

BAB II

TINJAUAN PUSTAKA

2.1. Pengertian Klasterisasi

Klasterisasi merupakan cara untuk mengelompokkan data. Lebih lengkapnya klasterisasi adalah proses pengelompokan data menjadi beberapa kelompok atau *cluster* sehingga dalam satu kelompok memiliki kemiripan yang maksimum dan data antar kelompok memiliki kemiripan yang minimum. Kemiripan yang dimaksud merupakan pengukuran secara numerik antara dua objek. Nilai kemiripan akan semakin tinggi jika kedua objek yang dibandingkan memiliki kemiripan yang tinggi (Tan, Steinbach, & Kumar, 2006).

Clustering sangat berguna untuk prediksi dan analisa masalah bisnis tertentu. Misalnya, marketing, segmentasi pasar dan pemetaan zonasi wilayah. Syarat dalam melakukan klasterisasi adalah skalabilitas, kemampuan analisa dalam bentuk data, menemukan kelompok dengan bentuk yang tidak terduga, mampu menangani data yang rusak, sensitifitas terhadap perubahan data, interpretasi dan kegunaan. Kualitas dari hasil klasterisasi bergantung pada metode yang dipakai (Han, Kamber, & Pei, 2012).

2.2. Metode *K-Means*

Metode *K-Means* merupakan metode pengelompokkan yang berbasis titik pusat. Metode ini memiliki kemampuan mengelompokkan data dalam jumlah besar dengan waktu yang cepat dan efisien. *K-Means* dijalankan secara independen dengan inisialiasi yang berbeda sehingga menghasilkan kelompok akhir yang berbeda (Jain, 2009). Berikut ini merupakan langkah-langkah dalam metode *K-Means*:

1. Tentukan berapa banyak *cluster*.
2. Tetapkan data k secara acak sebagai pusat *cluster*.
3. Hitung masing-masing jarak data ke pusat *cluster*.
4. Setiap data memilih pusat *cluster* terdekat.

5. Tentukan titik pusat *cluster* terbaru dengan menghitung rata-rata dari data yang memilih pusat *cluster* yang sama.
6. Ulangi langkah 3 jika posisi pusat *cluster* baru dan lama tidak sama.

Berikut ini merupakan contoh perhitungan *K-Means*.

- a. Jumlah kelompok : 2
- b. Jumlah data : 12
- c. Jumlah atribut : 8

Kriteria yang digunakan untuk parameter sebagai berikut:

P1 = Pedagogik

P2 = Kepribadian

P3 = Sosial

P4 = Profesional

P5 = Kehadiran

P6 = Disiplin

P7 = Kerajinan

P8 = IQ

Tabel 2.1 Struktur Tabel Data Nilai

No.	NIP	Nama	Hasil Nilai Parameter							
			P1	P2	P3	P4	P5	P6	P7	P8
1.	A-1	Hendra S	79	75	75	85	76	78	76	80
2.	A-2	Roni Purba	84	76	79	77	76	77	75	81
3.	A-3	Joni Darmawan	77	84	78	85	92	89	77	82
4.	A-4	Dewi Roslina	78	86	84	77	78	77	75	75
5.	A-5	Haris Siregar	82	82	81	91	90	82	79	91
6.	A-6	Indra Wijaya	75	75	70	82	75	75	79	75
7.	A-7	Yuyun Sundari	77	75	75	89	80	80	75	75
8.	A-8	Safrizal	77	76	70	77	77	80	75	80
9.	A-9	Gilang R	79	76	75	84	77	81	76	77
10.	A-10	Kiki Angreni	80	75	75	75	75	78	77	79
11.	A-11	Sandra W	76	71	75	75	77	81	79	84
12.	A-12	Robert Sitorus	80	77	75	78	78	77	77	80

Iterasi ke-1

1. Penentuan pusat awal *cluster*

Untuk penentuan awal pusat *cluster* di asumsikan sebagai berikut :

Diambil data ke - 2 sebagai pusat *Cluster* Ke - 1 :

$$C1 = 84; 76; 79; 77; 76; 77; 75; 81$$

Diambil data ke - 5 sebagai pusat *Cluster* Ke - 2 :

$$C2 = 82; 82; 81; 91; 90; 82; 79; 91$$

2. Perhitungan matrik jarak pusat *cluster*

Untuk mengukur jarak antara data dengan pusat *cluster* digunakan Euclidian distance, kemudian akan didapatkan matrik jarak sebagai berikut :

