

**IMPLEMENTASI *DATA MINING* DALAM PENGELOMPOKAN
DATA KRIMINALITAS UMUM DI PROVINSI SUMATRA
UTARA MENGGUNAKAN ALGORITMA
*K-MEANS CLUSTERING***

SKRIPSI

OLEH:

**EVIMAI INDRI SITORUS
208160018**



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN
2024**

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 9/1/25

Access From (repository.uma.ac.id)9/1/25

IMPLEMENTASI *DATA MINING* DALAM PENGELOMPOKAN
DATA KRIMINALITAS UMUM DI PROVINSI SUMATRA
UTARA MENGGUNAKAN ALGORITMA
K-MEANS CLUSTERING

SKRIPSI

Diajukan Sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana (S1) di Fakultas Teknik
Universitas Medan Area



OLEH:
EVIMAI INDRI SITORUS
208160018

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN
2024

LEMBAR PENGESAHAN


Judul Skripsi : Implementasi *Data Mining* dalam Pengelompokan Data
Kriminalitas Umum di Provinsi Sumatera Utara menggunakan
Algoritma *K-Means Clustering*

Nama : Evimai Indri Sitorus

NPM : 208160018

Fakultas : Teknik

Disetujui Oleh
Komisi Pembimbing


Nanda Novita, S.Kom, M.Kom

Pembimbing I


Dr. Drg. Supriatno, ST., MT
Dekan


Rizki Muliono, S.Kom, M.Kom
Kaprodi

Tanggal Lulus : 30 Agustus 2024

HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya susun, sebagai syarat memperoleh gelar sarjana merupakan hasil karya tulis saya sendiri. Adapun bagian-bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah.

Saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksi- sanksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari ditemukan adanya plagiat dalam skripsi ini.

Medan, 30 Agustus 2024



Evimai Indri Sitorus

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS
AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan di bawah ini:

Nama : Evimai Indri Sitorus
NPM : 208160018
Program Studi : Teknik Informatika
Fakultas : Teknik
Jenis Karya : Skripsi

demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty-Free Right) atas karya ilmiah saya yang berjudul :

Implementasi Data Mining dalam Pengelompokan Data Kriminalitas Umum di Provinsi Sumatera Utara menggunakan Algoritma K-Means Clustering

berserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Medan Area berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir/skripsi/tesis saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan

Pada tanggal: 30 Agustus 2024

Yang Menyatakan



(Evimai Indri Sitorus)

ABSTRAK

Penelitian ini bertujuan untuk mengimplementasikan teknik data mining dalam pengelompokan data kriminalitas umum di Provinsi Sumatra Utara dengan menggunakan *algoritma K-Means Clustering*. Metode K-Means adalah algoritma non-hierarki yang memilih titik pusat cluster secara acak dari data populasi awal. Algoritma ini kemudian mengevaluasi dan mengelompokkan setiap komponen data ke pusat cluster terdekat berdasarkan jarak minimum. Posisi pusat cluster dihitung ulang hingga semua data dikelompokkan dengan benar, menghasilkan pusat *cluster* baru setelah iterasi selesai. Hasil perhitungan menggunakan metode K-Means *Clustering* kemudian dibandingkan dengan data aktual untuk mengetahui tingkat akurasi hasil. Berdasarkan pengujian terhadap 29 daerah, diperoleh rata-rata persentase yaitu adalah 50%. Hal ini menunjukkan bahwa *algoritma K-Means Clustering* berhasil mengelompokkan daerah berdasarkan tingkat ke dengan presentasi cukup tinggi. *K-Means Clustering* dapat melakukan pengelompokan daerah berdasarkan tingkat kriminalitas umum secara akurat. Dimana hasil pengelompokan yang diperoleh yaitu pengelompokan untuk pusat *cluster* pertama (tinggi) ada 1 data sedangkan untuk pusat *cluster* kedua (sedang) ada 1 data dan pada pusat cluster ketiga (rendah) ada 27 data. Sehingga dapat disimpulkan bahwa ada 1 daerah yang memiliki tingkat kriminalitas paling tinggi, 1 daerah memiliki tingkat kriminalitas sedang dan 27 daerah yang memiliki tingkat kriminalitas paling rendah.

Kata Kunci: *Data Mining; K-means; Kriminal; Cluster.*

ABSTRACT

This research aimed to implement 10 data mining techniques in grouping general crime data in North Sumatra Province using the K-Means Clustering algorithm. The K-Means method is a non-hierarchical algorithm that selects cluster centers randomly from the initial population data. This algorithm then evaluated and groups each data component to the nearest cluster center based on the minimum distance. The cluster center positions are recalculated until all data was grouped correctly, resulting in a new cluster center once the iteration is complete. The calculation results using the K-Means Clustering method are then compared with actual data to determine the level of accuracy of the results. Based on testing of 29 regions, the average percentage obtained is 50%. This showed that the K-Means Clustering algorithm has succeeded in grouping regions based on levels with a fairly high presentation. K-Means Clustering can group areas based on general crime levels accurately. Where the grouping results obtained were that for the first cluster center (high) there was 1 data, while for the second cluster center (medium) there is 1 data and for the third cluster center (low) there are 27 data. So it can be concluded that there is 1 area that has the highest crime rate, 1 area has a medium crime rate and 27 areas have the lowest crime rate.

Keywords: Data Mining; K-means; Criminal; Clusters.



KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul **“Implementasi Data Mining Dalam Pengelompokan Data Krminalitas Umum di Provinsi Sumatera Utara Menggunakan Algoritma K-Means Clustering”** ini dengan baik. Skripsi ini disusun untuk memenuhi salah satu syarat menyelesaikan pendidikan Program Strata-1 pada Fakultas Teknik Program Studi Teknik Informatika di Universitas Medan Area. Dalam proses menyelesaikan skripsi ini, penulis menyadari bahwa tulisan ini masih jauh dari kata sempurna dan juga terdapat banyak kekurangan. Penulis juga mengharapkan kritik dan saran yang sifatnya membangun dari para pembaca. Kemudian penulis ingin mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Dadan Ramdan, M.Eng., M.Sc., selaku Rektor Universitas Medan Area.
2. Bapak Dr.Eng. Supriatno,S.T., M.T. selaku Dekan Fakultas Teknik Universitas Medan Area.
3. Bapak Rizki Muliono, S.Kom., M.Kom., selaku Kepala Program Studi Teknik Informatika Universitas Medan Area.
4. Ibu Nanda Novita, S.Kom., M.Kom., selaku dosen pembimbing yang telah membimbing, mengarahkan, dan memberikan motivasi kepada penulis dalam proses menyelesaikan skripsi ini.
5. Seluruh Dosen Teknik Informatika Universitas Medan Area yang selama ini telah membekali penulis dengan ilmu yang sangat bermanfaat.

6. IT Support Teknik Informatika yang baik hati Bang Robby Kurniawan Sari Damanik, ST yang telah banyak membantu penulis dalam menyiapkan pemberkasan yang diperlukan dari awal sampai selesai.
7. Kedua Orang Tua saya Alm. Guntur Sitorus dan Alm. Rianta Nadeak telah melahirkan saya dan membuat saya kuat saat ini sampai saya menyelesaikan pendidikan ini.
8. Teristimewa kepada saudara kandung saya Gunaro Tuani Sitorus yang telah memberikan saya pendidikan yang baik dengan penuh kasih dan selalu mendoakan penulis ini sehingga dapat menyelesaikan pendidikan hingga perguruan tinggi saat ini.
9. Saudara kandung saya Roris Indah Dwiana Sitorus, Ellysa Gusrin Sitorus, Jefri Gunawan Sitorus, abang ipar Ferdinan Purba dan kakak ipar Rohani Siska yang memberikan semangat dan dukungan serta doa untuk dapat menyelesaikan pendidikan ini.
10. Teman dekat saya Mhd. Qori Ramadhan Nasution S.Kom dan Meniati Zebua S.Kom, Rohany Cahaya S.Kom yang selalu mendukung dan mendoakan serta sangat banyak membantu penulis hingga berhasil menyelesaikan pendidikan saat ini.
11. Sahabat SMP saya Audia Theresya Sitepu dan Sherina Karina Pelawi yang selalu menemani dan menolong disaat saya mengerjakan penulisan ini.
12. Frensy Sipayung dan Yoni Masdwita sudah menjadi tempat berkeluh kesah dan memberi nasihat baik selama saya menyelesaikan pendidikan ini.

13. Keponakan saya Illona Fhelisha Purba, Ernest Geoffrey Purba, dan Gianna Aprillya Sitorus memberi hiburan sehingga berhasil menyelesaikan pendidikan ini.
14. Seseorang yang tidak dapat saya sebut, terima kasih atas memberikan waktunya dan menemani saya selama masa Pendidikan Sarjana saat ini.
15. Teman-teman Teknik Informatika 2020, terima kasih atas persahabatan dan persaudaraannya selama ini. Semoga Tuhan memudahkan untuk menyelesaikan study S-1 ini.
16. Dan yang paling terakhir dan paling penting saya mengucapkan terima kasih kepada diri saya sendiri yaitu Evimai Indri Sitorus S.Kom yang telah berusaha sejauh ini walaupun banyak ngeluhnya tetapi tidak berhenti berusaha dan berjuang. Semoga kedepannya menjadi seorang yang sukses dan menggapai cita – cita.

Medan, 30 Agustus 2024

Evimai Indri Sitorus

208160018

DAFTAR ISI

ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL.....	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	4
1.5 Batasan Penelitian	4
1.6 Sistematika Penulisan	5
BAB II LANDASAN TEORI	7
2.1 Data Mining Clustering	7
2.1.1 Pengertian Data Mining	8
2.1.2 Tahapan Data Mining.....	9
2.1.3 <i>Scraping Data</i>	10
2.1.4 Clustering.....	11
2.2 <i>Algoritma K-Means</i>	14
2.3 Kriminal Umum.....	16
2.4 <i>Flowchart</i>	17
2.5 <i>Davies-Bouldin Index (DBI)</i>	18
2.6 Penelitian Terdahulu.....	20
BAB III METODE PENELITIAN	23
3.1 Tahapan Penelitian	23
3.2 Pengumpulan Data.....	24
3.2.1 Sampel Data	24
3.2.2 Alat dan Bahan.....	25
3.3 Implementasi Metode K-Means	26
3.3.1 Preprocessing Data.....	26
3.3.2 Iterasi-1	27
3.3.3 Iterasi-2	32
3.3.4 Iterasi-3	36
3.4 Perancangan Sistem.....	40
3.4.1 Perancangan UML	40
3.4.2 Perancangan Basis Data	43
3.4.3 Perancangan User Interface (UI).....	45
BAB IV HASIL DAN PEMBAHASAN.....	50
4.1 Hasil.....	50
4.1.1 Analisa Hasil	50
4.1.2 Testing Mandiri.....	51
4.2 Pembahasan	51
BAB V KESIMPULAN DAN SARAN	56
5.1 Kesimpulan.....	56
5.2 Saran	57

DAFTAR PUSTAKA	58
LAMPIRAN.....	61
1. Source Code.....	61
2. Sample Data.....	69
3. Surat Keterangan Pembimbing.....	70
4. Surat Pengantar Riset	71
5. Surat Selesai Riset	72
6. Turnitin.....	73



DAFTAR GAMBAR

Gambar 2.1	Algoritma Sederhana K-Means.....	15
Gambar 3.1	Tahapan Penelitian.....	23
Gambar 3.2	Data asli penelitian.....	25
Gambar 3.3	<i>Use Case Diagram</i>	41
Gambar 3.4	<i>Activity Diagram</i>	42
Gambar 3.5	<i>Class Diagram</i>	43
Gambar 3.6	Rancangan Halaman <i>Login</i>	46
Gambar 3.7	Rancangan Halaman Utama.....	46
Gambar 3.8	Rancangan <i>Form</i> Daerah.....	47
Gambar 3.9	Rancangan <i>Form</i> Jenis Kasus.....	47
Gambar 3.10	Rancangan <i>Form</i> Kasus.....	48
Gambar 3.11	Rancangan <i>Form</i> Proses Pengelompokan.....	49
Gambar 3.12	Rancangan Laporan Pengelompokan.....	49
Gambar 4.1	Halaman <i>Login</i>	52
Gambar 4.2	Halaman Utama.....	53
Gambar 4.3	<i>Form</i> Daerah.....	53
Gambar 4.4	<i>Form</i> Jenis Kasus.....	54
Gambar 4.5	<i>Form</i> Kasus.....	54
Gambar 4.6	<i>Form</i> Proses Pengelompokan.....	55
Gambar 4.7	Laporan Pengelompokan.....	55

DAFTAR TABEL

Tabel 2.1	Tabel Flowchart.....	17
Tabel 2.2	Tabel Penelitian Terahulu	20
Tabel 3.1	Tabel Kriminalitas.....	24
Tabel 3.2	Data	26
Tabel 3.3	Pusat (<i>Centroid</i>) Awal <i>Cluster</i>	27
Tabel 3.4	Jarak data terhadap <i>cluster</i> pada iterasi ke – 1	28
Tabel 3.5	Data Kelompok C1.....	30
Tabel 3.6	Data Kelompok C2.....	30
Tabel 3.7	Data Kelompok C3.....	31
Tabel 3.8	Pusat (<i>Centroid</i>) <i>Cluster</i> Baru Iterasi 2.....	32
Tabel 3.9	Jarak data terhadap <i>cluster</i> pada iterasi ke – 2	32
Tabel 3.10	Data Kelompok C1.....	34
Tabel 3.11	Data Kelompok C2.....	35
Tabel 3.12	Data Kelompok C3.....	35
Tabel 3.13	Pusat (<i>Centroid</i>) <i>Cluster</i> Baru Iterasi 3.....	36
Tabel 3.14	Jarak data terhadap <i>cluster</i> pada iterasi ke – 3	36
Tabel 3.15	Hasil Pengelompokan.....	38
Tabel 3.16	Admin.....	44
Tabel 3.17	Daerah	44
Tabel 3.18	Jenis Kasus	44
Tabel 3.19	Nilai.....	44
Tabel 3.20	Pengelompokan	45
Tabel 4.1	Akurasi Hasil Pengelompokan	50
Tabel 4.2	<i>Testing</i> Mandiri	51

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi dalam kehidupan saat ini sangat berkembang pesat berjalannya waktu ke waktu. Pengaruh adanya pendorong pertumbuhan kebudayaan dan pesatnya perkembangan ilmu pengetahuan dan teknologi telah menciptakan persaingan dalam berbagai bidang kehidupan manusia, baik ideologi, kemasyarakatan, ekonomi, seni, etika, dan moralitas. Banyak terjadi perubahan terhadap nilai-nilai yang terlibat, seperti *materialisme* dan *hedonisme*. Hal ini juga menyebabkan perubahan nilai-nilai masyarakat. Perubahan yang bersifat positif sangat bermanfaat bagi masyarakat, namun perubahan yang bersifat negatif dapat menimbulkan kegelisahan dalam kehidupan masyarakat karena akan berperilaku negatif seperti melakukan kejahatan (Putra dkk., 2021).

