

Analisa Pemetaan Kriminalitas Kabupaten Bangkalan Menggunakan Metode *K-Means* dan *K-Means++*

Chindy Ayudia Sri Fastaf¹, Yuni Yamasari²

^{1,2}Jurusan Teknik Informatika, Prodi Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya

Chindy.18012@mhs.unesa.ac.id

yuniyamasari@unesa.ac.id

Abstrak— Kriminalitas merupakan suatu permasalahan umum di kehidupan sehari-hari, tak terkecuali di Kabupaten Bangkalan. Bangkalan merupakan kabupaten yang terdiri dari 18 kecamatan, yang mana tindakan kriminalitas semakin meningkat di setiap tahun khususnya pencurian dengan pemberatan (curat) dan kasus pencurian kendaraan bermotor (curanmor). Maka dari itu perlu dilakukan pengelompokan daerah rawan kriminalitas dengan tujuan agar dapat membantu berupa pemberian informasi kepada pihak kepolisian setempat dalam upaya meningkatkan keamanan di Kabupaten Bangkalan. Dalam penelitian ini dengan menggunakan 10 dataset jenis kriminalitas dari 18 kecamatan di Kabupaten Bangkalan, dilakukan perbandingan antara dua metode *clustering* untuk memperoleh metode yang terbaik dalam pemetaan daerah kriminalitas. *K-Means* dan *K-Means++* merupakan dua metode yang digunakan dalam penelitian ini. Dataset yang digunakan sebanyak 492 dari total kasus kriminalitas tahun 2021 di Kabupaten Bangkalan. Sebelum implementasi *clustering*, dilakukan validasi *cluster* dengan menentukan jumlah *cluster* optimum menggunakan metode *Elbow*. Hasil *clustering* pada 10 dataset jenis kriminalitas dengan menggunakan kedua metode terdapat perbedaan pemetaan pada 3 jenis kriminalitas yaitu penganiayaan, penipuan, dan perampokan. Selanjutnya dilakukan validitas dari kedua metode dengan menggunakan *Silhouette Coefficient*. Pada hasil validitas terdapat perbedaan nilai *Silhouette* pada 3 jenis kriminalitas yaitu penganiayaan, penipuan, dan perampokan. Hasil uji metode *K-Means* dan *K-Means++* dengan *Silhouette* pada Penganiayaan sebesar 0,1683 dan 0,2314 secara berturut-turut, sedangkan pada hasil uji pada Penipuan masing-masing sebesar 0,2243 dan 0,2534, dan hasil uji pada Perampokan sebesar 0,4898 dan 0,4057. Berdasarkan hasil uji dengan *Silhouette Coefficient*, metode *K-Means++* memberikan hasil uji yang lebih baik pada 2 jenis kriminalitas sedangkan metode *K-Means* lebih baik dalam 1 jenis kriminalitas.

Kata Kunci— Kriminalitas, *Clustering*, *K-Means*, *K-Means++*, *Elbow*, *Silhouette Coefficient*,.

I. PENDAHULUAN

Kriminalitas merupakan salah satu permasalahan yang umum terjadi di kehidupan sehari-hari, tak terkecuali di Kabupaten Bangkalan. Kriminalitas didefinisikan sebagai tindakan yang dapat merugikan orang lain dan salah satu bentuk pelanggaran yang telah diatur dalam Undang-Undang.

Tindakan ini tidak memandang status sosial, usia bahkan jenis kelamin[6]. Hal yang menjadi pemicu terjadinya tindak kriminalitas yaitu pendidikan, lingkungan, ekonomi, bahkan pekerjaan[7]. Mengingat Kabupaten Bangkalan terdiri dari 18 kecamatan, yang mana tindak kriminalitas semakin meningkat di setiap tahun khususnya pencurian dengan pemberatan (curat) dan tindak pencurian kendaraan bermotor (curanmor) [13]. Untuk meminimalisir terjadinya tindak kriminalitas dan untuk meningkatkan keamanan di Kabupaten Bangkalan maka perlu dilakukan pengelompokan daerah rawan kriminalitas. Penggunaan *data mining* merupakan salah satu metode yang dapat diimplementasikan dalam melakukan pengelompokan.

Data mining didefinisikan sebagai proses dalam menemukan pengetahuan dari berbagai kumpulan data yang dimiliki. Kumpulan data tersebut dapat mencakup diantaranya data *warehouse*, *database*, atau repositori informasi lainnya[5]. Pada *data mining*, terdapat sub kategori yang dikenal dengan *clustering*. *Clustering* diartikan sebagai proses dimana membagi data sehingga membentuk himpunan tertentu yang didalamnya terdapat kesamaan karakteristik data [15]. *Clustering* terdiri dari beberapa metode, diantaranya metode *K-Means* dan *K-Means++*. Cara kerja metode *K-Means* clustering yakni dengan melakukan partisi set data ke dalam sejumlah *K cluster* yang telah diimplementasikan di awal. Metode ini termasuk pada salah satu metode yang cepat dalam *clustering* dan mudah dilakukan. Dengan demikian, metode *K-Means* dikategorikan sebagai salah satu algoritma *clustering* yang umum dalam data mining [18]. Karena pemilihan *centroid* pada *K-Means* bersifat acak, hal ini menjadi kelemahan dari proses *clustering*. Metode *K-Means++* digunakan untuk mengatasi masalah dengan memilih secara acak pusat *cluster* awal yang meningkatkan jumlah iterasi. Tujuan utama dari algoritma ini adalah untuk mengambil titik-titik data tersebut sebagai pusat awal yang berjarak sejauh mungkin satu sama lain [19]. Pada metode *K-Means++* memastikan bahwa inialisasi *centroid* dilakukan lebih cerdas untuk meningkatkan kualitas pengelompokan.

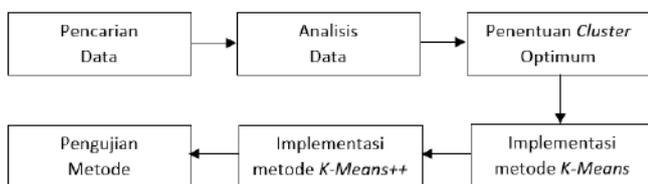
Beberapa penelitian sebelumnya telah dilakukan pengelompokan kriminalitas menggunakan metode *clustering* yang berbeda. Referensi [11] melakukan penelitian dengan menerapkan dua metode *clustering* yakni *K-Means* dan *Fuzzy C-Means*. Penelitian ini dapat memberikan hasil berupa perbandingan antar kedua metode yang mana jika dilihat pada

hasil *dunn index*, metode *clustering* yang memiliki hasil terbaik adalah *Fuzzy C-Means*. Sedangkan jika dilihat pada nilai *connectivity* dan *silhouette*, metode *clustering* yang lebih unggul adalah *K-Means*. Referensi [2] melakukan penelitian dengan membandingkan 2 metode yang berbeda yaitu *K-Means* dan *K-Means++* pada data segmentasi pelanggan mall, dimana dataset diperoleh dari *Kaggle*. Hasil analisis menunjukkan bahwa dalam hal waktu rata-rata, *K-Means++* mencapai sedikit lebih cepat daripada metode *K-Means*. Rata-rata waktu *KMeans++* dan *K-Means* masing-masing adalah 938.279 ms dan 935.134 ms. Kualitas *clustering* (segmentasi) pada metode *K-Means++* dan *K-Means* diuji dengan metode *Silhouette*. Dari segi skor *silhouette*, *K-Means++* memperoleh hasil yang lebih baik dibanding metode *K-Means*. Skor *Silhouette cluster K-Means++* dan *K-Means* masing-masing adalah 0,5539 dan 0,5532. Hasil *cluster* dianalisis untuk mendapatkan segmentasi pelanggan berdasarkan nilai pengeluaran dan pendapatan. Referensi [14] melakukan penelitian mengenai klasifikasi data kriminalitas Kepulauan Mentawai dengan metode *K-Means*. Hasil yang diperoleh yaitu terbaginya data tersebut menjadi 3 *cluster* diantaranya C1/paling rawan terdiri dari 1 kecamatan, C2/cukup rawan terdiri dari 2 kecamatan, dan C3/tidak rawan terdiri dari 7 kecamatan.

Hal yang menjadi pembeda pada penelitian ini yakni dilakukan analisis dengan implementasi 2 metode yaitu metode *K-Means* dan *K-Means++* dalam menentukan *clustering* pada 10 jenis kriminalitas dari 18 kecamatan di Kabupaten Bangkalan. Selanjutnya dilakukan uji pengolahan hasil *clustering* menggunakan *Silhouette Coefficient* untuk menemukan metode *clustering* yang terbaik. Dengan hasil pengelompokan berdasarkan metode *clustering* yang terbaik diharapkan dapat membantu pihak Kepolisian Kabupaten dalam meningkatkan keamanan dan menurunkan angka tindak kriminalitas yang terdapat di seluruh Kecamatan di Kabupaten Bangkalan.