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

Rumus 2.1 Rumus *Euclidian Distance*

Dimana :x = pusat *cluster*

y = data

Sebagai contoh, perhitungan jarak ke A-1 terhadap pusat *cluster* adalah :

$$C1 = \sqrt{\frac{(84-79)^2 + (76-75)^2 + (79-75)^2 + (77-85)^2 + (76-76)^2 + (77-78)^2 + (75-76)^2 + (81-80)^2}{8}} = 10.44$$

$$C2 = \sqrt{\frac{(82-79)^2 + (82-75)^2 + (81-75)^2 + (91-85)^2 + (90-76)^2 + (82-78)^2 + (79-76)^2 + (91-80)^2}{8}} = 21,7$$

Dan seterusnya dilanjutkan untuk data ke 2, 3, ... n

Kemudian akan didapatkan matrik jarak sebagai berikut :

Tabel 2.2 Struktur Tabel Matrik Jarak

A-1	A-2	A-3	A-4	A-5	A-6	A-7	A-8	A-9	A-10	A-11	A-12	
10.44	0	24.15	14.18	24.02	15.65	16.46	11.87	11.14	6.86	12.29	6.56	C1
21.73	24.02	14.56	26.04	0	28.81	22.14	26.23	22.47	27.09	25.85	22.98	C2

Setiap kolom pada matrik menunjukkan nilai jarak data terhadap pusat *cluster*.

Baris pertama pada matrik menunjukkan nilai jarak data terhadap titik pusat *cluster* pertama, baris kedua pada matrik menunjukkan nilai jarak data terhadap titik pusat *cluster* kedua dan seterusnya.

3. Matrik Pengelompokan data

Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat *cluster*, jarak ini menunjukkan bahwa data tersebut

berada dalam satu kelompok dengan pusat *cluster* terdekat. Berikut hasil matriks pengelompokan grup, nilai 1 berarti data tersebut berada dalam grup.

Tabel 2.3 Struktur Tabel Hasil Matrik Pengelompokan Grup

A-1	A-2	A-3	A-4	A-5	A-6	A-7	A-8	A-9	A-10	A-11	A-12	
1	1	0	1	0	1	1	1	1	1	1	1	C1
0	0	1	0	1	0	0	0	0	0	0	0	C2

Dan dari hasil matrix pengelompokan itu kita mendapatkan pembagian kelompok alternatif sebagai berikut :

Tabel 2.4 Struktur Tabel Hasil Pembagian Kelompok

No.	NIP	Nama	Hasil
1.	A-1	Hendra Setiawan	Berada pada kelompok : C1
2.	A-2	Roni Purba	Berada pada kelompok : C1
3.	A-3	Joni Darmawan	Berada pada kelompok : C2
4.	A-4	Dewi Roslina	Berada pada kelompok : C1
5.	A-5	Haris Siregar	Berada pada kelompok : C2
6.	A-6	Indra Wijaya	Berada pada kelompok : C1
7.	A-7	Yuyun Sundari	Berada pada kelompok : C1
8.	A-8	Safrizal	Berada pada kelompok : C1
9.	A-9	Gilang Ramadhan	Berada pada kelompok : C1
10.	A-10	Kiki Angreni	Berada pada kelompok : C1
11.	A-11	Sandra Wahyuni	Berada pada kelompok : C1
12.	A-12	Robert Sitorus	Berada pada kelompok : C1

4. Penentuan pusat *cluster* baru

Setelah diketahui anggota tiap-tiap *cluster*, kemudian pusat *cluster* baru dihitung berdasarkan data anggota tiap-tiap *cluster* sesuai dengan rumus pusat anggota *cluster*.

Karena C1 memiliki 10 anggota maka *cluster* baru untuk C1 menjadi :

$$C1 = (78.5; 76.2; 75.3; 79.9; 76.9; 78.4; 76.4; 78.6)$$

Dan begitu juga untuk C2. Karena C2 hanya mempunyai 2 anggota, maka *cluster* baru untuk C2 menjadi :

$$C2 = (79.5; 83; 79.5; 88; 91; 85.5; 78; 86.5)$$

Iterasi ke-2

1. Untuk proses Iterasi ke-2, ulangi lagi langkah ke 2 (kedua) seperti Iterasi Ke-1 yaitu untuk mencari perhitungan matrix jarak pusat *cluster* hingga posisi data

tidak mengalami perubahan, dan untuk Iterasi ke-2 matrik jarak pusat *cluster* nya menjadi seperti ini :

Tabel 2.5 Struktur Tabel Jarak Pusat Iterasi ke-2

A-1	A-2	A-3	A-4	A-5	A-6	A-7	A-8	A-9	A-10	A-11	A-12	
5.5	7.9	21.1	14.08	23.3	9	10.6	6.7	5.1	5.6	10	3.4	C2
20.5	22.9	7.2	23.13	7.2	26.6	19.5	23.3	19.8	24.9	23.8	21	C1

Langkah selanjutnya sama dengan langkah pada nomer 3 iterasi ke-1 diatas, jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat *cluster*, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat *cluster* terdekat

Berikut ini akan ditampilkan data matriks pengelompokkan group, nilai 1 berarti data tersebut berada dalam grup.