Tindakan kriminalitas berpengaruh terhadap keamanan masyarakat serta mengancam kenyamanan. Jika masyarakat merasa terancam keamanan dan kenyamanan maka kemungkinan besar pengaruh terhadap ketentraman masyarakat tersebut. Kriminalitas adalah segala bentuk tindakan dan perbuatan yang merugikan secara ekonomis dan psikologis yang melanggar hukum yang berlaku pada Indonesia serta norma nilai sosila dan agama.

Masalah kriminalitas umum merupakan isu sosial yang rumit dan memerlukan strategi yang tepat untuk menanggulangi dan mencegahnya., sebagai suatu daerah di Indonesia, juga menghadapi tantangan kriminalitas yang dapat mengancam ketertiban masyarakat. Dengan memanfaatkan kemajuan teknologi

informasi dan komunikasi, data kriminalitas dapat dikumpulkan dan dianalisis dengan lebih efisien, memberikan wawasan yang mendalam mengenai pola kejadian criminal (Gultom & Lebang, 2024)

Perilaku kejahatan di kategorikan dalam suatu kriminalitas, berbagai macam kasus kriminalitas dapat diketahui dari sumber media, dan sistem informasi keamanan publik dapat dijadikan dasar untuk analisis lebih mendalam (Wibawa, 2020). Disebabkan banyaknya kasus kejahatan terkhususnya di daerah Sumatera Utara membuat kesulitan dalam mengetahui kejahatan yang terdapat berbagai macam kasus. kemajuan pesat dalam teknologi informasi dan komunikasi telah membuka peluang baru dalam pengelompokan berbagai kasus kriminalitas di Sumatera Utara, penerapan pengelompokan dalam kasus kriminalitas salah satunya menggunakan *algoritma K-Means*.

Algoritma K-Means merupakan metode dalam bidang *data mining* yang mampu mengelompokkan data atau melakukan clustering terhadap suatu *dataset* menjadi satu atau lebih *cluster*. Dengan menggunakan metode ini, data yang memiliki karakteristik serupa akan dikelompokkan ke dalam satu *cluster*, sementara data yang memiliki karakteristik berbeda akan ditempatkan dalam kelompok yang berbeda (Dinata dkk., 2020).

Penelitian ini akan fokus pada implementasi *algoritma K-Means Clustering* untuk mengelompokkan data kriminalitas umum di Provinsi Sumatra Utara. *Algoritma* ini dipilih karena kemampuannya dalam membentuk kelompok secara otomatis berdasarkan kesamaan karakteristik data. Dengan menggunakan algoritma ini, diharapkan dapat diidentifikasi pola-pola kriminalitas yang muncul secara

alami dan memberikan dasar untuk pengambilan keputusan yang lebih baik dalam rangka pencegahan dan penanggulangan kriminalitas (Wahyuddin dkk., 2023).

Pemahaman lebih mendalam terhadap pola kriminalitas ini akan memberikan kontribusi signifikan dalam merancang kebijakan keamanan yang lebih efektif dan responsif terhadap kondisi aktual di lapangan (Zaidan, 2022). Oleh karena itu, penelitian ini memiliki nilai penting untuk pengembangan strategi keamanan di tingkat provinsi dan dapat menjadi referensi bagi peneliti dan praktisi keamanan dalam mengoptimalkan upaya penanggulangan kriminalitas. Dengan adanya penelitian ini, diharapkan dapat memberikan kontribusi nyata dalam perbaikan situasi keamanan di Provinsi Sumatra Utara, sekaligus memberikan pandangan baru terhadap pemanfaatan teknologi *Data Mining* dalam konteks keamanan masyarakat secara umum.

1.2 Rumusan Masalah

Berdasarkan Latar Belakang di atas, rumusan masalah pada penelitian ini adalah bagaimana mengimplementasikan data mining dalam pengelompokan data kriminalitas umum di Provinsi Sumatra Utara menggunakan *algoritma K-Means Clustering*.

1. Sejauh mana keberhasilan *algoritma K-Means Clustering* dalam membentuk kelompok yang representatif dan bermakna terkait pola kriminalitas umum?
2. Bagaimana *algoritma K-Means Clustering* dapat diimplementasikan untuk mengelompokkan data kriminalitas umum di Provinsi Sumatra Utara?

1.3 Tujuan Penelitian

Adapun tujuan dilakukannya penelitian ini adalah sebagai berikut:

1. Mengimplementasikan *algoritma K-Means Clustering* untuk mengelompokkan data kriminalitas umum berdasarkan kasus yang memiliki kesamaan karakteristik.
2. Mengidentifikasi faktor-faktor yang memiliki korelasi signifikan dengan tingkat kriminalitas umum di Provinsi Sumatra Utara berdasarkan analisis *Data Mining*.

1.4 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini sebagai berikut:

1. Menyediakan pemahaman yang mendalam terkait dengan pola dan karakteristik kriminalitas umum di Provinsi Sumatra Utara berdasarkan analisis data historis.
2. Menunjukkan potensi dan keefektifan penerapan teknologi Data Mining, khususnya *algoritma K-Means Clustering*, dalam menganalisis dan mengelompokkan data kriminalitas umum.

1.5 Batasan Penelitian

Batasan masalah pada penelitian ini adalah:

1. Fokus utama penelitian ini adalah pada kriminalitas umum, termasuk tindak pidana seperti pencurian, perampokan, dan kejahatan lain yang masuk dalam kategori kriminalitas umum. Penelitian ini tidak mencakup analisis tindak pidana khusus atau kategori lain di luar ruang lingkup kriminalitas umum.
2. Teknik yang digunakan adalah *data mining clustering*.
3. Data yang digunakan data dari Polda Sumut.

1.6 Sistematika Penulisan

Sistematika penulisan yang diajukkan dalam penelitian ini adalah sebagai berikut:

BAB I : PENDAHULUAN

Bab ini menguraikan tentang latar belakang, rumusan masalah yang didetailkan dengan research question, tujuan dan manfaat penelitian, batasa masalah, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Bab ini akan membahas seluruh dasar-dasar teori yang digunakan dalam penelitian, mulai dari data mining, tindakan kriminal umum, gambaran kriminal di Sumatra Utara, *Algoritma K-Means Clustering* dan tahapan analisis yang digunakan serta implementasi dalam penelitian.

BAB III : METODOLOGI PENELITIAN

Bab ini akan membahas mengenai tahap-tahap yang akan diterapkan pada penelitian. Setiap rencana dari tahapan penelitian dideskripsikan secara rinci.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini membahas tentang langkah-langkah implementasi dan hasil program yang terdiri dari tampilan program, alur program, dan penjelasan program.

BAB V : KESIMPULAN DAN SARAN

Bab ini menjelaskan tentang kesimpulan dari semua uraian-uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan

berguna dalam penerapan dan penelitian menggunakan data mining clustering pada sistem pengelompokan data selanjutnya.



BAB II

LANDASAN TEORI

2.1 Data Mining Clustering

Data Mining Clustering ialah metode di dalam bidang data mining yang memiliki tujuan untuk mengelompokkan data ke dalam kumpulan berdasarkan kesamaan ciri tertentu (Alhapizi dkk., 2020). Fokus utama dari clustering adalah mengungkap pola atau struktur tersembunyi dalam data tanpa memerlukan informasi label sebelumnya. Dengan kata lain, algoritma clustering berusaha mengelompokkan data sedemikian rupa sehingga objek-objek dalam satu kelompok memiliki tingkat kemiripan yang tinggi, sementara objek-objek di antara kelompok memiliki perbedaan yang signifikan (R. F. Putra dkk., 2023)

Pada dasarnya, clustering membantu mengidentifikasi keterkaitan mendasar antara data dan memudahkan pemahaman terhadap struktur yang mungkin ada dalam i Dengan melakukan pengelompokan data, kita dapat mengidentifikasi pola-pola, atau informasi penting lainnya yang mungkin sulit ditemukan jika melihat keseluruhan dataset tanpa adanya pembagian kelompok (Wahyuddin dkk., 2023).

Contoh penerapan *Data Mining Clustering*, sering terdapat pada segmentasi pelanggan berdasarkan perilaku belanja, pengelompokan dokumen berita ke dalam topik-topik tertentu, atau pembagian kelompok genom dalam penelitian biologi. Algoritma-algoritma yang biasa digunakan untuk mencapai tujuan tersebut ialah *K-Means*, *Hierarchical Clustering*. *Data Mining Clustering* memberikan dampak besar dalam mendukung proses pengambilan keputusan, analisis pola, dan

pemahaman yang lebih mendalam terhadap data yang bersifat mutlak (Sachrrial & Iskandar, 2023)

2.1.1 Pengertian Data Mining

Data mining merupakan analisis kumpulan-kumpulan data obsevasi yang besar, menumukan suatu hubungan yang tak terduga serta dapat meringkas data dengan cara terbaru sehingga mudah untuk dimengerti dan dipahami (Gustientiedina dkk., 2019). Data mining proses dengan kecerdasan buatan, Teknik statistic, matematika, dan *machine learning* untuk mengidentifikasi dan mengekstrasi dari suatu informasi yang memiliki tujuan dan pengetahuan terkait dari beberapa database yang besar (Setyaningtyas dkk., 2022).

Tujuan utama dari *data mining* adalah untuk menyatakan hubungan, pola, dan tren yang secara tidak langsung untuk mendapatkan dukungan pengambil keputusan yang tepat dan baik (Leto dkk., 2023). Proses ini mencakup tranformasi data mentah menjadi informasi yang lebih berarti dan dapat digunakan untuk memprediksi, menyimpulkan, memaksimalkan dan memahami suatu. Istilah data mining dapat dipandang dari beberapa sudut, seperti *knowledge discovery* atau pengenalan pola. Pemakaian istilah *knowledge discovery*, atau penemuan pengetahuan, sangat tepat karena tujuan utama dari data mining .Di sisi lain, istilah *pattern recognition* atau pengenalan pola juga sesuai karena bertujuan menemukan pola yang tersembunyi dalam set data (Nabila dkk., 2021).

Metode yang penting dari sebuah data mining pengelompokan adalah kemampuan mengambil keputusan pengguna depan tepat dengan menganalisis perilaku pengguna untuk menghasilkan pengelompokan data banyak yang tepat. Saran dalam penggunaan penggelompokan data banyak adalah metode *clustering*

untuk melakukan pengelompokan pola data tersembunyi dan data tujuan untuk kepentingan tertentu. Data mining memiliki beberapa metode, yaitu *Clustering*, *Classification*, *Regression*, *Association* (Rani Rotul Muhima dkk., 2022).

2.1.2 Tahapan Data Mining

Beberapa tahapan data mining terdiri dari 7 tahapan, ialah, (Vitali dkk., 2021):

1. Membersihkan data (*Data Cleaning*)

Data dibersihkan untuk menghilangkan ketidak konsistenan dan ketidak relevan yang mungkin berasal dari database yang tidak sempurna atau memiliki data yang hilang. Pembersihan bertujuan mengidentifikasi data tidak valid akibat kesalahan penulisan atau atribut tidak sesuai, mempermudah proses data mining.

2. Integrasi data (*Data Integration*)

Integrasi data mining dilakukan dengan menggabungkan berbagai jenis database atau file. Tujuannya adalah mengambil atribut yang diperlukan untuk proses data mining, memudahkan pemahaman data yang sebenarnya, termasuk data yang tidak ada.

3. Seleksi Data (*Data Selection*)

Dalam tahap ini, data mengalami seleksi di mana hanya atribut yang dianggap penting yang dipertahankan. Sebagai contoh, hanya `id_nama` yang diambil dari data nama.

4. Transformasi Data (*Data Transformation*)

Transformasi berperan dalam mengubah dan mengintegrasikan skema serta struktur yang berbeda ke dalam skema dan struktur yang didefinisikan dalam data warehouse.

5. Proses *Mining*

Data mining, menggunakan kecerdasan buatan, statistik, dan matematika, digunakan untuk menggali pengetahuan baru dari data yang melimpah dalam data warehouse. Teknologi ini diharapkan dapat memfasilitasi komunikasi antara data dan pengguna.

6. Evaluasi Pola (*Pattern Evaluation*)

Dalam evaluasi pola di data mining, pola model prediksi dievaluasi untuk mencapai tujuan. Jika hasil belum memenuhi, metode data mining lain atau yang sesuai dapat digunakan. Hasil yang tak terduga juga dapat diterima.

7. Presentasi Pengetahuan (*Knowledge Presentation*)

Proses evaluasi metode dalam data mining penting untuk menghasilkan keputusan yang dapat dipahami oleh berbagai orang. Presentasi dan visualisasi diperlukan untuk menjelaskan hasil proses data mining kepada mereka yang mungkin tidak familiar dengan konsep tersebut.

2.1.3 *Scraping Data*

Scraping data adalah cara mengumpulkan informasi secara otomatis dari situs web atau sumber online menggunakan program komputer (Flores dkk., 2020). Tujuannya adalah mengekstraksi data dengan efisien dan menghindari upaya manual dalam mengumpulkan informasi dari berbagai halaman web. Dengan menggunakan teknik ini, pengguna dapat mengakses, menarik, dan menyimpan

data dalam format yang dapat diolah lebih lanjut, memungkinkan analisis mendalam atau penggunaan data untuk keperluan bisnis dan penelitian (Fahrudin dkk., 2023).