II. METODE PENELITIAN

Tahapan metode untuk menghasilkan analisis *cluster* daerah rawan kriminalitas di Kabupaten Bangkalan mengacu pada diagram alur yang terdapat dalam Gbr 1. Alur penelitian diawali dari pencarian data, kemudian analisis data, selanjutnya penentuan *cluster* optimum, lalu implementasi metode *K-Means*, implementasi metode *K-Means++*, dan pengujian metode.



Gbr 1. Alur Penelitian Analisis Cluster

A. Pencarian Data

Pada penelitian ini, sumber data yang akan digunakan

berasal dari Data Kriminalitas Polres Kabupaten Bangkalan. Data yang diolah merupakan data kriminalitas Kabupaten Bangkalan selama tahun 2021 dari 18 kecamatan di Kabupaten Bangkalan. Pada Tabel 1 dapat dilihat jumlah data masing-masing jenis kriminalitas yang telah diperoleh.

Tabel I
 Jenis Kriminalitas

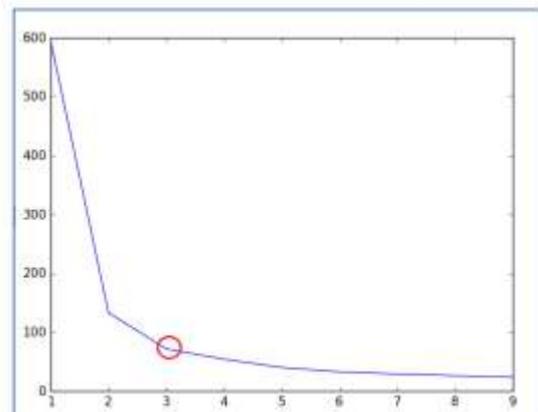
No	Jenis Kriminalitas	Jumlah Kasus
1	Pembunuhan	7 Kasus
2	Curanmor/Pencurian Motor	62 Kasus
3	Pemerksaan	9 Kasus
4	Perampokan/Curas	33 Kasus
5	Perjudian	10 Kasus
6	Penipuan	64 Kasus
7	Curat/Curian dengan Pemberatan	215 Kasus
8	Pengrusakan	4 Kasus
9	Penganiayaan	75 Kasus
10	Pengroyokan	13 Kasus

B. Analisis Data

Setelah data penelitian telah diperoleh, dilanjutkan dengan analisis data. Analisis data bertujuan untuk mengelompokkan data yang telah diperoleh berdasar pada kolom kecamatan, jenis kriminalitas, dan angka kriminalitas tiap bulan. Hasil analisis data tersebut nantinya akan digunakan sebagai dataset pada penelitian ini.

A. Penentuan Cluster Optimum

Untuk menentukan jumlah *cluster* optimum pada penelitian ini menggunakan metode *Elbow*. Cara menentukan titik jumlah *cluster* optimum pada grafik *Elbow* yakni dengan cara melihat persentase hasil perbandingan dari jumlah K yang membentuk sudut siku di suatu titik dalam grafik *Elbow* [8].



Gbr 2. Grafik Metode Elbow

Persentase hasil perhitungan yang telah didapatkan selanjutnya dijadikan pembandingan antara jumlah *cluster* yang ditambah. Perbedaan pada hasil persentase masing-masing nilai *cluster* dapat dilihat pada grafik yang ditampilkan. Hal yang perlu dipahami yaitu jika nilai pada *cluster* pertama dengan nilai pada *cluster* berikutnya membentuk sudut siku dalam

grafik yang ditampilkan atau nilai pada grafik mengalami penurunan yang signifikan maka nilai *cluster* tersebut merupakan jumlah *cluster* yang optimum [3].

Lalu untuk memperoleh hasil perbandingan tersebut yakni dengan menghitung nilai *Sum Of Square Error (SSE)* tiap *cluster*. Semakin besar jumlah *cluster* yang didapatkan, maka nilai *Sum Of Square Error (SSE)* akan semakin kecil [6]. Untuk menentukan persamaan *SSE K-Means* dapat menggunakan formula di bawah ini.

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \| X_i - C_k \|_2^2 \quad (1)$$

Keterangan :

X_i = nilai atribut data ke- i

C_k = nilai *centroid cluster* ke- i

B. Implementasi Metode *K-Means* pada Dataset

Setelah K *cluster* optimum telah diperoleh, tahap selanjutnya yaitu dilakukan *clustering* pada seluruh dataset menggunakan metode *K-Means*. Cara kerja pada algoritma *K-Means clustering* yakni dengan mempartisi data ke dalam beberapa *cluster* dengan berbasis jarak. Algoritma *K-Means* ini mempartisi data berdasarkan karakteristik yang dimiliki oleh tiap data [11]. Metode ini dapat disimpulkan sebagai salah satu *clustering* yang sederhana dengan alasan dapat melakukan pengelompokan data dalam waktu yang relatif cepat dengan jumlah yang cukup besar [13].

Data yang digunakan adalah pengelompokan data masing-masing jenis kriminalitas dari tiap kecamatan berdasarkan jumlah kasus per bulannya, kemudian dilakukan *clustering* berdasarkan langkah-langkah berikut:

1. Menentukan jumlah K *cluster* pada dataset.
2. Menginisialisasikan K pusat *cluster* secara acak/random.
3. Untuk menentukan pusat *cluster (centroid)* selanjutnya menggunakan formula dibawah ini.

$$v = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (2)$$

Keterangan :

v = *centroid* (titik pusat) pada *cluster*

n = banyak anggota *cluster*

4. Menentukan jarak dari setiap objek ke setiap *centroid* menggunakan *Euclidean Distance* menggunakan formula dibawah ini.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

Keterangan :

x_i = objek x ke- i

y_i = objek y ke- i

n = banyaknya objek

5. Menggolongkan tiap objek ke titik pusat/*centroid* terdekat.
6. Melakukan iterasi, selanjutnya menentukan posisi titik pusat *cluster* baru dengan menerapkan rumus pada persamaan (1).
7. Perhitungan akan selesai jika tidak terdapat perpindahan objek pada *cluster*, sedangkan jika masih terdapat objek yang berpindah maka ulangi dari langkah 2-4.

C. Implementasi Metode *K-Means++* pada Dataset

Pada metode *K-Means* pemilihan *centroidnya* bersifat acak, hal ini menjadi kelemahan pada proses *clustering* menggunakan *K-Means*. Oleh karena itu, untuk mengatasi kelemahan tersebut dibuatlah metode *K-Means++*. Pada metode ini memastikan bahwa inisialisasi *centroid* dilakukan lebih optimum untuk meningkatkan kualitas *clustering*. Perbedaan kedua metode, baik *K-Means* dan *K-Means++* hanya terletak pada pemilihan titik pusat/*centroid* awal. Dalam *K-Means*, pemilihan titik pusat *cluster* dilakukan secara random/acak sebanyak nilai K yang ditentukan tetapi pada *K-Means++*, 1 *centroid* dipilih secara acak kemudian dilakukan perhitungan jarak *centroid* dengan data yang lainnya. Nilai terbesar berarti memiliki jarak terjauh dan probabilitas tertinggi untuk menjadi *centroid* baru.

Dataset yang akan digunakan adalah dataset yang sama pada *clustering* dengan metode *K-Means* yaitu pengelompokan data masing-masing jenis kriminalitas dari tiap kecamatan berdasarkan jumlah kasus per bulannya, selanjutnya diolah menggunakan metode *K-Means++* berdasarkan langkah-langkah berikut:

1. Menentukan titik *centroid* pertama secara acak
2. Menghitung jarak semua titik dalam dataset dari *centroid* yang dipilih. Jarak titik x_i dari *centroid* terjauh dapat dihitung menggunakan formula dibawah ini.

$$d_i = \max_{(j:1 \rightarrow m)} \| x_i - C_j \|_2^2 \quad (4)$$

3. Menjadikan titik x_i sebagai *centroid* baru yang memiliki probabilitas maksimum.
4. Ulangi langkah 2-3 sampai menemukan *K-Centroid*.
5. Setelah mendapatkan *centroid*, langkah selanjutnya sama dengan metode *K-Means*.