Tabel 2.6 Struktur Tabel Pengelompokkan Grup Iterasi ke-2

A-1	A-2	A-3	A-4	A-5	A-6	A-7	A-8	A-9	A-10	A-11	A-12	
1	1	0	1	0	1	1	1	1	1	1	1	C1
0	0	1	0	1	0	0	0	0	0	0	0	C2

Dan dari hasil matrix pengelompokkan itu kita mendapatkan pembagian kelompok alternatif sebagai berikut :

Tabel 2.7 Struktur Tabel Hasil Pengelompokkan Iterasi ke-2

No.	NIP	Nama	Hasil
1.	A-1	Hendra Setiawan	Berada pada kelompok : C1
2.	A-2	Roni Purba	Berada pada kelompok : C1
3.	A-3	Joni Darmawan	Berada pada kelompok : C2
4.	A-4	Dewi Roslina	Berada pada kelompok : C1
5.	A-5	Haris Siregar	Berada pada kelompok : C2
6.	A-6	Indra Wijaya	Berada pada kelompok : C1
7.	A-7	Yuyun Sundari	Berada pada kelompok : C1
8.	A-8	Safrizal	Berada pada kelompok : C1
9.	A-9	Gilang Ramadhan	Berada pada kelompok : C1
10.	A-10	Kiki Angreni	Berada pada kelompok : C1
11.	A-11	Sandra Wahyuni	Berada pada kelompok : C1
12.	A-12	Robert Sitorus	Berada pada kelompok : C1

Karena pengelompokkan group iterasi ke-1 dan iterasi ke-2 memiliki anggota yang sama maka tidak perlu dilakukan iterasi/perulangan lagi. Hasil *clustering*

telah mencapai stabil dan konvergen. Begitulah perhitungan data mining menggunakan algoritma *K-Meansclustering* untuk mengelompokkan data.

2.3. *K-Nearest Neighbor*

Metode *K-Nearest Neighbor* adalah metode yang melakukan penerapan dan penggalian data untuk menentukan klasterisasi. Metode ini sangat sederhana dan memiliki tingkat kesalahan yang rendah. Hasil dari metode KNN adalah klasterisasi dari objek baru berdasarkan atribut dan *training sample*. Ada titik *query* yang bisa menentukan titik *training* yang paling dekat dengan titik *query*.

Metode KNN memiliki kelebihan yaitu kuat terhadap data *training* yang memiliki banyak *noise* dan sangat efektif meski memiliki data *training* yang besar. Meski memiliki kelebihan, ternyata KNN juga memiliki kekurangan. KNN perlu menentukan nilai *k* dan menentukan atribut mana yang menghasilkan nilai terbaik. Berikut ini merupakan contoh perhitungan metode KNN (Lestari, 2014).

Tabel 2.8 Struktur Tabel Dataset

X	Y	Kategori
7	6	Bad
6	6	Bad
6	5	Bad
1	3	Good
2	4	Good
2	2	Good
3	5	?

Langkah penyelesaian

Pertama, Kita tentukan parameter *K*. Misalnya kita buat jumlah tertangga terdekat $K = 3$.

Kedua, kita hitung jarak antara data baru dengan semua data training. Kita menggunakan Euclidean Distance. Kita hitung seperti pada table berikut :

Tabel 2.9 Struktur Tabel Perhitungan

X	Y	Euclidean Distance (3,5)
7	6	$\sqrt{(7-3)^2 + (6-5)^2} = \sqrt{(4)^2 + (1)^2} = \sqrt{17} = 4.12$
6	6	$\sqrt{(6-3)^2 + (6-5)^2} = \sqrt{(3)^2 + (1)^2} = \sqrt{10} = 3.16$
6	5	$\sqrt{(6-3)^2 + (5-5)^2} = \sqrt{(3)^2 + (0)^2} = \sqrt{9} = 3$
1	3	$\sqrt{(1-3)^2 + (3-5)^2} = \sqrt{(-2)^2 + (-2)^2} = \sqrt{8} = 2.82$
2	4	$\sqrt{(2-3)^2 + (4-5)^2} = \sqrt{(-1)^2 + (-1)^2} = \sqrt{2} = 1.41$
2	2	$\sqrt{(2-3)^2 + (2-5)^2} = \sqrt{(-1)^2 + (-3)^2} = \sqrt{10} = 3.16$

Ketiga, kita urutkan jarak dari data baru dengan data training dan menentukan tetangga terdekat berdasarkan jarak minimum K .