Dalam proses *scraping data*, elemen-elemen khusus seperti teks, gambar, atau tabel diidentifikasi dan ditarik dari halaman web sesuai kebutuhan pengguna. Penggunaan bot atau program komputer memungkinkan pengumpulan data dalam jumlah besar dengan cepat dan otomatis. Meski demikian, perlu diingat bahwa praktek *scraping data* harus mematuhi aturan privasi dan etika penggunaan data untuk menghindari pelanggaran hukum dan norma-norma yang berlaku dalam pemanfaatan informasi online (Setianto, 2020).

Keberhasilan *scraping data* tergantung pada keterampilan dan pemahaman penggunaannya dalam mengelola alat scraping serta memastikan bahwa kegiatan ini dilakukan dengan etika dan sesuai dengan hukum. Walaupun scraping data menawarkan cara yang efisien untuk mendapatkan informasi, tetap penting untuk memperhatikan hak cipta, peraturan penggunaan data, dan kebijakan privasi agar penggunaan teknik ini sesuai dengan standar dan peraturan yang berlaku (Santoso, 2023).

2.1.4 Clustering

Clustering adalah suatu metode dalam analisis data dan pembelajaran mesin yang dipergunakan dalam mengelompokkan beberapa kelompok objek atau data yang sejenis berdasarkan jenis jenis tertentu (Pratiwi et al., 2020). Metode clustering mempartisi data ke dalam kelompok sehingga data yang memiliki karakteristik yang sama di kelompokkan ke dalam satu kelompok. Proses clustering bertujuan untuk mengidentifikasi kelas-kelas taksonomi atau batryologi, atau

melakukan analisis topologi terhadap data yang ada. Jika dilihat dari perspektif data mining, *clustering* bukanlah suatu proses klarifikasi. Ini disebabkan oleh perbedaan mendasar dengan proses klarifikasi, di mana data dikelompokkan ke dalam kelas-kelas yang sudah diketahui sebelumnya. (Priyatman dkk., 2019).

Berdasarkan hasil penelitian yang sudah dilakukan, telah dibuat sistem pengelompokan data biomedis dengan terinci dan mengoptimalkan pola-pola yang tersembunyi dalam dataset. Landasan teori yang telah dikemukakan oleh Jain mengenai algoritma clustering dan metrik evaluasi, serta kerangka dari pengelompokan dari Han tentang data mining, merupakan beberapa gabungan dari berbagai clustering yang mencakup *K-Means*, *Hierarchical*, dan *DBSCAN*. Hasil penelitian ini menunjukkan bahwa penggunaan metode *clustering* dapat memberikan hasil pemahaman yang lebih akurat terhadap struktur internal data *biomedis*. Penelitian ini berhasil mengidentifikasi kelompok-kelompok pasien berdasarkan kelas medis yang serupa, perawatan, dan kelompok resiko tertentu. Selain itu, eksperimen dengan parameter *metode clustering* telah memberikan pengetahuan tambahan tentang penyesuaian parameter dapat mempengaruhi hasil clustering dalam hal spesifik data *biomedis* (Rahutomo dkk., 2022).

Berikut adalah beberapa kelebihan dari pendekatan ini:

1. Segmentasi Data Otomatis

Dapat segmentasi data secara otomatis berdasarkan kesamaan atau pola yang ada di dalamnya. Hal ini membantu mengidentifikasi kelompok data yang mungkin sulit ditemukan secara manual.

2. Skalabilitas

Membuat dataset yang besar dengan skalabel dan dapat digunakan untuk analisis data dalam jumlah besar.

3. Penemuan Struktur Internal

Memungkinkan pemahaman lebih dalam terhadap hubungan dan pola yang mungkin tersembunyi di dalamnya.

4. Fleksibilitas dalam Pemilihan Algoritma

Berdasarkan sifat data dan tujuan analisis fleksibilitas dapat memberikan mengenai mengadaptasi metode clustering sesuai kebutuhan spesifik.

5. Pembuatan Keputusan yang Informatif

Membantu dalam pembuatan keputusan yang lebih informatif dan terarah.

Metode clustering menjadi alat yang sangat berharga dalam memahami struktur dan pola data, serta mendukung pengambilan keputusan berdasarkan informasi yang dihasilkan dari pengelompokan data (Han dkk., 2022).

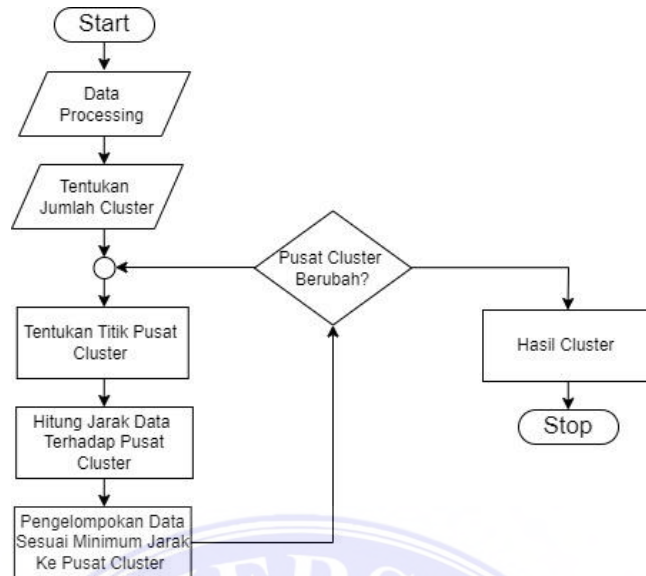
Metode clustering memiliki kelemahan, termasuk sensitivitas terhadap pemilihan parameter dan bentuk data, ketidakcocokan untuk data dengan noise atau outlier, serta keterbatasan dalam menangani pola non-linear. Interpretasi hasil clustering dapat bersifat subjektif, dan metode ini cenderung tidak efektif dalam menangani perubahan dinamis pada data seiring waktu. Menentukan jumlah cluster yang optimal juga bisa menjadi tantangan, dan beberapa algoritma clustering memiliki asumsi yang dapat membatasi kemampuannya untuk menangani pola hierarki yang kompleks. Oleh karena itu, pemilihan *metode clustering* dan

interpretasi hasilnya perlu dilakukan dengan hati-hati, dengan mempertimbangkan sifat data dan tujuan analisis secara menyeluruh (Ezugwu dkk., 2022).

2.2 *Algoritma K-Means*

Algoritma K-Means merupakan metode dalam bidang data mining yang mampu mengelompokkan data atau melakukan clustering terhadap suatu dataset menjadi satu atau lebih cluster. Dengan menggunakan metode ini, data yang memiliki karakteristik serupa akan dikelompokkan ke dalam satu cluster, sementara data yang memiliki karakteristik berbeda akan ditempatkan dalam kelompok yang berbeda (Dinata dkk., 2020).

Metode K-Means adalah salah satu algoritma non-hierarki yang awalnya memilih sebagian struktur populasi sebagai titik pusat *cluster* awal. Pada tahap awal, pusat cluster dipilih secara acak dari sejumlah data populasi. Kemudian, *K-Means* mengevaluasi setiap komponen dalam populasi data, mengidentifikasi komponen tersebut sebagai salah satu pusat cluster yang paling sesuai berdasarkan jarak minimum antara komponen tersebut dengan setiap *cluster*. Posisi pusat *cluster* kemudian dihitung ulang hingga semua komponen data dikelompokkan ke dalam pusat *cluster* masing-masing. Akhirnya, terbentuklah posisi pusat cluster baru setelah iterasi di dalam *cluster* selesai (Norshahlan dkk., 2023)



Gambar 2.1 Algoritma Sederhana K-Means

K-Means memiliki kecepatan clustering yang tinggi dan lebih efektif pada pengumpulan data yang besar, namun memiliki ketepatan *clustering* yang buruk, rentan terhadap *noise* dan data yang tersembunyi, dan nilai perlu dihitung dahulu (Hartatik dkk., 2023). Proses pengelompokan dataset bertujuan untuk mengidentifikasi pola yang signifikan dalam kumpulan data tanpa label. Dalam kelompok yang sama, sampel-sampel memiliki karakteristik serupa, sementara setiap kelompok memiliki perbedaan karakteristik dengan kelompok lainnya. Sebagai metode yang efektif untuk tugas ini, *K-Means* digunakan untuk mengelompokkan dataset (Muflikhah dkk., 2023)

Algoritma K-Means dijelaskan sebagai berikut (Hutagalung, 2022) :

1. Memasukkan data.
2. Menentukan total klaster.
3. Mengambil data random untuk *centroid*.
4. Menghitung jarak terdekat, menggunakan rumus berikut.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

5. Menghitung pusat kluster dengan anggota kluster yang baru.
6. Proses kluster dianggap selesai jika pusat kluster tidak mengalami perubahan, namun, jika terjadi perubahan pada pusat kluster, langkah-langkah perhitungan jarak diulangi kembali hingga pusat kluster tidak mengalami perubahan lagi.

2.3 Kriminal Umum

Kriminal Umum adalah tindakan yang dilakukan oleh individu, kelompok, atau komunitas yang melanggar norma hukum atau mencuri barang yang bukan haknya, dapat disebut sebagai kriminalitas. Kejadian tindak kriminal sering kali dipengaruhi oleh berbagai faktor kondisi, seperti penyebaran kemiskinan yang meningkat, lingkungan yang memberikan dukungan untuk perilaku kriminal, ketidaksetaraan sosial, tekanan mental, dan adanya rasa benci terhadap seseorang. (Efendi dkk, 2024).




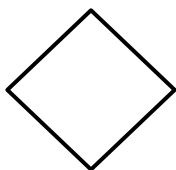
Kriminalitas merupakan suatu perbuatan yang berpotensi merugikan baik jiwa pelaku maupun orang lain. Dampak dari tindakan kejahatan ini sangat signifikan terhadap kehidupan sosial, karena dapat memicu timbulnya trauma dan konsekuensi negatif lainnya pada korban (Isnawan, 2023)


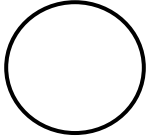
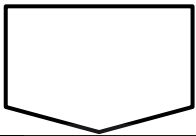

Dampak dari tindak kriminalitas dapat memengaruhi tingkat keamanan masyarakat dan mengancam kesejahteraan baik secara fisik maupun psikologis. Oleh karena itu, penelitian ini mengadopsi pendekatan data mining untuk memproses dan mengkategorikan data kriminalitas. Tujuan utama adalah untuk menilai jumlah kejadian tindak kriminal di Indonesia dengan memanfaatkan teknik data mining, khususnya algoritma K-Means (Wibowo dkk., 2023).

2.4 Flowchart

Flowchart biasanya disebut dengan diagram alir merupakan jenis diagram yang sering dipertunjukkan pada algoritma atau langkah-langkah pengarahan yang berurutan didalam sistem. Seorang peneliti sistem sering menggunakan *flowchart* untuk bukti lampiran sebagai penjelasan gambaran logis sebuah sistem yang membangun kepada *programmer*. Begitu pula *flowchart* dapat membuat sistem. *Flowchart* sering digambarkan dengan menggunakan simbol-simbol. Simbol yang diwakili merupakan suatu proses yang sering terjadi. Tetapi untuk menyatukan satu proses ke proses berikutnya yang digambarkan dengan menerapkan garis penghubung. Setiap rangkaian proses sering digambar menjadi lebih jelas. Selain itu, saat ada penambahan proses baru sering dilakukan akan dapat dengan mudah menggunakan flowchart (Rosaly & Prasetyo, 2019).

Tabel 2.1 Tabel Flowchart

Simbol	Nama	Keterangan
	Terminator	Awal atau akhir konsep (prosedur)
	Process	Proses Operasional
	Document	Dokumen atau laporan berupa print out
	Decision	Keputusan atau sub-point. Garis yang terhubung dengan bentuk <i>decision</i> merujuk pada situasi-situasi yang berbeda sesuai dengan keputusan yang digambarkan

Simbol	Nama	Keterangan
	Data	Input dan Output
	On-Page Reference/Connector	Penghubung alur dalam halaman yang sama
	Off-Page Reference Off-Page Connector	Penghubung alur dalam halaman yang berbeda
	Flow	Arah alur dalam konsep (prosedur)

2.5 Davies-Bouldin Index (DBI)

Metode ini diperkenalkan oleh *D.L. Davies* dan *D.W. Bouldin* dan nama metode ini menggabungkan kedua nama yaitu *Davies-Bouldin Index (DBI)*. *Davies-Bouldin Index* adalah salah satu metode evaluasi cluster yang di dapat pada suatu metode pengelompokan didasarkan pada niat kohosi dan separasi (Ferdiansyah & Chotijah, 2024). Dalam mengevaluasi kualitas *cluster* peran penting dalam memahami kinerja algoritma clustering adalah salah satu tujuan DBI yang mengukur rata-rata anantara pengelompokan dan membantu dalam pemisahan dan kompakness *cluster* (Umagapi dkk., 2023). Berikut adalah langkah-langkah perhitungan *Davies-Bouldin Index*:

1. Sum Of Square Within-Cluster(SSW)

Untuk mengetahui kohesi dalam sebuah cluster ke-*I* salah satunya adalah dengan menghitung nilai dari Sum Of Square Within-Cluster (SSW).

Dengan rumus sebagai berikut:

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(X_j, C_j) \tag{2.1}$$

Dimana:

m_i = jumlah data dalam *cluster* ke- i

c_i = *centroid cluster* ke- i

$d(X_j, C_j)$ = jarak setiap data ke *centroid* i yang dihitung menggunakan jarak jarak *euclidance*

2. Sum Of Square Between-Cluster (SSB)

Perhitungan *Sum Of Square Between-Cluster (SSB)* bertujuan untuk mengetahui separasi atau jarak antar *cluster*. Dengan rumus perhitungan sebagai berikut:

$$SSB_{ij} = d(X_i, X_j) \quad (2.2)$$

Dimana :

$d(X_i, X_j)$ = jarak antara data ke i dengan data ke j di *cluster* lain.