D. Pengujian Metode dengan *Silhouette Coefficient*

Cara kerja *Silhouette Coefficient* dalam melakukan pengujian kualitas metode berdasarkan pada hasil *cluster* yang telah diperoleh [3]. Metode pengujian ini menggabungkan antara metode *cohesion* dan *separation*. Range nilai yang dihasilkan oleh metode pengujian ini yakni antara -1 sampai 1. Jika nilai yang dihasilkan metode tersebut mendekati angka 1 maka dapat disimpulkan bahwa pengelompokan *cluster* tersebut semakin membaik. Adapaun langkah-langkah untuk menghitung nilai *Silhouette Coefficient* sebagai berikut:

1. Menghitung rata-rata jarak suatu data dengan data lain dalam suatu *cluster* yang sama menggunakan formula:

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (5)$$

Keterangan :

$a(i)$ = Perbedaan rata-rata pada data (i) ke semua data lain di cluster A

$d(i, j)$ = Jarak antara data i dan data j

2. Menghitung rata-rata jarak data tersebut dengan semua data di cluster lain menggunakan formula dibawah ini.

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (6)$$

Keterangan :

$d(i, C)$ = Perbedaan rata-rata data i ke seluruh data pada C

C = Cluster lain selain A

3. Memilih nilai jarak yang paling minimum:

$$b(i) = \min_{C \neq A} d(i, j) \quad (7)$$

Keterangan :

$b(i)$ = Nilai minimum jarak rata-rata data ke-i dengan semua data di cluster berbeda.

4. Menghitung nilai *Silhouette Coefficient* dengan menggunakan formula berikut:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (8)$$

Keterangan :

$a(i)$ = Perbedaan rata-rata pada data ke-i dengan semua data lain pada cluster A

$b(i)$ = Perbedaan rata-rata pada data ke-i dengan semua data pada cluster berbeda.

$s(i)$ = Nilai *Silhouette Coefficient*

Range nilai $s(i)$ yakni antara -1 dan 1, interpretasi nilai tersebut yaitu:

$s(i) = -1$: data ke-i digolongkan lemah (dekat pada cluster B daripada A)

$s(i) = 0$: data ke-i berada di tengah dua klaster (A dan B)

$s(i) = 1$: data ke-i digolongkan baik

Interpretasi nilai *Silhouette coefficient* juga ditunjukkan dalam Tabel II [1].

Tabel II
Interpretasi Nilai *Silhouette Coefficient*

<i>Silhouette Coefficient</i>	Interpretasi
≤ 0.25	Tidak terstruktur
0.26 – 0.50	Hasil struktur lemah
0.51 – 0.70	Hasil struktur baik
0.71 – 1.00	Hasil struktur kuat

III. HASIL DAN PEMBAHASAN

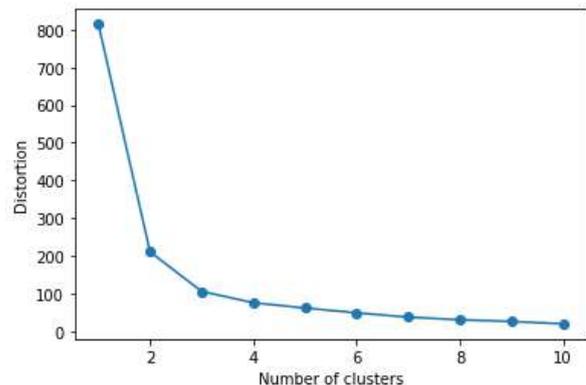
Terdapat 3 bagian yang akan dijabarkan pada hasil dan pembahasan yaitu hasil penentuan *cluster optimum*, hasil *clustering* seluruh dataset dan hasil uji dengan *Silhouette Coefficient* serta visualisasinya.

A. Hasil Penentuan *Cluster Optimum* pada Seluruh Dataset

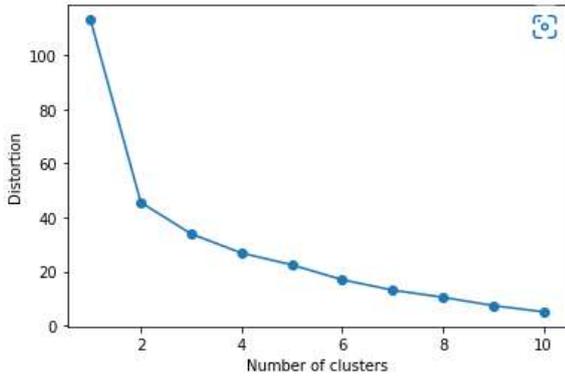
Sebelum implementasi metode *clustering* yakni *K-Means* dan *K-Means++*, agar memperoleh nilai K yang optimum maka diperlukan validasi *cluster*. Pada penelitian ini, untuk validasi *cluster* menerapkan metode *Elbow*. Metode *Elbow* di implementasikan pada setiap dataset jenis kriminalitas. Jika terdapat titik *Elbow* yang tidak jelas (ambigu), peneliti menentukan titik *cluster optimum* berdasarkan perkiraan manual. Berdasarkan Gbr 3 – Gbr 12 dan Tabel III dapat dilihat nilai K *cluster optimum* pada masing-masing jenis kriminalitas.

Tabel III
Cluster Optimum

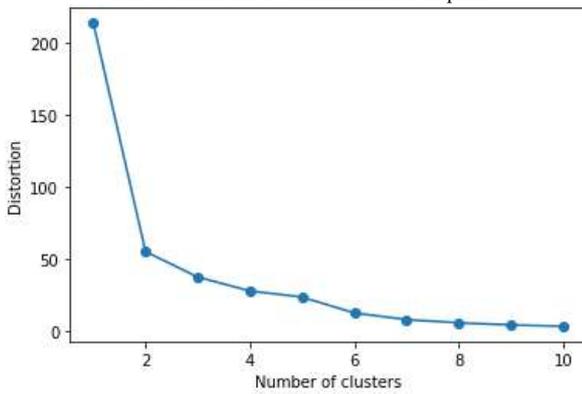
Jenis Kriminalitas	<i>Cluster Optimum</i>
Curat/ Curian dengan Pemberatan	K=3
Penipuan	K=9
Curanmor	K=5
Pemeriksaan	K=5
Pembunuhan	K=7
Penganiayaan	K=8
Perampokan/Curas	K=8
Perjudian	K=6
Pengrusakan	K=4
Pengroyokan	K=7



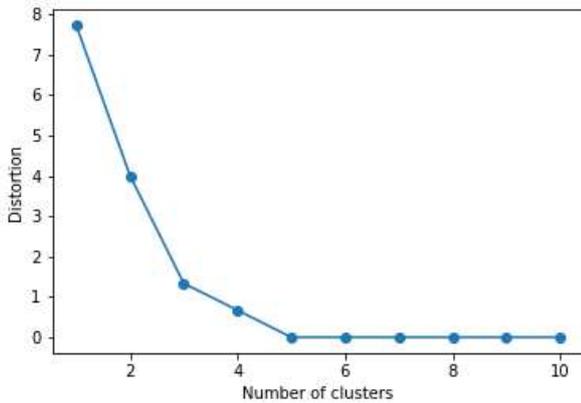
Gbr 3. Grafik Elbow Pada Dataset Curat



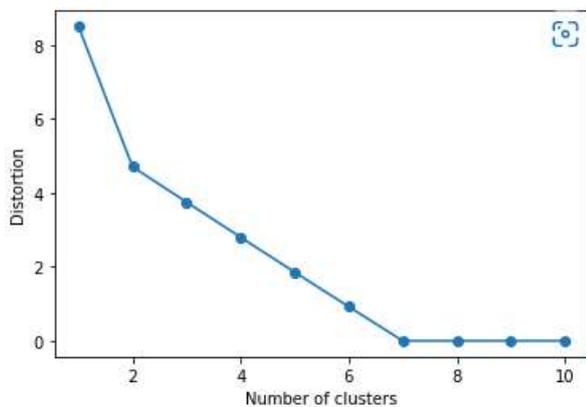
Gbr 4. Grafik Elbow Pada Dataset Penipuan



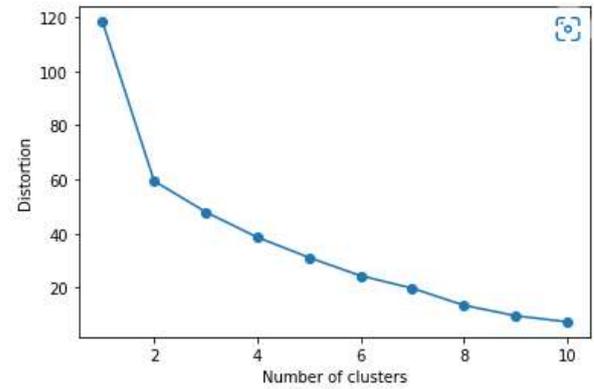
Gbr 5. Grafik Elbow Pada Dataset Curanmor



