

**IMPLEMENTASI *K-MEANS CLUSTERING* DALAM
PEMETAAN DAERAH RAWAN PENCURIAN KENDARAAN
BERMOTOR DAN PENCURIAN DENGAN KEKERASAN
BERBASIS SISTEM INFORMASI GEOGRAFIS
(STUDI KASUS DI KABUPATEN PROBOLINGGO)**

PROPOSAL SKRIPSI



Oleh :

Daffa Fauzi Rahman

NIM E41211408

**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI JEMBER
2024**

**IMPLEMENTASI K-MEANS CLUSTERING DALAM
PEMETAAN DAERAH RAWAN PENCURIAN KENDARAAN
BERMOTOR DAN PENCURIAN DENGAN KEKERASAN
BERBASIS SISTEM INFORMASI GEOGRAFIS
(STUDI KASUS DI KABUPATEN PROBOLINGGO)**

PROPOSAL SKRIPSI



Sebagai salah satu syarat untuk memperoleh gelar Sarjana Terapan (S.Tr.)
di Program Studi Teknik Informatika Jurusan Teknologi Informasi

Oleh :

Daffa Fauzi Rahman

NIM E41211408

**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI JEMBER
2024**

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET DAN TEKNOLOGI
POLITEKNIK NEGERI JEMBER
JURUSAN TEKNOLOGI INFORMASI**

HALAMAN PENGESAHAN PROPOSAL

1. Judul : Implementasi *K-Means Clustering* Dalam Pemetaan Daerah Rawan Pencurian Kendaraan Bermotor dan Pencurian Dengan Kekerasan Berbasis Sistem Informasi Geografis (Studi Kasus di Kabupaten Probolinggo)
2. Identitas Pelaksanaan
 - a. Nama Lengkap : Daffa Fauzi Rahman
 - b. NIM : E41211408
 - c. Jurusan/ Program Studi : Teknologi Informasi / Teknik Informatika
3. Lokasi : Politeknik Negeri Jember
4. Identitas Dosen Pembimbing
 - a. Nama Lengkap : Bety Etikasari, S.Pd, M.Pd
 - b. NIP : 19920528 201803 2 001
 - c. Jurusan/ Program Studi : Teknologi Informasi / Teknik Informatika
5. Lama Kegiatan : Sepuluh (10) Bulan

Jember, 1 Juli 2024

Menyetujui,

Dosen Pembimbing

Dosen Pembahas

Pelaksana

Bety Etikasari, S.Pd, M.Pd

NIP. 19920528 201803 2 001

Dia Bitari Mei Yuana, S.ST., M.Tr.Kom.

NIP. 19930508 202203 2 013

NIM. E41211408

Mengetahui,

Koordinator Program Studi D-IV Teknik Informatika

Bety Etikasari, S.Pd, M.Pd
NIP. 19920528 201803 2 001

DAFTAR ISI

	Halaman
HALAMAN PENGESAHAN PROPOSAL	iii
DAFTAR ISI.....	iv
DAFTAR GAMBAR.....	vi
DAFTAR TABEL	vii
DAFTAR LAMPIRAN	viii
BAB 1. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	5
1.3 Tujuan Penelitian	5
1.4 Manfaat	5
1.5 Batasan Masalah.....	6
BAB 2. TINJAUAN PUSTAKA.....	7
2.1 State Of The Art.....	7
2.2 Landasan Teori.....	9
2.2.1 Curanmor dan Curas	9
2.2.2 Sistem Informasi Geografis (SIG)	10
2.2.3 K-Means Clustering	13

2.2.4 Davies-Bouldin Index (DBI) Error!

Bookmark not defined.

BAB 3. METODE PENELITIAN	17
3.1 Tempat dan Waktu Pelaksanaan	17
3.2 Alat dan Bahan	17
3.2.1 Alat Penelitian.....	17
3.2.2 Bahan Penelitian.....	18
3.3 Tahapan Penelitian	18
3.3.1 Studi Literatur	19
3.3.2 Pengumpulan Data	19
3.3.3 Pengolahan Data.....	19
3.3.4 Pengembangan Sistem	20
3.3.5 Pengujian.....	23
3.3.6 Analisis dan Pembahasan.....	24
DAFTAR PUSTAKA	25
LAMPIRAN	74

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Gambaran Data Spasial.....	13
Gambar 2.2 Tahapan Algoritma <i>K-Means Clustering</i>	14
Gambar 3.1 Tahapan Penelitian.....	18
Gambar 3.2 Use Case Sistem	Error!
Bookmark not defined.	
Gambar 3.3 Flowchart Sistem Admin	Error!
Bookmark not defined.	
Gambar 3.4 Flowchart Sistem User.....	Error!
Bookmark not defined.	

DAFTAR TABEL

	Halaman
Tabel 2.1 State Of The Art	7
Tabel 3.1 Tahapan Kegiatan	17

DAFTAR LAMPIRAN

	Halaman
LAMPIRAN 1 Kenaikan Kasus Kriminal di Probolinggo	74
LAMPIRAN 2 Data Kasus Curas dan Curanmor Tahun 2022	75

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Kriminalitas atau kejahatan merupakan suatu kejadian umum yang mempunyai dampak luas terhadap seluruh lapisan masyarakat, dan kriminalitas atau kejahatan ini bisa terjadi dimana saja dan tidak mengenal waktu (Apriliana & Haris R, 2022). Seperti yang telah disampaikan sebelumnya, Riswandi juga berpendapat yang sama dalam (Rohman, 2023) tindakan kriminal merugikan seluruh lapisan masyarakat baik dalam segi ekonomis. Psikologi, dan juga merupakan tindakan yang melanggar hukum dan norma-norma agama maupun sosial yang ada pada masyarakat. Tindak kriminal terjadi tentunya disebabkan oleh beberapa faktor, seperti : kemiskinan, kesempatan kerja, karakter pelaku yang melakukan kasus kejahatan, kepadatan penduduk, keadaan lingkungan, dan jumlah patroli polisi. Jenis-jenis tindakan kriminal ada beberapa, seperti : pencurian, pembunuhan, tindak asusila, dan lain sebagainya (Apriliana & Haris R, 2022).

Badan Pusat Statistik (BPS) telah merilis hasil perhitungannya tentang kriminalitas di Indonesia pada publikasinya di tahun 2023. Pada publikasi tersebut dinyatakan bahwa tingkat kriminalitas di Indonesia mengalami kenaikan pada tahun 2022 dengan jumlah tindak kriminal 372.965 kasus. Pada tahun sebelumnya yaitu pada tahun 2020 – 2021 angka Tingkat kriminalitas Indonesia mengalami penurunan, pada tahun 2020 terjadi 247.218 dan pada tahun 2021 terjadi 239.481 kasus. Pada hal tersebut tampak peningkatan yang cukup signifikan terjadinya beberapa tindak kriminalitas di tahun 2022. Jika dalam interval waktu (*crime clock*), dapat disampaikan bahwa setiap 2 menit 11 detik terjadi satu tindak kriminal pada tahun 2021, sedangkan pada tahun 2022 tindak kriminal terjadi dalam rentang waktu yang lebih cepat, yaitu 1 menit 24 detik terjadi satu tindak kriminalitas di wilayah Indonesia. Jika dilihat lebih detail lagi pada publikasi BPS tersebut, provinsi dengan tingkat kriminalitas tertinggi terjadi pada provinsi Jawa Timur dengan jumlah kasus sebesar 51.905 kasus. (BPS, 2023)

Salah satu kabupaten pada provinsi dengan tingkat kriminalitas tertinggi di tahun 2022 (Jawa Timur) adalah Kabupaten Probolinggo. Kabupaten Probolinggo memiliki luas $1.696,17 \text{ km}^2$ dengan 24 kecamatan, 5 kelurahan dan 325 desa. Menurut data dari Badan Pusat Statistik (BPS) Kabupaten Probolinggo pada tahun 2022, jumlah penduduk Kabupaten Probolinggo tercatat sebanyak 1.159.965 jiwa (BPS Kab Probolinggo, 2024). Selain itu letak geografis Kabupaten Probolinggo yang berbatasan langsung dengan selat Madura dan ada beberapa tempat yang termasuk dalam dataran tinggi, membuat Kabupaten Probolinggo memiliki beberapa wisata yang terkenal, baik di tingkat nasional maupun mancanegara. Salah satu wisata di kabupaten Probolinggo yang terkenal hingga ke mancanegara yaitu Gunung Bromo. Dengan adanya banyak wisata yang terkenal di tingkat nasional maupun mancanegara, membuat Kabupaten Probolinggo banyak dikunjungi wisatawan. Berdasarkan data BPS Kabupaten Probolinggo, di tahun 2023 terdapat 854.956 wisatawan domestik dan 37.094 wisatawan mancanegara (BPS Kab Probolinggo, 2024). Dengan banyaknya jumlah penduduk Kabupaten Probolinggo dan banyaknya wisatawan yang berkunjung ke Kabupaten Probolinggo tentu dapat membuat ancaman terjadinya tindak kriminalitas semakin tinggi.

Pada Kabupaten sendiri tingkat kejahatan dari tahun 2021 - 2022 mengalami peningkatan. Menurut Badan Pusat Statistik Kabupaten Probolinggo tindak kriminalitas yang terjadi di wilayah hukum Polres Probolinggo pada tahun 2021 tercatat sebanyak 399 kasus, sedangkan pada tahun 2022 tindak kriminalitas yang tercatat sebanyak 442 kasus (BPS Kab Probolinggo, 2023). Dari data tersebut dapat dinyatakan terjadi kenaikan sebesar 9,7 %. Jenis tindak kriminalitas yang paling banyak terjadi pada jenis tindak kriminal pencurian. Tindak kriminal pencurian ini terbagi lagi dalam dua kategori, yaitu pencurian kendaraan bermotor (curanmor) dan pencurian dengan kekerasan (curas).

Dapat terlihat pada data BPS tentang kriminalitas di Kabupaten Probolinggo mengalami kenaikan dan tindak kriminal terbanyak adalah pencurian dengan kekerasan (curas) dan pencurian kendaraan bermotor (curanmor). Dengan peningkatan tersebut dapat memberikan kesan bahwa Kabupaten Probolinggo kurang aman, sehingga secara tidak langsung dapat mempengaruhi jumlah

wisatawan yang ingin berkunjung ke Kabupaten Probolinggo. Salah satu upaya *preventif* dalam memberikan rasa aman bagi penduduk asli Kabupaten Probolinggo maupun wisatawan yang berkunjung ke Kabupaten Probolinggo atau tindakan untuk meminimalkan ancaman tindak kriminal dapat dengan memberikan pengamanan atau patroli pada beberapa titik daerah yang dapat dibilang rawan untuk terjadi tindakan-tindakan curas dan curanmor. Dalam menentukan suatu daerah tersebut termasuk ke dalam daerah rawan atau aman perlu dilakukan pemetaan untuk memastikannya, dan hasil dari pemetaan tersebut bisa dipublikasikan agar setiap orang dapat mengetahuinya. Pada Kabupaten Probolinggo sendiri belum ada pemetaan terkait tingkat kerawanan suatu kecamatan terhadap kasus tindak curas dan curanmor.

Sebelum ini terdapat beberapa penelitian yang membahas tentang pemetaan. Seperti penelitian yang dilakukan oleh Ruziq Nawaf Zulfahmi, dkk pada tahun 2023 yang menyatakan hasil dari pemetaan tingkat kriminalitas pada Kab/Kota di Jatim, Jateng, dan Yogyakarta menghasilkan dua *cluster* dengan rincian *cluster* 1 terdiri dari 15 kabupaten/kota, dan *cluster* 2 terdiri dari 63 kabupaten/kota, kemudian dengan melakukan pengujian *profiling cluster* menggunakan nilai rata - rata setiap *cluster*, Klaster 1 dapat dikatakan tinggi dibandingkan klaster 2 karena memiliki nilai dominan yang lebih besar untuk kategori dengan skor tinggi dari masing-masing variabel (Ruziq Nawaf Zulfahmi et al., 2023). Selain itu Preddy Marpaung, dkk melakukan penelitian pada tahun 2023 dan menyimpulkan bahwa dengan menerapkan algoritma *K-Means*, dapat mengelompokkan kepadatan penduduk di Kabupaten Deli Serdang ke dalam tiga kelompok (*cluster*) ,yaitu *Cluster* 1 daerah penduduk sangat padat sebanyak 3 kecamatan, *Cluster* 2 daerah Penduduk padat terdapat 4 kecamatan. *Cluster* 3 daerah Penduduk sedang 15 kecamatan (Preddy et al., 2023). Kemudian pada tahun 2022 Andrea Santana Adzani melakukan penelitian dan menyimpulkan bahwa tingkat akurasi perhitungan jarak antar data pada setiap klaster daerah rawan kecelakaan di Jember dalam sistem tersebut sebesar 83,87%. Kemudian berdasarkan hasil pengujian UAT dan blackbox yang telah dilakukan kepada 61 responden dan kepada pihak unit laka Satlantas Jember serta ahli IT, diperoleh hasil yang telah sesuai dengan target atau

hasil yang diharapkan. Andrea pada penelitian tersebut menggunakan metode K-Means dengan perhitungan jarak antar data menggunakan persamaan *euclidean distance*. (Andrea Santana Adzani, 2022). Kemudian pada tahun 2021 Nofita Safira Anggraini juga melakukan penelitian mendapatkan hasil yaitu. analisis menggunakan metode clustering k-means terdapat 3 cluster daerah rawan begal, yaitu hijau untuk daerah aman dengan nilai rata rata cluster 2.889, orange untuk tingkat keamanan sedang dengan nilai rata rata cluster 1.333, dan merah untuk daerah rawan dengan nilai rata rata cluster 0.270833333. Pada penelitian itu Nofita juga menggunakan metode K-Means dengan perhitungan jarak antar data *euclidean distance* (Anggraini, 2021).

Berdasarkan pernyataan dan data-data yang telah di paparkan, dapat disimpulkan bahwa Kabupaten Probolinggo perlu menerapkan pemetaan daerah rawan curas dan curanmor pada setiap kecamatan yang berbasis Sistem Informasi Geografis, agar lebih mudah dipahami daerah mana saja yang termasuk kategori rawan, sehingga para penduduk Kabupaten Probolinggo dan wisatawan bisa lebih berhati hati ketika melewati suatu daerah, dan hasil dari pemetaan tersebut bisa digunakan aparat kepolisian sebagai acuan dalam melakukan patroli atau pengamanan. Selain itu dari beberapa penelitian terdahulu yang telah disebutkan, banyak dari penelitian tersebut menggunakan metode *K-Means clustering* yang telah dinyatakan oleh (Andrea Santana Adzani, 2022) memiliki akurasi 87,86 % yang dapat dinyatakan baik. Namun untuk perhitungan jarak antar datanya menggunakan persamaan *manhattan*, karena menurut Dinata dalam (Alifah & Fauzan, 2023) menyatakan bahwa persamaan *manhattan* lebih akurat dari pada persamaan *euclidean distance*.. Jadi Sistem Informasi Geografis yang akan dikembangkan dapat menggunakan metode K – Means *Clustering*, namun dengan perhitungan jarak datanya menggunakan persamaan *manhattan*.

1.2 Rumusan Masalah

Dari latar belakang di atas, terdapat beberapa rumusan masalah dalam penelitian ini, sebagai berikut :

1. Bagaimana menerapkan algoritma *K - Means* dalam suatu website dan memvisualisasikan hasilnya ke dalam suatu peta Kabupaten Probolinggo ?
2. Bagaimana mendapatkan nilai *k* yang optimal untuk masing-masing data Curas dan Curanmor ?
3. Bagaimana hasil *clustering* untuk masing-masing data curas dan curanmor ?

1.3 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah di atas terdapat beberapa tujuan dalam penelitian ini, sebagai berikut:

1. Mampu menerapkan algoritma *K – Means Clustering* ke dalam suatu website dan bisa memvisualisasikan hasilnya ke dalam suatu peta Kabupaten Probolinggo.
2. Mengetahui bagaimana mendapatkan nilai *k* yang optimal untuk masing-masing data Curas dan Curanmor.
3. Mengetahui hasil *clustering* final untuk masing masing data curas dan curanmor.

1.4 Manfaat

Dengan adanya penelitian ini, diharapkan dapat memberikan manfaat kepada beberapa pihak, sebagai berikut :

1. Memberikan informasi bagi Kepolisian Resort Kabupaten Probolinggo tentang kecamatan yang berpotensi tinggi atau rendah untuk terjadi tindak kriminal pencurian kendaraan bermotor (curanmor) dan pencurian dengan kekerasan (curas) agar dapat dipertimbangkan untuk patroli rutin atau Upaya preventif lainnya.
2. Memberikan informasi bagi Masyarakat umum ataupun penduduk Kabupaten probolinggo tentang kecamatan yang memiliki Tingkat kerawanan tinggi untuk terjadi tindak kriminal pencurian kendaraan bermotor (curanmor) dan pencurian dengan kekerasan (curas) agar dapat lebih berhati hati ketika memasuki kecamatan tersebut.

1.5 Batasan Masalah

1. Data yang digunakan pada penelitian ini merupakan data kasus curas dan curanmor pada tahun 2024 yang didapatkan dari Kepolisian Resort Kabupaten Probolinggo.
2. Tindak kriminal yang digunakan dalam penelitian ini hanya pada kategori pencurian kendaraan bermotor (curanmor) dan pencurian dengan kekerasan (curas)

BAB 2. TINJAUAN PUSTAKA

2.1 State Of The Art

Tabel 2.1 State Of The Art

No.	Judul	Penulis	Tahun	Hasil
1	Pemetaan Kerentanan Tingkat Kriminalitas Menggunakan Metode <i>Self Organizing Map</i>	Ruziq Nawaf Zulfahmi , Maria Kristiana Daul , Muhammad Al Ayyubi, I Wayan Juliantha Pradnyana, Rokhana Dwi Bekti	2023	<p><i>Cluster</i> 1 terdiri dari 15 kabupaten/kota, dan <i>cluster</i> 2 terdiri dari 63 kabupaten/kota. Berdasarkan pengujian profiling cluster menggunakan nilai ratarata setiap cluster, Klaster 1 dapat dikatakan tinggi dibandingkan klaster 2 karena memiliki nilai dominan yang lebih besar untuk kategori dengan skor tinggi dari masingmasing variabel.</p>
2	Penerapan Data Mining Untuk Pengelompokan Kepadatan Penduduk Kabupaten Deli Serdang Menggunakan Algoritma K-Means	Preddy Marpaung , Ibnu Pebrian, Widia Putri	2023	<p>Dengan menerapkan algoritma K-Means, dapat mengelompokkan kepadatan penduduk ke dalam tigah kelompok (cluster), yaitu Cluster 1/ daerah Penduduk sangat padat sebanyak 3 kecamatan, Cluster 2/ daerah Penduduk padat terdapat 4 kecamatan. Cluster 3/ daerah Penduduk sedang 15 kecamatan. Dengan adanya</p>

				pengetahuan pengelompokan kepadatan penduduk ini, akan meminimalkan dampak akibat kepdenan penduduk bagi setiap orang yang mau bertempat tinggal di daerahdeli Serdang Dari hasil analisa peneliti, untuk kedepanya perlu dilakukan penelitian untuk pengelompokan kepadatan penduduk berdasarkan luas wilayah
3	Klastering Pemetaan Daerah Rawan Kecelakaan Menggunakan Metode K-Means Berbasis Sistem Informasi Geografis (Studi Kasus Di Kabupaten Jember)	Andrea Santana Adzani	2022	Tingkat akurasi perhitungan jarak antar data pada setiap klaster dalam sistem tersebut sebesar 83,87%. Kemudian berdasarkan hasil pengujian UAT dan blackbox yang telah dilakukan kepada 61 responden dan kepada pihak unit laka Satlantas Jember serta ahli IT, diperoleh hasil yang telah sesuai dengan target atau hasil yang diharapkan diawal sehingga sistem dapat dikatakan dapat dioperasikan dengan baik.

