1 = (7,150)$

$$[1] d(x1,c1) = \sqrt{(a1 - c1a)^2 + (b1 - c1b)^2} = \sqrt{(7 - 7)^2 + (150 - 150)^2} = 0$$

$$[2] d(x2,c1) = \sqrt{(a2 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(8 - 7)^2 + (370 - 150)^2} = 220,0022727$$

$$[3] d(x3,c1) = \sqrt{(a3 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(9 - 7)^2 + (180 - 150)^2} = 30,06659276$$

Jarak data dengan Centroid C2 adalah

$C2 = (8,250)$

$$[1] d(x1,c1) = \sqrt{(a1 - c1a)^2 + (b1 - c1b)^2} = \sqrt{(7 - 8)^2 + (150 - 250)^2} = 100,0049999$$

$$[2] d(x2,c1) = \sqrt{(a2 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(8 - 8)^2 + (370 - 250)^2} = 120$$

$$[3] d(x3,c1) = \sqrt{(a3 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(9 - 8)^2 + (180 - 250)^2} = 70,00714249$$

Jarak data dengan Centroid C3 adalah

$C3 = (9,350)$

$$[1] d(x1,c1) = \sqrt{(a1 - c1a)^2 + (b1 - c1b)^2} = \sqrt{(7 - 9)^2 + (150 - 350)^2} = 200,0099998$$

$$[2] d(x2,c1) = \sqrt{(a2 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(8 - 9)^2 + (370 - 350)^2} = 20,02498439$$

$$[3] d(x3,c1) = \sqrt{(a3 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(9 - 9)^2 + (180 - 350)^2} = 170$$

Untuk seterusnya, hitung jarak pada setiap baris data, dan hasilnya seperti pada tabel dibawah.



Tabel.3. Tabel Proses I

Jarak ke pusat cluster		Parameter				
	Nama	Kapasitas	Watt	C1	C2	C3
1	Sanken TW-7750	7	150	0	100,0049999	200,0099998
2	Denpoo DW 8907 B	8	370	220,0022727	120	20,02498439
3	Toshiba VH - H95MN	9	180	30,06659276	70,00714249	170
4	DENPOO DWF-112	7	400	250	150,0033333	50,03998401
5	SANKEN TW-1127656	9	250	100,019998	1	100
6	SANKEN TWIN TUB-TW-882N	7	230	80	20,02498439	120,0166655
7	SHARP EST 95CR	9	255	105,0190459	5,099019514	95
8	TOSHIBA HV-J140MN	9	220	70,0285656	30,01666204	130
9	DENPOO DW 898	8	370	220,0022727	120	20,02498439
10	SAMSUNG TOP LOADING WA 70H420056	7	350	200	100,0049999	2
11	SANYO AQUA 880XT	8	120	30,01666204	130	230,0021739
12	SHARP T65	7	330	180	80,00624976	20,09975124
13	DENPOO DW 976	8	400	250,002	150	50,009999
14	AQUA AQW-77D	7	380	230	130,0038461	30,06659276
15	SHARP ES-T75NT-PK	7	330	180	80,00624976	20,09975124
16	SHARP ES T97CM-BL	9	225	75,02666193	25,01999201	125
17	SHARP MW	9	130	20,09975124	120,0041666	220
18	PANASONIC NA-WG-FCI	9	250	100,019998	1	100
19	DENPOO DW 901	8	370	220,0022727	120	20,02498439
20	SANKEN TW-8660-EPK	8	250	100,0049999	0	100,0049999
21	POLYTRON PAW 90617	9	330	180,0111108	80,00624976	20
22	SANKEN TW-1127656	9	250	100,019998	1	100
23	SHARP T65	7	140	10	110,0045454	210,0095236
24	L6 P9050R	9	350	200,0099998	100,0049999	0
25	AQUA QW-881XT	8	350	200,0025	100	1

Pengelompokkan data pada pengulangan ke-1

Tabel.4. Tabel Proses II

Pusat Cluster Baru					
C1		C2		C3	
Penyimpanan	Baterai	Penyimpanan	Baterai	Penyimpanan	Baterai
7	150				
				8	370
9	180				
				7	400
		9	250		
		7	230		
		9	255		
		9	220		
				8	370
				7	350
8	120				
				7	330
				8	400
				7	380
				7	330
		9	225		
9	130				



		9	250		
				8	370
		8	250		
				9	330
		9	250		
7	140				
				9	350
				8	350
Pengelompokkan data pada pengulangan ke-1					

Kelompokan data sesuai dengan cluster-nya, yaitu data yang memiliki jarak terpendek.

Pada tabel diatas, data N = 1 masuk ke dalam cluster 3 karena $dc3 < dc2 < dc1$, sedangkan N = 5,6,7,8 masuk ke dalam cluster 2 karena $dc2 < dc3 < dc1$.

Setelah mendapatkan label cluster untuk masing-masing data n=1,2,3,4,5 – 29 maka dicari nilai rata-ratanya dengan menjumlahkan seluruh anggota masing-masing cluster dan dibagi jumlah anggotanya.