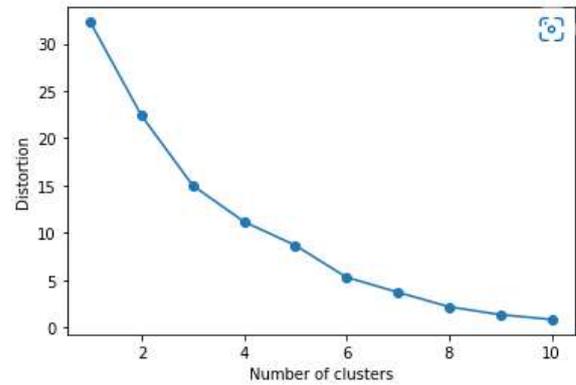
Gbr 6. Grafik Elbow Pada Dataset Pemerksaan



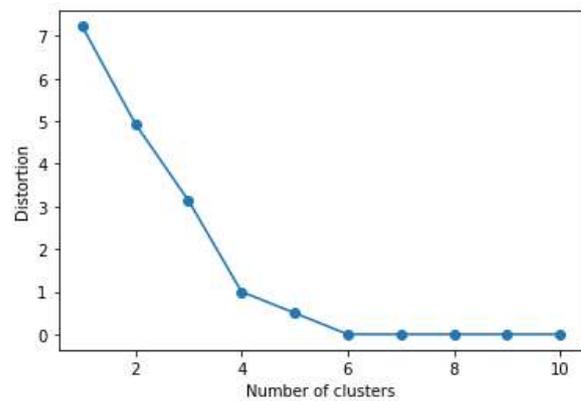
Gbr 7. Grafik Elbow Pada Dataset Pembunuhan



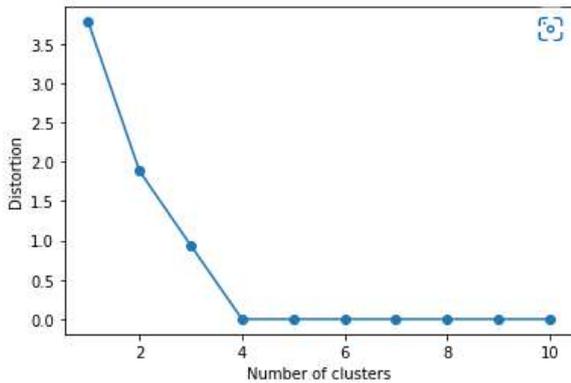
Gbr 8. Grafik Elbow Pada Dataset Penganiayaan



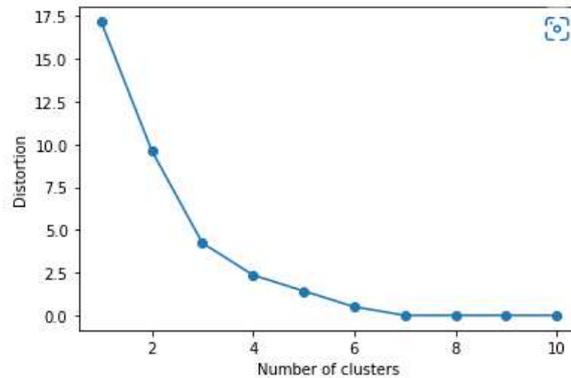
Gbr 9. Grafik Elbow Pada Dataset Perampokan



Gbr 10. Grafik Elbow Pada Dataset Perjudian



Gbr 11. Grafik Elbow Pada Dataset Pengrusakan



Gbr 12. Grafik Elbow Pada Dataset Pengroyokan

B. Hasil *clustering* pada jenis kriminalitas Curian dengan Pemberatan menggunakan *K-Means* dan *K-Means++*

Setelah diperoleh nilai *cluster* optimum $K = 3$, selanjutnya dilakukan penerapan metode *K-Means* dan *K-Means++* pada jenis kriminalitas Curat. Hasil *clustering* dengan kedua metode dapat dilihat pada Tabel IV dimana tidak terdapat perbedaan hasil *cluster* pada kedua metode.

Tabel IV
Hasil *Clustering* Jenis Kriminalitas Curat dengan *K-Means* dan *K-Means++*

Level Rawan	Metode <i>K-Means</i>	Metode <i>K-Means++</i>
0	Arosbaya, Burneh, Konang, Galis, Kwanyar, Socah, Labang, Kokop, Tanjung Bumi, Sepulu, Klampis, Blega, Geger, Modung, Tanah Merah, Tragah	Arosbaya, Burneh, Konang, Galis, Kwanyar, Socah, Labang, Kokop, Tanjung Bumi, Sepulu, Klampis, Blega, Geger, Modung, Tanah Merah, Tragah
1	Kamal	Kamal
2	Bangkalan	Bangkalan

C. Hasil *clustering* pada jenis kriminalitas Penipuan menggunakan *K-Means* dan *K-Means++*

Berdasarkan Tabel III yakni $K=9$ sebagai jumlah *cluster* optimum pada jenis kriminalitas Penipuan, selanjutnya

dilakukan implementasi metode *K-Means* dan *K-Means++*. Hasil *clustering* kedua metode ditunjukkan pada Tabel V.

Tabel V
Hasil *Clustering* Jenis Kriminalitas Penipuan dengan *K-Means* dan *K-Means++*

Level Rawan	Metode <i>K-Means</i>	Metode <i>K-Means++</i>
0	Arosbaya, Klampis, Labang, Galis, Kwanyar, Modung, Kokop, Geger, Sepulu	Arosbaya, Geger, Kokop, Galis, Klampis, Labang, Kwanyar, Modung, Sepulu
1	Blega, Geger	Blega
2	Modung	Kamal
3	Burneh	Tanah Merah
4	Tanah Merah	Tragah
5	Tragah	Burneh, Tanjung Bumi
6	Kamal, Socah	Socah
7	Konang, Tanjung Bumi	Konang
8	Bangkalan	Bangkalan

Untuk hasil *clustering* jenis kriminalitas Penipuan, terdapat perbedaan anggota *cluster* yakni pada hasil metode *K-Means*, kecamatan Konang dan Tanjung Bumi berada dalam satu *cluster* yang sama yaitu *cluster* 7. Sedangkan hasil *clustering* menggunakan *K-Means++*, kecamatan Konang dan Tanjung Bumi terdapat pada *cluster* yang berbeda yakni kecamatan Konang terdapat pada *cluster* 7 dan Tanjung Bumi terdapat dalam satu *cluster* bersama kecamatan Burneh yakni pada *cluster* 5. Selanjutnya hasil *clustering* dengan *K-Means* untuk kecamatan Geger terdapat pada *cluster* 2 bersama kecamatan Blega. Sedangkan hasil *K-Means++*, kecamatan Geger terdapat pada *cluster* 0 bersama kecamatan Arosbaya, Galis, Modung, Kokop, Klampis, Labang, Kwanyar, dan Sepulu. Untuk kecamatan Socah, pada hasil *clustering* dengan *K-Means* terdapat pada *cluster* 6 bersama kecamatan Kamal. Sedangkan pada hasil *K-Means++*, kecamatan Socah terdapat pada *cluster* 6 dan merupakan anggota tunggal.

D. Hasil *clustering* pada jenis kriminalitas Curanmor menggunakan *K-Means* dan *K-Means++*

Setelah didapatkan nilai $K=5$ sebagai jumlah *cluster* optimum, kemudian dilakukan *clustering* dengan menerapkan metode *K-Means* dan *K-Means++* pada jenis kriminalitas Curanmor. Hasil *clustering* dengan kedua metode pada Tabel VI menunjukkan bahwa tidak terdapat perbedaan hasil *cluster*.

Tabel VI
Hasil *Clustering* Jenis Kriminalitas Curanmor dengan *K-Means* dan *K-Means++*

Level Rawan	Metode <i>K-Means</i>	Metode <i>K-Means++</i>
0	Blega	Blega
1	Modung, Tragah	Modung, Tragah
2	Arosbaya, Galis, Klampis, Sepulu,	Arosbaya, Galis, Klampis, Sepulu, Konang, Tanah

	Konang, Tanah Merah, Kokop, Kwanyar, Kamal, Socah, Labang, Tanjung Bumi	Merah, Kokop, Kwanyar, Kamal, Socah, Labang, Tanjung Bumi
3	Burneh, Geger	Burneh, Geger
4	Bangkalan	Bangkalan

E. Hasil *clustering* pada jenis kriminalitas Pemerksosaan menggunakan *K-Means* dan *K-Means++*

Hasil *clustering* pada jenis kriminalitas Pemerksosaan dengan menerapkan metode *K-Means* dan *K-Means++* dan $K=5$ sebagai jumlah *cluster* optimum ditunjukkan pada Tabel VII.