4	Pemetaan Daerah Rawan Pencurian Dengan Kekerasan (Begal) Di Kabupaten Lumajang	Nofita Safira Anggraini	2021	Dari hasil menggunakan clustering k-means terdapat 3 cluster yaitu Hijau untuk daerah aman dengan nilai rata rata cluster 2.889, Orange untuk tingkat keamanan sedang dengan nilai rata rata cluster 1.333, dan Merah untuk daerah rawan dengan nilai rata rata cluster 0.270833333.	analisis metode
---	--------------------------------------------------------------------------------	-------------------------	------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------

Dari beberapa penelitian yang telah dipaparkan pada *state of the art* di atas peneliti menyimpulkan bahwa penelitian ini sejalan dengan penelitian yang dilakukan oleh Nofita Safira Anggraini di tahun 2021, namun lebih melengkapi kategori tindak kriminalnya, tidak hanya terfokus pada kasus begal atau pencurian dengan kekerasan (curas) saja, melainkan menambahkan pada kasus pencurian kendaraan bermotor juga (curanmor). Selain itu peneliti juga menggunakan metode *K-Means Clustering* untuk pemetaannya, namun untuk persamaan pengukuran jarak antar data peneliti menggunakan persamaan *manhattan* guna mengetahui keakurasiannya.

2.2 Landasan Teori

2.2.1 Curanmor dan Curas

KUHP (Kitab Undang-undang Hukum Pidana) pada pasal 362 menyatakan bahwa perbuatan pelaku kejahatan dengan mengambil suatu barang berupa kendaraan bermotor yang seluruhnya atau sebagian kepunyaan orang lain dengan maksud untuk memiliki kendaraan bermotor tersebut secara melawan hukum. Perbuatan kejahatan dengan mengambil kendaraan bermotor atau yang sering disebut curanmor berdasarkan pasal tersebut termasuk ke dalam tindak pidana pencurian. Sama seperti pencurian dengan kekerasan atau yang sering

disebut curas merupakan bagian dari tindak pidana pencurian yang disertai kekerasan ataupun ancaman kekerasan. Peraturan terkait curas ini juga diatur pada KUHP (Kitab Undang-undang Hukum Pidana) pada pasal 365. Kasus curas sering dianggap lebih serius dan berbahaya daripada jenis kasus pencurian lainnya, karena sudah melibatkan ancaman fisik dan psikologis korban, bahkan juga mengancam nyawa.

Kasus pencurian secara umum, baik curas maupun curanmor terjadi tentunya disebabkan oleh faktor internal dan eksternal. Faktor internal yang mempengaruhi terjadinya curas dan curanmor disebabkan oleh kemiskinan, kesempatan kerja, karakter pelaku yang melakukan kasus kejahatan. Sedangkan faktor eksternal yang mempengaruhi terjadinya curas dan curanmor disebabkan oleh kepadatan penduduk, keadaan lingkungan, dan jumlah patroli polisi (Apriliana & Haris R, 2022).

2.2.2 Sistem Informasi Geografis (SIG)

Nugroho menjelaskan tentang pengertian dan tujuan Sistem Informasi geografis dalam (Umar, 2021) bahwa Geographic Information System atau Sistem Informasi Geografis (SIG) merupakan aplikasi pengolahan data spasial dengan menggunakan sistem terkomputerisasi dengan menggabungkan antara data grafis dengan data atribut objek menggunakan peta dasar digital (basic map) geoerensi bumi. SIG adalah sebuah sistem atau teknologi berbasis komputer yang dibangun dengan tujuan untuk mengumpulkan, menyimpan, mengolah dan menganalisa, serta menyajikan data informasi dari suatu objek atau fenomena yang berkaitan dengan letak atau keberadaannya di permukaan bumi yang bertujuan untuk membantu masyarakat mencari lokasi yang sedang dicari. SIG juga dapat diartikan sebagai sistem informasi yang digunakan untuk memasukkan, menyimpan, memanggil kembali, mengolah, menganalisis dan menghasilkan data bereferensi geografis atau data geospasial, untuk mendukung pengambilan keputusan dalam perencanaan dan pengelolaan penggunaan lahan, sumber daya alam, lingkungan transportasi, fasilitas kota, dan pelayanan umum lainnya. Teknologi SIG mengintegrasikan operasi-operasi umum database, seperti query, dan analisa statistic,

dengan kemampuan visualisasi dan analisa yang unik yang dimiliki oleh pemetaan. Kemampuan inilah yang membedakan SIG dengan Sistem Informasi lainnya (Umar, 2021).

Sistem Informasi Geografis (GIS) memiliki ciri ciri khusus seperti yang disampaikan oleh Susianto dan Guntoro dalam (Andrea Santana Adzani, 2022), sebagai berikut :

- a. Masukan data yang mampu memuat dan memproses data spasial dari berbagai sumber merupakan subsistem dari SIG, sub sistem ini juga mampu memproses perubahan data spasial yang memiliki perbedaan jenis, seperti dari peta kontur menjadi titik ketinggian.
- b. Subsistem SIG mampu menyimpan dan memanggil data yang memungkinkan data spasial untuk ditampilkan, diubah, dan dihapus.
- c. Subsistem lain yang dimiliki oleh SIG yakni mampu memanipulasi dan menganalisis peran data, pengelompokan dan pemisahan, perkiraan parameter dan hambatan, serta fungsi permodelan dari data yang dimuat.
- d. Pelaporan yang dimiliki subsistem SIG berbentuk peta, grafis, dan tabel.

Menurut Susianto dan Guntoro juga dalam (Andrea Santana Adzani, 2022) suatu sistem informasi geografis memiliki beberapa sub sistem, sebagai berikut :

a. *Data Input*

Data *input* berfungsi untuk mengumpulkan lalu mempersiapkan suatu data *spasial* beserta atributnya dari berbagai sumber. Subsistem ini juga bertanggung jawab dalam mengkonversi atau merepresentasikan format data yang asli ke dalam format SIG.

b. *Data Output*

Data *output* berfungsi untuk menampilkan atau menghasilkan luaran hasil dari suatu proses, baik dalam bentuk *softcopy* maupun *hardcopy* seperti: tabel, grafik, peta dan lain – lain.

c. *Data Management*

Data *management* berfungsi untuk memanajemen data, baik data *spasial* maupun atribut ke dalam penyimpanan seperti basis data dengan sedemikian rupa agar dipanggil dan diubah dengan mudah.

d. Data Manipulasi dan Analisis

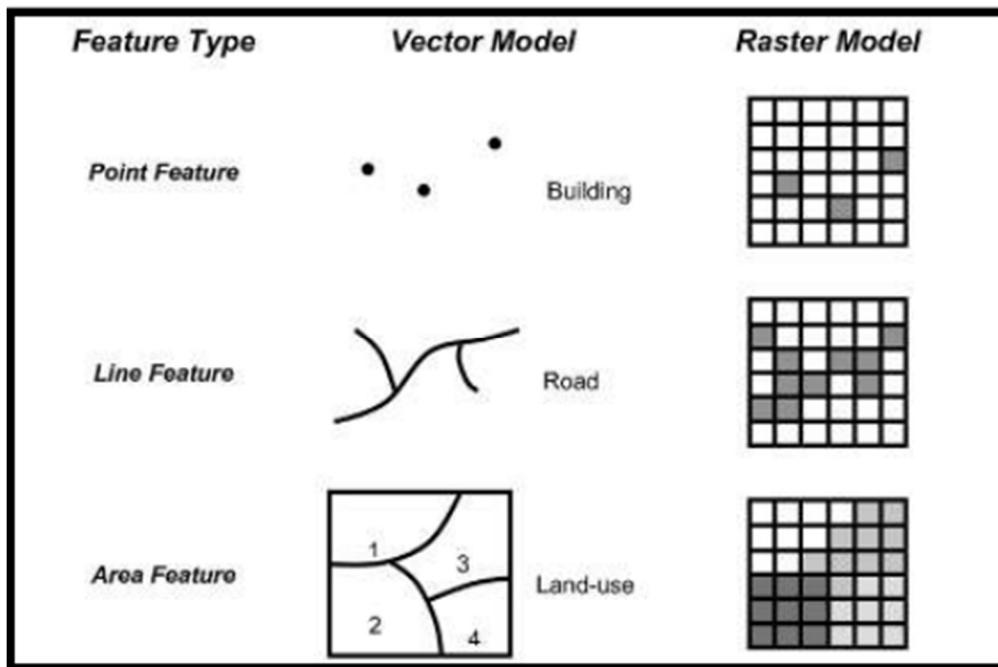
Data manipulasi dan analisis berfungsi untuk menentukan informasi mana saja yang dapat dihasilkan oleh SIG. Subsistem ini memiliki fungsi lain, yakni mampu memanipulasi dan melakukan permodelan data untuk menghasilkan luaran yang diharapkan.

Sistem Informasi Geografis tentunya membutuhkan data untuk diolah. Data yang diolah dan dihasilkan oleh suatu Sistem Informasi Geografis (SIG) terdapat dua model data, yaitu :

a. Data *Spasial*

Data *spasial* merupakan data yang memuat gambaran permukaan bumi. Model data ini dibagi menjadi dua, model data *raster* dan model data *vektor*,

- 1) Model data *raster* merupakan data yang sederhana, dimana setiap data atau informasi disimpan di *grid*, yang berupa bidang. *Grid* tersebut biasa disebut dengan *pixel*. Data tersebut merupakan hasil dari scanning seperti citra satelit digital
- 2) Model data *vektor* berupa simbol – simbol atau lebih dikenal dengan istilah *feature*, seperti *feature garis (line)*, *feature daerah (area)*, dan *feature titik (point)*.



Gambar 2.1 Gambaran Data Spasial

- b. Data *Atribut / Data Non Spasial* : merupakan data yang menyimpan suatu *atribut* dari gambaran yang ada di permukaan bumi.

2.2.3 K-Means Clustering

Menurut Fina dalam (Rahayu, 2022) Algoritma K-Means *clustering* merupakan suatu metode penganalisaan data atau metode data mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode K-Means berusaha mengelompokkan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain. Algoritma K-Means *Clustering* ini pertama kali diterbitkan pada tahun 1955 dan terus digunakan sampai sekarang (Preddy et al., 2023). Rahmat juga menjelaskan dalam (Andrea Santana Adzani, 2022) bahwa algoritma K – Means merupakan algoritma pengelompokan data berdasarkan titik pusat cluster (centroid) paling dekat dengan data. Tujuan K – Means adalah pengelompokan data yang memaksimalkan kesamaan data yang dikelompokkan

dan meminimalkan kesamaan data antara cluster. Persamaan fungsi jarak digunakan dalam cluster. Maksimalkan kesamaan data berdasarkan jarak terpendek antara data ke titik pusat. Sedangkan pengertian *clustering* yang dijelaskan oleh Rahmat dalam (Andrea Santana Adzani, 2022) menyatakan bahwa *clustering* merupakan proses dalam membagi data yang semulanya tidak berlabel menjadi sekumpulan data yang membentuk kelompok berdasarkan kemiripan yang dimiliki oleh data tersebut dengan data lainnya.

Dalam tahapan algoritma *K-Means Clustering* terdapat tahap menghitung jarak dari masing-masing data yang ada terhadap masing-masing pusat *cluster* (*centroid*). Ada beberapa persamaan yang digunakan seperti *euclidean distance* dan *manhattan*. Menurut Dinata dalam (Alifah & Fauzan, 2023) persamaan *manhattan* lebih akurat daripada persamaan *euclidean distance*.

Menurut Ramadhani dalam (Alifah & Fauzan, 2023) untuk menerapkan algoritma K-Means *Clustering* ada beberapa tahapan yang digambarkan pada Gambar 2.2.



Gambar 2.2 Tahapan Algoritma *K-Means Clustering*

Tahapan Algoritma *K-Means Clustering* berdasarkan Gambar 2.1 dapat dijelaskan sebagai berikut :

- a. Menentukan nilai k sebagai jumlah kluster yang akan dibentuk
- b. Tentukan titik pusat (*centroid*) awal secara acak dari setiap kluster yang telah ditentukan
- c. Hitung jarak dari setiap objek terhadap masing masing titik pusat (*centroid*) dari masing masing kluster dengan persamaan *Manhattan*. yang tertera pada persamaan (2.1) :

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \dots \dots \dots (2.1)$$

Keterangan :

d = jarak antar x dan y

x = data pada atribut

y = pusat klaster (*centroid*)

- d. Kelompokkan masing masing objek ke dalam titik pusat (*centroid*) terdekat
- e. Lakukan iterasi, kemudian hitung pusat cluster (*centroid*) terbaru, dengan persamaan (2.2) :

$$v = \frac{\sum_{i=1}^n x_i}{n}; 1, 2, 3, \dots, n \dots \dots \dots (2.2)$$

- f. Ulangi langkah c – e , hingga data tidak berpindah lagi ke *cluster* yang lain.

2.2.4 Metode *Elbow*

Metode *Elbow* menurut Jollyta dalam (Riani et al., 2023) merupakan suatu teknik yang digunakan untuk menentukan jumlah klaster (c) yang optimal dengan menghitung nilai *Sum of Square Error* (SSE) untuk setiap klaster. Semakin besar perbedaan nilai SSE antara klaster yang satu dengan klaster berikutnya, yang membentuk sebuah titik sudut siku, maka jumlah klaster yang dipilih dianggap semakin baik. Sejalan dengan Jollyta, (Maori & Evanita, 2023) menyatakan bahwa penggunaan metode elbow untuk menentukan nilai k yang tegolong kecil, tetapi punya nilai withinss atau nilai kedekatan data pada setiap klaster yang rendah juga. Selain itu, metode elbow agar lebih mudah dipahami titik siku yang terbentuk biasanya di ditampilkan dalam suatu grafik. Sumbu x pada grafik tersebut

merupakan rentang nilai k yang di uji. Sedangkan untuk sumbu y merupakan nilai SSE untuk setiap nilai k. Untuk menghitung nilai *Sum of Square Error* (SSE) bisa menggunakan persamaan *CEK

$$SSE = \sum_{i=1}^k \sum_{x \in S_k} ||x - \mu_i||^2$$

Keterangan :

k : jumlah klaster

C_i : klaster ke-i

x : data yang termasuk dalam klaster C_i

μ_i : pusat (centroid) dari klaster C_i

$||x - \mu_i||^2$: kuadrat jarak antara data xxx dengan centroid-nya

2.2.5 User Acceptance Testing (UAT)

User Acceptance Testing (UAT) merupakan salah satu pengujian sistem yang berfokus menguji interaksi antara *user* atau pengguna dengan sistem secara langsung yang berfungsi untuk memverifikasi bahwa fitur telah berjalan sesuai dengan kebutuhan user tersebut (Rumariana & Arifin, 2022). Hady menyatakan dalam (Andrea Santana Adzani, 2022) bahwa *User Acceptance Testing* (UAT) memiliki tujuan untuk mengetahui, apakah sistem yang telah dirancang telah memenuhi harapan pengguna, sehingga dapat mempermudah peneliti untuk mengetahui, bagian mana yang masih dirasa kurang. Output dari pengujian *User Acceptance Testing* (UAT) yaitu dokumen hasil uji *software* dengan nilai yang baik dan sudah memenuhi kebutuhan yang diminta atau memenuhi kriteria (*acceptance criteria*) yang dibutuhkan pengguna.

BAB 3. METODE PENELITIAN

3.1 Tempat dan Waktu Pelaksanaan

Penelitian dengan judul “Implementasi K-Means Clustering Dalam Pemetaan Daerah Rawan Curanmor dan Curas Berbasis Sistem Informasi Geografis (Studi Kasus di Kabupaten Probolinggo) “ dilaksanakan di Politek Negeri Jember, Kepolisian Resort Probolinggo, Badan Pusat Statistik Kabupaten Probolinggo. Waktu yang dibutuhkan untuk penelitian ini 10 bulan, dimulai dari bulan Juni 2024 sampai bulan Mei. Detai waktu yang diperlukan pada penelitian ini digambarkan pada table berikut.:

Tabel 3.1 Tahapan Kegiatan

Kegiatan	2024						2025			
	7	8	9	10	11	12	1	2	3	4
Studi Literatur										
Pengumpulan Data										
Pengolahan Data										
Pengembangan Sistem										
Pengujian										
Analisis dan pembahasan										

3.2 Alat dan Bahan

3.2.1 Alat Penelitian

Pada penelitian ini menggunakan alat penelitian berupa perangkat keras dan perangkat lunak, sebagai berikut :

- a. Perangkat Keras
 - 1) Laptop Asus AMD E2
 - 2) Smartphone Oppo F9
 - 3) WiFi

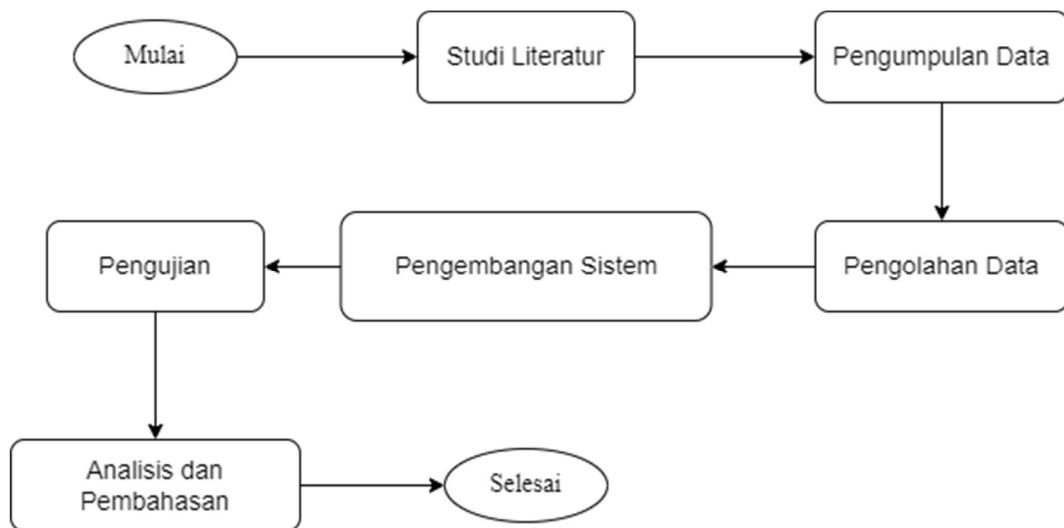
- b. Perangkat Lunak
 - 1) OS Windows 10
 - 2) *Visual Studio Code*
 - 3) *MySQL*
 - 4) *Framework Laravel*
 - 5) *Library JS (Leaflet)*
 - 6) Microsoft Office

3.2.2 Bahan Penelitian

Bahan yang digunakan pada penelitian ini adalah data pencurian kendaraan bermotor (curanmor) dan pencurian dengan kekerasan (curas) pada Kabupaten Probolinggo pada tahun 2024. Data tersebut didapat Kepolisian Resort Kabupaten Probolinggo

3.3 Tahapan Penelitian

Pada tahapan penelitian ini terdapat beberapa tahapan yang digambarkan pada gambar 3.1.