3. Ratio (Rasio)

Perhitungan rasio ($R_{i,j}$) ini bertujuan untuk mengetahui nilai perbandingan antara *cluster* ke- i dan *cluster* ke- j untuk menghitung nilai rasio yang dimiliki oleh masing-masing *cluster*. Untuk menentukan nilai rasio dengan rumus sebagai berikut:

$$R_{ij, \dots, n} = \frac{SSW_i + SSW_j + \dots + SSW_n}{SSB_{i,j} + \dots + SSB_{n,i,j}} \quad (2.3)$$

Dimana :

SSW_i = *Sum Of Square Within-Cluster* pada *centroid* i

$SSB_{i,j}$ = *Sum of Square Between Cluster* data ke i dengan j pada *cluster*

yang berbeda

4. *Davies Bouldin Index (DBI)*

Nilai rasio yang diperoleh dari persamaan 2.5 digunakan untuk mencari nilai DBI dengan menggunakan perhitungan sebagai berikut :

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (Ri, j, \dots k) \tag{2.4}$$

Dimana, Ri, j merupakan *ratio* dari nilai SSW dan SSB melalui perhitungan 2.3 dan perhitungan 2.4 maka dapat diketahui k adalah jumlah *cluster*. Dari perhitungan *Davies Bouldin Index (DBI)* dapat disimpulkan bahwa jika semakin kecil nilai *Davies Bouldin Index (DBI)* yang diperoleh (non negatif ≥ 0) maka *cluster* tersebut semakin baik.

2.6 Penelitian Terdahulu

Penelitian sebelumnya merujuk pada studi yang telah dilaksanakan oleh peneliti sebelum penulis, yang memiliki relevansi dengan proyek penelitian yang akan dijalankan. Penelitian sebelumnya menjadi sumber informasi penting, memberikan referensi yang berguna dalam penyusunan dan penelaahan studi yang akan dilakukan oleh penulis. Sebagai contoh, penelitian sebelumnya mengenai sistem rekomendasi menggunakan *algoritma K-Means* menjadi pedoman dan referensi yang membimbing penulis dalam merancang penelitian ini.

Tabel 2.2 Tabel Penelitian Terahulu

No	Penulis dan Tahun	Judul	Hasil
1	Fitria Rahmadayanti dan Rika Rahayu (2023).	Penerapan Metode Data Mining pada Kasus Kriminalitas Indonesia	Algoritma k-means mampu mengetahui pengelompokan provinsi mana yang memiliki tingkat kriminalitas di Indonesia dari data tahun 2019 sampai dengan data tahun 2021 berdasarkan jenis kejahatan

No	Penulis dan Tahun	Judul	Hasil
2	Ronal Watrianthos, Sudi Suryadi, Kusmanto, Samsi (2023)	Pemetaan Tingkat Kriminalitas di Indonesia Analisis Spasial dengan Pendekatan SIG pada Tingkat Provinsi	Dapat memastikan konfigurasi yang terukur dan ideal untuk pemetaan berbasis provinsi.
3	Anis Hoerunnisa, dkk (2024)	Komparasi <i>Algoritma K-Means</i> dan <i>K-Medoids</i> dalam analisis pengelempokan daerah Rawan Kriminalitas di Indonesia	<i>Algoritma K-Means</i> dan <i>K-Medoids Clustering</i> dapat mengelompokkan daerah rawan tindak kriminal di Indonesia.
4	Yosi Mayona, Relita Buaton, dan Magdalena Simanjatak (2022)	<i>Data Mining Clustering</i> Tingkat Kejahatan Dengan <i>Metode Algoritma K-Means</i> (Studi Kasus : Kejaksaan Negeri Binjai)	Dalam pengelompokan metode <i>K-Means</i> tindakan kejahatan dalam kejaksaan Negeri Binjai memberikan performa yang baik.
5	Nur Kholidin, Nana Suarna, Willy Prihartono (2024)	Segmentasi Gaya Hidup Mahasiswa Menggunakan <i>Algoritma K-Means Clustering</i>	Dalam penelitian ini, ditemukan bahwa cluster terbaik dan paling optimal adalah $K=2$. Hasil ini mencerminkan kecenderungan bahwa mahasiswa dapat dikelompokkan menjadi dua kategori utama berdasarkan gaya hidup mereka.
6	L.Suriani (2020)	Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan <i>Data Mining Algoritma K-Means Clustering</i>	Melakukan pembagian kategori berdasarkan data-data JTP, JPTP, kemudian akan dikategorikan menjadi daerah rawan, sedang, tidak rawan, dengan nilai-nilai yang didapat pada setiap cluster
7	Mohammad Ferdiansyah, Umi Chotijah (2024)	Implementasi <i>Algoritme K-Means++</i> Untuk <i>Clustering</i> Penjualan Bahan Bangunan	Penentuan titik awal atau centroid sangat berpengaruh terhadap hasil dari clustering, peneliti sendiri menggunakan <i>Algoritme KMeans++</i> untuk mengatasi masalah pemilihan

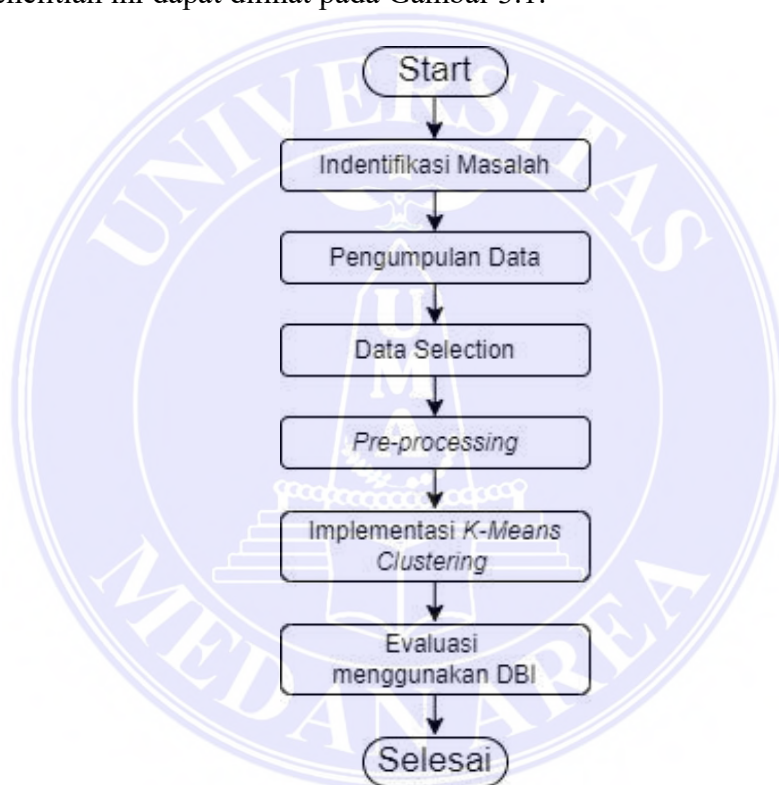
No	Penulis dan Tahun	Judul	Hasil
			pusat <i>cluster</i> awal agar hasil menjadi optimal. dan juga dilakukan evaluasi cluster untuk menentukan jumlah <i>centroid</i> terbaik.
8	Lia Awaliyah , Nining Rahaningsih , Raditya Danar Dana (2024)	Implementasi Algoritma <i>K-Means</i> dalam Analisis <i>Cluster</i> Korban Kekerasan di Provinsi Jawa Barat	<i>Algoritma k-means</i> dapat diimplementasikan untuk mengelompokkan korban kekerasan di Provinsi Jawa Barat menggunakan tools <i>Rapidminer</i> . Pengujian dilakukan dengan membentuk model <i>cluster</i> K=2 sampai K=10 menggunakan tipe pengukuran <i>Numerical Measures</i> dan metrik jarak <i>Euclidean Distance</i> . Hasil pengujian menggunakan metrik evaluasi <i>Davies Bouldin Index</i> (DBI) dengan nilai validitas sebesar 0.157, diperoleh pada K=2, dengan karakteristik, cluster 0 didominasi oleh korban dengan tingkat pendidikan NA, tidak sekolah, SD, SLTA, dan TK pada jenis kelamin laki-laki, dan tingkat pendidikan NA, SD, SLTA, Perguruan Tinggi, TK, dan PAUD pada korban dengan jenis kelamin perempuan

BAB III

METODE PENELITIAN

3.1 Tahapan Penelitian

Berikut tahapan penelitian yang akan menjadi pedoman untuk mengelompokkan kriminalitas umum di Sumatera Utara. Tahapan yang digunakan pada penelitian ini dapat dilihat pada Gambar 3.1.



Gambar 3.1 Tahapan Penelitian

Tujuan dari penelitian ini adalah mengidentifikasi masalah mengenai pengelompokan kriminalitas umum di Provinsi Sumatera Utara. Untuk mengetahui hasil pengelompokan tindakan kriminal diperlukannya penerapan *data mining* yaitu menggunakan algoritma *k-means* menggunakan Bahasa pemrograman *Python*.

3.2 Pengumpulan Data

Tahap pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan dalam mencapai tujuan penelitian. Teknik pengumpulan data dalam penelitian ini dilakukan dengan cara wawancara dan observasi langsung di Polda Sumatera Utara.

1. Wawancara

Wawancara dilakukan untuk mencari informasi mengenai proses pengelompokan tindakan kriminal di Polda Provinsi Sumatera Utara.

2. Observasi

Observasi penelitian melakukan pengamatan secara langsung kegiatan pengolahan data tindakan kriminal .

3. Dokumentasi

Pengambilan data yang di perlakukn untuk mengelompokan tindak kriminal, yakni data input tindakan.

3.2.1 Sampel Data

Tabel berikut digunakan untuk penelitian:

Tabel 3.1 Tabel Kriminalitas

No	Daerah	JTP	JPTP
1	DIT RESKRIMUM	970	623
2	RES BATU BARA	1.015	714
3	RESTABES MEDAN	8.718	5.664
4	RESTA DELI SERDANG	1.412	1.093
5	REST. TINGGI	997	904
6	RES SERGE	1.435	1.299
7	RES LANGKAT	1.573	1.116
8	RES BINJAI	1.594	1.131
9	RES TANAH KARO	789	419
10	RES SIMALUNGUN	1.548	941

Keterangan:

JTP : Jumlah Tindak Pidana

JPTP : Jumlah Penyelesaian Tindak Pidana

1	Bulan	Tempat	JENIS KEJAHATAN	CT	CC
4	Februari	POLRESTABES MEDAN.xlsx	PENGGELAPAN	99	58
5	Septembe	POLRES LAB. BATU.xlsx	PENCURIAN RINGAN	98	54
6	Februari	POLRESTABES MEDAN.xlsx	LAIN-LAIN (AGAR DIISI TINDAK PIDANANYA)	92	98
7	Juni	POLRES LAB. BATU.xlsx	PENCURIAN RINGAN	90	47
8	Septembe	POLRES PS.xlsx	PENIPUAN / PERBUATAN CURANG	9	4
9	Oktober	POLRES LAB. BATU.xlsx	PEMERASAN DAN PENGANCAMAN	9	6
10	Novembe	POLRES PSP.xlsx	PENCURIAN DENGAN PEMBERATAN	9	
11	Mei	POLRES TOBA.xlsx	PENGANIAYAAN BIASA	9	12
12	Mei	POLRES TAPSEL.xlsx	PENCABULAN	9	4
13	Mei	POLRES T. BALAI.xlsx	PENCURIAN DENGAN PEMBERATAN	9	7
14	Mei	POLRES SIBOLGA.xlsx	PENCURIAN BIASA	9	5
15	Mei	POLRES SERGAI.xlsx	PENGANIAYAAN BERAT	9	7
16	Mei	POLRES PALAS.xlsx	PENGANIAYAAN	9	5
17	Mei	POLRES BELAWAN.xlsx	PENGHAPUSAN KEKERASAN DALAM RUMAH T	9	8
19	Juli	POLRES ASAHAN.xls	PERADILAN ANAK	9	0
20	Februari	POLRES TOBA.xlsx	PENGANIAYAAN BIASA	9	11
21	Februari	POLRES TEBING.xlsx	PENCURIAN DENGAN PEMBERATAN	9	8
22	Februari	POLRES T. KARO.xlsx	CURANMOR	9	1
23	Februari	POLRESTABES MEDAN.xlsx	PENIPUAN / PERBUATAN CURANG	88	51
24	Agustus	POLRES LAB. BATU.xlsx	PENCURIAN DENGAN PEMBERATAN	85	59

Gambar 3.2 Data asli penelitian

3.2.2 Alat dan Bahan

Dalam menyelesaikan penelitian ini, beberapa alat atau perangkat lunak yang mungkin digunakan untuk mendukung penelitian tersebut adalah sebagai berikut:

1. Spesifikasi Kebutuhan *Hardware*
 - a. Processor : Intel Core i3-1115G4
 - b. Ram : 8 GB
 - c. *SSD* : 256 GB
 - d. *System tipe* : 64-bit *Operating System*
2. Spesifikasi Kebutuhan *Software*
 - a. Sistem Operasi : *Microsoft windows 11*
 - b. Bahasa Pemrograman : *Visual Studio 2012*
 - c. *SAP Crystal Report*

3.3 Implementasi Metode K-Means

Metode yang akan digunakan dalam pengelompokan data kriminalitas umum di provinsi Sumatera Utara adalah metode *K-Means Clustering*. Variabel yang digunakan adalah jumlah tindakan pidana dan jumlah penyelesaian tindakan pidana.

3.3.1 Preprocessing Data

Data yang akan dikelompokkan adalah data daerah di provinsi Sumatera Utara. Dimana data ini terdiri dari nama daerah, jumlah tindakan pidana dan jumlah penyelesaian tindakan pidana yang dapat dilihat pada Tabel 3.2.