Tabel 2.10 Struktur Tabel Perhitungan Jarak Terdekat

X	Y	Euclidean Distance (3,5)	Urutan Jarak	Apakah termasuk 3-NN ?
7	6	4.12	5	Tidak ($K > 3$)
6	6	3.16	4	Tidak ($K > 3$)
6	5	3	3	Ya ($K = 3$)
1	3	2.82	2	Ya ($K < 3$)
2	4	1.41	1	Ya ($K < 3$)
2	2	3.16	4	Tidak ($K > 3$)

Dari kolom 4 (urutan jarak) kita mengurutkan dari yang terdekat ke terjauh antara jarak data baru dengan data training. ada 2 jarak yang sama (yaitu 4) pada data baris 2 dan baris 6, sehingga memiliki urutan yang sama. Pada kolom 5 (Apakah termasuk 3-NN?) maksudnya adalah K-NN menjadi 3-NN, karena nilai K ditentukan sama dengan 3.

Keempat, tentukan kategori dari tetangga terdekat. Kita perhatikan baris 3, 4, dan 5 pada gambar sebelumnya (diatas). Kategori Ya diambil jika nilai $K \leq 3$. Jadi baris 3, 4, dan 5 termasuk kategori Ya dan sisanya Tidak.

Tabel 2.11 Struktur Tabel Penentuan Kategori

X	Y	Euclidean Distance (3,5)	Urutan Jarak	Apakah termasuk 3-NN ?	Kategori Ya untuk KNN
7	6	4.12	5	Tidak ($K > 3$)	-
6	6	3.16	4	Tidak ($K > 3$)	-
6	5	3	3	Ya ($K = 3$)	Bad
1	3	2.82	2	Ya ($K < 3$)	Good
2	4	1.41	1	Ya ($K < 3$)	Good
2	2	3.16	4	Tidak ($K > 3$)	-

Kategori ya untuk K-NN pada kolom 6, mencakup baris 3,4, dan 5. Kita berikan kategori berdasarkan tabel awal. baris 3 memiliki kategori Bad, dan 4,5 memiliki kategori Good.

Kelima, gunakan kategori mayoritas yang sederhana dari tetangga yang terdekat tersebut sebagai nilai prediksi data yang baru.

Tabel 2.12 Struktur Tabel Hasil Klasterisasi

X	Y	Kategori
7	6	Bad
6	6	Bad
6	5	Bad
1	3	Good
2	4	Good
2	2	Good
3	5	Good

Data yang kita miliki pada baris 3, 4 dan 5 kita punya 2 kategori Good dan 1 kategori Bad. Dari jumlah mayoritas ($\text{Good} > \text{Bad}$) tersebut kita simpulkan bahwa data baru ($X=3$ dan $Y=5$) termasuk dalam kategori Good.

2.4. Virus Corona

Virus corona masih satu keluarga dengan SARS (*Severe Acute Respiratory Syndrome*). Virus ini menyerang saluran pernafasan. Gejala yang dialami adalah batuk, pilek, demam, gangguan pernafasan, sakit tenggorokan, letih dan lesu. Pencegahan yang harus dilakukan dengan sering mencuci tangan menggunakan sabun, memakai masker jika batuk atau pilek, makan makanan yang bergizi, rajin olahraga dan mengunjungi fasilitas kesehatan jika merasakan gejala batuk atau pilek.

Virus corona memiliki tingkat fatalitas yang rendah, namun bisa menjadi fatalitas yang tinggi jika menyerang anak-anak dan orang lanjut usia dengan penyakit yang diderita sebelumnya. Virus ini memiliki tingkat fatalitas 2,20 persen. Penyebaran virus yang tidak dapat terdeteksi menyebabkan penyebarluasan yang cepat sehingga manusia bisa tertular tanpa terdeteksi. Hal tersebut mengakibatkan sulit untuk ditangani (Garnesia, 2020).