Gambar 3.1 Tahapan Penelitian

Pada gambar 3.1 gambar dari tahapan pada penelitian ini yang dilakukan oleh penulis. Tahapan penelitian mencangkup studi literatur, pengumpulan data,

pengolahan data, pengembangan sistem, pengujian, analisis dan pembahasan. Penjelasan lebih detail tentang masing-masing tahapan penelitian sebagai berikut :

3.3.1 Studi Literatur

Studi literatur dilakukan agar peneliti memperoleh referensi maupun teori yang relevan dan sesuai dengan topik penelitian. Referensi yang telah penulis pelajari, antara lain :

- a. Jurnal atau paper yang berkaitan dengan algoritma *K-Means Clustering*
- b. Jurnal atau paper yang berkaitan dengan *elbow methods*
- c. Jurnal atau paper yang berkaitan dengan sistem informasi geografis
- d. Data sekunder yang diperoleh dari BPS Kabupaten Probolinggo

3.3.2 Pengumpulan Data

Untuk mendukung berjalannya penelitian ini, diperlukan data terkait kasus pencurian dengan kekerasan (curas) dan pencurian kendaraan bermotor (curanmor) pada tahun 2024. Pengumpulan data curas dan curanmor pada penelitian ini dilakukan dengan mengajukan permohonan untuk mendapatkan data dari Kepolisian Resort Kabupaten Probolinggo. Pengajuan untuk mendapatkan data yang dilakukan kepada Polres Kabupaten Probolinggo diawali dengan mengirimkan Surat Ijin Survei dan Pengambilan Data, yang terlampir pada Lampiran *CEK.

3.3.3 Pengolahan Data

Dalam tahapan ini menjelaskan bagaimana alur atau tahapan data sekunder mentah yang didapat dari publikasi Badan Pusat Statistik (BPS) Kabupaten Probolinggo diolah sebelum memasuki penerapan algoritma *K-Means Clustering* hingga setelah penerapan algoritma *K-Means Clustering* yang menghasilkan 3 *cluster* dengan masing-masing anggotanya, Berikut ini tahapan pengolahan data pada penelitian ini :

- a. Menghitung jumlah banyaknya data
- b. Menentukan nilai k sebagai jumlah kluster yang akan dibentuk

- c. Tentukan titik pusat (*centroid*) awal secara acak dari setiap kluster yang telah ditentukan
- d. Hitung jarak dari setiap objek terhadap masing masing titik pusat (*centroid*) dari masing masing kluster dengan persamaan *Manhattan*. yang tertera pada persamaan (3.1) :

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \dots \dots \dots (3.1)$$

Keterangan :

d = jarak antar x dan y

x = data pada pusat kluster

y = data pada atribut

- e. Kelompokkan masing masing objek ke dalam titik pusat (*centroid*) terdekat
- f. Lakukan iterasi, kemudian hitung pusat cluster (*centroid*) terbaru, dengan persamaan (3.2) :

$$v = \frac{\sum_{i=1}^n x_i}{n}; 1, 2, 3, \dots \dots \dots \dots \dots (3.2)$$

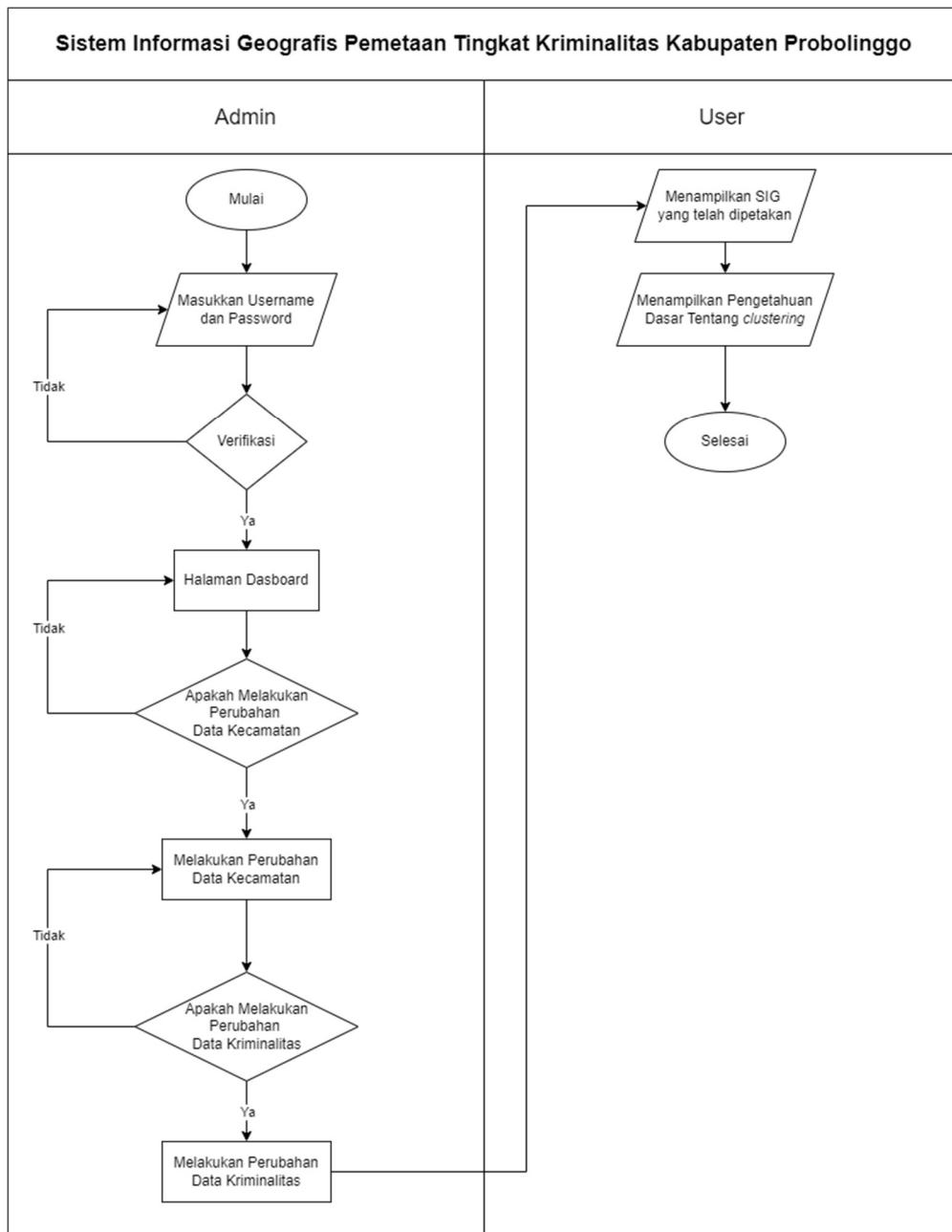
- g. Ulangi langkah f – h, hingga data tidak berpindah lagi ke *cluster* yang lain

3.3.4 Pengembangan Sistem

Pada tahap ini peneliti melakukan pengembangan Sistem Informasi Geografis (SIG) yang di dalamnya terdapat pemetaan daerah rawan curas dan curanmor dengan mengimplementasikan metode K-Means *clustering* dan pengukuran jarak antar datanya menggunakan persamaan *manhattan*. Sistem Informasi Geografis ini outputnya akan berbasis *website*, jadi secara umum bahasa pemrograman yang digunakan yaitu php dengan menggunakan juga *framework laravel*. Perhitungan K – Means yang digunakan untuk memetakan tingkat kerawanan suatu kecamatan akan di implementasikan menggunakan bahasa php dengan *function* tersendiri. Kemudian, hasil pemetaan yang telah diperoleh dari perhitungan K – Means akan ditampilkan dalam warna di setiap kecamatan menggunakan salah satu *library JS* yaitu *leaflet*.

Rancangan tentang sistem yang akan dikembangkan dalam penelitian ini digambarkan dalam beberapa diagram sebagai berikut :

a. Flowchart Sistem

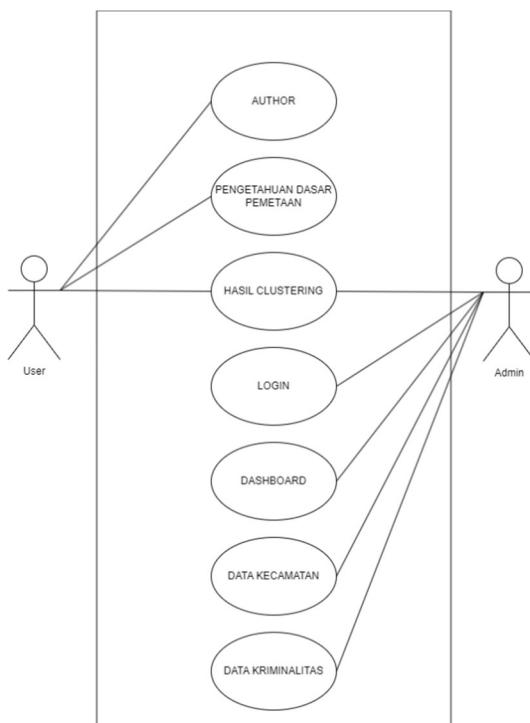


Gambar 3.2 Flowchart Sistem

Pada gambar 3.2 di atas dijelaskan bahwa admin harus masuk atau *login* ke sistem dengan memasukkan username dan password dengan benar. Setelah berhasil masuk ke sistem, admin dapat melakukan perubahan atau input data kecamatan dan juga melakukan perubahan data atau input data kriminalitas (curas dan curanmor)

pada setiap kecamatan. Dari dua data utama tersebut, sistem akan memproses klusterisasi daerah dengan tingkat kerawannya. Setelah proses *clustering* selesai, maka hasil *clustering* tersebut akan di visualisasikan dalam bentuk peta Kabupaten Probolinggo dengan warna di setiap kecamatannya sebagai kategori atas tingkat kerawanan kasus curas dan curanmor pada kecamatan tersebut. Peta Hasil *clustering* tersebut dapat dilihat secara langsung oleh user, tanpa harus *login* pada sistem.

b. Use Case Sistem



Gambar 3.3 Use Case Sistem

Pada gambar 3.3 dijelaskan terdapat dua pengguna dalam system yang akan dikembangkan. Dua pengguna tersebut merupakan admin dan user yang memiliki hak akses yang berbeda beda dalam system. Admin mempunyai akses pada halaman dashboard, halaman data kecamatan, dan halaman data kriminalitas melalui fitur login terlebih dahulu.. Dapat dinyatakan bahwa admin mempunyai akses yang lebih tinggi terhadap system, karena dengan memiliki akses ke halaman data kecamatan, dan halaman data kriminalitas, maka dapat juga mengubah data data tersebut yang

dapat mempengaruhi hasil *clustering* yang terjadi. Kemudian untuk *user* memiliki hak akses pada halaman hasil *clustering*, halaman pengetahuan dasar pemetaan, dan halaman author, maka dapat dinyatakan bahwa *user* hanya dapat melihat data yang ditampilkan dan tidak dapat merubahnya.

3.3.5 Pengujian

Pengujian adalah proses uji coba sistem, Tujuannya yaitu untuk memastikan kesesuaian perangkat lunak dengan rancangan yang telah ditetapkan serta untuk mengevaluasi kinerja fungsionalitas sistem agar dapat menentukan apakah semuanya berjalan dengan baik atau tidak. Proses pengujian dalam penelitian ini menggunakan dua metode yaitu:

a. *User Acceptance Testing* (UAT)

User Acceptance Testing (UAT) merupakan salah satu pengujian sistem yang berfokus menguji interaksi antara *user* atau pengguna dengan sistem secara langsung yang berfungsi untuk memverifikasi bahwa fitur telah berjalan sesuai dengan kebutuhan *user* tersebut (Rumariana & Arifin, 2022). Hady menyatakan dalam (Andrea Santana Adzani, 2022) bahwa *User Acceptance Testing* (UAT) memiliki tujuan untuk mengetahui, apakah sistem yang telah dirancang telah memenuhi harapan pengguna, sehingga dapat mempermudah peneliti untuk mengetahui, bagian mana yang masih dirasa kurang. Output dari pengujian *User Acceptance Testing* (UAT) yaitu dokumen hasil uji *software* dengan nilai yang baik dan sudah memenuhi kebutuhan yang diminta atau memenuhi kriteria (*acceptance criteria*) yang dibutuhkan pengguna.

b. Pengujian Metode *Elbow*

Metode *Elbow* menurut Jollyta dalam (Riani et al., 2023) merupakan suatu teknik yang digunakan untuk menentukan jumlah klaster (*c*) yang optimal dengan menghitung nilai *Sum of Square Error* (SSE) untuk setiap klaster. Semakin besar perbedaan nilai SSE antara klaster yang satu dengan klaster berikutnya, yang membentuk sebuah titik sudut siku, maka jumlah klaster yang dipilih dianggap semakin baik. Sejalan dengan Jollyta, (Maori & Evanita, 2023) menyatakan bahwa penggunaan metode elbow untuk menentukan nilai *k* yang tegolong kecil, tetapi

punya nilai withinss atau nilai kedekatan data pada setiap klaster yang rendah juga. Selain itu, metode elbow agar lebih mudah dipahami titik siku yang terbentuk biasanya di ditampilkan dalam suatu grafik. Sumbu x pada grafik tersebut merupakan rentang nilai k yang di uji. Sedangkan untuk sumbu y merupakan nilai SSE untuk setiap nilai k. Untuk menghitung nilai *Sum of Square Error* (SSE) bisa menggunakan persamaan *CEK

$$SSE = \sum_{i=1}^k \sum_{x \in S_k} ||x - \mu_i||^2$$

Keterangan :

k : jumlah klaster

C_i : klaster ke-i

x : data yang termasuk dalam klaster C_i

μ_i : pusat (centroid) dari klaster C_i

$||x - \mu_i||^2$: kuadrat jarak antara data xxx dengan centroid-nya

3.3.6 Analisis dan Pembahasan

Pada tahap penelitian ini, akan memberikan sebuah penjelasan akhir dari pembahasan pada tahap tahap sebelumnya. Kemudian pada tahap ini juga dilakukan analisis terhadap hasil *clustering* akhir dan juga hasil pengujian tingkat akurasi algoritma K – Means dalam penerapannya pada sistem informasi geografis tingkat kerawanan kasus curas dan curanmor di wilayah hukum Polres Kabupaten Probolinggo.

BAB 4. HASIL DAN PEMBAHASAN

4.1 Studi Literatur

Studi literatur yang dilakukan oleh penulis meliputi beberapa referensi berupa jurnal ilmiah yang relevan dengan masalah pada penelitian ini. Berikut beberapa referensi yang digunakan pada penelitian ini

- a. Jurnal ilmiah yang ditulis (Preddy et al., 2023) dengan judul Penerapan Data Mining Untuk Pengelompokan Kepadatan Penduduk Kabupaten Deli Serdang Menggunakan Algoritma K-Means. Pada penelitian tersebut membahas bagaimana algoritma k-means clustering bisa digunakan untuk mengelompokan kepadatan penduduk menggunakan algoritma *k-means*
- b. Penelitian dengan judul Klastering Pemetaan Daerah Rawan Kecelakaan Menggunakan Metode K-Means Berbasis Sistem Informasi Geografis (Studi Kasus Di Kabupaten Jember) yang ditulis oleh (Andrea Santana Adzani, 2022). Pada penelitian tersebut juga membahas penerapan algoritma k-means untuk mengelompokkan daerah rawan kecelakaan dan divisualisasikan ke dalam peta
- c. Penelitian yang berjudul Pemetaan Daerah Rawan Pencurian Dengan Kekerasan (Begal) Di Kabupaten Lumajang yang ditulis oleh (Anggraini, 2021). Pada penelitian tersebut juga membahas penerapan algoritma k-means untuk kasus curas

4.2 Pengumpulan Data

Sesuai dengan perencanaan, penelitian ini menggunakan data pencurian dengan kekerasan (curas) dan pencurian kendaraan bermotor (curanmor) di tahun 2024 yang didapatkan dari Kepolisian Resort Kabupaten Probolinggo. Proses pengajuan untuk mendapatkan data Kepada Polres Kabupaten Probolinggo dilakukan selama 6 bulan yang terhitung dari tanggal 25 Juni 2024 hingga 31 Januari 2025. Proses pengambilan data dari Polres Kabupaten Probolinggo dilampirkan pada Lampiran *CEK

Data yang didapatkan dari Polres Kabupaten Probolinggo berupa data curas dan curanmor pada masing-masing kecamatan di wilayah Kabupaten Probolinggo. Data yang dikumpulkan terlihat pada gambar *CEK

Kecamatan	2024		Total Kasus
	Curas	Curanmor	Per - Kecamatan
Bantaran	0	5	5
Banyuanyar	0	4	4
Besuk	0	2	2
Dringu	0	22	22
Gading	1	4	5
Gending	1	18	19
Kotaanyar	0	0	0
Kraksaan	0	37	37
Krenjengan	0	9	9
Krucil	0	3	3
Kuripan	0	2	2
Leces	0	13	13
Lumbang	0	1	1
Maron	0	21	21
Paiton	0	14	14
Pakuniran	0	4	4
Pajarakan	0	10	10
Sukapura	0	0	0
Sumber	0	1	1
Sumberasih	1	10	11
Tegalsiwalan	0	1	1
Tiris	0	2	2
Tongas	3	15	18
Wonomerto	2	4	6
Total	8	202	210

Pada tabel *CEK merupakan data kasus curas dan curanmor yang terjadi pada setiap kecamatan di Kabupaten Probolinggo sepanjang tahun 2024. Untuk total kasus curas (pencurian dengan kekerasan) yang terjadi di seluruh wilayah Kabupaten Probolinggo pada tahun 2024 sebanyak 8 kasus. Kecamatan yang memiliki kasus curas tertinggi yaitu Kecamatan Tongas dengan 3 kasus curas. Kemudian, untuk total kasus curanmor (pencurian kendaraaan bermotor) yang terjadi di seluruh wilayah Kabupaten Probolinggo pada tahun 2024 sebanyak 202 kasus. Kecamatan yang memiliki kasus curanmor tertinggi pada tahun 2024 yaitu Kecamatan Kraksaan dengan 37 kasus.

4.3 Pengolahan Data

Tahap pengolahan data pada penelitian ini, menjelaskan bagaimana implementasi algoritma *k-means* agar bisa mengolah data mentah, yang berupa data curas dan curanmor dari Polres, agar bisa memberikan kategori untuk setiap kecamatan berdasarkan kasus curas dan curanmor yang terjadi pada kecamatan tersebut. Pengolahan data yang dilakukan, tentunya mengikuti tahapan-tahapan dari algoritma *k-means*. Dalam hal ini penulis memaparkan proses perhitungan algoritma *k-means* dengan nilai $k=3$ untuk data curas dan curanmor. Alasan penentuan nilai $k = 3$, karena nilai tersebut merupakan nilai yang optimal untuk data curas dan curanmor, berdasarkan pengujian yang telah dilakukan pada poin 4.5. Berikut tahapan dari perhitungan algoritma *k-means* dengan nilai $k = 3$ untuk masing-masing data curas dan curanmor :

4.3.1 Menghitung Jumlah Data

Pada tahap ini dilakukan perhitungan jumlah data yang didapatkan dari Polres. Jumlah data ini perlu dihitung karena juga sebagai dasar untuk menentukan nilai k atau jumlah klaster nanti. Pada data yang didapat dari Polres, menjelaskan data curas dan curanmor dari masing-masing kecamatan, dan ada 24 kecamatan di wilayah Kabupaten Probolinggo. Jadi dapat dinyatakan jumlah data untuk masing-masing kasus curas dan curanmor ada 24 data.

4.3.2 Menentukan nilai k

Dalam menentukan nilai k atau jumlah klaster yang akan dibentuk, perlu memperhatikan jumlah data yang ada. Karena penentuan nilai k harus lebih kecil

atau sama dengan jumlah data. Dalam studi kasus ini penulis menentukan nilai $k = 3$ untuk kedua data curas dan curanmor, dikarenakan nilai tersebut merupakan nilai k yang paling optimal berdasarkan hasil pengujian pada poin 4.5.

4.3.3 Menentukan Centroid Awal Pada Setiap Klaster

Setelah mengetahui atau menentukan nilai k yang optimal untuk masing-masing data curas dan curanmor, selanjutnya perlu menentukan nilai centroid awal untuk setiap klaster. Pada penelitian ini nilai k yang telah ditetapkan yaitu 3, baik untuk data curas maupun data curanmor, sehingga perlu 3 nilai centroid untuk masing-masing data. Dalam menentukan nilai centroid awal, tidak ada aturan yang mengatur bagaimana pemilihan centroid awal, atau dengan kata lain, penentuan nilai centroid awal dipilih secara acak. Berikut nilai centroid awal yang telah dipilih secara acak untuk masing-masing data.

a. Centroid awal untuk data curas

Centroid	Nilai Centroid
C1	0
C2	1
C3	3

Sesuai pada tabel *CEK, nilai centroid awal yang ditetapkan untuk data curas Kabupaten Probolinggo yaitu C1 = 0, C2 = 1, dan untuk C3 = 3.

b. Centroid awal untuk data curanmor

Centroid	Nilai Centroid
C1	10
C2	20
C3	30

Sesuai pada tabel *CEK, nilai centroid awal yang ditetapkan untuk data curanmor Kabupaten Probolinggo yaitu C1 = 0, C2 = 1, dan untuk C3 = 3.

4.3.4 Menghitung Jarak Setiap Data Terhadap Centroid

Setelah menentukan nilai centroid awal pada setiap klaster, sesuai dengan tahapan algoritma k-means langkah selanjutnya yaitu menghitung jarak dari masing masing data terhadap masing-masing centroid. Pada tahap perhitungan jarak ini

dilakukan berulang kali hingga tidak ada perubahan pada anggota dari masing-masing klaster pada iterasi selanjutnya. Perhitungan jarak berulang tersebut disebut sebagai iterasi. Dalam perhitungan jarak ini digunakan persamaan *Euclidean Distance*, yang tertuang pada persamaan nomor 4.1

$$d(K_i, C_i) = \sqrt{(x_i - C_i)^2} \dots \dots \dots (4.1)$$

Keterangan :

$d(K_i, C_i)$: Jarak data curas pada kecamatan i terhadap centroid ke i

K_i : Data curas atau curanmor pada kecamatan ke i

C_i : Centroid ke i

Jadi pada persamaan *euclidean distance* penulis menghitung jarak antar data pada masing-masing kecamatan di setiap kasus terhadap masing-masing centroid pada setiap kasusnya.

d. Perhitungan Jarak Data Kasus Curas

Dalam iterasi pertama ini penulis menggunakan data curas pada kecamatan Bantaran sebagai contoh perhitungan jarak data terhadap setiap centroid. Pada Kecamatan Bantaran terjadi kasus curas sebanyak 0 kasus.

$$\begin{aligned} d(K_1, C_1) &= \sqrt{(0 - 0)^2} \\ &= 0 \end{aligned}$$

$$\begin{aligned} d(K_1, C_2) &= \sqrt{(0 - 1)^2} \\ &= 1 \end{aligned}$$

$$\begin{aligned} d(K_1, C_3) &= \sqrt{(0 - 3)^2} \\ &= 3 \end{aligned}$$

Pada iterasi pertama ini seluruh data curas pada setiap kecamatan dihitung jaraknya terhadap masing-masing nilai centroid awal. Setelah setiap data curas dihitung jaraknya, kemudian dicari nilai minimum dari jarak data terhadap masing-masing centroid yang ada. Jarak minimum tersebut merupakan acuan dalam menentukan data tersebut masuk dalam klaster apa. Semakin kecil jarak data terhadap salah satu centroid pada suatu klaster, maka data tersebut merupakan

anggota dari klaster tersebut. Hasil perhitungan jarak data curas terhadap masing-masing centroid bisa dilihat pada tabel *CEK

Kecamatan	C1	C2	C3	Klaster
Bantaran	0,000	1,000	3,000	C1
Banyuanyar	0,000	1,000	3,000	C1
Besuk	0,000	1,000	3,000	C1
Dringu	0,000	1,000	3,000	C1
Gading	1,000	0,000	2,000	C2
Gending	1,000	0,000	2,000	C2
Kotaanyar	0,000	1,000	3,000	C1
Kraksaan	0,000	1,000	3,000	C1
Krenjengan	0,000	1,000	3,000	C1
Krucil	0,000	1,000	3,000	C1
Kuripan	0,000	1,000	3,000	C1
Leces	0,000	1,000	3,000	C1
Lumbang	0,000	1,000	3,000	C1
Maron	0,000	1,000	3,000	C1
Paiton	0,000	1,000	3,000	C1
Pakuniran	0,000	1,000	3,000	C1
Pajarakan	0,000	1,000	3,000	C1
Sukapura	0,000	1,000	3,000	C1
Sumber	0,000	1,000	3,000	C1
Sumberasih	1,000	0,000	2,000	C2
Tegalsiwalan	0,000	1,000	3,000	C1
Tiris	0,000	1,000	3,000	C1
Tongas	3,000	2,000	0,000	C3
Wonomerto	2,000	1,000	1,000	C2

Pada iterasi pertama proses klasterisasi data curas, jumlah data yang masuk ke masing-masing klaster adalah : C1 sebanyak 19 data, C2 terdiri dari 4 data, dan untuk C3 berjumlah 1 data. Meskipun demikian, jumlah ini masih dapat berubah pada iterasi berikutnya karena proses *optimasi* akan memperbarui posisi *centroid*,

sehingga pembagian data ke dalam klaster bisa mengalami penyesuaian berdasarkan perhitungan jarak yang terbaru. Untuk mempermudah dalam mengetahui hasil *clustering* pada iterasi pertama, dapat dilihat pada tabel *CEK

C1	C2	C3
Bantaran	Gading	Tongas
Banyuanyar	Gending	
Besuk	Sumberasih	
Dringu	Wonomerto	
Kotaanyar		
Kraksaan		
Krenjengan		
Krucil		
Kuripan		
Leces		
Lumbang		
Maron		
Paiton		
Pakuniran		
Pajarakan		
Sukapura		
Sumber		
Tegalsiwalan		
Tiris		

Hasil *clustering* pada tabel *CEK akan menjadi acuan dalam perhitungan *centroid* untuk iterasi selanjutnya.

e. Perhitungan Jarak Data Kasus Curanmor

Dalam iterasi pertama ini penulis menggunakan data curanmor pada kecamatan Bantaran sebagai contoh perhitungan jarak data terhadap setiap centroid. Pada Kecamatan Bantaran terjadi kasus curanmor sebanyak 0 kasus.

$$d(K1, C1) = \sqrt{(5 - 10)^2}$$

$$= 5$$

$$d(K1, C2) = \sqrt{(5 - 20)^2}$$

$$= 15$$

$$d(K1, C2) = \sqrt{(5 - 30)^2}$$

$$= 25$$

Pada iterasi pertama ini seluruh data curanmor pada setiap kecamatan dihitung jaraknya terhadap masing-masing nilai centroid awal. Setelah setiap data curanmor dihitung jaraknya, kemudian dicari nilai minimum dari jarak data terhadap masing-masing centroid yang ada. Jarak minimum tersebut merupakan acuan dalam menentukan data tersebut masuk dalam klaster apa. Semakin kecil jarak data terhadap salah satu centroid pada suatu klaster, maka data tersebut merupakan anggota dari klaster tersebut. Hasil perhitungan jarak data curanmor terhadap masing-masing centroid bisa dilihat pada tabel *CEK

Kecamatan	C1	C2	C3	Klaster
Bantaran	5,000	15,000	25,000	C1
Banyuanyar	6,000	16,000	26,000	C1
Besuk	8,000	18,000	28,000	C1
Dringu	12,000	2,000	8,000	C2
Gading	6,000	16,000	26,000	C1
Gending	8,000	2,000	12,000	C2
Kotaanyar	10,000	20,000	30,000	C1
Kraksaan	27,000	17,000	7,000	C3
Krenjengan	1,000	11,000	21,000	C1
Krucil	7,000	17,000	27,000	C1
Kuripan	8,000	18,000	28,000	C1
Leces	3,000	7,000	17,000	C1
Lumbang	9,000	19,000	29,000	C1
Maron	11,000	1,000	9,000	C2

Paiton	4,000	6,000	16,000	C1
Pakuniran	6,000	16,000	26,000	C1
Pajarakan	0,000	10,000	20,000	C1
Sukapura	10,000	20,000	30,000	C1
Sumber	9,000	19,000	29,000	C1
Sumberasih	0,000	10,000	20,000	C1
Tegalsiwalan	9,000	19,000	29,000	C1
Tiris	8,000	18,000	28,000	C1
Tongas	5,000	5,000	15,000	C1
Wonomerto	6,000	16,000	26,000	C1

Pada iterasi pertama proses klasterisasi data curanmo, jumlah data yang masuk ke masing-masing klaster adalah : C1 sebanyak 20 data, C2 terdiri dari 3 data, dan untuk C3 berjumlah 1 data. Meskipun demikian, jumlah ini masih dapat berubah pada iterasi berikutnya karena proses *optimasi* akan memperbarui posisi *centroid*, sehingga pembagian data ke dalam klaster bisa mengalami penyesuaian berdasarkan perhitungan jarak yang terbaru. Untuk mempermudah dalam mengetahui hasil *clustering* pada iterasi pertama, dapat dilihat pada tabel *CEK

C1	C2	C3
Bantaran	Dringu	Kraksaan
Banyuanyar	Gending	
Besuk	Maron	
Gading		
Kotaanyar		
Krenjengan		
Krucil		
Kuripan		
Leces		
Lumbang		

Paiton

Pakuniran

Pajarakan

Sukapura

Sumber

Sumberasih

Tegalsiwalan

Tiris

Tongas

Wonomerto

Hasil *clustering* pada tabel *CEK akan menjadi acuan dalam perhitungan *centroid* untuk iterasi selanjutnya.

4.3.5 Menentukan Nilai Centroid Baru

Guna mengetahui apakah hasil *clustering* dari iterasi pertama sudah benar, perlu dilakukan iterasi kedua. Iterasi kedua digunakan untuk memvalidasi, apakah anggota klaster yang telah dihasilkan oleh iterasi pertama masih sama pada iterasi kedua ataukah ada data yang berubah klasternya. Proses dalam perhitungan iterasi kedua sama dengan iterasi pertama. Namun centroidnya tidak menggunakan nilai acak lagi, melainkan berpedoman pada hasil *clustering* dari iterasi pertama. Proses menentukan centroid yang baru dengan menghitung rata-rata data pada setiap klaster yang sudah terbentuk di iterasi pertama atau dapat dirumuskan dalam persamaan *CEK

$$C_i \text{ baru} = \frac{\sum_{i=1}^n xci}{nci} \dots \dots \dots$$

Keterangan :

$C_i \text{ baru}$: Nilai Centroid pada klaster ke-i yang baru.

$\sum_{i=1}^n xci$: Jumlah seluruh data pada anggota klaster ke-i

nci : Jumlah anggota klaster ke-i

Berikut merupakan perhitungan centroid baru dari masing-masing kasus curas dan curanmor.

a. Perhitungan Centroid Baru Data Curas

$$C1 \text{ baru} = \frac{0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0}{19} \\ = 0$$

$$C2 \text{ baru} = \frac{1+1+1+2}{4} = 1,25$$

$$C3 \text{ baru} = \frac{3}{1} = 3$$

Dari perhitungan tersebut sudah menghasilkan nilai centroid baru yang berbeda dari centroid acak di awal. Hasil perhitungan yang berupa centroid baru dapat dilihat pada tabel *CEK

Centroid	Nilai Centroid
C1	0
C2	1,25
C3	3

Centroid baru untuk kasus curas pada tabel*CEK akan digunakan untuk perhitungan jarak data pada iterasi kedua.

b. Perhitungan Centroid Baru Data Curanmor

Selanjutnya perlu juga menghitung centroid baru untuk kasus curanmor dengan berdasarkan hasil *clustering* dari ietrasasi pertama.

$C1 \text{ baru}$

$$= \frac{5 + 4 + 2 + 4 + 0 + 9 + 3 + 2 + 13 + 1 + 14 + 4 + 10 + 0 + 1 + 10 + 1 + 2 + 15 + 4}{20} \\ = 5,2$$

$$C1 \text{ baru} = \frac{22 + 18 + 21}{3} = 20,333$$

$$C1 \text{ baru} = \frac{37}{1} = 37$$

Dari perhitungan tersebut sudah menghasilkan nilai centroid baru yang berbeda dari centroid acak di awal. Hasil perhitungan yang berupa centroid baru dapat dilihat pada tabel *CEK

Centroid	Nilai Centroid
C1	5,2
C2	20,333
C3	37

Centroid baru untuk kasus curanmor pada tabel*CEK akan digunakan untuk perhitungan jarak data pada iterasi kedua.

4.3.6 Melakukan Iterasi Selanjutnya

Setelah mendapatkan centroid baru dari hasil iterasi pertama, langkah selanjutnya melakukan iterasi kedua dengan menggunakan centroid baru tersebut. Pada iterasi kedua juga mengelompokkan data ke klaster berdasarkan jarak terdekat dari data tersebut ke masing-masing centroid.

Setelah mengetahui anggota klaster yang baru dari iterasi kedua, dilakukan pengecekan apakah anggota klaster iterasi kedua sama dengan anggota kaster dari iterasi pertama tadi. Jika tidak ada perubahan anggota klaster dari iterasi pertama ke iterasi kedua, maka proses *clustering* sudah bisa dianggap konvergen, sehingga hasil akhir dari iterasi kedua merupakan hasil akhir dari proses *K-Means Clustering*.

Namun ketika ada perbedaan anggota klaster antara hasil iterasi pertama dan iterasi kedua, maka proses iterasi akan dilanjutkan. Proses tersebut mengulangi tahap penentuan centroid baru dari iterasi sebelumnya, perhitungan jarak untuk iterasi selanjutnya, dan pengecekan anggota klaster dari iterasi sebelumnya dengan iterasi selanjutnya hingga mencapai konvergen atau tidak ada perubahan anggota klaster pada dua iterasi terakhir.

4.3.7 Hasil Akhir K-Means Clustering

Setelah melanjutkan iterasi kedua dan seterusnya untuk masing-masing data kasus, proses iterasi *k-means clustering* berhenti sesuai ketentuan ketika tidak ada perubahan anggota klaster lagi. Iterasi terakhir pada masing-masing kasus berbeda. Berikut iterasi terakhir dan juga hasil final dari *k-means clustering* pada masing-masing kasus.

- a. Iterasi Terakhir dan Hasil Final *Clustering* Kasus Curas

Proses iterasi pada kasus curas berhenti pada iterasi kedua, karena tidak ada lagi perubahan anggota klaster antara iterasi pertama ke iterasi kedua. Selain itu untuk centroid terakhirnya merupakan centroid yang dihasilkan dari hasil klasterisasi pada iterasi pertama. Centroid terakhir yang digunakan dan juga perhitungan jarak di iterasi kedua (iterasi terakhir) dicantumkan pada tabel *CEK

Centroid	Nilai Centroid
C1	0
C2	1,25
C3	3

Nilai centroid terakhir pada tabel *CEK merupakan hasil perhitungan dari iterasi sebelumnya (iterasi pertama). Nilai centroid tersebut yang kemudian digunakan untuk menghitung jarak pada iterasi selanjutnya (iterasi kedua) pada tabel *CEK.

Kecamatan	C1	C2	C3	Klaster
Bantarhan	0,000	1,250	3,000	C1
Banyuanyar	0,000	1,250	3,000	C1
Besuk	0,000	1,250	3,000	C1
Dringu	0,000	1,250	3,000	C1
Gading	1,000	0,250	2,000	C2
Gending	1,000	0,250	2,000	C2
Kotaanyar	0,000	1,250	3,000	C1
Kraksaan	0,000	1,250	3,000	C1
Krenjengan	0,000	1,250	3,000	C1
Krucil	0,000	1,250	3,000	C1
Kuripan	0,000	1,250	3,000	C1
Leces	0,000	1,250	3,000	C1
Lumbang	0,000	1,250	3,000	C1

Maron	0,000	1,250	3,000	C1
Paiton	0,000	1,250	3,000	C1
Pakuniran	0,000	1,250	3,000	C1
Pajarakan	0,000	1,250	3,000	C1
Sukapura	0,000	1,250	3,000	C1
Sumber	0,000	1,250	3,000	C1
Sumberasih	1,000	0,250	2,000	C2
Tegalsiwalan	0,000	1,250	3,000	C1
Tiris	0,000	1,250	3,000	C1
Tongas	3,000	1,750	0,000	C3
Wonomerto	2,000	0,750	1,000	C2

Pada tabel *CEK merupakan perhitungan jarak atau iterasi terakhir dalam proses *k-means clustering* karena sudah tidak ada perubahan anggota klaster antara iterasi sebelumnya (iterasi pertama) dengan iterasi selanjutnya (iterasi kedua). Hasil *clustering* dari iterasi kedua atau bisa disebut hasil akhir proses *k-means clustering* dapat dijelaskan pada tabel *CEK

C1	C2	C3
Bantaran	Gading	Tongas
Banyuanyar	Gending	
Besuk	Sumberasih	
Dringu	Wonomerto	
Kotaanyar		
Kraksaan		
Krenjengan		
Krucil		
Kuripan		
Leces		
Lumbang		
Maron		

Paiton
Pakuniran
Pajarakan
Sukapura
Sumber
Tegalsiwalan
Tiris

Untuk mempermudah pembaca dalam memahami hasil akhir clustering atau untuk mempermudah mengetahui kecamatan mana yang termasuk kategori rawan., maka hasil akhir clustering pada tabel *CEK di ubah nama kategorinya yang sebelumnya C1, C2, C3 menjadi Rawan, Sedang, dan Tidak Rawan.

Nama klaster yang baru yaitu Rawan, Sedang, dan Tidak Rawan memiliki karakteristik masing-masing. Nama klaster rawan memiliki arti bahwa kecamatan pada klaster tersebut sering terjadi kasus curas dari pada kecamatan di klaster yang lain. Sedangkan klaster aman memiliki karakteristik kecamatan yang masuk pada klaster ini memiliki kejadian curas lebih sedikit dari kecamatan pada klaster lain. Kemudian untuk klaster sedang, memiliki karakteristik kecamatan yang masuk menjadi anggotanya merupakan kecamatan yang kejadian curasnya tidak lebih banyak kasus curasnya daripada kecamatan yang ada di klaster rawan dan tidak lebih sedikit dari kecamatan yang ada di klaster aman, atau dengan kata lain klaster yang dikategorikan sedang merupakan klaster yang letaknya berada di tengah-tengah dalam diagram persebaran klaster.

Dalam perubahan nama kategori tersebut perlu memperhatikan karakteristik dari masing-masing kategori dengan data yang masuk dalam kategori tersebut. Jadi penamaan kategori yang baru mengikuti centroid terakhir yang digunakan, karena centroid merupakan cerminan dari data yang ada dalam klaster tersebut. Dalam studi kasus ini klaster C3 bisa berubah nama menjadi klaster rawan, karena nilai centroidnya lebih tinggi dari klaster lainnya. Kemudian untuk klaster C1 bisa berubah nama menjadi klaster aman, karena nilai centroidnya lebih kecil dari klaster lainnya. Selanjutnya untuk C2 bisa berubah menjadi klaster sedang, karena

nilai centroidnya tidak lebih besari dari C3 dan tidak lebih kecil dari C1. Jadi hasil *clustering* menggunakan nama klaster yang baru dapat dilihat pada tabel *CEK

Aman	Sedang	Rawan
Bantaran	Gading	Tongas
Banyuanyar	Gending	
Besuk	Sumberasih	
Dringu	Wonomerto	
Kotaanyar		
Kraksaan		
Krenjengan		
Krucil		
Kuripan		
Leces		
Lumbang		
Maron		
Paiton		
Pakuniran		
Pajarakan		
Sukapura		
Sumber		
Tegalsiwalan		
Tiris		

Dari tabel *CEK dapat disimpulkan bahwa klaster rawan hanya memiliki 1 anggota yaitu kecamatan Tongas. Untuk klaster sedang memiliki 4 anggota, yaitu kecamatan Gading, Gending, Sumberasih, dan Wonomerto. Sedangkan untuk klaster aman memiliki 19 anggota yang terdiri dari Kecamatan Bantaran, Banyuanyar, Besuk, Dringu, Kotaanyar, Kraksaan, Krenjengan, Krucil, Kuripan, Leces Lumbang Maron, Paiton, Pakuniran, Pajarakan, Sukapura, Sumber, Tegalsiwalan, dan Tiris.

b. Iterasi Terakhir dan Hasil Final *Clustering* Kasus Curanmor

Proses iterasi pada kasus curanmor berhenti pada iterasi ketiga, karena tidak ada lagi perubahan anggota klaster antara iterasi kedua ke iterasi ketiga. Selain itu untuk centroid terakhirnya merupakan centroid yang dihasilkan dari hasil klasterisasi pada iterasi kedua. Centroid terakhir yang digunakan dan juga perhitungan jarak di iterasi ketiga (iterasi terakhir) dicantumkan pada tabel *CEK

Centroid	Nilai Centroid
C1	3,647
C2	17,167
C3	37,000

Nilai centroid terakhir pada tabel *CEK merupakan hasil perhitungan dari iterasi sebelumnya (iterasi kedua). Nilai centroid tersebut yang kemudian digunakan untuk menghitung jarak pada iterasi selanjutnya (iterasi ketiga) pada tabel *CEK.

Kecamatan	C1	C2	C3	Klaster
Bantaran	1,353	12,167	32,000	C1
Banyuanyar	0,353	13,167	33,000	C1
Besuk	1,647	15,167	35,000	C1
Dringu	18,353	4,833	15,000	C2
Gading	0,353	13,167	33,000	C1
Gending	14,353	0,833	19,000	C2
Kotaanyar	3,647	17,167	37,000	C1
Kraksaan	33,353	19,833	0,000	C3
Krenjengan	5,353	8,167	28,000	C1
Krucil	0,647	14,167	34,000	C1
Kuripan	1,647	15,167	35,000	C1
Leces	9,353	4,167	24,000	C2
Lumbang	2,647	16,167	36,000	C1
Maron	17,353	3,833	16,000	C2

Paiton	10,353	3,167	23,000	C2
Pakuniran	0,353	13,167	33,000	C1
Pajarakan	6,353	7,167	27,000	C1
Sukapura	3,647	17,167	37,000	C1
Sumber	2,647	16,167	36,000	C1
Sumberasih	6,353	7,167	27,000	C1
Tegalsiwalan	2,647	16,167	36,000	C1
Tiris	1,647	15,167	35,000	C1
Tongas	11,353	2,167	22,000	C2
Wonomerto	0,353	13,167	33,000	C1

Pada tabel *CEK merupakan perhitungan jarak atau iterasi terakhir dalam proses *k-means clustering* untuk kasus curanmor, karena sudah tidak ada perubahan anggota klaster antara iterasi sebelumnya (iterasi kedua) dengan iterasi saat ini (iterasi ketiga). Hasil *clustering* dari iterasi ketiga atau bisa disebut hasil akhir proses *k-means clustering* dapat dijelaskan pada tabel *CEK

C1	C2	C3
Bantaran	Dringu	Kraksaan
Banyuanyar	Gending	
Besuk	Leces	
Gading	Maron	
Kotaanyar	Paiton	
Krenjengan	Tongas	
Krucil		
Kuripan		
Lumbang		
Pakuniran		
Pajarakan		
Sukapura		

Sumber

Sumberasih

Tegalsiwalan

Tiris

Wonomerto

Untuk mempermudah pembaca dalam memahami hasil akhir *clustering* kasus curanmor atau untuk mempermudah mengetahui kecamatan mana yang termasuk kategori rawan., maka hasil akhir clustering pada tabel *CEK di ubah nama kategorinya yang sebelumnya C1, C2, C3 menjadi Rawan, Sedang, dan Tidak Rawan.

Nama klaster yang baru yaitu Rawan, Sedang, dan Tidak Rawan memiliki karakteristik masing-masing. Nama klaster rawan memiliki arti bahwa kecamatan pada klaster tersebut sering terjadi kasus curanmor dari pada kecamatan di klaster yang lain. Sedangkan klaster aman memiliki karakteristik kecamatan yang masuk pada klaster ini memiliki kejadian curanmor lebih sedikit dari kecamatan pada klaster lain. Kemudian untuk klaster sedang, memiliki karakteristik kecamatan yang masuk menjadi anggotanya merupakan kecamatan yang kasus curanmornya tidak lebih banyak daripada kecamatan yang ada di klaster rawan dan juga tidak lebih sedikit dari kecamatan yang ada di klaster aman, atau dengan kata lain, klaster yang dikategorikan sedang merupakan klaster yang letaknya berada di tengah-tengah dalam diagram persebaran klaster.

Dalam perubahan nama kategori tersebut perlu memperhatikan karakteristik dari masing-masing kategori dengan data yang masuk dalam kategori tersebut. Jadi penamaan kategori yang baru mengikuti centroid terakhir yang digunakan, karena centroid merupakan cerminan dari data yang ada dalam klaster tersebut. Dalam studi kasus ini klaster C3 bisa berubah nama menjadi klaster rawan, karena nilai centroidnya lebih tinggi dari klaster lainnya. Kemudian untuk klaster C1 bisa berubah nama menjadi klaster aman, karena nilai centroidnya lebih kecil dari klaster lainnya. Selanjutnya untuk C2 bisa berubah menjadi klaster sedang, karena

nilai centroidnya tidak lebih besari dari C3 dan tidak lebih kecil dari C1. Hasil *clustering* menggunakan nama klaster yang baru dapat dilihat pada tabel *CEK

Aman	Sedang	Rawan
Bantaran	Dringu	Kraksaan
Banyuanyar	Gending	
Besuk	Leces	
Gading	Maron	
Kotaanyar	Paiton	
Krenjengan	Tongas	
Krucil		
Kuripan		
Lumbang		
Pakuniran		
Pajarakan		
Sukapura		
Sumber		
Sumberasih		
Tegalsiwalan		
Tiris		
Wonomerto		

Dari tabel *CEK dapat disimpulkan bahwa klaster rawan untuk kasus curanmor hanya memiliki 1 anggota yaitu kecamatan Kraksaan. Untuk klaster sedang pada kasus curanmor memiliki 6 anggota, yaitu kecamatan Dringu, Gending, Leces, Maron, Paiton, dan Tongas. Sedangkan untuk klaster aman memiliki 17 anggota yang terdiri dari Kecamatan Bantaran, Banyuanyar, Besuk, Gading, Kotaanyar, Krenjengan, Krucil, Kuripan, Lumbang, Pakuniran, Pajarakan, Sukapura, Sumber, Sumberasih, Tegalsiwalan, Tiris, dan Wonomerto.

4.4 Pengembangan Sistem

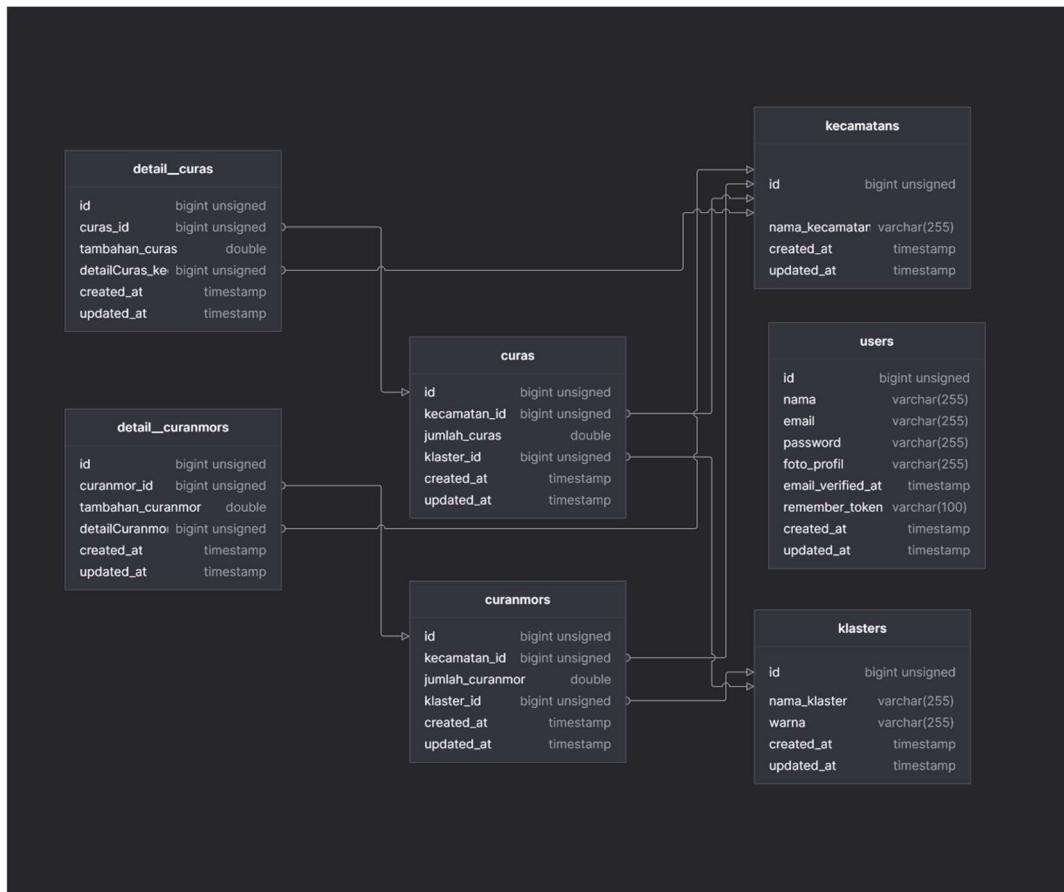
Pada tahapan pengembangan sistem, penulis mulai membuat produk untuk penelitian ini. Produk yang dibuat oleh penulis untuk penelitian ini berupa web sistem informasi geografis yang menampilkan pemetaan dari kasus curas dan juga

kasus curanmor pada wilayah Kabupaten Probolinggo. Dalam web tersebut, juga terdapat implementasi dari algoritma k-means clustering untuk melakukan klasterisasi pada data curas dan data curanmor yang ada agar data tersebut lebih mudah untuk bisa dipahami, karena sudah terbagi atas beberapa klaster. Sesuai dengan poin 4.3.7 Hasil Akhir *K-Means Clustering*, yang menyatakan bahwa hasil klasterisasi terdiri dari 3 klaster, yaitu rawan, sedang, dan aman.

Dalam melakukan pengembangan sistem, penulis berpedoman dengan flowchart di gambar *CEK dan *use case diagram* di gambar *CEK, sehingga sistem yang dikembangkan sesuai dengan perencanaan. Hasil dari pengembangan sistem ini, penulis jelaskan dalam beberapa bagian utama sebagai berikut.

4.4.1 Database Sistem

Pembuatan *database* dilakukan berdasarkan analisis yang telah dirancang pada tahap sebelumnya. Penulis merujuk pada *flowchart* dan diagram *use case* yang telah disusun pada Gambar *CEK dan Gambar *CEK sebagai acuan utama dalam merancang struktur *database*. *Flowchart* menggambarkan alur proses sistem secara keseluruhan, sedangkan *use case* menunjukkan interaksi antara aktor dan sistem. Informasi dari kedua diagram tersebut membantu penulis dalam mengidentifikasi kebutuhan data serta relasi antar entitas yang dibutuhkan dalam sistem. Setiap entitas dan atribut yang terdapat pada *database* dirancang agar sesuai dengan kebutuhan fungsional yang telah diuraikan pada *flowchart* dan *use case*. Selain itu, penulis juga memperhatikan normalisasi agar struktur *database* efisien dan terhindar dari redundansi. Hasil dari perancangan ini kemudian dibuat menjadi suatu *database* yang memiliki relasi antar tabelnya. *Database* sistem yang dibuat dapat dilihat pada Gambar *CEK



Pada Gambar *CEK yang merupakan struktur database yang digunakan pada Web SIG. Pada database tersebut terdiri dari tujuh tabel dengan 3 tabel diantaranya merupakan tabel master. Tiga tabel yang dimaksud dalam tabel master yaitu, tabel users yang menyimpan data pengguna, tabel klasters yang menyimpan data klaster, dan tabel kecamatans, yang menyimpan data kecamatan. Pada tabel tersebut juga menggunakan teori normalisasi 2nf, yang terbukti dengan adanya tabel detail_curas dan detail_curanmors yang menyimpan tambahan data kasus curas dan curanmor pada setiap tanggalnya. Penjelasan lebih rinci terkait masing-masing tabel pada database akan diuraikan sebagai berikut.

a. Tabel Users

Name	users	Primary	id		
#	column_name	data_type	character_set	collation	is_nullable
1	<i>id</i>	bigint unsigned	▼ NULL	NULL	NO
2	<i>nama</i>	varchar(255)	▼ utf8mb4	utf8mb4_unicode_ci	NO
3	<i>email</i>	varchar(255)	▼ utf8mb4	utf8mb4_unicode_ci	NO
4	<i>password</i>	varchar(255)	▼ utf8mb4	utf8mb4_unicode_ci	NO
5	<i>foto_profil</i>	varchar(255)	▼ utf8mb4	utf8mb4_unicode_ci	YES
6	<i>email_verified_at</i>	timestamp	▼ NULL	NULL	YES
7	<i>remember_token</i>	varchar(100)	▼ utf8mb4	utf8mb4_unicode_ci	YES
8	<i>created_at</i>	timestamp	▼ NULL	NULL	YES
9	<i>updated_at</i>	timestamp	▼ NULL	NULL	YES

Penulis melampirkan struktur tabel *users* pada Gambar *CEK untuk memberikan gambaran rinci mengenai desain tabel yang digunakan dalam sistem. Tabel *users* memiliki sembilan *field*, yaitu *id*, *nama*, *email*, *password*, *foto_profil*, *email_verified_at*, *remember_token*, *created_at*, dan *updated_at*. *Field id* berperan sebagai *primary key* yang digunakan untuk mengidentifikasi setiap data secara unik. Selain itu, *field id* dan *email* masing-masing diberi aturan *unique* agar tidak terjadi duplikasi data pengguna. *Field password* menyimpan kata sandi yang telah dienkripsi guna menjaga keamanan akun admin. *Field email_verified_at* berfungsi mencatat waktu ketika alamat surel berhasil diverifikasi oleh sistem. Sementara itu, *field remember_token*, *created_at*, dan *updated_at* digunakan dalam proses autentikasi serta pelacakan waktu pembuatan dan pembaruan data pengguna dalam sistem. Tabel *users* ini berfungsi untuk menyimpan data akun admin yang memiliki wewenang dalam mengelola dan melakukan perubahan terhadap data yang tersedia di dalam aplikasi web.

b. Tabel Kecamatans

Name	kecamatans	Primary	id		
#	column_name	data_type	character_set	collation	is_nullable
1	<i>id</i>	bigint unsigned	▼ NULL	NULL	NO
2	<i>nama_kecamatan</i>	varchar(255)	▼ utf8mb4	utf8mb4_unicode_ci	NO
3	<i>created_at</i>	timestamp	▼ NULL	NULL	YES
4	<i>updated_at</i>	timestamp	▼ NULL	NULL	YES

Penulis melampirkan struktur tabel *kecamatans* pada Gambar *CEK sebagai bagian dari dokumentasi perancangan *database*. Tabel ini terdiri atas empat *field*, yaitu *id*, *nama_kecamatan*, *created_at*, dan *updated_at*. *Field id* berperan sebagai *primary key* yang berfungsi untuk mengidentifikasi setiap data kecamatan secara unik. Tabel *kecamatans* digunakan untuk menyimpan data seluruh kecamatan yang

berada di wilayah Kabupaten Probolinggo. *Field nama_kecamatan* berisi nama-nama kecamatan, sedangkan *created_at* dan *updated_at* mencatat waktu pembuatan serta pembaruan data pada tabel tersebut. Dengan adanya tabel ini, sistem dapat mengelola data wilayah secara terstruktur dan efisien.

c. Tabel Klasters

Name	klasters	Primary	id		
#	column_name	data_type	character_set	collation	is_nullable
1	<i>id</i>	bigint unsigned	NULL	NULL	NO
2	<i>nama_klaster</i>	varchar(255)	utf8mb4	utf8mb4_unicode_ci	NO
3	<i>warna</i>	varchar(255)	utf8mb4	utf8mb4_unicode_ci	NO
4	<i>created_at</i>	timestamp	NULL	NULL	YES
5	<i>updated_at</i>	timestamp	NULL	NULL	YES

Struktur tabel *klasters* ditampilkan pada Gambar *CEK sebagai bagian dari penjabaran rancangan tabel yang diterapkan dalam sistem. Tabel ini terdiri atas lima *field*, yaitu *id*, *nama_klaster*, *warna*, *created_at*, dan *updated_at*. *Field id* berperan sebagai *primary key* yang berfungsi untuk membedakan setiap entri klaster secara unik. Tabel *klasters* dibuat untuk menyimpan informasi klaster yang menjadi dasar penamaan klaster hasil proses *K-Means*. *Field nama_klaster* digunakan untuk mencatat nama dari masing-masing klaster, sedangkan *field warna* berfungsi sebagai indikator visual dalam proses pemetaan. Adapun *field created_at* dan *updated_at* mencatat waktu saat data klaster dibuat dan diperbarui. Tabel ini memegang peran penting dalam mendukung penyajian hasil klasterisasi secara visual, terutama pada tampilan peta dalam antarmuka sistem.

d. Tabel Curas

Name	curas	Primary	id		
#	column_name	data_type	character_set	collation	is_nullable
1	<i>id</i>	bigint unsigned	NULL	NULL	NO
2	<i>kecamatan_id</i>	bigint unsigned	NULL	NULL	NO
3	<i>jumlah_curas</i>	double	NULL	NULL	NO
4	<i>klaster_id</i>	bigint unsigned	NULL	NULL	YES
5	<i>created_at</i>	timestamp	NULL	NULL	YES
6	<i>updated_at</i>	timestamp	NULL	NULL	YES

Penulis melampirkan struktur tabel *curas* pada Gambar *CEK sebagai bagian dari dokumentasi perancangan *database* dalam sistem. Tabel ini memiliki enam *field*, yaitu *id*, *klaster_id*, *kecamatan_id*, *jumlah_curas*, *created_at*, dan

updated_at. Field *id* berfungsi sebagai *primary key* yang digunakan untuk mengidentifikasi setiap entri data kasus curas secara unik. Sementara itu, *klaster_id* merupakan *foreign key* yang menghubungkan data pada tabel *curas* dengan tabel *klasters*, dengan relasi satu klaster dapat memiliki banyak data kasus curas (*one to many*). Di sisi lain, *kecamatan_id* juga berperan sebagai *foreign key* yang menghubungkan tabel *curas* dengan tabel *kecamatans*, dengan relasi satu kecamatan hanya memiliki satu entri kasus curas (*one to one*).

Tabel *curas* dirancang khusus untuk menyimpan data mengenai jumlah tindak pidana pencurian dengan kekerasan yang terjadi di setiap kecamatan. Data ini menjadi komponen utama dalam proses pengelompokan menggunakan metode *K-Means clustering*. Melalui tabel ini, sistem dapat mengelola dan menganalisis pola sebaran kasus curas berdasarkan wilayah serta klaster yang terbentuk. Informasi yang tersimpan juga memungkinkan visualisasi data secara akurat pada antarmuka pengguna, seperti peta dan tabel analisis. Dengan struktur dan relasi yang telah dirancang, tabel *curas* berperan penting dalam mendukung sistem pengambilan keputusan berbasis data spasial.

e. Tabel Detail_Curas

Name	detail_curas	Primary	id			
#	column_name	data_type	character_set	collation	is_nullable	
1	<i>id</i>	bigint unsigned	▼ NULL	NULL	NO	
2	<i>curas_id</i>	bigint unsigned	▼ NULL	NULL	NO	
3	<i>tambahan_curas</i>	double	▼ NULL	NULL	NO	
4	<i>detailCuras_kecamatan_Id</i>	bigint unsigned	▼ NULL	NULL	NO	
5	<i>created_at</i>	timestamp	▼ NULL	NULL	YES	
6	<i>updated_at</i>	timestamp	▼ NULL	NULL	YES	

Penulis melampirkan struktur tabel *detail_curas* pada Gambar *CEK sebagai bagian dari perancangan sistem untuk mencatat rincian pembaruan data kasus curas. Tabel ini memiliki enam *field*, yaitu *id*, *curas_id*, *tambahan_curas*, *detailCuras_kecamatan_Id*, *created_at*, dan *updated_at*. Field *id* berperan sebagai *primary key* yang mengidentifikasi setiap entri secara unik. Field *curas_id* merupakan *foreign key* yang merepresentasikan relasi satu data pada tabel *curas* dapat memiliki banyak data pada tabel *detail_curas* (*one to many*). Selain itu, field *detailCuras_kecamatan_Id* juga berperan sebagai *foreign key* yang menunjukkan bahwa satu kecamatan dapat memiliki banyak entri pada tabel ini (*one to many*).

Tabel *detail_curas* dirancang untuk menyimpan data pembaruan jumlah kasus curas yang terjadi di kecamatan tertentu pada waktu atau tanggal tertentu, sehingga riwayat perkembangan kasus dapat tercatat dan dianalisis secara kronologis.

f. Tabel Curanmors

Name	curanmors	Primary	id		
#	column_name	data_type	character_set	collation	is_nullable
1	<i>id</i>	bigint unsigned	✓	NULL	NO
2	<i>kecamatan_id</i>	bigint unsigned	✓	NULL	NO
3	<i>jumlah_curanmor</i>	double	✓	NULL	NO
4	<i>klaster_id</i>	bigint unsigned	✓	NULL	YES
5	<i>created_at</i>	timestamp	✓	NULL	YES
6	<i>updated_at</i>	timestamp	✓	NULL	YES

Gambar 4.8 menampilkan struktur tabel *curanmors* yang disertakan penulis sebagai bagian dari rancangan *database* sistem. Tabel ini memiliki enam *field*, yakni *id*, *klaster_id*, *kecamatan_id*, *jumlah_curanmor*, *created_at*, dan *updated_at*. *Field id* berperan sebagai *primary key* yang berfungsi untuk membedakan setiap entri data secara unik. Adapun *klaster_id* merupakan *foreign key* yang menunjukkan keterkaitan antara tabel *curanmors* dan *klasters*, di mana satu klaster dapat memiliki banyak data kasus curanmor (*one to many*). Di sisi lain, *kecamatan_id* juga berperan sebagai *foreign key* yang menggambarkan bahwa satu kecamatan hanya memiliki satu data curanmor (*one to one*), sehingga satu baris data merepresentasikan satu kecamatan.

Tabel *curanmors* difungsikan untuk merekam jumlah kasus pencurian kendaraan bermotor di setiap kecamatan di wilayah Kabupaten Probolinggo. Informasi yang dihimpun dalam tabel ini menjadi komponen penting dalam pelaksanaan metode *K-Means clustering*, yang digunakan untuk mengelompokkan wilayah berdasarkan tingkat kerawanan. Melalui data tersebut, sistem dapat mengungkap pola penyebaran kasus curanmor dan menghasilkan visualisasi dalam bentuk peta atau tampilan analisis lainnya. Dengan desain struktur dan hubungan antar tabel yang terencana, keberadaan tabel *curanmors* sangat vital dalam menunjang analisis spasial serta pengambilan keputusan berbasis data.

g. Tabel Detail_Curanmors

Name	detail_curanmors	Primary	id			
#	column_name		data_type	character_set	collation	is_nullable
1	<i>id</i>		bigint unsigned	✓ NULL	NULL	NO
2	<i>curanmor_id</i>		bigint unsigned	✓ NULL	NULL	NO
3	<i>tambahan_curanmor</i>		double	✓ NULL	NULL	NO
4	<i>detailCuranmor_kecamatan_Id</i>		bigint unsigned	✓ NULL	NULL	NO
5	<i>created_at</i>		timestamp	✓ NULL	NULL	YES
6	<i>updated_at</i>		timestamp	✓ NULL	NULL	YES

Gambar 4.7 menampilkan struktur tabel *detail_curanmor* yang dilampirkan oleh penulis sebagai bagian dari dokumentasi perancangan sistem. Tabel ini memuat enam *field*, yakni *id*, *curanmor_id*, *tambahan_curanmor*, *detailCuranmor_kecamatan_id*, *created_at*, dan *updated_at*. *Field id* berperan sebagai *primary key* yang berfungsi untuk membedakan setiap data secara unik. Sementara itu, *field curanmor_id* merupakan *foreign key* yang menjalin relasi dengan tabel *curanmor*, di mana satu entri pada tabel *curanmor* dapat memiliki banyak catatan pada tabel *detail_curanmor* (*one to many*). Selain itu, *field detailCuranmor_kecamatan_id* juga bertindak sebagai *foreign key* yang menghubungkan data kecamatan, memungkinkan satu kecamatan memiliki banyak entri pembaruan kasus curanmor (*one to many*). Tabel ini digunakan untuk mencatat perkembangan data kasus pencurian kendaraan bermotor yang terjadi di masing-masing kecamatan berdasarkan tanggal tertentu. Dengan demikian, informasi peningkatan jumlah kasus dapat direkam dan dianalisis secara terstruktur dari waktu ke waktu.

4.4.2 Penerapan Algoritma K-Means

Sesuai dengan tujuan penelitian ini, perlu menerapkan algoritma *k-means clustering* kepada produk web SIG yang dibuat. Karena pada pengembangan sistem ini menggunakan *framework Laravel*, maka penulisan *code* dari algoritma *k-means clustering* diletakan pada folder *Service* agar bisa digunakan pada beberapa fitur lainnya. Algoritma *k-means clustering* akan dijalankan ketika terdapat perubahan data, terutama perubahan pada data di tabel curas, tabel kecamatans, dan tabel klasters. Dalam penelitian ini penulis mengimplementasikan algoritma *k-means clustering* menjadi dua bagian, yaitu *k-means* untuk *clustering* kasus curas dan *k-means* untuk *clustering* kasus curanmor. Namun pada tahapan ini penulis akan

menampilkan *code* untuk penerapan algoritma *k-means* pada kasus curas. Penjelasan code akan dibagi dalam beberapa bagian sebagai berikut.

a. Persiapan data



```

1
2 $data = Curas::select('id', 'kecamatan_id', 'klaster_id', 'jumlah_curas')->orderBy('jumlah_curas', 'asc')->get();
3 $k = Klaster::count('id');
4 $maxIterasi = 100;

```

Kode ini bertujuan untuk mempersiapkan data yang akan digunakan dalam proses pengelompokan (klasterisasi) kasus pencurian dengan kekerasan (curas) menggunakan metode K-Means. Data yang diperlukan mencakup informasi tentang jumlah kasus curas di setiap kecamatan, serta klaster yang akan digunakan untuk mengelompokkan data berdasarkan tingkat kerawanan. Jumlah klaster yang akan dibentuk ditentukan berdasarkan data yang tersedia, dan proses klasterisasi ini dibatasi hingga 100 iterasi untuk memastikan efisiensi perhitungan. Dengan persiapan data ini, proses klasterisasi dapat dilakukan untuk mengidentifikasi pola kerawanan yang terjadi di berbagai kecamatan.

b. Validasi Data Unik



```

1 $uniqueCount = $data->unique('jumlah_curas')->count();
2 if ($uniqueCount < $k) {
3     throw new \Exception("Jumlah nilai unik pada 'jumlah_curas' ($uniqueCount) kurang dari jumlah klaster ($k). Pastikan data memiliki variasi yang cukup.");
4 }

```

Kode ini berfungsi untuk memeriksa apakah jumlah nilai unik dari data *jumlah_curas* cukup untuk membentuk jumlah klaster yang telah ditentukan. Jika jumlah nilai unik yang ada pada *jumlah_curas* lebih sedikit daripada jumlah klaster yang diinginkan, maka akan muncul pengecualian (error) yang memberi tahu bahwa data tidak memiliki variasi yang cukup untuk membentuk klaster sesuai jumlah yang diinginkan. Hal ini penting agar proses klasterisasi dapat berjalan dengan baik, mengingat setiap klaster membutuhkan data yang cukup berbeda untuk menghasilkan kelompok yang valid.

c. Inisiasi Centroid Awal



```

1 $centroids = $data->unique('jumlah_curas')
2             ->shuffle()
3             ->take($k)
4             ->values()
5             ->map(function ($item) {
6                 return [
7                     'jumlah_curas' => $item->jumlah_curas,
8                 ];
9             });
10            $centroidAwal = $centroids->toArray();

```

Kode ini berfungsi untuk memilih centroid (titik pusat) awal yang akan digunakan dalam proses K-Means. Pertama, data berdasarkan jumlah_curas diurutkan dan nilai uniknya dipilih, kemudian data tersebut diacak (shuffle) untuk memastikan pemilihan centroid yang lebih acak dan tidak terpengaruh oleh urutan data. Setelah itu, sebanyak jumlah klaster (\$k), data tersebut diambil dan dipetakan untuk hanya mengambil nilai jumlah_curas sebagai atribut centroid. Hasilnya kemudian disimpan dalam variabel \$centroidAwal sebagai array untuk digunakan pada iterasi selanjutnya dalam proses klasterisasi. Pemilihan centroid awal yang acak ini penting untuk memastikan bahwa klasterisasi tidak terjebak pada solusi lokal yang kurang optimal.

d. Iterasi K-Means



```

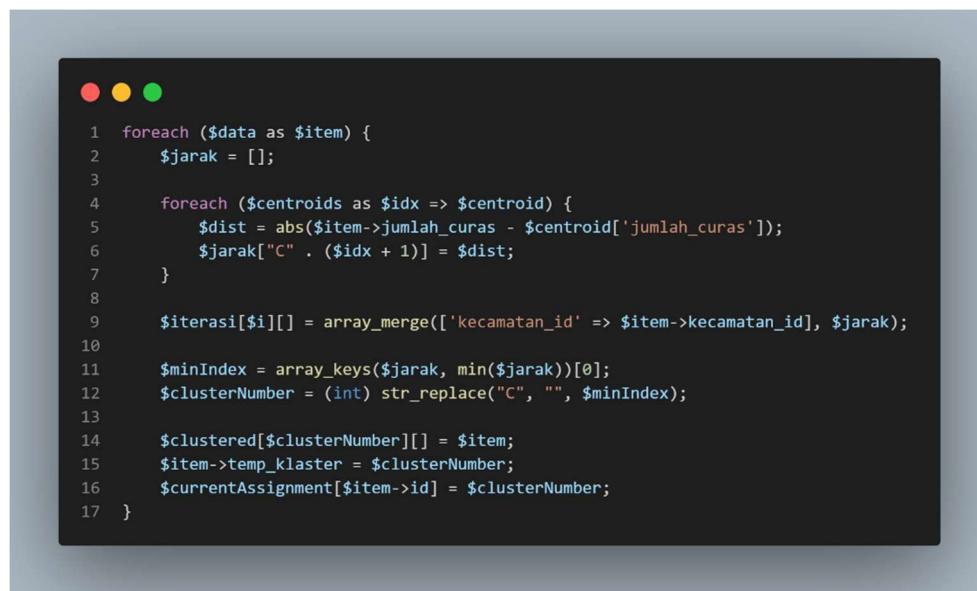
1 $iterasi = [];
2 $prevAssignment = [];
3
4 for ($i = 0; $i < $maxIterasi; $i++) {
5     ...
6 }
7

```

Kode ini mempersiapkan struktur data yang diperlukan untuk menjalankan iterasi dalam algoritma K-Means. Variabel \$iterasi digunakan untuk menyimpan

hasil dari setiap iterasi, sementara `$prevAssignment` menyimpan pengelompokan klaster sebelumnya. Kemudian, dalam loop yang akan berjalan hingga mencapai batas maksimum iterasi (`$maxIterasi`), algoritma akan mengelompokkan data berdasarkan kedekatannya dengan centroid yang ada, dan hasil dari pengelompokan tersebut akan disimpan pada setiap iterasi. Proses iterasi ini akan terus dilakukan sampai konvergensi tercapai, yaitu ketika penugasan data ke klaster tidak berubah lagi antara iterasi berturut-turut, atau sampai mencapai jumlah iterasi maksimum yang telah ditentukan.

e. Perhitungan Jarak dan Pengelompokan



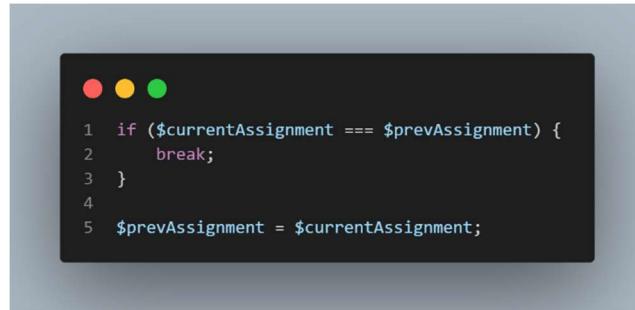
```

1 foreach ($data as $item) {
2     $jarak = [];
3
4     foreach ($centroids as $idx => $centroid) {
5         $dist = abs($item->jumlah_curas - $centroid['jumlah_curas']);
6         $jarak["C" . ($idx + 1)] = $dist;
7     }
8
9     $iterasi[$i][] = array_merge(['kecamatan_id' => $item->kecamatan_id], $jarak);
10
11    $minIndex = array_keys($jarak, min($jarak))[0];
12    $clusterNumber = (int) str_replace("C", "", $minIndex);
13
14    $clustered[$clusterNumber][] = $item;
15    $item->temp_klaster = $clusterNumber;
16    $currentAssignment[$item->id] = $clusterNumber;
17 }