Tabel VII
Hasil *Clustering* Jenis Kriminalitas Pemerksosaan dengan *K-Means* dan *K-Means++*

Level Rawan	Metode <i>K-Means</i>	Metode <i>K-Means++</i>
0	Blega, Klampis, Konang, Galis, Socah, Labang, Kokop, Burneh, Tanah Merah, Sepulu, Kwanyar, Tanjung Bumi	Blega, Klampis, Konang, Galis, Socah, Labang, Kokop, Burneh, Tanah Merah, Sepulu, Kwanyar, Tanjung Bumi
1	Arosbaya, Geger	Arosbaya, Geger
2	Modung	Modung
3	Tragah	Tragah
4	Kamal, Bangkalan	Kamal, Bangkalan

Hasil implementasi metode *K-Means* dan *K-Means++* pada Tabel VII dapat disimpulkan bahwa tidak terdapat perbedaan anggota pada 5 *cluster* yang telah diperoleh.

F. Hasil *clustering* pada jenis kriminalitas Pembunuhan menggunakan *K-Means* dan *K-Means++*

Berdasarkan Tabel III yakni $K=7$ sebagai jumlah *cluster* optimum pada jenis kriminalitas Pembunuhan, selanjutnya dilakukan implementasi metode *K-Means* dan *K-Means++*. Hasil *clustering* pada Tabel VIII menunjukkan tidak terdapat perbedaan hasil *cluster*.

Tabel VIII
Hasil *Clustering* Jenis Kriminalitas Pembunuhan dengan *K-Means* dan *K-Means++*

Level Rawan	Metode <i>K-Means</i>	Metode <i>K-Means++</i>
0	Blega, Kokop, Burneh, Kamal, Socah, Klampis, Tanjung Bumi, Konang, Tanah Merah, Labang, Tragah	Blega, Kokop, Burneh, Kamal, Socah, Klampis, Tanjung Bumi, Konang, Tanah Merah, Labang, Tragah
1	Geger	Geger
2	Arosbaya	Arosbaya
3	Sepulu	Sepulu
4	Galis	Galis
5	Kwanyar	Kwanyar
6	Bangkalan	Bangkalan

G. Hasil *clustering* pada jenis kriminalitas Penganiayaan menggunakan *K-Means* dan *K-Means++*

Berdasarkan pada Tabel III yakni $K=8$ sebagai jumlah *cluster* optimum pada Penganiayaan, selanjutnya dilakukan penerapan metode *K-Means* dan *K-Means++*. Hasil *clustering* kedua metode ditunjukkan pada Tabel IX.

Tabel IX
Hasil *Clustering* Jenis Kriminalitas Penganiayaan dengan *K-Means* dan *K-Means++*

Level Rawan	Metode <i>K-Means</i>	Metode <i>K-Means++</i>
0	Arosbaya, Blega, Geger, Kokop, Tanah Merah, Kwanyar, Tanjung Bumi, Konang, Tragah	Geger, Kokop, Konang, Kwanyar, Tanah Merah, Tanjung Bumi, Tragah
1	Sepulu, Socah	Sepulu, Socah
2	Burneh, Labang	Burneh, Labang
3	Modung	Modung
4	Kamal	Blega, Arosbaya
5	Klampis	Kamal, Klampis
6	Galis	Galis
7	Bangkalan	Bangkalan

Pada hasil *clustering* jenis kriminalitas Penganiayaan terdapat beberapa perbedaan yakni pada *K-Means*, kecamatan Blega dan Arosbaya berada dalam satu *cluster* yang sama dengan kecamatan Geger, Tanah Merah, Konang, Kwanyar, Kokop, Tanjung Bumi, dan Tragah yakni pada *cluster* 0. Sedangkan pada *K-Means++* Arosbaya dan Blega berada dalam satu *cluster* yakni *cluster* 4 dan tidak ada kecamatan lain pada *cluster* tersebut. Sementara pada *K-Means*, kecamatan Kamal dan Klampis berada dalam *cluster* yang berbeda yakni dalam *cluster* 4 dan *cluster* 5 secara berturut-turut. Sedangkan pada *K-Means++*, Kamal dan Klampis berada dalam satu *cluster* yang sama yaitu pada *cluster* 5.

H. Hasil *clustering* pada jenis kriminalitas Perampokan/Curas menggunakan *K-Means* dan *K-Means++*

Setelah didapatkan $K=8$ sebagai jumlah *cluster* optimum, selanjutnya dilakukan penerapan metode *K-Means* dan *K-Means++* pada jenis kriminalitas Perampokan. Hasil *clustering* pada jenis kriminalitas Perampokan ditunjukkan pada Tabel X.

Tabel X
Hasil *Clustering* Jenis Kriminalitas Perampokan dengan *K-Means* dan *K-Means++*

Level Rawan	Metode <i>K-Means</i>	Metode <i>K-Means++</i>
0	Galis, Geger, Klampis, Modung, Tragah, Konang	Galis, Geger, Klampis, Modung, Tragah, Konang
1	Blega, Kwanyar	Blega, Kwanyar
2	Tanjung Bumi	Kokop, Sepulu, Tanjung Bumi
3	Burneh, Kokop, Sepulu	Burneh
4	Arosbaya, Kamal, Tanah	Arosbaya, Kamal, Tanah

	Merah	Merah
5	Labang	Labang
6	Socah	Socah
7	Bangkalan	Bangkalan

Untuk hasil *clustering* pada jenis kriminalitas Perampokan, terdapat beberapa perbedaan anggota *cluster* yakni pada *K-Means*, Kecamatan Burneh, Kokop, dan Sepulu terdapat dalam *cluster* yang sama yaitu *cluster* 3, dan Tanjung Bumi terdapat dalam *cluster* 2. Sedangkan pada *K-Means++*, Kokop, Sepulu, dan Tanjung Bumi terdapat dalam satu *cluster* yang sama yaitu *cluster* 2 dan Burneh berada dalam *cluster* 3.

I. Hasil *clustering* pada jenis kriminalitas Perjudian menggunakan *K-Means* dan *K-Means++*

Setelah didapatkan K=6 sebagai jumlah *cluster* optimum, selanjutnya pada Tabel XI dapat dilihat hasil *clustering* dengan *K-Means* dan *K-Means++* yang menunjukkan tidak terdapat perbedaan hasil *cluster* dari kedua metode.

Tabel XI
Hasil *Clustering* Jenis Kriminalitas Perjudian dengan *K-Means* dan *K-Means++*

Level Rawan	Metode <i>K-Means</i>	Metode <i>K-Means++</i>
0	Arosbaya, Galis, Sepulu, Kokop, Blega, Klampis, Burneh, Tanah Merah, Labang, Socah, Tanjung Bumi	Arosbaya, Galis, Sepulu, Kokop, Blega, Klampis, Burneh, Tanah Merah, Labang, Socah, Tanjung Bumi
1	Kwanyar	Kwanyar
2	Modung, Geger	Modung, Geger
3	Tragah	Tragah
4	Kamal	Kamal
5	Bangkalan	Bangkalan

J. Hasil *clustering* pada jenis kriminalitas Pengrusakan menggunakan *K-Means* dan *K-Means++*

Berdasarkan pada Tabel III diperoleh K=4 sebagai jumlah *cluster* optimum. Selanjutnya dilakukan penerapan metode *K-Means* dan *K-Means++* pada jenis kriminalitas Pengrusakan. Hasil *clustering* dengan kedua metode ditunjukkan pada Tabel XII.

Tabel XII
Hasil *Clustering* Jenis Kriminalitas Pengrusakan dengan *K-Means* dan *K-Means++*

Level Rawan	Metode <i>K-Means</i>	Metode <i>K-Means++</i>
0	Arosbaya, Galis, Konang, Blega, Geger, Burneh, Labang, Kokop, Sepulu, Klampis, Bangkalan, Galis, Modung, Konang, Tanah Merah	Arosbaya, Galis, Konang, Blega, Geger, Burneh, Labang, Kokop, Sepulu, Klampis, Bangkalan, Galis, Modung, Konang, Tanah Merah
1	Kamal	Kamal
2	Tanjung Bumi	Tanjung Bumi
3	Socah	Socah

Berdasarkan Tabel XII, hasil *clustering* menunjukkan tidak terdapat perbedaan hasil *cluster* dari kedua metode baik metode *K-Means* dan *K-Means++*.

K. Hasil *clustering* pada jenis kriminalitas Pengroyokan menggunakan *K-Means* dan *K-Means++*

Berdasarkan Tabel III, setelah diperoleh K=7 sebagai jumlah *cluster* optimum, selanjutnya dilakukan penerapan metode *K-Means* dan *K-Means++* pada jenis kriminalitas Pengroyokan. Hasil *clustering* dengan kedua metode ditunjukkan pada Tabel XIII.

Tabel XIII
Hasil *Clustering* Jenis Kriminalitas Pengroyokan dengan *K-Means* dan *K-Means++*

Level Rawan	Metode <i>K-Means</i>	Metode <i>K-Means++</i>
0	Arosbaya, Galis, Geger, Klampis, Tanah Merah, Kokop, Konang, Kwanyar, Sepulu, Tanjung Bumi, Socah, Tragah	Arosbaya, Galis, Geger, Klampis, Tanah Merah, Kokop, Konang, Kwanyar, Sepulu, Tanjung Bumi, Socah, Tragah
1	Blega	Blega
2	Modung	Modung
3	Labang	Labang
4	Kamal	Kamal
5	Bangkalan	Bangkalan
6	Burneh	Burneh

Hasil *clustering* pada jenis kriminalitas Pengroyokan pada Tabel XIII dapat dilihat bahwa tidak ada perbedaan baik dari *cluster* 0-6, seluruh kecamatan terdapat pada *cluster* yang sama.

L. Hasil Uji Metode *K-Means* dan *K-Means++* dengan *Silhouette Coefficient* dan Visualisasi Hasil *Clustering*

Setelah dilakukan *clustering* pada seluruh dataset jenis kriminalitas dengan menerapkan kedua metode yakni *K-Means* dan *K-Means++*, selanjutnya dilakukan pengujian dengan menerapkan metode *Silhouette Coefficient*. Pengujian tersebut bertujuan untuk mencari metode yang lebih baik. Adapun hasil pengujian kedua metode ditunjukkan pada Tabel XIV.

Tabel XIV
Hasil Nilai Pengujian dengan *Silhouette Coefficient*

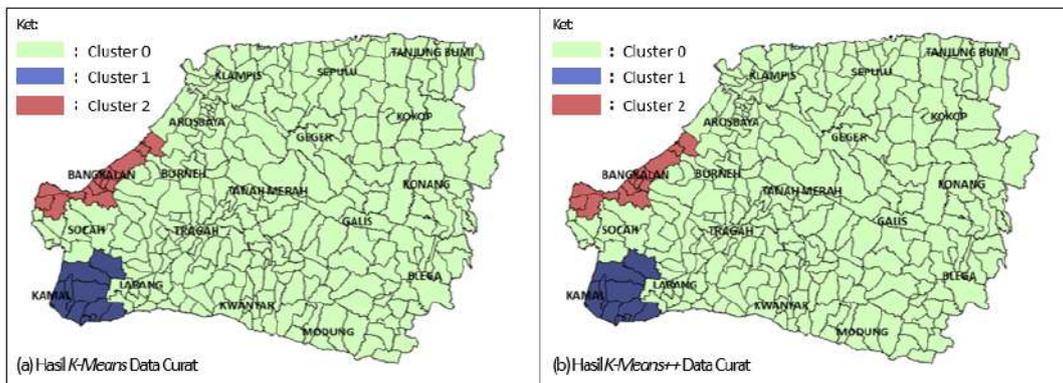
<i>Silhouette Coefficient</i>		
Jenis Kriminalitas	<i>K-Means</i> dan Interpretasi	<i>K-Means++</i> dan Interpretasi
Pembunuhan	0.6667 (Baik)	0.6667 (Baik)
Curanmor	0.7769 (Kuat)	0.7769 (Kuat)
Pemerksaan	0.8889 (Kuat)	0.8889 (Kuat)
Perampokan/Curas	0.4898 (Lemah)	0.4057 (Lemah)
Perjudian	0.7778 (Kuat)	0.7778 (Kuat)

Penipuan	0.2243 (Tidak Terstruktur)	0.2534 (Tidak Terstruktur)
Curat	0.5908 (Baik)	0.5908 (Baik)
Pengrusakan	0.8333 (Kuat)	0.8333 (Kuat)
Penganiayaan	0.1683 (Tidak Terstruktur)	0.2314 (Tidak Terstruktur)
Pengroyokan	0.6667 (Baik)	0.6667 (Baik)

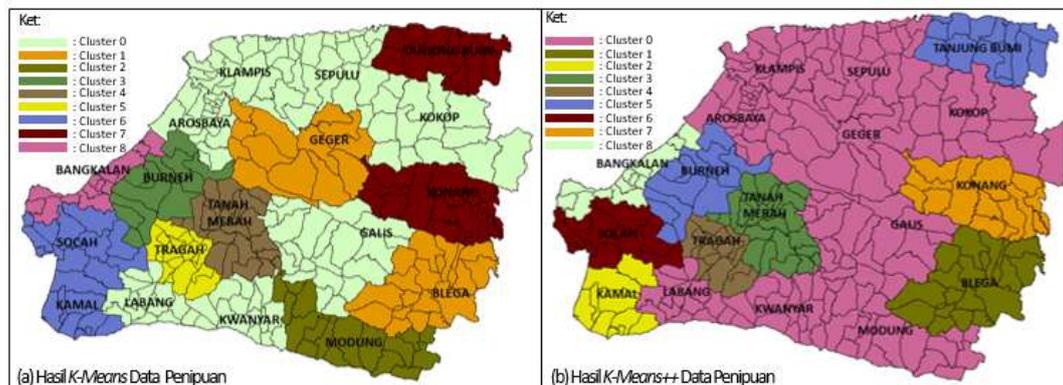
Perampokan.

Pada jenis kriminalitas Penganiayaan, hasil uji *Silhouette Coefficient* pada metode *K-Means++* lebih unggul daripada nilai uji metode *K-Means* yakni 0.1683 untuk metode *K-Means* dan 0.2314 untuk *K-Means++*. Pada jenis kriminalitas Penipuan, hasil uji metode *K-Means++* juga lebih unggul daripada nilai uji metode *K-Means* yaitu 0.2243 untuk metode *K-Means* dan 0.2534 untuk *K-Means++*. Sedangkan pada Perampokan, metode *K-Means* memiliki hasil uji *Silhouette Coefficient* yang lebih unggul daripada nilai uji metode *K-Means++* yaitu 0.4898 untuk metode *K-Means* dan 0.4057 untuk *K-Means++*. Hal ini mengindikasikan bahwa metode *K-Means++* memiliki hasil uji *Silhouette Coefficient* yang lebih baik pada 2 jenis kriminalitas yaitu Penganiayaan dan Penipuan. Sedangkan metode *K-Means* memberikan hasil nilai yang lebih baik pada 1 jenis kriminalitas yaitu Perampokan.

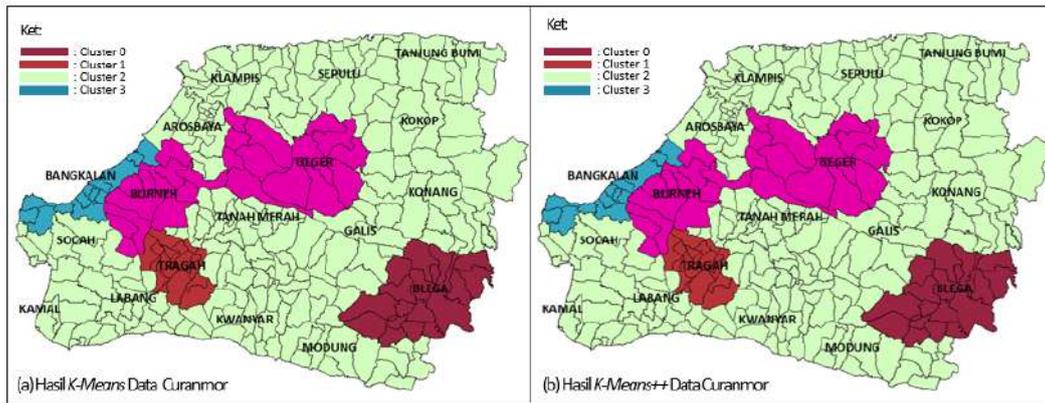
Pada Tabel XIV dapat dilihat bahwa dari 10 dataset terdapat 7 jenis kriminalitas yaitu Pembunuhan, Curanmor, Pemerkosaan, Perjudian, Curat/Pencurian dengan Pemberatan, Pengrusakan, dan pengroyokan memperoleh hasil uji *Silhouette Coefficient* yang sama, baik itu metode *K-Means* dan *K-Means+*. Sedangkan 3 jenis kriminalitas lainnya memiliki hasil pengujian *Silhouette Coefficient* yang berbeda pada kedua metode yaitu jenis kriminalitas Penganiayaan, Penipuan, dan



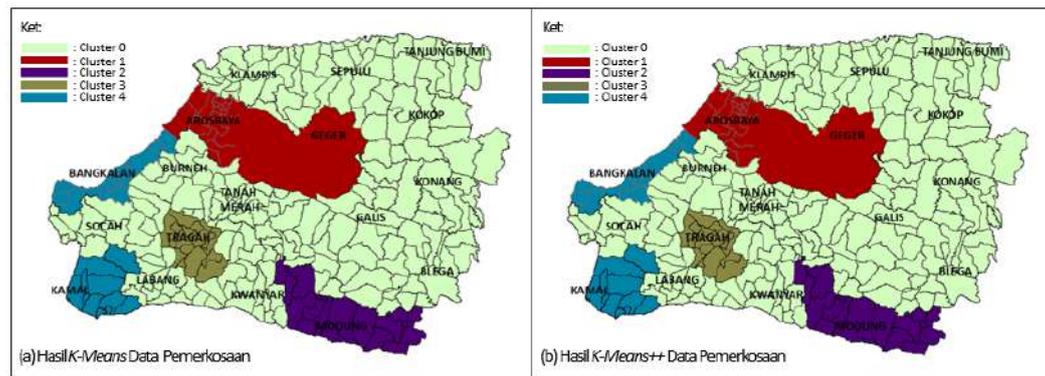
Gbr 13. Visualisasi Hasil *Clustering* Jenis Kriminalitas Curat



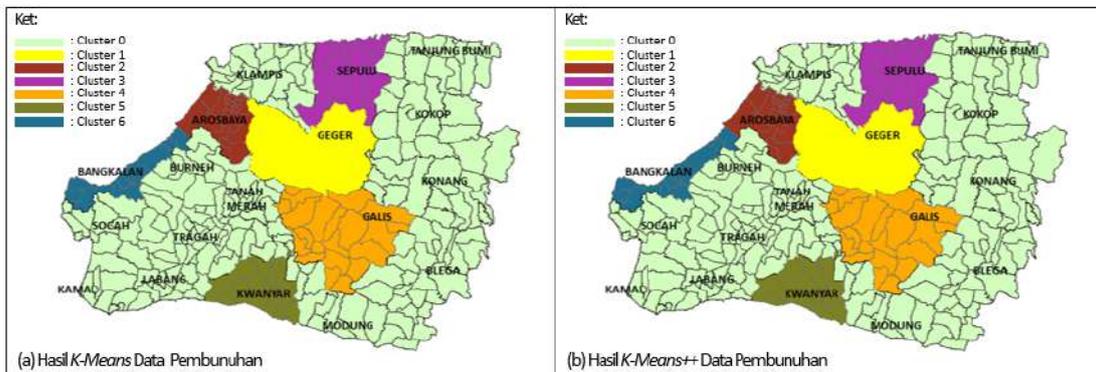
Gbr 14. Visualisasi Hasil *Clustering* Jenis Kriminalitas Penipuan



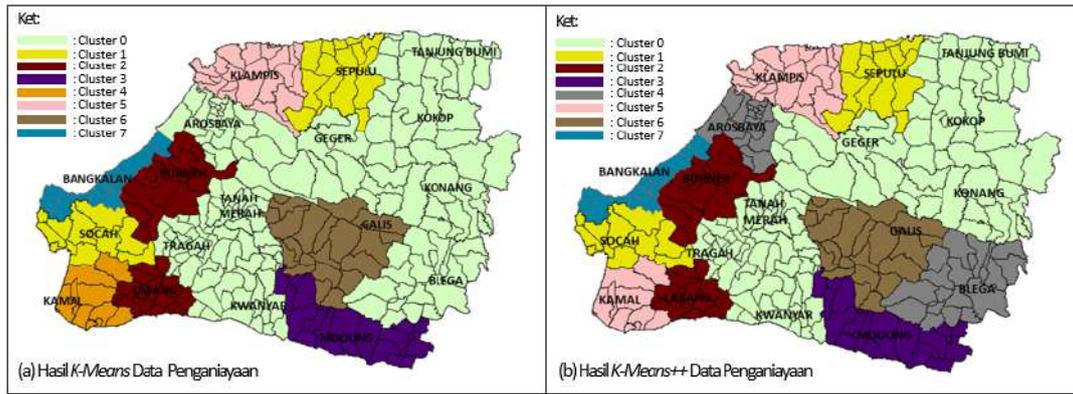
Gbr 15. Visualisasi Hasil Clustering Jenis Kriminalitas Curanmor



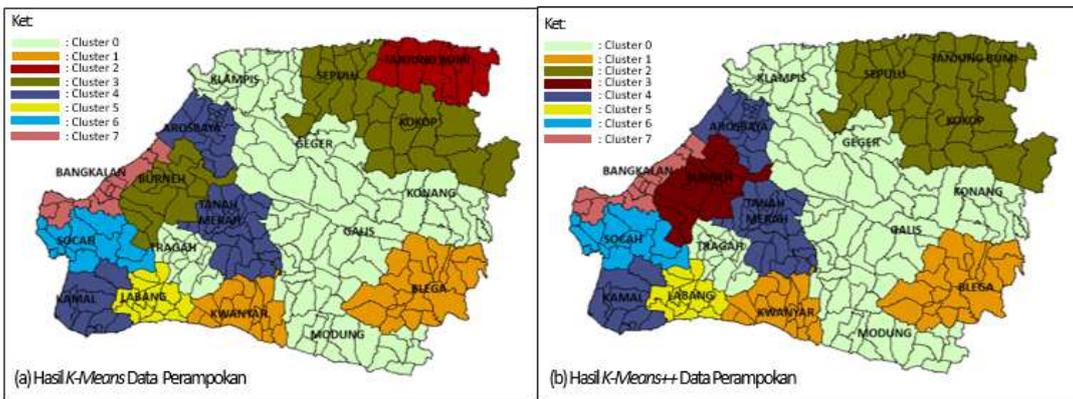
Gbr 16. Visualisasi Hasil Clustering Jenis Kriminalitas Pemerksaan



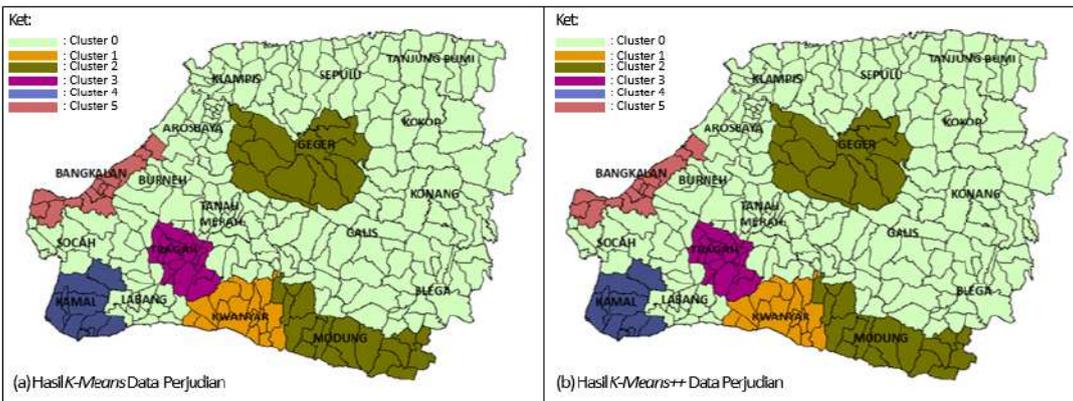
Gbr 17. Visualisasi Hasil Clustering Jenis Kriminalitas Pembunuhan



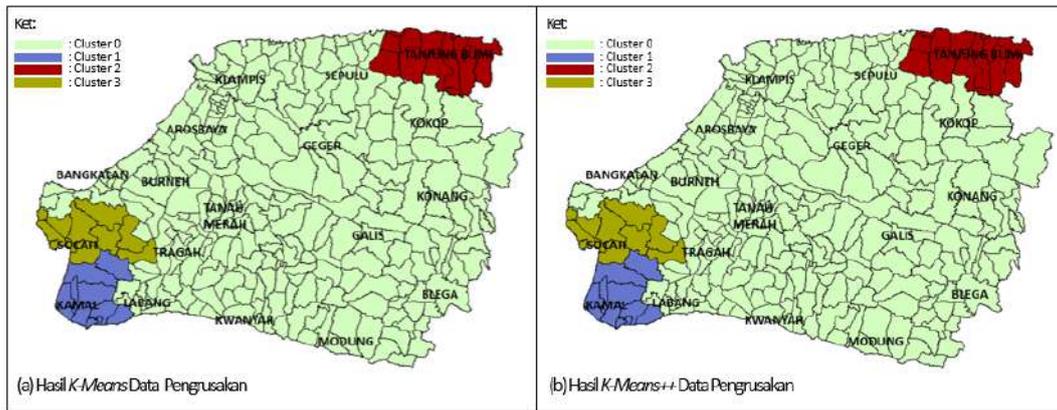
Gbr 18. Visualisasi Hasil *Clustering* Jenis Kriminalitas Penganiayaan