Tabel.5. Tabel Cluster

Pusat Cluster		
C1	8	144
C2	8,625	241,25
C3	7,75	360,8333333

Dari Tabel hasil iterasi 1 dapat dilihat centroid baru yang dihasilkan dari iterasi 1 dengan titik pusat awal tidak sama, maka perlu dilakukan iterasi kedua. Berikut ini beberapa hasil perhitungan jarak Euclidean iterasi ke-2

Jarak data dengan Centroid C1 adalah

$$C1 = (8,144)$$

$$[1] \quad d(x1,c1) = \sqrt{(a1 - c1a)^2 + (b1 - c1b)^2} = \sqrt{(7 - 8)^2 + (150 - 144)^2} = 6,08276253$$

$$[2] \quad d(x2,c1) = \sqrt{(a2 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(8 - 8)^2 + (370 - 144)^2} = 226$$

$$[3] \quad d(x3,c1) = \sqrt{(a3 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(9 - 8)^2 + (180 - 144)^2} = 36,01388621$$

Jarak data dengan Centroid C2 adalah

$$C2 = (8.625,241.25)$$

$$[1] \quad d(x1,c1) = \sqrt{(a1 - c1a)^2 + (b1 - c1b)^2} = \sqrt{(7 - 8.625)^2 + (150 - 241.25)^2} = 91,26446803$$

$$[2] \quad d(x2,c1) = \sqrt{(a2 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(8 - 8.625)^2 + (370 - 241.25)^2} = 128,751517$$

$$[3] \quad d(x3,c1) = \sqrt{(a3 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(9 - 8.625)^2 + (180 - 241.25)^2} = 61,25114795$$

Jarak data dengan Centroid C3 adalah

$$C3 = (7.75,360,83)$$

$$[1] \quad d(x1,c1) = \sqrt{(a1 - c1a)^2 + (b1 - c1b)^2} = \sqrt{(7 - 7.75)^2 + (150 - 360.83)^2} = 210,8346673$$

$$[2] \quad d(x2,c1) = \sqrt{(a2 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(8 - 7.75)^2 + (370 - 360.83)^2} = 9,170075124$$

$$[3] \quad d(x3,c1) = \sqrt{(a3 - c1a)^2 + (b2 - c1b)^2} = \sqrt{(9 - 7.75)^2 + (180 - 360.83)^2} = 180,8376536$$



Tabel.6. Pusat Cluster Baru

Pusat Cluster baru					
C1		C2		C3	
Kapasitas	WAT T	Kapasitas	WAT T	Kapasitas	WAT T
7	150				
				8	370
9	180			7	400
		9	250		
		7	230		
		9	255		
		9	220		
				8	370
				7	350
8	120				
				7	330
				8	400
				7	380
				7	330
		9	225		
9	130				
		9	250		
				8	370
		8	250		
				9	330
		9	250		
7	140				
				9	350
				8	350

Setelah mendapatkan label cluster untuk masing-masing data $n=1,2,3,4,5 - 25$ maka dicari nilai rata-ratanya dengan menjumlahkan seluruh anggota masing-masing cluster dan dibagi jumlah anggotanya.

Tabel.7. Pusat cluster baru

Pusat Cluster		
C1	8	144
C2	8,625	241,25
C3	7,75	360,8333333
Nilai rata-rata centroid pada pengulangan ke-2		

Karna Centroid tidak mengalami perubahan (sama dengan centroid sebelumnya) maka proses clustering selesai.

Hasil :

Tabel.8. Table Hasil

No	Nama	Kelompok Cluster
1	Sanken TW-7750	C1
2	Denpoo DW 8907B	C3
3	Toshiba VH - H95MN	C1
4	DENPOO DWF-112	C3
5	SANKEN TW-1127656	C2
6	SANKEN TWIN TUB- TW-	C2



	882N	
7	SHARP EST 95CR	C2
8	TOSHIBA HV-J140MN	C2
9	DENPOO DW 898	C3
10	SAMSUNG TOP LOADING WA 70H420056	C3
11	SANYO AQUA 880XT	C1
12	SHARP T65	C3
13	DENPOO DW 976	C3
14	AQUA AQW-77D	C3
15	SHARP ES-T75NT-PK	C3
16	SHARP ES T97CM-BL	C2
17	SHARP MW	C1
18	PANASONIC NA-WG-FCI	C2
19	DENPOO DW 901	C3
20	SANKEN TW-8660-EPK	C2
21	POLYTRON PAW 90617	C3
22	SANKEN TW-1127656	C2
23	SHARP T65	C1
24	L6 P9050R	C3
25	AQUA QW-881XT	C3

V. KESIMPULAN

Berdasarkan penelitian ini didapat kesimpulan sebagai berikut:

1. Telah berhasil dikelompokkan data berbagai jenis merk Mesin cuci menggunakan algoritma K-means Clustering menjadi 3 kelompok yaitu untuk kelompok 1 berjumlah 5 data, kelompok 2 berjumlah 8 data, dan kelompok 3 berjumlah 12 data.
2. Penentuan titik pusat awal (centroid) sangat berpengaruh terhadap jumlah iterasi yang akan dihitung.
3. Berdasarkan penelitian ini dapat dijadikan acuan dalam memilih merk mesin cuci yang tepat dan sesuai dengan kebutuhan juga keinginan.

DAFTAR PUSTAKA

- [1] Kusrini dan Luthfi Emha Taufiq, Algoritma Data Mining, Yogyakarta: Andi, 2009.
- [2] Ridlo, M.R., Defiyanti, S., dan Primajaya, A. Implementasi Algoritma K-Means untuk Pemetaan Produktivitas Panen Padi di Kabupaten Karawang. Yogyakarta. 2017.
- [3] Santosa, B. (2007). Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu.