

BAB II

LANDASAN TEORI

2.1 Sistem Informasi Geografis

Sistem Informasi Geografis (SIG) atau dikenal dengan istilah *Geographical Information System* didefinisikan sebagai suatu alat/media untuk memasukan, menyimpan, mengambil, memanipulasi, menganalisa, dan menampilkan data-data beratribut Geografis (data spasial) yang berguna untuk mendukung proses pengambilan keputusan dalam perencanaan dan manajemen sumber daya alam, lingkungan, transportasi, masalah perkotaan dan administrative (*Burrough, 1986*).

Data spasial adalah data-data yang memiliki sistem koordinat geografis. Dengan kata lain SIG merupakan suatu sistem *database* yang memiliki kemampuan spesifik untuk melakukan operasi tertentu pada data. Teknologi SIG biasanya telah terintegrasi dengan teknologi *database* seperti *query* dan analisa statistik dengan tampilan yang unik, serta analisis geografis yang menguntungkan data peta. Kemampuan ini membuat SIG berbeda dengan sistem informasi lainnya, sehingga SIG menjadi barang berharga bagi masyarakat umum dan perusahaan untuk menjelaskan peristiwa, memprediksi pendapat dan perencanaan strategis (ESRI, 1996).

SIG merupakan sebuah sistem informasi yang didesain untuk bekerja dengan sumber data spasial. SIG merupakan suatu media yang sangat handal untuk merepresentasikan data *Remote Sensing* (RS) menjadi informasi yang berguna bagi banyak pihak untuk berbagai keperluan (Indarto, 2013)

2.2 Daerah Rawan Kriminalitas

Secara hukum, kriminalitas/kejahatan didefinisikan sebagai tindakan atau kelalaian yang dilarang oleh hukum yang dapat dihukum dengan pidana penjara dan atau denda. Pembunuhan, perampokan, pencurian, pemerkosaan, mengemudi mabuk, pembuangan anak, dan tidak membayar pajak adalah semua contoh umum kriminalitas.

Kriminalitas termasuk pada ilmu hukum pidana dan kriminologi, dimana kriminologi adalah orang yang melakukan kejahatan (pelaku) itu sendiri. Adapun tujuannya adalah agar menjadi mengerti apa sebab-sebabnya sehingga sampai berbuat jahat itu. Apakah memang karena bakatnya adalah jahat, ataukah didorong oleh keadaan masyarakat di sekitarnya (milieu) baik keadaan sosiologi maupun ekonomis. Ataukah ada sebab-sebab lain lagi. Jika sebab-sebab itu sudah diketahui, maka di samping pemidanaan, dapat diadakan tindakan-tindakan yang tepat, agar orang tadi tidak lagi berbuat demikian, atau agar orang-orang lain tidak akan melakukannya (Moeljatno, 2009).

Daerah rawan kriminalitas adalah daerah yang memiliki potensi terhadap adanya tindak kriminalitas disuatu daerah. Daerah ini biasanya tergantung dari mata pencaharian penduduk, kondisi ekonomi dan tingkat pendidikan masyarakat tersebut.

2.3 K-means

K-means adalah salah satu metode penganalisaan data. Metode k-means berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok lain. Dengan kata lain, metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam satu *cluster* dan memaksimalkan variasi dengan data yang ada di *cluster* lainnya.

Dikutip dari jurnal (Andri Syafrianto, 2012) algoritma clustering ke dalam kelompok besar seperti berikut :

1. *Partitioning algorithms*. algoritma dalam kelompok ini membentuk bermacam partisi dan kemudian mengevaluasinya dengan berdasarkan beberapa kriteria.
2. *Hierarchy algorithms*. pembentukan dekomposisi hirarki dari sekumpulan data menggunakan beberapa kriteria.

3. *Density-based*. pembentukan cluster berdasarkan pada koneksi dan fungsi densitas.
4. *Grid-based*. pembentukan *cluster* berdasarkan pada struktur *multiple level granularity*.
5. *Model-based*. sebuah model dianggap sebagai hipotesa untuk masing-masing cluster dan model yang baik dipilih diantara model hipotesa tersebut.

Algoritma k-means merupakan algoritma yang membutuhkan parameter input sebanyak k dan membagi sekumpulan n objek ke dalam k cluster sehingga tingkat kemiripan antar anggota dalam satu cluster tinggi sedangkan tingkat kemiripan dengan anggota pada cluster lain sangat rendah. Kemiripan anggota terhadap cluster diukur dengan kedekatan objek terhadap nilai *mean* pada cluster atau dapat disebut sebagai *centroid cluster* atau pusat massa.

Dikutip dari jurnal (Ernie Kurniawan. dkk, 2013) berikut ini adalah rumus untuk menentukan jumlah cluster :

$$k \approx \sqrt{n\sqrt{2}} \quad (1)$$

Berikut rumus pengukuran jarak :

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

Adapun rumus perhitungan jarak lainnya didefinisikan sebagai berikut :

$$d_{(x,y)} = \sqrt{(x_i - y_i)^2 + (x_i - y_i)^2} \quad (3)$$

Keterangan :

d = titik dokumen

x = data *record*

y = data *centroid*

Berikut rumus dari penentuan centroid.

$$C(i) = \frac{1}{|G_i|} \sum x_{ec} d\bar{x} \quad (4)$$

Adapun rumus iterasi lainnya didefinisikan sebagai berikut :

$$C(i) = \frac{x_1+x_2+x_3+\dots+x_n}{\sum x} \quad (5)$$

Keterangan :

x_1 = nilai data record ke-1

x_2 = nilai data record ke-2

$\sum x$ = jumlah data record

Pada Tabel 2.1 merupakan contoh data awal sebelum dilakukan perhitungan dengan menggunakan metode k-means untuk mencari nilai hasil klasterisasi yang akan dikelompokkan menjadi 3 *cluster* yaitu C1, C2, C3.

Tabel 2.1 Tabel Data Pengujian

No.	Kecamatan	JUMLAH KRIMINALITAS				
		pencurian	perjudian	pemerksaan	penganiayaan	penipuan
1	cilacap utara	2	3	1	1	1
2	cilacap selatan	5	3	3	0	0
3	cilacap tengah	1	2	0	2	0
4	kesugihan	1	3	1	2	1
5	jeruklegi	2	0	0	0	0
6	maos	2	3	1	0	0
7	adipala	2	1	0	4	2
8	kroya	1	4	1	1	1
9	sampang	3	2	0	0	1
10	binangun	0	1	0	0	1
11	nusawungu	1	0	0	1	0
12	kawunganten	1	0	0	0	0
13	kampung laut	0	0	1	0	0
14	bantarsari	1	1	2	1	0
15	patimuan	0	2	0	0	0
16	gandrungmangu	1	1	0	0	1
17	kedungreja	0	2	1	0	0
18	sidareja	0	1	3	0	0
19	cipari	2	1	1	0	0
20	karangpucung	0	1	1	1	0
21	cimanggu	0	1	0	0	0
22	wanareja	1	0	0	1	0
23	majenang	0	0	0	0	1
24	dayeuhluhur	0	0	0	0	1

Selanjutnya dari data pengujian diatas, dilakukan proses normalisasi yang bertujuan untuk standarisasi data dengan menempatkan data dalam range 0 sampai

1. Normalisasi data pengujian dapat dilihat pada tabel 2.2

$$\text{Normalisasi} = (x - \min) / (\max - \min)$$

Tabel 2.2 Tabel Normalisasi

No.	Kecamatan	NORMALIASASI				
		pencurian	perjudian	pemerkosaan	penganiayaan	penipuan
1	cilacap utara	0.4	0.75	0.333333333	0.25	0.5
2	cilacap selatan	1	0.75	1	0	0
3	cilacap tengah	0.2	0.5	0	0.5	0
4	kesugihan	0.2	0.75	0.333333333	0.5	0.5
5	jeruklegi	0.4	0	0	0	0
6	maos	0.4	0.75	0.333333333	0	0
7	adipala	0.4	0.25	0	1	1
8	kroya	0.2	1	0.333333333	0.25	0.5
9	sampang	0.6	0.5	0	0	0.5
10	binangun	0	0.25	0	0	0.5
11	nusawungu	0.2	0	0	0.25	0
12	kawunganten	0.2	0	0	0	0
13	kampung laut	0	0	0.333333333	0	0
14	bantarsari	0.2	0.25	0.666666667	0.25	0
15	patimuan	0	0.5	0	0	0
16	gandrungmangu	0.2	0.25	0	0	0.5
17	kedungreja	0	0.5	0.333333333	0	0
18	sidareja	0	0.25	1	0	0
19	cipari	0.4	0.25	0.333333333	0	0
20	karangpucung	0	0.25	0.333333333	0.25	0
21	cimanggung	0	0.25	0	0	0
22	wanareja	0.2	0	0	0.25	0
23	majenang	0	0	0	0	0.5
24	dayeuhluhur	0	0	0	0	0.5