2.5. Pemrograman Web

Pemrograman web merupakan suatu kegiatan penulisan bahasa pemrograman atau sintak-sintak untuk menghasilkan suatu tampilan web tertentu. Pemrograman web erat kaitannya dengan bahasa pemrograman, yakni bahasa

yang digunakan untuk membuat web. Banyak sekali bahasa pemrograman yang bisa dipelajari mulai dari HTML, PHP, dll. Setiap bahasa pemrograman memiliki ciri dan tingkat kesulitan yang berbeda namun masih tetap memiliki fungsi yang sama (Winantu & Saputro, 2010).

Pemrograman web juga diartikan sebagai proses pembuatan web menggunakan bahasa pemrograman tertentu yang nantinya ditampilkan melalui bantuan web browser. Web browser memiliki peranan penting dalam pemrograman web. Seiring perkembangan zaman yang membuat yang semua kegiatan menjadi terkomputerisasi membuat web semakin diminati oleh masyarakat pebisnis. Web juga semakin banyak diperjualkan untuk mempermudah kegiatan yang sebelumnya dilakukan secara manual (Satria, 2016).

2.6. HTML

HTML merupakan bahasa pemrograman dasar untuk membuat suatu laman website sederhana. HTML ini merupakan bahasa pemrograman yang mudah dipahami oleh seseorang yang ingin belajar coding. Sebuah data yang memiliki script HTML merupakan data yang disajikan berbentuk website. HTML juga bisa dipelajari dengan mudah, karena merupakan awal dari bahasa pemrograman (Winantu & Saputro, 2010).

HTML digunakan untuk membuat struktur halaman website yang berarti HTML juga mendesain halaman website. Meskipun dalam pemrogramannya HTML tidak berdiri sendiri melainkan digabungkan dengan CSS atau script lain, seperti javascript. HTML menjadi desain dasar dalam pembuatan web. Untuk memperindah tampilan bisa ditambahkan dengan CSS (Entreprise, 2018).

2.7. PHP

PHP memiliki pengertian sebagai bahasa pemrograman untuk pembuatan web. Bahasa pemrograman ini lebih diminati daripada bahasa pemrograman yang lain, karena PHP mudah dipahami dan dipelajari. PHP juga bisa menjadikan tampilan web menjadi dinamis. Web dinamis pastinya lebih diminati daripada

web statis yang sudah ketinggalan jaman dan fitur yang ditawarkan pada web dinamis tentunya memudahkan user(Entreprise, 2018).

PHP adalah bahasa pemrograman yang bersifat server side. Memiliki tujuan untuk menghasilkan script yang akan digenerate dalam kode HTML yang merupakan bahasa standar web. Sehingga sebelum mempelajari PHP diharuskan sudah mengerti tentang HTML. Karena HTML merupakan dasar untuk mendesain web (Winarno & Zaky, 2013).

2.8. Xampp

Xampp merupakan *database* dari phpMyAdmin. Xampp juga software yang biasa digunakan untuk menjalankan website berbasis PHP dan menggunakan pengolah data MySQL di komputer lokal. Xampp berperan sebagai web server yang berdiri sendiri (localhost). Xampp juga disebut sebagai CPanel server virtual, yang bisa membantu melakukan preview sehingga bisa memodifikasi website tanpa harus online(Haqi & Setiawan, 2019).

Karena instalasi yang mudah dan sifatnya yang open source membuat Xampp diminati banyak orang. Xampp juga mudah digunakan dan versi Xampp terbaru sudah mendukung PHP7. Versi terbaru Xampp membuat Xampp semakin disukai oleh masyarakat. Pembaruan-pembaruan yang dilakukan Xampp senantiasa menjadi perkembangan dalam *database*phpMyAdmin (Wicaksono, 2010).

2.9. MySQL

MySQL merupakan *database* dari phpMyAdmin. MySQL memiliki model client-server yang sangat simpel. Mysql ini digunakan *programmer* untuk menyimpan basik datanya. Sifatnya *yang open source* menjadikan lebih diminati. MySQL memiliki fungsi untuk mengolah *database* menggunakan bahasa SQL(Entreprise, 2018).

Pemrograman PHP juga sangat mendukung MySQL ini. Program yang digunakan untuk mengolah dan mengelola data adalah MySQL. Memiliki sekumpulan prosedur dan struktur sedemikian rupa sehingga mempermudah

dalam menyimpan, mengatur dan menampilkan data. Hal ini menjadi kelebihan MySQL yang membuat diminati masyarakat (Anhar, 2010).