Tabel 3.2 Data

No	Daerah	JTP	JPTP
1	DIT RESKRIMUM	970	623
2	RES BATU BARA	1.015	714
3	RESTABES MEDAN	8.718	5.664
4	RESTA DELI SERDANG	1.412	1.093
5	REST. TINGGI	997	904
6	RES SERGE	1.435	1.299
7	RES LANGKAT	1.573	1.116
8	RES BINJAI	1.594	1.131
9	RES TANAH KARO	789	419
10	RES SIMALUNGUN	1.548	941
11	RES P.SIANTAR	867	494
12	RES ASAHAN	1.802	1.486
13	RES TJ. BALAI	666	481
14	RES LAB. BATU	4.490	3.333
15	RES TAPUT	399	260
16	RES TOBASA	483	397
17	RES SAMOSIR	229	194
18	RES HUMBAHAS	154	112
19	RES DAIRI	692	689
20	RES P. BARAT	80	53
21	RES TAPSEL	1.029	744

No	Daerah	JTP	JPTP
22	RES P. SIDIMPUAN	510	398
23	RES MADINA	625	318
24	RES TAPTENG	724	436
25	RES SIBOLGA	344	246
26	RES NIAS	769	326
27	RES NISEL	343	157
28	RES PEL. BELAWAN	1.839	1.354
29	RES PALAS	411	255

3.3.2 Iterasi-1

Iterasi 1 adalah langkah awal yang digunakan dalam menyelesaikan suatu permasalahan. Dimana iterasi akan dilakukan berulang sampai masalah dapat dipecahkan.

3.3.2.1 Inisialisasi Pusat Cluster Awal

Setelah data dinormalisasi maka dilanjutkan dengan telah diperoleh maka dapat melakukan perhitungan manual metode *K-Means* dimana pada kasus ini telah ditentukan jumlah *centroid* yang digunakan adalah 3.

Untuk menentukan pusat (*centroid*) awal ditentukan secara acak. Pada kasus ini pusat (*centroid*) awalnya diambil dari data dengan total nilai transformasi tertinggi (8.718 ; 5.664), sedang (4.399 ; 2.859) dan terendah (80 ; 53). Adapun pusat *cluster* baru dapat dilihat pada Tabel 3.3.

Tabel 3.3 Pusat (*Centroid*) Awal Cluster

No	Cluster	K1	K2	Keterangan
1	Centroid 1	8.718	5.664	Tinggi
2	Centroid 2	4.399	2.859	Sedang
3	Centroid 3	80	53	Rendah

3.3.2.2 Jarak Euclidian Distance

Perhitungan jarak data yang telah dinormalisasi dengan pusat *cluster* menggunakan rumus *Euclidian Distante*. Perhitungan jarak dari data ke - 1 terhadap pusat *cluster* adalah:

$$d(1,1) = \sqrt{(970 - 8718)^2 + (623 - 5664)^2} = 9243,5483$$

$$d(1,2) = \sqrt{(970 - 4399)^2 + (623 - 2858,5)^2} = 4093,3484$$

$$d(1,3) = \sqrt{(970 - 80)^2 + (623 - 53)^2} = 1056,8822$$

Dan Seterusnya dilakukan jarak untuk data ke - 2 sampai data ke - 29. Kemudian akan didapatkan hasil perhitungan jarak setiap data terhadap pusat *cluster* baru yang dapat dilihat pada Tabel 3.4.

Tabel 3.4 Jarak data terhadap *cluster* pada iterasi ke - 1

No	C1	C2	C3	Jarak Terpendek	Cluster	JT ²
1	9243,548	4093,348	1056,882	1056,882	C3	1117000
2	9156,348	4006,287	1145,053	1145,053	C3	1311146
3	0	5150,203	10300,406	0	C1	0
4	8618,102	3469,749	1689,918	1689,918	C3	2855824
5	9070,361	3923,477	1251,036	1251,036	C3	1565090
6	8490,896	3349,229	1840,799	1840,799	C3	3388541
7	8469,671	3320,027	1832,762	1832,762	C3	3359018
8	8443,901	3294,280	1858,569	1858,569	C3	3454280
9	9506,791	4356,978	797,895	797,895	C3	636637
10	8585,781	3435,842	1715,683	1715,683	C3	2943568
11	9400,378	4250,398	902,136	902,136	C3	813850
12	8080,021	2937,374	2240,262	2240,262	C3	5018773

No	C1	C2	C3	Jarak Terpendek	Cluster	JT ²
13	9575,917	4425,810	725,658	725,658	C3	526580
14	4827,996	483,147	5496,044	483,147	C2	233431,250
15	9920,130	4769,927	380,276	380,276	C3	144610
16	9775,301	4625,369	529,854	529,854	C3	280745
17	10098,714	4948,582	205,139	205,139	C3	42082
18	10206,214	5056,015	94,641	94,641	C3	8957
19	9442,844	4295,181	882,632	882,632	C3	779040
20	10300,406	5150,203	0	0	C3	0
21	9128,369	3978,443	1173,917	1173,917	C3	1378082
22	9752,026	4601,998	551,294	551,294	C3	303925
23	9699,297	4549,419	606,012	606,012	C3	367250
24	9551,755	4401,606	749,283	749,283	C3	561425
25	9973,896	4823,710	327,024	327,024	C3	106945
26	9575,011	4426,111	741,114	741,114	C3	549250
27	10023,356	4873,319	282,816	282,816	C3	79985
28	8117,681	2969,364	2187,849	2187,849	C3	4786682
29	9912,796	4762,600	387,769	387,769	C3	150365
					SSW	36763081,25

Perhitungan jarak antar pusat cluster dapat dilakukan dengan menerapkan persamaan *Euclidian Distance* terhadap c1 dengan c2, c1 dengan c3 dan c2 dengan c3.

$$m(1,2) = \sqrt{(8718 - 4399)^2 + (5664 - 2858,5)^2} = 5150,203$$

$$m(1,3) = \sqrt{(8718 - 80)^2 + (5664 - 53)^2} = 10300,406$$

$$m(2,3) = \sqrt{(4399 - 80)^2 + (2858,5 - 53)^2} = 5150,203$$

$$SSB = 5150,203 + 10300,406 + 5150,203 = 20600,812$$

$$SSW = 36763081,25$$

$$DBI = \frac{20600,812}{36763081,25} = 0,000560366852272866$$

Karena perhitungan masih iterasi 1 maka dilanjutkan ke iterasi 2.

3.3.2.3 Perhitungan Pusat Cluster Baru

Untuk mendapatkan pusat *cluster* yang baru diperlukan untuk mengelompokkan data berdasarkan jarak terdekat dengan pusat *cluster*. kemudian pusat *cluster* baru di hitung berdasarkan data anggota tiap – tiap *cluster* dan membagikan dengan jumlah anggota masing-masing *cluster*.

1. Pusat Cluster C1

Data yang telah dinormalisasi di pisahkan sesuai dengan kelompok pada iterasi 1. Setelah dipisahkan, maka data dijumlahkan dan dibagi dengan jumlah data dalam kelompok atau biasa disebut nilai rata-rata. Pusat cluster C1 dapat dilihat pada Tabel 3.5.

Tabel 3.5 Data Kelompok C1

Data Ke	K1	K2
3	8718	5664
C1	8718	5664

2. Pusat Cluster C2

Pusat cluster C2 dapat dilihat pada Tabel 3.6.

Tabel 3.6 Data Kelompok C2

Data ke	K1	K2
14	4490	3333
C2	4490	3333

3. Pusat Cluster C3

Pusat cluster C3 dapat dilihat pada Tabel 3.7.

Tabel 3.7 Data Kelompok C3

Data Ke	K1	K2
1	970	623
2	1015	714
4	1412	1093
5	997	904
6	1435	1299
7	1573,000	1116,000
8	1594,000	1131,000
9	789,000	419,000
10	1548,000	941,000
11	867,000	494,000
12	1802,000	1486,000
13	666,000	481,000
15	399,000	260,000
16	483,000	397,000
17	229,000	194,000
18	154,000	112,000
19	692,000	689,000
20	80,000	53,000
21	1029,000	744,000
22	510,000	398,000
23	625,000	318,000
24	724,000	436,000
25	344,000	246,000
26	769,000	326,000
27	343,000	157,000
28	1839,000	1354,000
29	411,000	255,000
C3	862,926	616,296

Berdasarkan perhitungan diatas maka diperoleh pusat cluster baru yang didapat dari rata-rata data tranformasi yang dikelompokkan berdasarkan cluster iterasi-1. Pusat cluster baru dapat dilihat pada Tabel 3.8.

Tabel 3.8 Pusat (Centroid) Cluster Baru Iterasi 2

No	Cluster	K1	K2
1	Centroid 1	8718	5664
2	Centroid 2	4490	3333
3	Centroid 3	862,926	616,296

3.3.3 Iterasi-2

Iterasi 2 adalah yang dilakukan setelah iterasi 1. Dimana iterasi akan dilakukan berulang sampai masalah dapat dipecahkan.

3.3.3.1 Jarak Euclidian Distance

Perhitungan jarak data yang telah dinormalisasi dengan pusat *cluster* menggunakan rumus *Euclidian Distante*. Perhitungan jarak dari data ke – 2 terhadap pusat *cluster* adalah:

$$d(1,1) = \sqrt{(970 - 8718)^2 + (623 - 5664)^2} = 9243,5483$$

$$d(1,2) = \sqrt{(970 - 4490)^2 + (623 - 3333)^2} = 4442,3530$$

$$d(1,3) = \sqrt{(970 - 862,926)^2 + (623 - 616,296)^2} = 107,2837$$

Dan Seterusnya dilakukan jarak untuk data ke – 2 sampai data ke – 29. Kemudian akan didapatkan hasil perhitungan jarak setiap data terhadap pusat *cluster* baru yang dapat dilihat pada Tabel 3.9.

Tabel 3.9 Jarak data terhadap cluster pada iterasi ke – 2

No	C1	C2	C3	Jarak Terpendek	Cluster	JT^2
1	9243,548	4442,353	107,284	107,284	C3	11509,797
2	9156,348	4351,412	180,755	180,755	C3	32672,538
3	0,000	4827,996	9337,103	0,000	C1	0,000
4	8618,102	3806,794	727,137	727,137	C3	528728,760

No	C1	C2	C3	Jarak Terpendek	Cluster	JT ²
5	9070,361	4254,538	317,410	317,410	C3	100749,278
6	8490,896	3670,175	890,704	890,704	C3	793353,093
7	8469,671	3663,875	868,279	868,279	C3	753908,982
8	8443,901	3638,079	894,086	894,086	C3	799389,204
9	9506,791	4710,499	210,691	210,691	C3	44390,871
10	8585,781	3791,705	758,129	758,129	C3	574758,982
11	9400,378	4602,831	122,364	122,364	C3	14972,982
12	8080,021	3261,404	1279,939	1279,939	C3	1638244,649
13	9575,917	4770,417	238,924	238,924	C3	57084,908
14	4827,996	0,000	4531,682	0,000	C2	0,000
15	9920,130	5116,601	584,957	584,957	C3	342174,316
16	9775,301	4967,509	438,674	438,674	C3	192434,575
17	10098,714	5292,395	761,706	761,706	C3	580196,241
18	10206,214	5401,457	869,995	869,995	C3	756890,723
19	9442,844	4627,693	185,746	185,746	C3	34501,501
20	10300,406	5496,044	964,508	964,508	C3	930275,723
21	9128,369	4322,203	209,497	209,497	C3	43888,834
22	9752,026	4945,162	414,982	414,982	C3	172209,982
23	9699,297	4901,882	381,562	381,562	C3	145589,427
24	9551,755	4751,354	227,612	227,612	C3	51807,167
25	9973,896	5169,031	637,498	637,498	C3	406403,464
26	9575,011	4784,129	305,113	305,113	C3	93094,019
27	10023,356	5223,465	693,741	693,741	C3	481276,056
28	8117,681	3308,208	1223,490	1223,490	C3	1496927,353
29	9912,796	5110,022	578,595	578,595	C3	334772,056
					SSW	11412205,48

Perhitungan jarak antar pusat cluster dapat dilakukan dengan menerapkan persamaan *Euclidian Distance* terhadap c1 dengan c2, c1 dengan c3 dan c2 dengan c3.

$$m(1,2) = \sqrt{(8718 - 4490)^2 + (5664 - 3333)^2} = 4827,996$$

$$m(1,3) = \sqrt{(8718 - 862,926)^2 + (5664 - 616,296)^2} = 4827,996$$

$$m(2,3) = \sqrt{(4490 - 862,926)^2 + (3333 - 616,296)^2} = 4827,996$$

$$SSB = 4827,996 + 4827,996 + 4827,996 = 18696,782$$

$$SSW = 11412205,481$$

$$DBI = \frac{18696,782}{11412205,481} = 0,001638314510601990$$

$$DBI \text{ Iterasi Ke-1} = 0,000560366852272866$$

Karena DBI pada iterasi-1 dan iterasi 2 belum sama atau jaraknya tidak sama dengan 0, maka lanjut iterasi ke-3.

3.3.3.2 Perhitungan Pusat Cluster Baru

Untuk mendapatkan pusat *cluster* yang baru diperlukan untuk mengelompokkan data berdasarkan jarak terdekat dengan pusat *cluster*. kemudian pusat *cluster* baru di hitung berdasarkan data anggota tiap – tiap *cluster* dan membagikan dengan jumlah anggota masing-masing *cluster*.

1. Pusat Cluster C1

Data yang telah dinormalisasi di pisahkan sesuai dengan kelompokan pada iterasi 2. Setelah dipisahkan, maka data dijumlahkan dan dibagikan dengan jumlah data dalam kelompok atau biasa disebut nilai rata-rata. Pusat cluster C1 dapat dilihat pada Tabel 3.10.

Tabel 3.10 Data Kelompok C1

Data Ke	K1	K2
3	8718	5664
C1	8718	5664

2. Pusat Cluster C2

Pusat cluster C2 dapat dilihat pada Tabel 3.11.

Tabel 3.11 Data Kelompok C2

Data ke	K1	K2
14	4490	3333
C2	4490	3333

3. Pusat Cluster C3

Pusat cluster C3 dapat dilihat pada Tabel 3.12.

Tabel 3.12 Data Kelompok C3

Data Ke	K1	K2
1	970,000	623,000
2	1015,000	714,000
4	1412,000	1093,000
5	997,000	904,000
6	1435,000	1299,000
7	1573,000	1116,000
8	1594,000	1131,000
9	789,000	419,000
10	1548,000	941,000
11	867,000	494,000
12	1802,000	1486,000
13	666,000	481,000
15	399,000	260,000
16	483,000	397,000
17	229,000	194,000
18	154,000	112,000
19	692,000	689,000
20	80,000	53,000
21	1029,000	744,000
22	510,000	398,000
23	625,000	318,000
24	724,000	436,000
25	344,000	246,000
26	769,000	326,000
27	343,000	157,000
28	1839,000	1354,000

Data Ke	K1	K2
29	411,000	255,000
C3	862,926	616,296

Berdasarkan perhitungan diatas maka diperoleh pusat cluster baru yang didapat dari rata-rata data tranformasi yang dikelompokkan berdasarkan cluster iterasi-2. Pusat cluster baru dapat dilihat pada Tabel 3.13.

Tabel 3.13 Pusat (Centroid) Cluster Baru Iterasi 3

No	Cluster	K1	K2
1	Centroid 1	8718	5664
2	Centroid 2	4490	3333
3	Centroid 3	862,926	616,296

3.3.4 Iterasi-3

Iterasi 3 adalah yang dilakukan setelah iterasi 2. Dimana iterasi akan dilakukan berulang sampai masalah dapat dipecahkan.

3.3.4.1 Jarak Euclidian Distance

Perhitungan jarak data yang telah dinormalisasi dengan pusat *cluster* menggunakan rumus *Euclidian Distante*. Perhitungan jarak dari data ke – 2 terhadap pusat *cluster* adalah:

$$d(1,1) = \sqrt{(970 - 8718)^2 + (623 - 5664)^2} = 9243,5483$$

$$d(1,2) = \sqrt{(970 - 4490)^2 + (623 - 3333)^2} = 4442,3530$$

$$d(1,3) = \sqrt{(970 - 862,926)^2 + (623 - 616,296)^2} = 107,2837$$

Dan Seterusnya dilakukan jarak untuk data ke – 2 sampai data ke – 29. Kemudian akan didapatkan hasil perhitungan jarak setiap data terhadap pusat *cluster* baru yang dapat dilihat pada Tabel 3.14.

Tabel 3.14 Jarak data terhadap cluster pada iterasi ke – 3

No	C1	C2	C3	Jarak Terpendek	Cluster	JT ²
1	9243,548	4442,353	107,284	107,284	C3	11509,797
2	9156,348	4351,412	180,755	180,755	C3	32672,538
3	0,000	4827,996	9337,103	0,000	C1	0,000
4	8618,102	3806,794	727,137	727,137	C3	528728,760
5	9070,361	4254,538	317,410	317,410	C3	100749,278
6	8490,896	3670,175	890,704	890,704	C3	793353,093
7	8469,671	3663,875	868,279	868,279	C3	753908,982
8	8443,901	3638,079	894,086	894,086	C3	799389,204
9	9506,791	4710,499	210,691	210,691	C3	44390,871
10	8585,781	3791,705	758,129	758,129	C3	574758,982
11	9400,378	4602,831	122,364	122,364	C3	14972,982
12	8080,021	3261,404	1279,939	1279,939	C3	1638244,649
13	9575,917	4770,417	238,924	238,924	C3	57084,908
14	4827,996	0,000	4531,682	0,000	C2	0,000
15	9920,130	5116,601	584,957	584,957	C3	342174,316
16	9775,301	4967,509	438,674	438,674	C3	192434,575
17	10098,714	5292,395	761,706	761,706	C3	580196,241
18	10206,214	5401,457	869,995	869,995	C3	756890,723
19	9442,844	4627,693	185,746	185,746	C3	34501,501
20	10300,406	5496,044	964,508	964,508	C3	930275,723
21	9128,369	4322,203	209,497	209,497	C3	43888,834
22	9752,026	4945,162	414,982	414,982	C3	172209,982
23	9699,297	4901,882	381,562	381,562	C3	145589,427
24	9551,755	4751,354	227,612	227,612	C3	51807,167
25	9973,896	5169,031	637,498	637,498	C3	406403,464
26	9575,011	4784,129	305,113	305,113	C3	93094,019
27	10023,356	5223,465	693,741	693,741	C3	481276,056
28	8117,681	3308,208	1223,490	1223,490	C3	1496927,353
29	9912,796	5110,022	578,595	578,595	C3	334772,056
					SSW	11412205,48

Perhitungan jarak antar pusat cluster dapat dilakukan dengan menerapkan persamaan *Euclidian Distance* terhadap c1 dengan c2, c1 dengan c3 dan c2 dengan c3.

$$m(1,2) = \sqrt{(8718 - 4490)^2 + (5664 - 3333)^2} = 4827,996$$

$$m(1,3) = \sqrt{(8718 - 862,926)^2 + (5664 - 616,296)^2} = 4827,996$$

$$m(2,3) = \sqrt{(4490 - 862,926)^2 + (3333 - 616,296)^2} = 4827,996$$

$$SSB = 4827,996 + 4827,996 + 4827,996 = 18696,782$$

$$SSW = 11412205,481$$

$$DBI = \frac{18696,782}{11412205,481} = 0,001638314510601990$$

$$DBI \text{ Iterasi Ke-2} = 0,001638314510601990$$

Karena DBI pada iterasi-2 dan iterasi 3 sama atau jaraknya sama dengan 0, maka iterasi selesai dan hasil pengelompokan diambil dari perhitungan jarak iterasi yang terakhir atau iterasi ke-3.

3.3.4.2 Hasil Pengelompokan

Hasil pengelompokan merupakan nilai terpendek dari perhitungan jarak pusat cluster dengan data yang diperoleh dari iterasi yang terakhir. Hasil pengelompokan yang diperoleh dapat dilihat pada Tabel 3.15.

Tabel 3.15 Hasil Pengelompokan

No	Nama Daerah	Kelompok	Keterangan
1	DIT RESKRIMUM	C3	Rendah
2	RES BATU BARA	C3	Rendah
3	RESTABES MEDAN	C1	Tinggi
4	RESTA DELI SERDANG	C3	Rendah
5	REST. TINGGI	C3	Rendah
6	RES SERGE	C3	Rendah

No	Nama Daerah	Kelompok	Keterangan
7	RES LANGKAT	C3	Rendah
8	RES BINJAI	C3	Rendah
9	RES TANAH KARO	C3	Rendah
10	RES SIMALUNGUN	C3	Rendah
11	RES P.SIANTAR	C3	Rendah
12	RES ASAHAN	C3	Rendah
13	RES TJ. BALAI	C3	Rendah
14	RES LAB. BATU	C2	Sedang
15	RES TAPUT	C3	Rendah
16	RES TOBASA	C3	Rendah
17	RES SAMOSIR	C3	Rendah
18	RES HUMBAHAS	C3	Rendah
19	RES DAIRI	C3	Rendah
20	RES P. BARAT	C3	Rendah
21	RES TAPSEL	C3	Rendah
22	RES P. SIDIMPUAN	C3	Rendah
23	RES MADINA	C3	Rendah
24	RES TAPTENG	C3	Rendah
25	RES SIBOLGA	C3	Rendah
26	RES NIAS	C3	Rendah
27	RES NISEL	C3	Rendah
28	RES PEL. BELAWAN	C3	Rendah
29	RES PALAS	C3	Rendah

Keterangan :

Pengelompokan untuk pusat *cluster* pertama (tinggi) ada 1 data sedangkan untuk pusat *cluster* kedua (sedang) ada 1 data dan pada pusat *cluster* ketiga (rendah) ada 27 data. Sehingga dapat disimpulkan bahwa ada 1 daerah yang memiliki tingkat kriminalitas paling tinggi, 1 daerah memiliki tingkat kriminalitas sedang dan 27 daerah yang memiliki tingkat kriminalitas paling rendah.

3.4 Perancangan Sistem

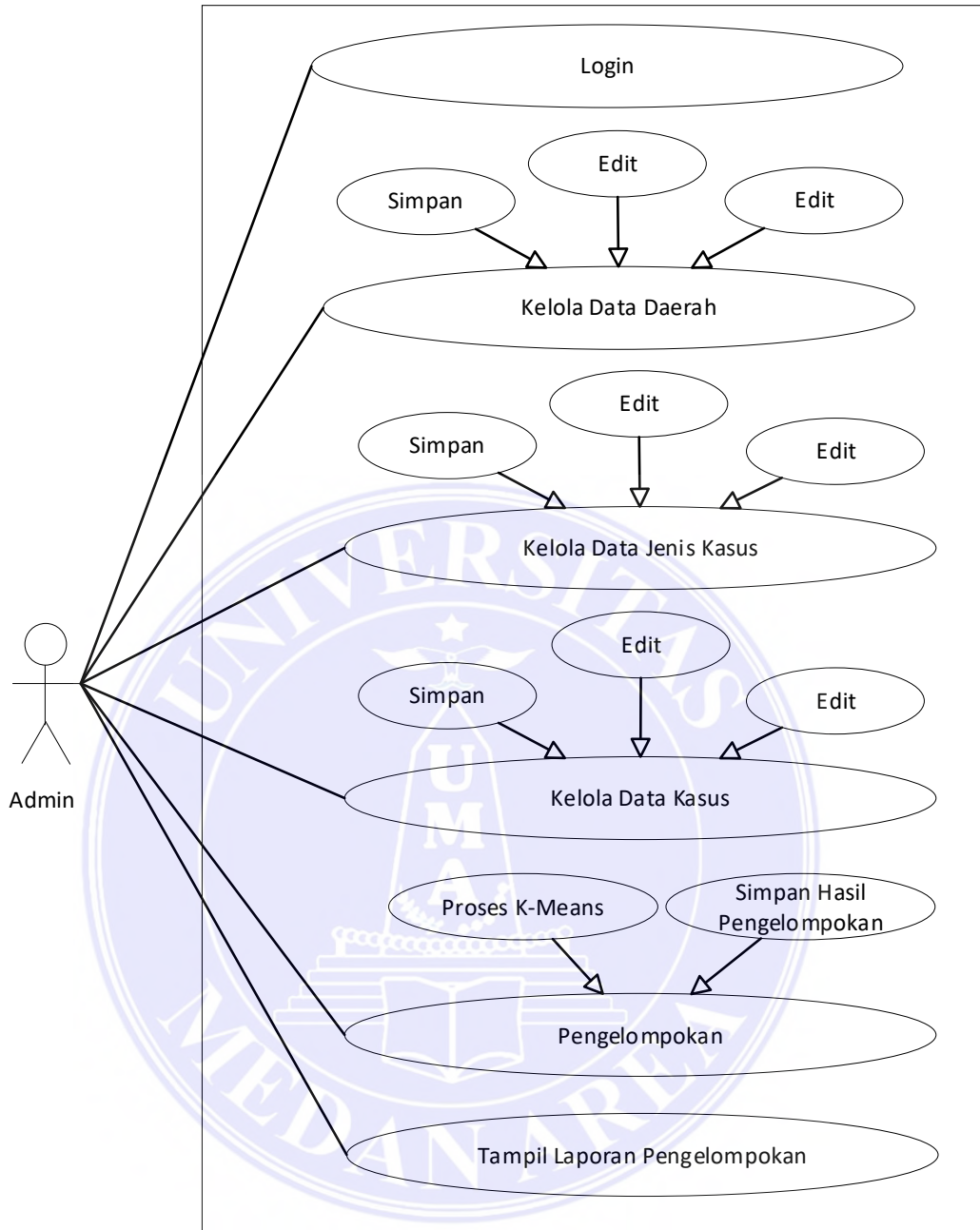
Berdasarkan bagan alur sistem yang berjalan, maka dapat dilakukan pengembangan sistem dengan membangun sebuah *web* untuk memudahkan dan bentuk transparansi secara universal dalam penyajian data yang terkait dengan pengelompokan data kriminalitas umum di provinsi Sumatera Utara, sehingga perlu disimpulkan pengembangan sistem. Perancangan sistem terdiri dari pemodelan UML yang berisi *use case diagram*, *activity diagram*, *class diagram*, perancangan tabel dan perancangan antarmuka.

3.4.1 Perancangan UML

Pemodelan aplikasi atau *website* dalam pengelompokan data kriminalitas umum dirancang menggunakan standarisasi UML. Adapun standarisasi UML yang digunakan memiliki urutan yaitu *use case diagram*, *activity diagram* serta *class diagram*.

1. Use case diagram

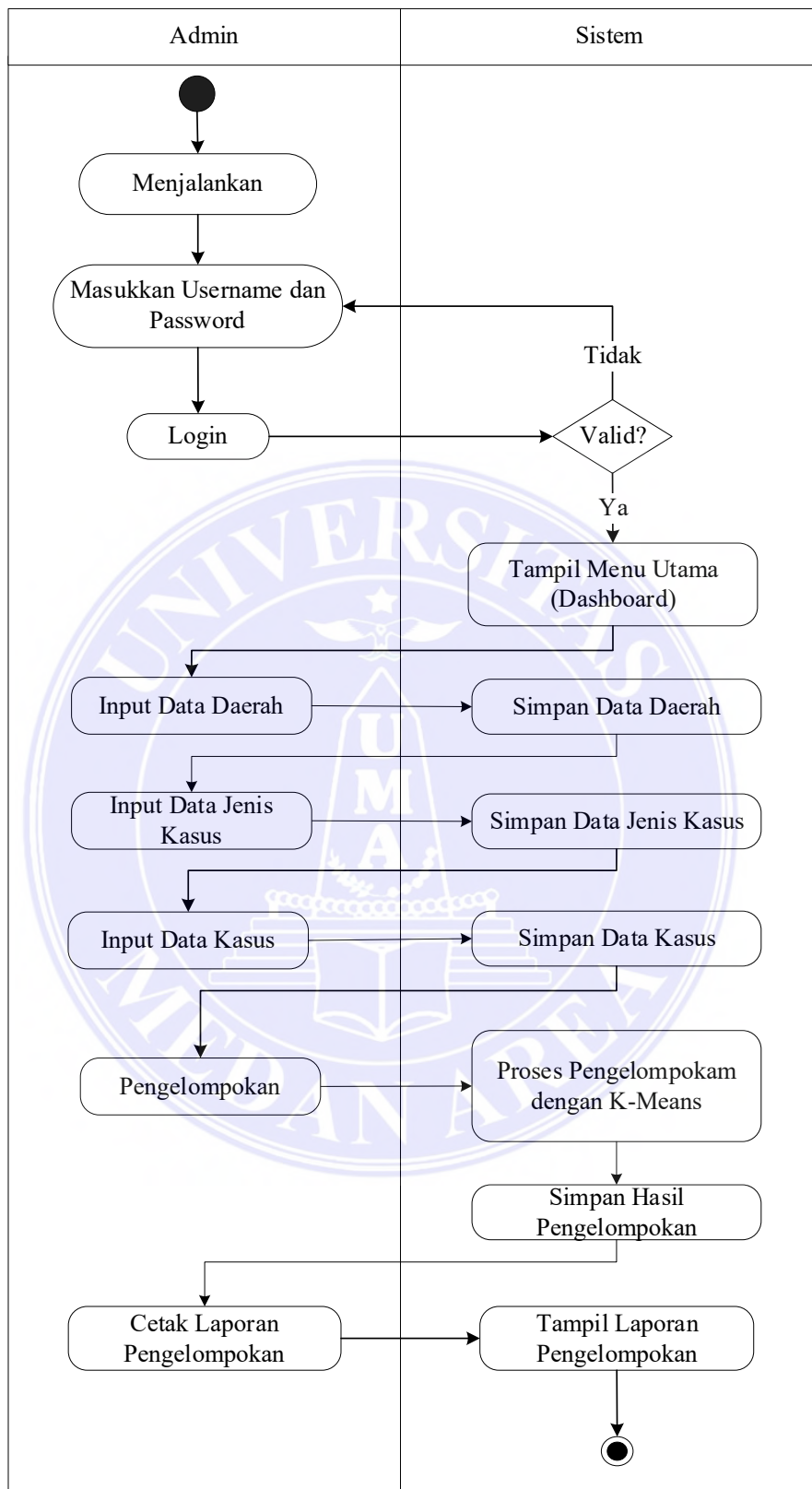
Pemodelan *Use Case Diagram* pengelompokan data kriminalitas umum yang diperoleh dapat dilihat pada Gambar 3.2.



Gambar 3.3 Use Case Diagram

2. Activity diagram

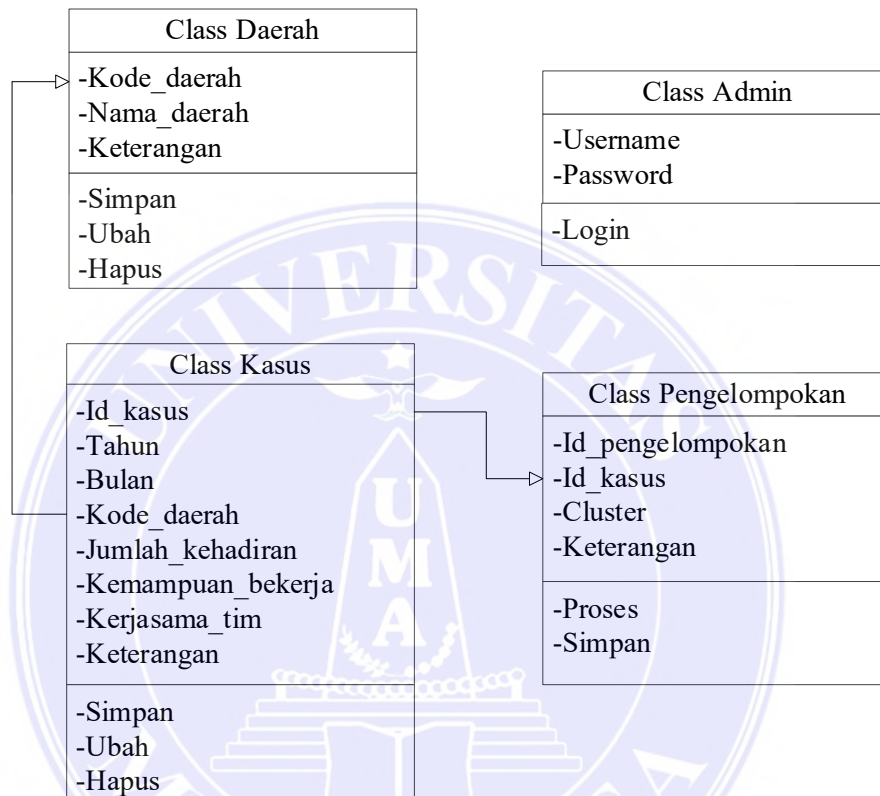
Pemodelan *Activity Diagram* lompkan data kriminalitas umum dapat dilihat pada Gambar 3.3.



Gambar 3.4 Activity Diagram

3. Class diagram

Class Diagram merupakan suatu diagram yang dapat menggambarkan seluruh hubungan dari setiap *class* pada suatu sistem. *Class Diagram* dari sistem yang dirancang dapat dilihat pada Gambar 3.4.



Gambar 3.5 Class Diagram

3.4.2 Perancangan Basis Data

Perancangan basis data digunakan untuk penyimpanan data yang dibutuhkan oleh sistem lompokan data kriminalitas umum. Berikut ini adalah rancangan dari tabel-tabel yang akan digunakan pada basis datanya yaitu:

1. Tabel Admin

Rancangan tabel pengguna dari sistem yang akan dibangun dapat dilihat pada Tabel 3.7.

Tabel 3.16 Admin

<i>No</i>	<i>Field Name</i>	<i>Type Field</i>	<i>Field Size</i>	<i>Description</i>
2	Username	<i>Text</i>	30	<i>Primary Key</i>
3	Password	<i>Text</i>	100	-

2. Tabel Daerah

Rancangan tabel daerah dari sistem yang ingin dibangun dapat dilihat pada Tabel 3.8.

Tabel 3.17 Daerah

<i>No</i>	<i>Field Name</i>	<i>Type Field</i>	<i>Field Size</i>	<i>Description</i>
1	Kode_daerah	<i>Text</i>	10	<i>Primary Key</i>
2	Nama_Daerah	<i>Text</i>	50	-
3	Keterangan	<i>Text</i>	500	-

3. Tabel Jenis Kasus

Rancangan tabel jenis kasus dari sistem yang ingin dibangun dapat dilihat pada Tabel 3.9.

Tabel 3.18 Jenis Kasus

<i>No</i>	<i>Field Name</i>	<i>Type Field</i>	<i>Field Size</i>	<i>Description</i>
1	Kode_jenis	<i>Text</i>	10	<i>Primary Key</i>
2	Kasus	<i>Text</i>	50	-
3	Keterangan	<i>Text</i>	500	-

4. Tabel Kasus

Rancangan tabel kasus dari sistem yang ingin dibangun dapat dilihat pada Tabel 3.10.

Tabel 3.19 Nilai

<i>No</i>	<i>Field Name</i>	<i>Type Field</i>	<i>Field Size</i>	<i>Description</i>
1	Kode_kasus	<i>Text</i>	10	<i>Primary Key</i>

<i>No</i>	<i>Field Name</i>	<i>Type Field</i>	<i>Field Size</i>	<i>Description</i>
2	Tahun	<i>Int</i>	4	
3	Bulan	<i>Int</i>	2	
4	Kode_daerah	<i>Text</i>	10	
5	Kode_kasus	<i>Text</i>	10	
6	JTP	<i>Int</i>	10	
7	JPTP	<i>Int</i>	10	

5. Tabel Pengelompokan

Rancangan tabel pengelompokan dari sistem yang ingin dibangun dapat dilihat pada Tabel 3.11.

Tabel 3.20 Pengelompokan

<i>No</i>	<i>Field Name</i>	<i>Type Field</i>	<i>Field Size</i>	<i>Description</i>
1	Kode_pengelompokan	<i>Text</i>	10	<i>Primary Key</i>
2	Kode_kasus	<i>Text</i>	10	
3	Cluster	<i>Text</i>	5	
4	Keterangan	<i>Text</i>	20	

3.4.3 Perancangan User Interface (UI)

Bentuk *user interface* dari *website* pengelompokan data kriminalitas umum adalah sebagai berikut:

1. Rancangan *Form Login*

Form Login merupakan *form* yang digunakan sebagai media untuk mengisi nama pengguna dan kata kunci. dimana nama pengguna dan kata kunci tersebut merupakan data rahasia untuk dapat menggunakan sistem. *Form login* dapat dilihat pada gambar 3.5.

Gambar 3.6 Rancangan Halaman *Login*

2. Rancangan Form Utama

Form utama berisi menu yang dapat digunakan untuk menampilkan halaman lain yang terkait dengan sistem yang dibangun. Rancangan *form* utama dapat dilihat pada Gambar 3.6.

Gambar 3.7 Rancangan Halaman Utama

3. Rancangan *Form* Daerah

Form daerah berisi tentang data daerah yang akan dikelompokkan berdasarkan kasus yang pernah ditangani. Rancangan *form* daerah dapat dilihat pada Gambar 3.7.

Daerah	-	<input type="checkbox"/>	X
Kode Daerah <input type="text"/> Nama Daerah <input type="text"/> Keterangan <input type="text"/>			
<input type="button" value="Simpan"/> <input type="button" value="Batal"/> <input type="button" value="Edit"/> <input type="button" value="Hapus"/>			
<input type="button" value="Cari"/> <input type="text"/>			
No	Kode Daerah	Nama Daerah	Keterangan
99	XXX	XXX	XXX,XXX,XXX
99	XXX	XXX	XXX,XXX,XXX
99	XXX	XXX	XXX,XXX,XXX
99	XXX	XXX	XXX,XXX,XXX

Gambar 3.8 Rancangan Form Daerah

4. Rancangan Form Jenis Kasus

Form jenis kasus berisi tentang data jenis kasus yang ada sesuai peraturan yang berlaku di Indonesia. Rancangan form jenis kasus dapat dilihat pada Gambar 3.8.

Jenis Kasus	-	<input type="checkbox"/>	X
Kode Jenis <input type="text"/> Nama Kasus <input type="text"/> Keterangan <input type="text"/>			
<input type="button" value="Simpan"/> <input type="button" value="Batal"/> <input type="button" value="Edit"/> <input type="button" value="Hapus"/>			
<input type="button" value="Cari"/> <input type="text"/>			
No	Kode Jenis	Nama Kasus	Keterangan
99	XXX	XXX	XXX,XXX,XXX
99	XXX	XXX	XXX,XXX,XXX
99	XXX	XXX	XXX,XXX,XXX
99	XXX	XXX	XXX,XXX,XXX

Gambar 3.9 Rancangan Form Jenis Kasus

5. Rancangan *Form* Kasus

Form kasus berisi tentang data kasus yang pernah ditangani pada setiap daerah di Sumatera Utara. Rancangan *form* kasus dapat dilihat pada Gambar 3.9.

Kasus	-	<input type="checkbox"/>	X
-------	---	--------------------------	---

Kode Kasus	<input type="text"/>			
Tahun	<input type="text" value="v"/>			
Bulan	<input type="text" value="v"/>			
Daerah	<table border="1"> <tr> <td>Kode Daerah</td> <td>...</td> <td>Nama Daerah</td> </tr> </table>	Kode Daerah	...	Nama Daerah
Kode Daerah	...	Nama Daerah		
Jenis Kasus	<table border="1"> <tr> <td>Kode Jenis</td> <td>...</td> <td>Nama Kasus</td> </tr> </table>	Kode Jenis	...	Nama Kasus
Kode Jenis	...	Nama Kasus		
JTP	<input type="text" value="Kode Daerah"/>			
JPTP	<input type="text" value="Kode Jenis"/>			

Simpan	Batal	Edit	Hapus	Import
--------	-------	------	-------	--------

Cari	<input type="text"/>
------	----------------------

No	Kode Kasus	Tahun	Bulan	Kode Daerah	Nama Daerah	Kode Jenis	Nama Kasus	JTP	JPTP
99	XXXX	XX	XXX	XX	XXX	XX	XXX	999	999
99	XXXX	XX	XXX	XX	XXX	XX	XXX	999	999
99	XXXX	XX	XXX	XX	XXX	XX	XXX	999	999
99	XXXX	XX	XXX	XX	XXX	XX	XXX	999	999

Gambar 3.10 Rancangan *Form* Kasus

6. Rancangan *Form* Proses Pengelompokan

Form proses pengelompokan berisi tentang proses pengelompokan K-Means.