Setelah proses normalisasi selesai, kemudian akan digunakan algoritma k-means untuk mengelompokkan data yang ada. Data akan dikelompokkan menjadi 3 kelompok. Adapun contoh langkah dari pengelompokkan data sebagai berikut :

1. Tentukan pusat *cluster* secara acak, misalkan C1= (1, 0.75, 1, 0, 0); C2= (0.2 0, 0, 0, 0, 0); C3= (0, 0, 0, 0, 0.5)
2. Hitung jarak setiap data terhadap pusat cluster. Misalkan untuk menghitung jarak data kriminalitas pertama (cilacap utara) dengan pusat C1:

$$d_{11} = \sqrt{(0.4 - 1)^2 + (0.75 - 0.75)^2 + (0.333333 - 1)^2 + (0.25 - 0)^2 + (0.5 - 0)^2}$$

$$= 1.056855924$$

Jarak data kriminalitas (cilacap utara) dengan pusat C2:

$$d_{12} = \sqrt{(0.4 - 0.2)^2 + (0.75 - 0)^2 + (0.333333 - 1)^2 + (0.25 - 0)^2 + (0.5 - 0)^2} = 1.012971427$$

Jarak data kriminalitas (cilacap utara) dengan pusat C3:

$$d_{13} = \sqrt{(0.4 - 0)^2 + (0.75 - 0)^2 + (0.333333 - 0)^2 + (0.25 - 0)^2 + (0.5 - 0.5)^2} = 0.946631455$$

Pada tabel 2.3 adalah contoh hasil perhitungan jarak data ke pusat *cluster*.

Tabel 2.3 Tabel Jarak Data Ke Pusat *Cluster*

No.	Kecamatan	Jenis Kriminalitas			CLUSTER
		C1	C2	C3	
1	cilacap utara	1.056855924	1.012971427	0.946631455	C3
2	cilacap selatan	0	1.484082208	1.677050983	C1
3	cilacap tengah	1.397318861	0.707106781	0.888819442	C2
4	kesugihan	1.258747173	1.083333333	0.981636955	C3
5	jeruklegi	1.386542462	0.2	0.640312424	C2
6	maos	0.89690827	0.84475506	1.040966431	C2
7	adipala	1.9	1.45	1.21346611	C3
8	kroya	1.208074685	1.193151755	1.101640191	C3
9	sampang	1.21346611	0.81240384	0.781024968	C3
10	binangun	1.58113883	0.593717104	0.25	C3
11	nusawungu	1.504991694	0.25	0.593717104	C2
12	kawunganten	1.484082208	0	0.538516481	C2
13	kampung laut	1.416666667	0.388730126	0.600925213	C2
14	bantarsari	1.031315234	0.754615428	0.927062266	C2
15	patimuan	1.436140662	0.538516481	0.707106781	C2
16	gandrungmangu	1.462873884	0.559016994	0.320156212	C3
17	kedungreja	1.227576655	0.633333333	0.78173596	C2
18	sidareja	1.118033989	1.05	1.145643924	C2
19	cipari	1.026861453	0.462180821	0.763944442	C2
20	karangpucung	1.32549781	0.525462759	0.697216689	C2
21	cimanggu	1.5	0.320156212	0.559016994	C2
22	wanareja	1.504991694	0.25	0.593717104	C2
23	majenang	1.677050983	0.538516481	0	C3
24	dayeuhluhur	1.677050983	0.538516481	0	C3