```

Pada bagian kode ini, setiap data dari tabel curas (jumlah kasus) dihitung jaraknya terhadap masing-masing centroid yang telah ditentukan sebelumnya. Jarak dihitung dengan mencari selisih antara nilai `jumlah_curas` dari data dan centroid yang ada, menggunakan rumus jarak *euclidean distance*. Setelah itu, data akan diberikan label klaster berdasarkan jarak terkecil yang ditemukan antara data tersebut dan centroid. Klaster terdekat ditentukan dengan memilih centroid yang memiliki jarak terkecil, dan data tersebut kemudian dimasukkan ke dalam klaster yang sesuai. Proses ini berlanjut untuk semua data, dan pada akhirnya akan membentuk kelompok (klaster) yang merepresentasikan kesamaan pola berdasarkan nilai `jumlah_curas` pada setiap kecamatan.

f. Pengecekan Konvergensi



```

1 if ($currentAssignment === $prevAssignment) {
2     break;
3 }
4
5 $prevAssignment = $currentAssignment;

```

Pada bagian ini, sistem memeriksa apakah penempatan data ke dalam klaster sudah stabil atau tidak. Jika penempatan data ke klaster tidak berubah lagi antara iterasi sebelumnya dan iterasi saat ini, maka proses klasterisasi dihentikan karena sudah mencapai hasil yang stabil. Jika masih ada perubahan, sistem akan melanjutkan iterasi untuk memperbaiki penempatan data ke klaster lebih lanjut.

g. Update Centroid



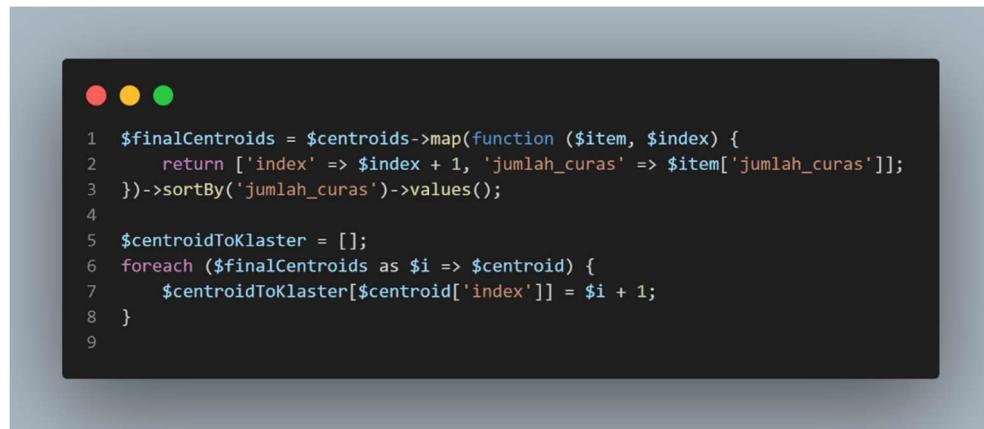
```

1 foreach ($clustered as $key => $group) {
2     $avg = collect($group)->avg('jumlah_curas');
3     $centroids = $centroids->map(function ($item, $index) use ($key, $avg) {
4         return $index === ($key - 1)
5             ? ['jumlah_curas' => $avg]
6             : $item;
7     });
8 }

```

Pada bagian ini, sistem menghitung rata-rata nilai *jumlah_curas* untuk setiap klaster yang terbentuk. Rata-rata ini digunakan untuk memperbarui posisi pusat klaster (*centroid*). Setiap klaster yang berisi beberapa data akan memiliki posisi baru berdasarkan nilai rata-rata tersebut, yang akan menjadi titik pusat klaster yang baru untuk iterasi selanjutnya. Dengan cara ini, sistem terus memperbaiki posisi klaster hingga hasilnya stabil.

h. Penyesuaian Klaster Final



```

1 $finalCentroids = $centroids->map(function ($item, $index) {
2     return ['index' => $index + 1, 'jumlah_curas' => $item['jumlah_curas']];
3 })->sortBy('jumlah_curas')->values();
4
5 $centroidToKlaster = [];
6 foreach ($finalCentroids as $i => $centroid) {
7     $centroidToKlaster[$centroid['index']] = $i + 1;
8 }
9

```

Pada bagian ini, sistem menyusun kembali daftar *centroid* yang telah diperbarui, menyusunnya berdasarkan urutan nilai *jumlah_curas* yang terkecil hingga terbesar. Kemudian, sistem membuat pemetaan antara setiap *centroid* dengan klaster yang sesuai, yang menunjukkan hubungan antara indeks *centroid* dan nomor klaster yang akan digunakan dalam penugasan akhir. Pemetaan ini penting untuk memastikan bahwa setiap data yang telah dikelompokkan akan diberikan klaster yang tepat sesuai dengan posisi pusat klaster terakhir.

i. Update Database



```

1 foreach ($data as $item) {
2     Curas::where('id', $item->id)->update([
3         'klaster_id' => $centroidToKlaster[$item->temp_klaster],
4     ]);
5 }

```

Pada bagian ini, sistem memperbarui data dalam database dengan menetapkan klaster yang sesuai untuk setiap entri. Untuk setiap data yang ada, sistem menggunakan informasi tentang klaster yang dihitung sebelumnya untuk mengupdate kolom *klaster_id* dalam tabel *Curas*. Dengan menggunakan pemetaan yang telah dibuat sebelumnya antara *centroid* dan klaster, data yang sebelumnya belum terklaster akan diberikan klaster yang tepat sesuai dengan hasil perhitungan K-Means. Ini memastikan bahwa setiap kasus curas akan dikelompokkan ke dalam klaster yang relevan, sesuai dengan hasil iterasi terakhir.

j. Output Data Perhitungan



```

 1 return [
 2   'centroid_awal' => $centroidAwalFormatted,
 3   'centroid_akhir' => $centroidAkhirFormatted,
 4   'iterasi' => $iterasi
 5 ];

```

Pada bagian ini, sistem mengembalikan hasil akhir dari proses K-Means clustering. Hasil yang dikembalikan mencakup beberapa elemen penting: pertama, centroid_awal, yang berisi nilai centroid pada awal iterasi sebelum pemrosesan dimulai; kedua, centroid_akhir, yang berisi nilai centroid setelah seluruh iterasi selesai, mencerminkan posisi rata-rata klaster yang sudah diperbarui; dan terakhir, iterasi, yang menyimpan informasi tentang setiap langkah iterasi yang terjadi selama proses K-Means. Data ini memberikan gambaran yang jelas tentang perubahan dan perkembangan selama proses pengelompokan data.

4.4.3 Fitur Fitur Pada Sistem

Fitur-fitur yang ada pada web GIS yang dibuat oleh penulis, mengikuti dan berpedoman dengan flowchart sistem, juga *use case* diagram, dan database yang telah dibuat. Begitu juga dengan hak akses yang dimiliki seluruh pengguna terhadap website tersebut. Sesuai dengan *use case* diagram, hak akses secara umum terbagi menjadi dua, yaitu pengunjung web dan juga admin. Dalam web GIS ini pengunjung web hanya bisa melihat informasi pada halaman utama saja, seperti informasi tentang curas, curanmor, k-means, dan fitur pemetaan kecamatan rawan curas dan curanmor di Kabupaten Probolinggo. Berbeda dengan pengunjung, admin bisa mengakses keseluruhan fitur web, seperti halaman utama dan halaman admin, sehingga admin bisa melakukan interaksi pada data-data pada web. Berikut macam-macam fitur yang ada pada web GIS.

a. Halaman Utama (*Landing Page*)

KProtect Probolinggo

Website Informasi Geografis yang dirancang untuk memetakan kasus Curas dan Curanmor di Kabupaten Probolinggo

Cek Daerahmu

Apa Sih Curas dan Curanmor Itu ?

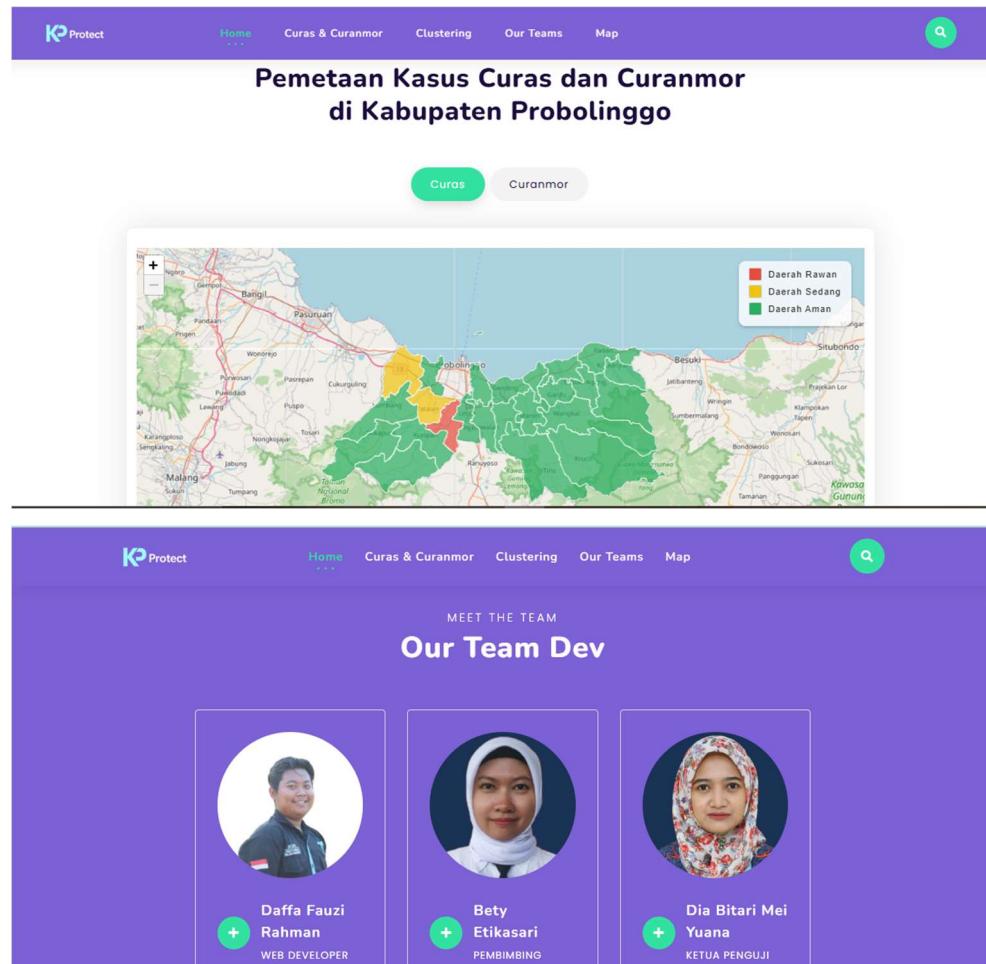
Ternyata banyak yang belum paham apa itu Curas dan Curanmor. Yuk Kita Bahas

Curas (Pencurian Dengan Kekerasan)

Pencurian dengan Kekerasan atau yang sering disebut Curas, merupakan salah satu tindak pidana pencurian yang dalam praktiknya melakukan kekerasan secara fisik maupun ancaman kepada korbananya. Salah satu contoh Curas yaitu Begal

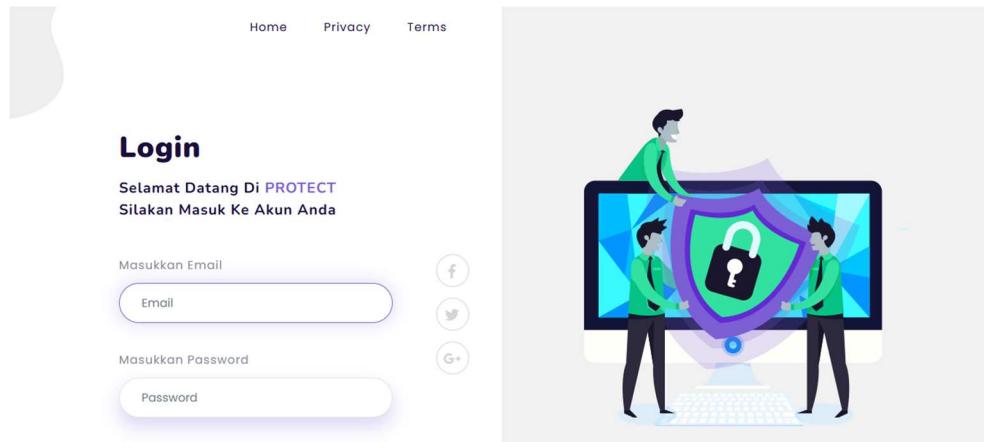
Curanmor (Pencurian Kendaraan Bermotor)

K - Means Clustering merupakan salah satu algoritma machine learning yang digunakan untuk memetakan data berdasarkan kemiripan dari masing - masing data. Terus bedanya K - Means dengan algoritma clustering yang lain apa ? Pada awal iterasi K - Means menggunakan nilai titik pusat klosternya secara acak.



Penulis mencantumkan gambar tampilan halaman utama pada Gambar *CEK untuk menggambarkan antarmuka sistem. Halaman utama ini menyajikan informasi singkat mengenai curas dan curanmor, serta penjelasan sederhana tentang K-Means. Selain itu, halaman ini juga menyertakan fitur pemetaan kasus curas yang dapat diakses oleh pengguna. Pada fitur pemetaan, pengguna dapat memilih jenis data yang ingin dipetakan, apakah itu kasus curas atau curanmor, melalui dropdown yang tersedia. Halaman ini bertujuan untuk memberikan gambaran umum serta mempermudah pengguna dalam memahami dan melihat data secara visual. Dengan adanya dropdown ini, pengguna bisa dengan mudah beralih antara kedua jenis kasus yang ingin ditampilkan pada peta.

b. Halaman Login

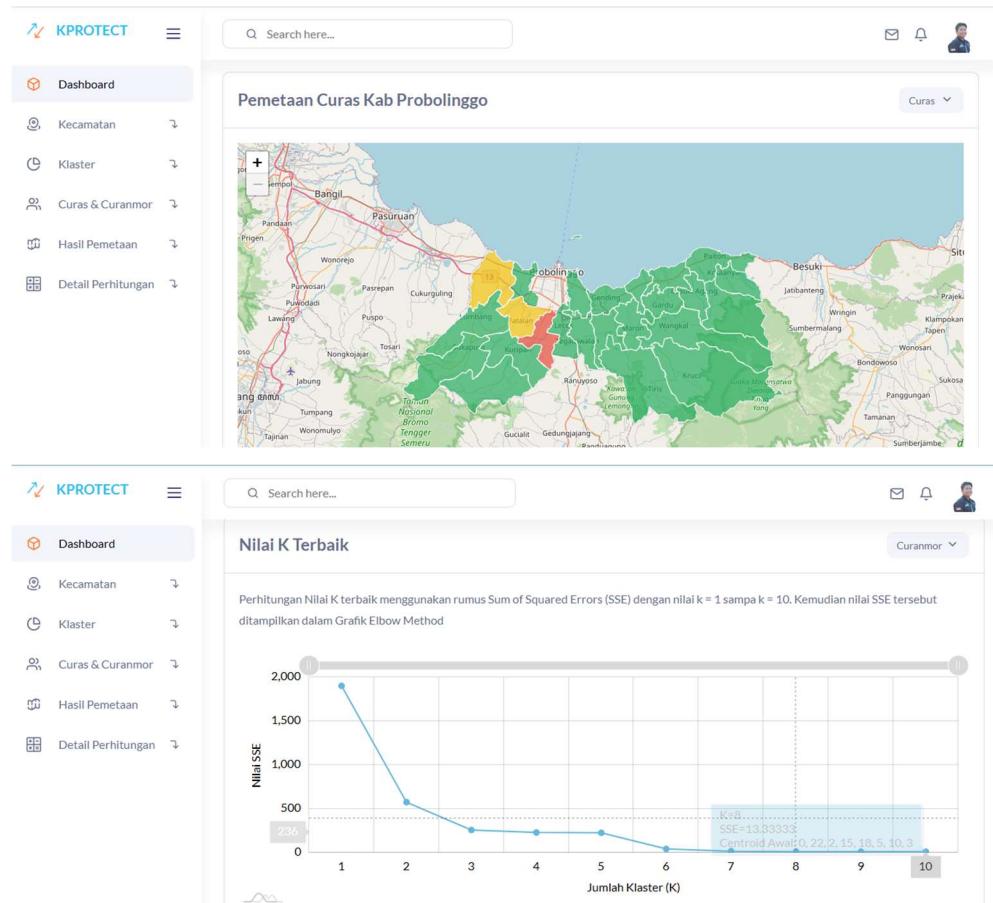


Penulis menampilkan gambar halaman login sebagai bagian dari dokumentasi sistem. Halaman login ini hanya dapat diakses oleh admin yang memiliki hak akses. Untuk melakukan login, admin harus memasukkan email yang terdaftar dan password yang sesuai. Setelah informasi yang dimasukkan valid, sistem akan memverifikasi dan memberikan akses. Jika login berhasil, admin akan diarahkan ke halaman dashboard. Halaman dashboard ini menyediakan fitur dan informasi yang hanya dapat diakses oleh admin yang telah terverifikasi.

c. Halaman Dashboard

Nama Kecamatan	Jumlah Curas	Klaster
Bantaran	11	Rawan

Nama Kecamatan	Jumlah Curanmor	Klaster
Krakaan	37	Rawan



Selanjutnya, penulis menampilkan tampilan halaman dashboard pada Gambar 4 yang hanya dapat diakses oleh admin. Halaman ini dirancang untuk memberikan ringkasan informasi penting secara cepat dan terstruktur. Terdapat daftar lima kecamatan teratas dengan jumlah kasus curas tertinggi yang ditampilkan secara otomatis berdasarkan data yang ada. Selain itu, juga ditampilkan lima kecamatan teratas dengan kasus curanmor tertinggi. Dashboard ini turut menyajikan hasil pemetaan yang menggambarkan distribusi kasus secara visual. Di samping itu, nilai k terbaik untuk proses clustering ditentukan menggunakan metode SSE dan divisualisasikan melalui grafik elbow pada halaman yang sama.

d. Halaman Master Data Kecamatan

No	Nama Kecamatan	Action
1	Bantaran	
2	Banyuanyar	
3	Besuk	
4	Dringu	
5	Gading	

Selanjutnya, penulis menampilkan gambar halaman data kecamatan dan halaman tambah data kecamatan. Halaman ini menampilkan daftar kecamatan yang berada di Kabupaten Probolinggo, di mana data tersebut diambil langsung dari tabel *kecamatans*. Admin memiliki akses penuh untuk mengelola data kecamatan yang ada pada sistem ini. Pengelolaan tersebut mencakup kemampuan untuk melihat, menambahkan, memperbarui, dan menghapus data kecamatan sesuai kebutuhan. Untuk menambahkan data baru, tersedia halaman khusus bernama tambah kecamatan yang dirancang secara sederhana dan mudah digunakan. Dengan fitur ini, admin dapat memastikan bahwa informasi kecamatan selalu terkini dan lengkap.

e. Halaman Master Data Klaster

The screenshots illustrate the 'Daftar Klaster' (List of Clusters) and 'Tambah Klaster Baru' (Add New Cluster) sections of the KPROTECT application.