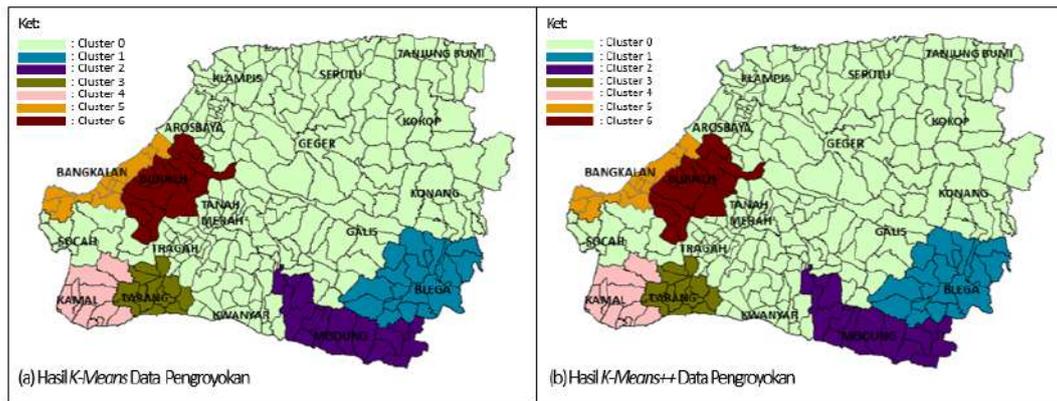
Gbr 19. Visualisasi Hasil *Clustering* Jenis Kriminalitas Perampokan



Gbr 20. Visualisasi Hasil *Clustering* Jenis Kriminalitas Perjudian



Gbr 21. Visualisasi Hasil Clustering Jenis Kriminalitas Pengrusakan



Gbr 22. Visualisasi Hasil Clustering Jenis Kriminalitas Pengroyokan

IV. KESIMPULAN

Perbedaan antara metode *K-Means* dan metode *K-Means++* hanya terletak pada pemilihan titik pusat *cluster*. Pada metode *K-Means*, pemilihan titik pusat *cluster* (*centroid*) dilakukan secara *random*, sedangkan pada metode *K-Means++* pemilihan titik pusat *cluster* (*centroid*) dilakukan dengan menentukan probabilitas optimum. Pada hasil *clustering* menggunakan kedua metode terdapat beberapa perbedaan, yakni pada jenis kriminalitas Penganiayaan, Penipuan, dan Perampokan. -SAWS Hasil uji pengolahan kedua metode dengan menggunakan *Silhouette Coefficient* juga terdapat perbedaan. Hasil uji *Silhouette* dengan metode *K-Means* dan *K-Means++* pada dataset Penganiayaan sebesar 0,1683 dan 0,2314 secara berturut-turut, sedangkan pada dataset Penipuan sebesar 0,2243 dan 0,2534 secara berturut-turut, dan hasil uji pada dataset Perampokan dengan *K-Means* dan *K-Means++* berturut-turut sebesar 0,4898 dan 0,4057. Dengan demikian dapat diketahui bahwa pada penelitian ini, metode *K-Means++* memberikan hasil uji *Silhouette* yang lebih baik pada 2 dataset yakni Penganiayaan dan Penipuan. Sedangkan metode *K-Means* memiliki hasil uji *Silhouette* yang lebih baik pada dataset Perampokan.

V. SARAN

Pada penelitian sejenis untuk kedepannya dapat dikembangkan dengan menggunakan dataset kriminalitas Kabupaten Bangkalan pada tahun selanjutnya dan juga dapat dikembangkan dengan membuat Sistem Informasi Geografis. Tidak hanya itu, peneliti selanjutnya juga dapat menambahkan metode *clustering* lainnya sebagai bahan untuk perbandingan.

UCAPAN TERIMA KASIH

Dengan mengucap Alhamdulillah atas Karunia Allah SWT yang telah diberikan kepada penulis, sehingga penelitian yang saya lakukan dapat diselesaikan dengan baik. Tidak lupa Sholawat tetap tercurahkan kepada Nabi Muhammad SAW. Kemudian terimakasih kepada kedua orang tua dan keluarga yang telah mendoakan dan memberi dukungan dengan segenap hati. Ucapan terimakasih juga saya sampaikan kepada Dr. Yuni Yamasari, S.Kom., M.Kom. selaku dosen pembimbing, yang tiada henti memberikan solusi dan arahan kepada penulis sehingga penelitian ini dapat terselesaikan dengan baik. Terimakasih kepada sahabat saya yakni Novita,

Aini, Dewi, dan Khonsa yang selalu siap meluangkan waktu untuk mendengarkan keluh kesah penulis. Terakhir, ucapan banyak terimakasih kepada *Soon to be* Prof Arief R. H., S.Mat., Ph.D. yang selalu memberikan solusi terhadap berbagai keraguan pada penelitian ini.

REFERENSI

- [1] A. Struyf, M. Hubert, and P. J. Rousseeuw. "Clustering in an Object-Oriented Environment". *Journal of Statistical Software*, vol 1, Issue 4, pp. 1-30, 1997.
- [2] Arief, Danang, Mahmud. "Implementation of K-Means++ Algorithm for Store Customers Segmentation Using Neo4J". *Ind. Journal on Computing*. ISSN: 2460-9056. 2021.
- [3] Bholowalia, Purnima & Kumar, Arvind, "EBK-Means: A Clustering Techniques based on Elbow Method and K-Means in WSN. *International Journal of Computer Application*" (0975-8887), IX(105), pp. 17-24. 2014.
- [4] D. F. Azuri, Zulhanif, and R. S. Pontoh. "Pengelompokan Kabupaten/Kabupaten Di Pulau Jawa Berdasarkan Pembangunan Manusia Berbasis Gender Menggunakan Bisecting K-Means," *Peran Penelit. Ilmu Dasar dalam Menunjang Pembang. Berkelanjutan*, pp. 78–83, 2016, [Online].
- [5] Han, J., Kamber, M., & Pei, J. "Data Mining: Concepts and Techniques Third Edition". Morgan Kaufmann. Massachusetts (US). 2012.
- [6] Hapsari, D. P., & Widodo, E. "Pengelompokan Daerah Rawan Kriminalitas di Indonesia Menggunakan Analisis K-Means Clustering. *Prosiding SI MaNis*, 147-153. 2017
- [7] Hidayatunnismah. "Analisis Korespondensi Beberapa Kejadian Tindak Kriminal (Studi Kasus: Wilayah Kabupaten madya Bogor). Bogor: Institut Pertanian Bogor". 2003.
- [8] Irwanto, et. al "Optimasi Kinerja Algoritma Klasterisasi K-Means untuk kuantisasi Warna Citra". *Jurnal Teknik ITS*, I(1), pp.197-202. 2012.
- [9] Kodinariya, T.M and Makwana, P.R. "Review on determining number of cluster in K-Means Clustering.". *Int. J. Adv. Res. Comput. Sci. Manag. Stud.*, vol. 1, no.6, pp. 90-95. 2013
- [10] Madhulatha, T. S. "An Overview on Clustering Methods. *IOSR Journal of Engineering*", 2(4). 2012.
- [11] Margareth Dwiyantri, Arie Wahyu. "Analisis Klaster Berdasarkan Tindakan Kriminalitas Di Indonesia Tahun 2019". *Jurnal Statistika Industri dan Komputasi*. Vol.6, No.1. 2021
- [12] Merliana, N. P. E., & Santoso, A. J. "Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering". 2015
- [13] Aida. (2020). "Tahun 2020, Angka Kriminalitas di Bangkalan Makin Tinggi". Available: <https://www.korek.id/2020/12/31/tahun-2020-angka-kriminalitas-di-bangkalan-makin-tinggi/>.
- [14] Prasetyo, Eko. "Data Mining: Konsep dan Aplikasi Menggunakan Matlab". Jakarta: Andi Publisher. 2013h
- [15] Prasetyo, Eko. "Mengolah Data Menjadi Informasi Menggunakan Matlab". Yogyakarta: ANDI. 2014
- [16] Wirawan, Y. A., Indwiarti, D., Si, M., Sibaroni, Y., dan Si, S. "Kombinasi Algoritma Agglomerative Clustering dan K-Means Untuk Segmentasi Pengunjung Website". *Jurnal Ilmu Komputasi Fakultas Informatika. Universitas Telkom*. 2015.
- [17] Yuni, Sarjon, & Gnadi. "Algoritma K-Means Clustering dalam Mengklasifikasi Data Daerah Rawan Tindak Kriminalitas (Polres Kepulauan Mentawai)". *Jurnal Sistim Informasi dan Teknologi*, Vol.3, No.4. 2021.
- [18] Wu, X. and Kumar, V. "The Top Ten Algorithms in Data Mining. London: CRC Press Taylor & Francis Group". 2009.
- [19] Z. Min, Kai-fei, "Improved research to K-Means initial cluster centers" in *Ninth International Conference on Frontier of Computer Science and Technology*, pp. 349-353. 2015.