- Suatu data akan menjadi anggota dari suatu *cluster* yang memiliki jarak terkecil dari pusat clusternya. Misalkan untuk data pertama, jarak terkecil diperoleh pada pusat C1, sehingga data pertama akan menjadi anggota dari *cluster* C1. Demikian juga untuk data kedua, jarak terkecil ada pada *cluster* C2, maka data tersebut akan masuk pada *cluster* C2.
- Hitung pusat *cluster* baru. Sebagai contoh untuk *cluster* ketiga pada atribut pemerksaan, ada 1 data yaitu data ke-2, sehingga :

Pusat *cluster* pertama untuk atribut pemerksaan $C_{11} = 1$

Untuk *cluster* kedua pada atribut pemerksaan ada 15 data yaitu ke-3, ke-5, ke-6, ke-11, ke-12, ke-13, ke-14, ke-15, ke-17, ke-18, ke-19, ke-20, ke-21, ke-22, sehingga:

$$C_{21} = (0+0+0.33333333+0+0+0.33333333+0.66666667+0+0.33333333+1+0.33333333+0.33333333+0+0)/14 = 0.238095238$$

Untuk *cluster* ketiga pada atribut pemerkosaan ada 8 data yaitu ke-1, ke-4, ke-7 ke-8, ke-9, ke-10, ke-16, ke-23, ke-24, sehingga:

$$C_{31} = (0.33333333+0.33333333+0+0.33333333+0+0+0+0)/9 = 0.1$$

Pada tabel 2.4 adalah contoh hasil perhitungan pusat *cluster* yang baru

Tabel 2.4 Tabel Pusat *Cluster* Baru

PUSAT YANG BARU ITERASI 1					
C1	1	0.75	1	0	0
C2	0.157143	0.25	0.238095	0.107143	0
C3	0.222222	0.416667	0.111111	0.222222	0.555556

5. Ulangi langkah 2 dan 4 hingga posisi pusat *cluster* sebelumnya dan sesudahnya tidak ada perubahan lagi. Tabel hasil akhir iterasi yang dimana pusat *cluster* sudah tidak berubah dapat dilihat pada tabel 2.5

Tabel 2.5 Tabel Hasil Iterasi Terakhir

No.	Kecamatan	Jenis Kriminalitas					CLUSTER
		pencurian	perjudian	pemerkosaan	penganiayaan	penipuan	
2	cilacap selatan	5	3	3	0	0	C1
3	cilacap tengah	1	2	0	2	0	C2
5	jeruklegi	2	0	0	0	0	C2
6	maos	2	3	1	0	0	C2
11	nusawungu	1	0	0	1	0	C2
12	kawunganten	1	0	0	0	0	C2
13	kampung laut	0	0	1	0	0	C2
14	bantarsari	1	1	2	1	0	C2
15	patimuan	0	2	0	0	0	C2
17	kedungreja	0	2	1	0	0	C2
18	sidareja	0	1	3	0	0	C2
19	cipari	2	1	1	0	0	C2
20	karangpucung	0	1	1	1	0	C2
21	cimanggu	0	1	0	0	0	C2
22	wanareja	1	0	0	1	0	C2
1	cilacap utara	2	3	1	1	1	C3
4	kesugihan	1	3	1	2	1	C3
7	adipala	2	1	0	4	2	C3
8	kroya	1	4	1	1	1	C3
9	sampang	3	2	0	0	1	C3
10	binangun	0	1	0	0	1	C3
16	gandrungmangu	1	1	0	0	1	C3
23	majenang	0	0	0	0	1	C3
24	dayeuhluhur	0	0	0	0	1	C3

Karena posisi *cluster* tidak berubah lagi atau stabil, maka iterasi akan berhenti dan hasil yang diperoleh adalah tiga *cluster*:

- a. *Cluster* pertama memiliki pusat (1; 0.75; 1; 0; 0) yang dapat diartikan sebagai kelompok daerah dengan tingkat kriminalitas tinggi.
- b. *Cluster* kedua memiliki pusat (0.153846154; 0.230769231; 0.256410256; 0.076923077; 0) yang dapat diartikan sebagai kelompok daerah dengan tingkat kriminalitas rendah.
- c. *Cluster* ketiga memiliki pusat (0.22; 0.425; 0.1; 0.25; 0.5) yang dapat diartikan sebagai kelompok daerah dengan tingkat kriminalitas sedang (Ayu, 2013).