Daftar Klaster:

Id	Nama Klaster	Warna	Action
1	Aman	#27AE60	[Edit, Delete]
2	Sedang	#F1C40F	[Edit, Delete]
3	Rawan	#E74C3C	[Edit, Delete]

Tambah Klaster Baru:

- Nama Klaster:** Input field for the cluster name.
- Pilih Warna:** A color picker or dropdown menu for selecting the cluster's color.
- Action Buttons:** 'Tambah Klaster' (Add Cluster) in blue and 'Reset' in pink.

Selanjutnya, penulis menampilkan gambar halaman data klaster dan halaman tambah data klaster. Halaman ini menampilkan daftar klaster yang diambil dari tabel *klasters* dan digunakan untuk mengelompokkan wilayah berdasarkan hasil analisis. Admin dapat mengelola data klaster yang ada, seperti melihat, mengubah, maupun menghapus informasi yang sudah tersimpan. Selain itu, tersedia juga halaman tambah klaster yang memungkinkan admin untuk menambahkan data klaster baru. Pada halaman tersebut, admin cukup mengisikan nama klaster serta memilih warna yang akan digunakan untuk merepresentasikan klaster tersebut. Fitur ini bertujuan untuk mempermudah proses pengelompokan dan visualisasi data pada sistem.

f. Halaman Master Data Curas

Daftar Kasus Pencurian Dengan Kekerasan (CURAS)

Berikut ini merupakan data kasus Pencurian Dengan Kekerasan (CURAS) pada masing masing kecamatan di Kabupaten Probolinggo

No	Nama Kecamatan	Jumlah Kasus Curas	Klaster
1	Bantaran	11	Rawan
2	Tongas	3	Sedang
3	Wonomerto	2	Sedang
4	Gading	1	Aman
5	Gending	1	Aman

Tambah Data Kasus CURAS Hari Ini

Detail Kasus Curas Per Tanggal

Berikut ini merupakan rincian kasus Pencurian Dengan Kekerasan (CURAS) yang dikelompokkan berdasarkan tanggal di inputkannya kasus CURAS yang terjadi

Tanggal	Nama Kecamatan	Tambahan Kasus Curas	Total Curas Per Kecamatan	Batalkan Update Kasus
06 May 2025	Bantaran	10	11	
	Bantaran	1	11	

Selanjutnya, penulis menampilkan halaman data kasus curas, halaman tambah data curas, serta halaman detail kasus curas. Pada halaman data curas, ditampilkan daftar kasus pencurian dengan kekerasan (curas) di setiap kecamatan yang datanya diambil dari tabel *curas*. Daftar tersebut ditampilkan dengan urutan berdasarkan kecamatan dengan jumlah kasus tertinggi, sehingga mempermudah dalam mengidentifikasi wilayah yang paling rawan. Informasi ini memberikan

gambaran umum tentang persebaran kasus curas di wilayah Kabupaten Probolinggo.

Pada halaman tambah kasus curas, admin dapat memasukkan data kasus curas terbaru dengan memilih nama kecamatan yang tersedia dan mengisikan jumlah kasus yang ingin ditambahkan. Sementara itu, halaman detail curas menampilkan informasi lebih rinci dari penambahan data yang dilakukan, yang dikelompokkan berdasarkan tanggal. Hal ini memungkinkan admin untuk melihat riwayat penambahan data curas, termasuk tanggal perubahan dan jumlah tambahan kasus pada setiap entri.

g. Halaman Master Data Curanmor

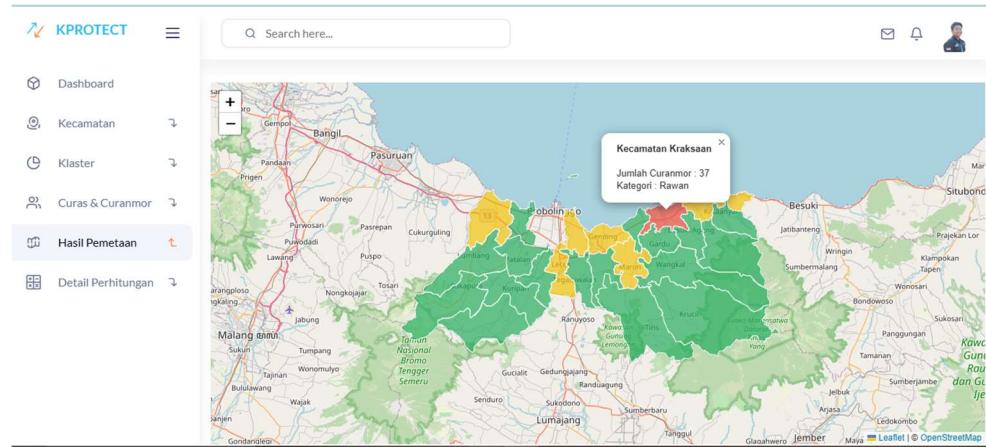
The screenshots illustrate the 'Curas & Curanmor' section of the KPROTECT application. The top screenshot displays a table titled 'Daftar Kasus Pencurian Kendaraan Bermotor (CURANMOR)' showing five entries for different kecamatans (Kraksaan, Dringu, Maron, Gending, Tongas) with their respective case counts (37, 22, 21, 18, 15) and risk levels (Rawan, Sedang, Sedang, Sedang, Sedang). The bottom screenshot shows a form titled 'Tambah Data Kasus Pencurian Kendaraan Bermotor (CURANMOR)' where users can input the name of a kecamatan and the number of cases, with buttons for 'Tambah Data Kasus Curanmor' and 'Reset'.

No	Nama Kecamatan	Jumlah Kasus Curanmor	Klaster
1	Kraksaan	37	Rawan
2	Dringu	22	Sedang
3	Maron	21	Sedang
4	Gending	18	Sedang
5	Tongas	15	Sedang

Berikutnya, penulis menampilkan tampilan halaman data curanmor, halaman untuk menambahkan data curanmor, serta halaman detail kasus curanmor. Di halaman data curanmor, ditampilkan informasi mengenai jumlah kasus pencurian kendaraan bermotor di tiap kecamatan, yang sumber datanya berasal dari tabel *curanmors*. Penyajian data dilakukan dengan mengurutkan kecamatan berdasarkan jumlah kasus terbanyak, sehingga memudahkan dalam mengenali daerah yang memiliki tingkat kerawanan curanmor tertinggi di wilayah Kabupaten Probolinggo.

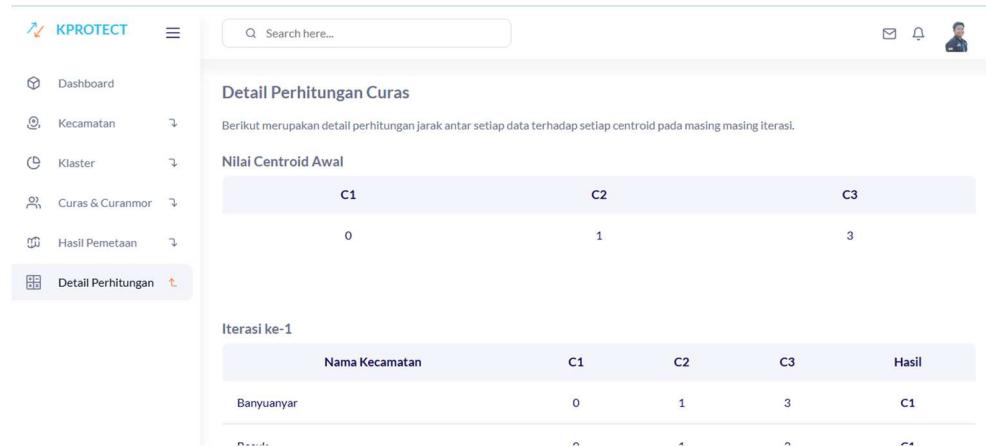
Pada halaman tambah data curanmor, admin dapat memasukkan data kasus baru dengan cara memilih kecamatan yang diinginkan dan menginput jumlah kasusnya. Sementara itu, pada halaman detail curanmor, ditampilkan rincian penambahan data kasus yang dikelompokkan menurut tanggal penambahan. Terdapat informasi waktu perubahan data serta jumlah penambahannya.

h. Halaman Hasil Pemetaan



Penulis menampilkan halaman hasil pemetaan yang terdiri dari pemetaan kasus curas dan pemetaan kasus curanmor. Tampilan pemetaan ini menggunakan warna berbeda sesuai dengan kategori yang ditentukan berdasarkan data yang terdapat pada tabel klaster. Warna-warna tersebut digunakan untuk membedakan tingkat kerawanan di masing-masing kecamatan. Saat pengguna mengklik salah satu wilayah pada peta, akan muncul informasi detail yang berisi nama kecamatan, jumlah kasus curas atau curanmor, serta kategori klaster yang dimiliki wilayah tersebut. Hal ini memudahkan pengguna dalam memahami sebaran dan tingkat kerawanan kejahatan di Kabupaten Probolinggo secara visual.

i. Halaman Detail Perhitungan



Nama Kecamatan	C1	C2	C3	Hasil
Kotaanyar	10	20	30	C1

Selanjutnya, penulis menampilkan halaman detail perhitungan K-Means untuk kasus curas dan curanmor. Halaman ini memberikan gambaran lengkap mengenai proses pengelompokan data berdasarkan algoritma K-Means. Di bagian awal halaman, ditampilkan centroid awal yang dipilih secara acak dari data yang tersedia. Informasi ini penting karena menjadi titik awal dalam menentukan kelompok atau klaster. Selanjutnya, pengguna dapat melihat proses perhitungan jarak antara setiap data terhadap semua centroid pada setiap iterasi. Dari proses ini, data akan dikelompokkan ke dalam klaster yang memiliki jarak terdekat. Setiap iterasi yang terjadi akan ditampilkan hingga hasilnya tidak mengalami perubahan lagi.

4.5 Pengujian

4.5.1 User Acceptance Testing

No.	Pertanyaan	Kesimpulan
1	Seberapa setuju Anda dengan pernyataan bahwa Ketika Anda pertama kali membuka website ini, Anda langsung bisa memahami tujuan dari web ini ?	
2	Setelah membaca penjelasan tentang Curas & Curanmor, Seberapa paham Anda memahami perbedaan antara curas dan curanmor?	
3	Menurut Anda seberapa penting atau relevan informasi tentang curas dan curanmor ini untuk diketahui	

- Setelah membaca penjelasan di website ini, Seberapa
4 paham Anda terhadap K-Means dan bagaimana K-
Means digunakan pada web ini?
- Seberapa setuju Anda dengan pernyataan bahwa
5 penjelasan k-means sudah cukup sederhana dan tidak
terlalu teknis
- Seberapa paham Anda memahami arti warna-warna
6 pada peta?
- Menurut Anda seberapa bermanfaat peta ini dalam
7 membantu untuk memahami tingkat kejahatan di
berbagai kecamatan
- Seberapa setuju anda bahwa peta ini mudah
8 digunakan, baik di desktop maupun di ponsel?
- Menurut Anda apakah informasi yang Anda
9 butuhkan mudah ditemukan dalam satu halaman?
- Seberapa setuju Anda dengan pernyataan bahwa
10 tombol-tombol seperti “Cek Daerahmu”, “Curas”,
atau “Curanmori” mudah ditemukan dan dimengerti
fungsinya?
- Seberapa setuju Anda dengan pernyataan
11 tampilannya tetap rapi dan mudah digunakan jika
diakses dari perangkat seluler (HP)
- Apakah menurut Anda tampilan halaman ini menarik
12 dan mudah dipahami?
- Apakah warna dan font yang digunakan membuat
13 Anda nyaman saat membaca?
- Apakah bahasa yang digunakan mudah dimengerti?
14
- Seberapa setuju Anda dengan pernyataan bahwa
15 secara keseluruhan website ini bermanfaat

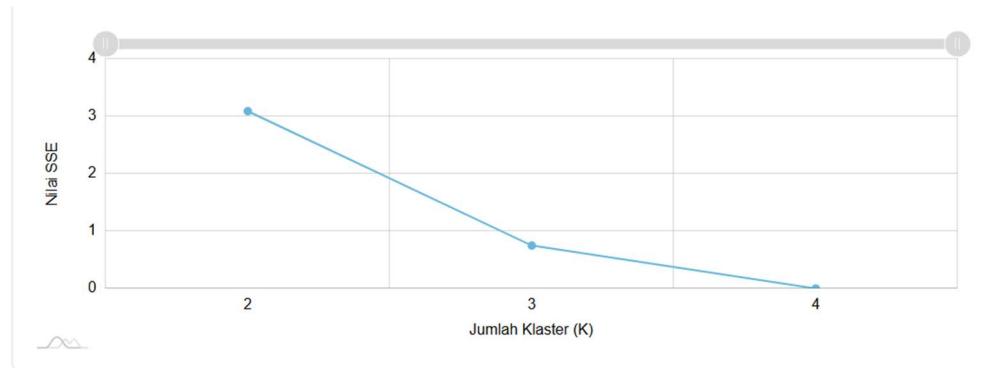
16 Seberapa puas Anda atas Website Gisbegal ini

4.5.2 Pengujian Metode *Elbow*

Tahapan ini bertujuan untuk menentukan jumlah klaster yang paling optimal dalam proses pengelompokan data curas dan curanmor menggunakan algoritma *K-Means*. Pemilihan nilai klaster yang tepat sangat penting agar hasil pengelompokan lebih akurat dan merepresentasikan pola data dengan baik. Dalam proses ini, penulis menggunakan metode Elbow untuk membantu menemukan nilai k yang ideal. Metode Elbow bekerja dengan cara menghitung nilai *Sum of Square Error* (SSE) pada setiap variasi jumlah klaster. Nilai SSE dihitung untuk setiap jumlah k, kemudian hasilnya digambarkan dalam bentuk grafik. Titik optimal ditandai dengan terbentuknya sudut siku pada grafik dan juga memiliki selisih nilai SSE tertinggi dari nilai SSE pada klaster sebelumnya. Titik inilah yang dianggap sebagai jumlah klaster terbaik untuk digunakan. Dengan pendekatan ini, hasil klasterisasi menjadi lebih efektif dalam membedakan tingkat kerawanan kasus curas dan curanmor di setiap wilayah. Dalam penelitian ini, penulis menentukan rentang nilai k yang digunakan yaitu $k = 2$ hingga $k = 10$. Berikut hasil perhitungan SSE dan Tampilan Grafik untuk rentang k yang sudah ditentukan pada masing-masing kasus

a. SSE dan Grafik Metode Elbow Kasus Curas

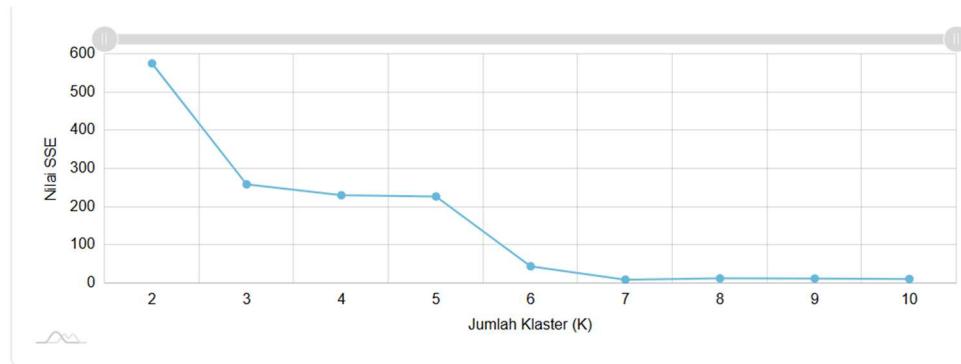
Nilai k	Nilai SSE	Selisih SSE
2	13,217	-
3	3,2	10,017
4	2,57	0,63



Berdasarkan Tabel 4, diperoleh bahwa selisih nilai SSE tertinggi adalah sebesar 10,017. Selisih tersebut muncul antara jumlah klaster $k = 2$ dan $k = 3$. Nilai ini menunjukkan penurunan signifikan dalam SSE ketika jumlah klaster ditingkatkan dari dua menjadi tiga. Penurunan yang besar ini menjadi indikator penting dalam analisis menggunakan metode Elbow. Selanjutnya, jika mengacu pada Grafik Elbow 4, terlihat jelas adanya bentuk siku pada nilai $k = 3$. Bentuk siku tersebut mengindikasikan bahwa setelah titik tersebut, penurunan nilai SSE tidak lagi signifikan. Berdasarkan hal tersebut, maka nilai $k = 3$ dianggap sebagai titik optimal untuk jumlah klaster. Oleh karena itu, dapat disimpulkan bahwa jumlah klaster terbaik untuk kasus curas adalah tiga.

b. SSE dan Grafik Metode Elbow Kasus Curanmor

Nilai k	Nilai SSE	Selisih SSE
2	575,88	-
3	259,309	316,571
4	230,97	28,339
5	226,84	4,13
6	43	183,84
7	17,366	25,643
8	13,333	4,033
9	12,083	1,25
10	11,416	0,667



Merujuk pada Tabel 4, nilai selisih SSE paling tinggi tercatat sebesar 316,571. Selisih ini muncul saat jumlah klaster meningkat dari $k = 2$ menjadi $k = 3$. Besarnya penurunan SSE tersebut mengindikasikan adanya perbaikan signifikan dalam pembentukan klaster saat ditambah menjadi tiga kelompok. Penurunan tajam ini menjadi petunjuk penting dalam menentukan jumlah klaster yang tepat. Selain itu, tampak pada Grafik Elbow 4 bahwa terbentuk sudut siku yang jelas pada titik $k = 3$. Sudut tersebut menunjukkan bahwa setelah $k = 3$, penurunan SSE menjadi lebih landai dan tidak terlalu berarti. Keadaan ini menunjukkan bahwa $k = 3$ merupakan titik optimal sebelum efektivitas pengelompokan berkurang. Maka, dapat disimpulkan bahwa tiga klaster adalah pilihan terbaik untuk pengelompokan data curanmor.

4.6 Analisis dan Pembahasan