Rancangan *form* pengelompokan dapat dilihat pada Gambar 3.10.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

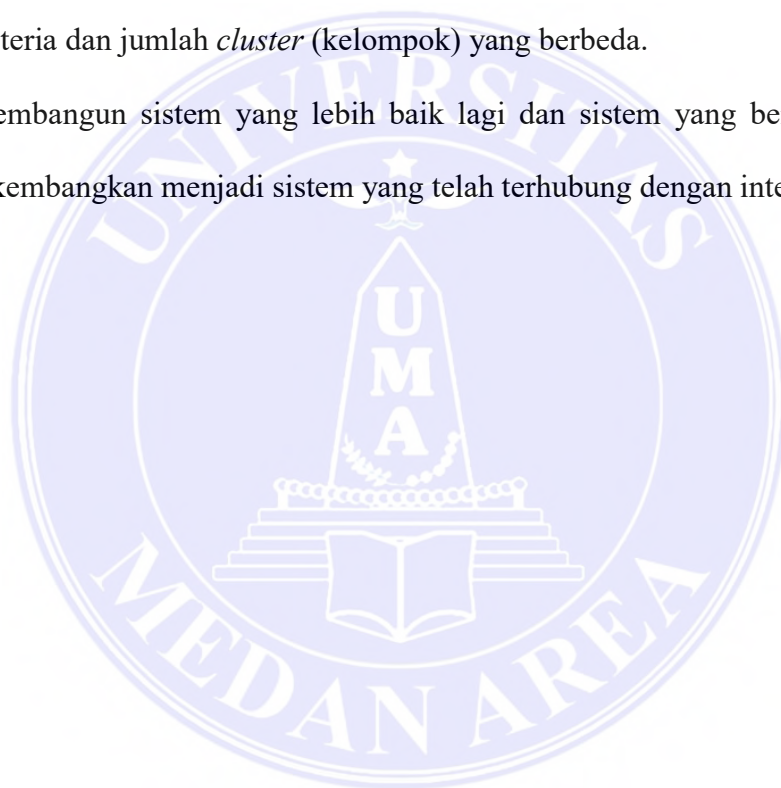
Berdasarkan hasil penelitian yang telah dilakukan, maka analisa dan pembahasan mengenai aplikasi pengelompokan data kriminalitas umum di Provinsi Sumatera Utara menggunakan metode K-Means *Clustering*, dapat ditarik kesimpulan bahwa :

1. K-Means Clustering dapat melakukan pengelompokan daerah berdasarkan tingkat kriminalitas umum secara akurat. Dimana hasil pengelompokan yang diperoleh yaitu pengelompokan untuk pusat *cluster* pertama (tinggi) ada 1 data sedangkan untuk pusat *cluster* kedua (sedang) ada 1 data dan pada pusat *cluster* ketiga (rendah) ada 27 data. Sehingga dapat disimpulkan bahwa ada 1 daerah yang memiliki tingkat kriminalitas paling tinggi, 1 daerah memiliki tingkat kriminalitas sedang dan 27 daerah yang memiliki tingkat kriminalitas paling rendah.
2. Sistem ini dapat di implementasikan untuk pengelompokan daerah berdasarkan kasus dan tingkat kriminalitas menggunakan metode K-Means Clustering dengan login kedalam system dan menambah data daerah, jenis kasus dan data kasus untuk dapat di proses melalui *form* proses pengelompokan. Hasil pengelompokan untuk data sampel 29 daerah memiliki tingkat akurasi sebesar 50%. Sehingga metode K-Means Clustering memiliki kemampuan yang cukup untuk melakukan pengelompokan tingkat kriminalitas umum di Provinsi Sumatera Utara.

5.2 Saran

Dari hasil penelitian yang dilakukan terkait pengelompokan data kriminalitas umum di Provinsi Sumatera Utara menggunakan metode *K-Means Clustering*, maka dapat diberikan beberapa saran sebagai berikut:

1. Mengembangkan aplikasi menggunakan metode lain sebagai studi banding dan pengembangan bidang keilmuan.
2. Aplikasi disempurnakan dengan sistem yang dapat beradaptasi dengan jumlah kriteria dan jumlah *cluster* (kelompok) yang berbeda.
3. Membangun sistem yang lebih baik lagi dan sistem yang berbasis *dektop* dikembangkan menjadi sistem yang telah terhubung dengan internet.



DAFTAR PUSTAKA

- Alhapizi, M. R., Nasir, M., & Effendy, I. (2020). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang. *Journal of Software Engineering Ampera*, 1(1), 1–14.
- Dinata, R. K., Safwandi, S., Hasdyna, N., & Azizah, N. (2020). Analisis k-means clustering pada data sepeda motor. *INFORMAL: Informatics Journal*, 5(1), 10–17.
- Efendi, M. F., & others. (2024). PENEGAKAN HUKUM DALAM MENANGANI PERKARA TINDAK PIDANA PENIPUAN. *Kultura: Jurnal Ilmu Hukum, Sosial, Dan Humaniora*, 2(3), 153–161.
- Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., Abualigah, L., Agushaka, J. O., Eke, C. I., & Akinyelu, A. A. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Engineering Applications of Artificial Intelligence*, 110, 104743.
- Fahrudin, T. M., Riyantoko, P. A., & Hindrayani, K. M. (2023). Implementation of Web Scraping on Google Search Engine for Text Collection Into Structured 2D List. *Telematika: Jurnal Informatika Dan Teknologi Informasi*, 20(2), 139–152.
- Ferdiansyah, M., & Chotijah, U. (2024). Implementasi Algoritme K-Means++ Untuk Clustering Penjualan Bahan Bangunan. *Jurnal Ilmiah Teknik Informatika Dan Komunikasi*, 4(1), 181–193.
- Flores, V. A., Permatasari, P. A., & Jasa, L. (2020). Penerapan web scraping sebagai media pencarian dan menyimpan artikel ilmiah secara otomatis berdasarkan keyword. *Maj. Ilm. Teknol. Elektro*, 19(2), 157.
- Gultom, Y., & Lebang, A. S. (2024). Analisis Yuridis Penegak Hukum Terhadap Pelanggaran Ham Oleh Kelompok Kriminal Bersenjata (KKB) Berdasarkan Perspektif Hukum Indonesia. *SCIENTIA JOURNAL: Jurnal Ilmiah Mahasiswa*, 6(3).
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24.
- Han, J., Pei, J., & Tong, H. (2022). *Data mining: concepts and techniques*. Morgan kaufmann.
- Hartatik, H., Kwintiana, B., Nengsih, T. A., Baradja, A., Harto, B., Sudipa, I. G. I., Handika, I. P. S., Adhicandra, I., Gugat, R. M. D., & others. (2023). *DATA*

SCIENCE FOR BUSINESS: Pengantar & Penerapan Berbagai Sektor. PT. Sonpedia Publishing Indonesia.

- Hutagalung, J. (2022). Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(1), 606–620.
- Isnawan, F. (2023). Pencegahan Tindak Pidana Kejahatan Jalanan Klitih Melalui Hukum Pidana dan Teori Kontrol Sosial. *KRTHA BHAYANGKARA*, 17(2), 349–378.
- Leto, C., Sujana, D., Windyasaki, V. S., Muhammad, R., & others. (2023). *Konsep Data Mining Dan Penerapan*
- Muflikhah, L., Mahmudy, W. F., & Kurnianingtyas, D. (2023). *Machine Learning*. Universitas Brawijaya Press.
- Nabila, Z., Isnain, A. R., Permata, P., & Abidin, Z. (2021). Analisis data mining untuk clustering kasus covid-19 di Provinsi Lampung dengan algoritma k-means. *Jurnal Teknologi Dan Sistem Informasi*, 2(2), 100–108.
- Norshahlan, M., Jaya, H., & Kustini, R. (2023). Penerapan Metode Clustering Dengan Algoritma K-means Pada Pengelompokan Data Calon Siswa Baru. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, 2(6), 1042–1053.
- Pratiwi, D. A., Awangga, R. M., & Setyawan, M. Y. H. (2020). *Seleksi Calon Kelulusan Tepat Waktu Mahasiswa Teknik Informatika Menggunakan Metode Naïve Bayes (Vol. 1)*. Kreatif.
- Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 5(1), 62.
- Putra, A. D., Martha, G. S., Fikram, M., & Yuhan, R. J. (2021). Faktor-Faktor yang Memengaruhi Tingkat Kriminalitas di Indonesia Tahun 2018. *Indonesian Journal of Applied Statistics*, 3(2), 123–131.
- Putra, R. F., Zebua, R. S. Y., Budiman, B., Rahayu, P. W., Bangsa, M. T. A., Zulfadhilah, M., Choirina, P., Wahyudi, F., & Andiyan, A. (2023). *DATA MINING: Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia.
- Rahutomo, F., Sutrisno, S., Sulisty, M. E., Sumantyo, J. T. S., & Harjito, B. (2022). Pendekatan Agile Scrum pada Pengembangan Aplikasi Analitik Akademik dan Kemahasiswaan. *INOVTEK Polbeng-Seri Informatika*, 7(2), 345–358.
- Rani Rotul Muhima, S. S. M. T., Muchamad Kurniawan, S. K. M. K., Septiyawan Rosetya Wardhana, S. K. M. K., Anton Yudhana, S. T. M. T., Sunardi, S. T. M. T., Weny Mistarika Rahmawati, S. K. M. K. M. S., & Gusti Eka Yuliasuti, S. K. M. K. (2022). *Kupas Tuntas Algoritma Clustering: Konsep, Perhitungan Manual, Dan Program*. Penerbit Andi.

<https://books.google.co.id/books?id=H55rEAAAQBAJ>

- Rosaly, R., & Prasetyo, A. (2019). *Pengertian Flowchart Beserta Fungsi dan Simbol-simbol Flowchart yang Paling Umum Digunakan*. Academia.
- Sachrrial, R. H., & Iskandar, A. (2023). Analisa Perbandingan Complete Linkage AHC dan K-Medoids Dalam Pengelompokan Data Kemiskinan di Indonesia. *Building of Informatics, Technology and Science (BITS)*, 5(2), 509–522.
- Santoso, J. T. (2023). Teknologi Keamanan Siber (Cyber Security). *Penerbit Yayasan Prima Agus Teknik*, 1–173.
- Setianto, M. F. (2020). *Pengembangan Aplikasi Alat Bantu Analisis Korelasi Data Faktual Dan Google Trends Untuk Monitoring/Surveillance*.
- Setyaningtyas, S., Nugroho, B. I., & Arif, Z. (2022). Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means. *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, 10(2), 52–61.
- Umagapi, I. T., Umaternate, B., Hazriani, H., & ... (2023). Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa. *Prosiding ...*, 303–308. <http://www.seminar.iaii.or.id/index.php/SISFOTEK/article/view/411>
- Vitali, G., Francia, M., Golfarelli, M., & Canavari, M. (2021). Crop management with the iot: An interdisciplinary survey. *Agronomy*, 11(1), 181.
- Wahyuddin, S., Sudipa, I. G. I., Putra, T. A. E., Wahidin, A. J., Syukrilla, W. A., Wardhani, A. K., Heryana, N., Indriyani, T., Santoso, L. W., & others. (2023). *Data Mining*. Global Eksekutif Teknologi.
- Wibawa, D. (2020). *Jurnalisme Warga Perlindungan, Pertanggungjawaban Etika dan Hukum*. CV. Mimbar Pustaka.
- Wibowo, A., Wangsajaya, Y., & Surahmat, A. (2023). *Pemolisian Digital dengan Artificial Intelligence*. PT. RajaGrafindo Persada-Rajawali Pers.
- Zaidan, M. A. (2022). *Menuju pembaruan hukum pidana*. Sinar Grafika.

LAMPIRAN

1. Source Code

```
Imports System.Data.OleDb
Public Class Form_Proses

    Dim rasio_lama As Double = 0

    Sub tampil()
        Call koneksi()
        str = "select Kode_Kasus, JTP, JPTP from query_kasus order by
kode_kasus asc"
        adp = New OleDbDataAdapter(str, con)
        Dim ds As New DataSet
        adp.Fill(ds)
        dgv1.DataSource = ds.Tables(0)
        con.Close()
        dgv1.Columns(0).Width = 80
        dgv1.Columns(1).Width = 50
        dgv1.Columns(2).Width = 50

        Dim c1k1 As Double = 0
        Dim c1k2 As Double = 0

        Dim c2k1 As Double = 0
        Dim c2k2 As Double = 0

        Dim c3k1 As Double = 999999999999999999
        Dim c3k2 As Double = 999999999999999999

        For a = 0 To dgv1.RowCount - 1
            Dim k1 As Double = dgv1.Rows(a).Cells(1).Value
            Dim k2 As Double = dgv1.Rows(a).Cells(2).Value

            'cari max [C1]
            If k1 >= c1k1 Then
                c1k1 = k1
            End If
            If k2 >= c1k2 Then
                c1k2 = k2
            End If

            'cari min [C3]
            If k1 <= c3k1 Then
                c3k1 = k1
            End If
            If k2 <= c3k2 Then
                c3k2 = k2
            End If
        Next
        'cari median [C2]
        c2k1 = c1k1 - ((c1k1 - c3k1) / 2)
        c2k2 = c1k2 - ((c1k2 - c3k2) / 2)

        dgv2.ColumnCount = 3
        dgv2.Columns(0).HeaderText = dgv1.Columns(0).HeaderText
```

```

dgv2.Columns(1).HeaderText = dgv1.Columns(1).HeaderText
dgv2.Columns(2).HeaderText = dgv1.Columns(2).HeaderText
dgv2.RowCount = 3
dgv2.Rows(0).Cells(0).Value = "C1"
dgv2.Rows(0).Cells(1).Value = c1k1
dgv2.Rows(0).Cells(2).Value = c1k2
dgv2.Rows(1).Cells(0).Value = "C2"
dgv2.Rows(1).Cells(1).Value = c2k1
dgv2.Rows(1).Cells(2).Value = c2k2
dgv2.Rows(2).Cells(0).Value = "C3"
dgv2.Rows(2).Cells(1).Value = c3k1
dgv2.Rows(2).Cells(2).Value = c3k2
dgv2.Refresh()
End Sub

Sub lebar_grid()
Dim y As Integer = 90
Dim z As Integer = 50
Try
    dgv1.Columns(0).Width = 100
    dgv1.Columns(1).Width = 70
    dgv1.Columns(2).Width = 70
Catch ex As Exception
End Try
Try
    For d1 As Integer = 1 To dgv1.ColumnCount - 1
        dgv1.Columns(d1).Width = z
    Next
Catch ex As Exception
End Try
Try
    For d2 As Integer = 1 To dgv2.ColumnCount - 1
        dgv2.Columns(d2).Width = z
    Next
Catch ex As Exception
End Try
Try
    For d3 As Integer = 1 To dgv3.ColumnCount - 1
        dgv3.Columns(d3).Width = z
    Next
Catch ex As Exception
End Try
Try
    For d4 As Integer = 1 To dgv4.ColumnCount - 1
        dgv4.Columns(d4).Width = z
    Next
Catch ex As Exception
End Try
Try
    For d5 As Integer = 1 To dgv4.ColumnCount - 1
        dgv4.Columns(d5).Width = z
    Next
Catch ex As Exception
End Try
Try
    For d6 As Integer = 1 To dgv5.ColumnCount - 1
        dgv5.Columns(d6).Width = z
    Next
Catch ex As Exception

```

```

End Try
Try
    For d7 As Integer = 1 To dgv6.ColumnCount - 1
        dgv6.Columns(d7).Width = z
    Next
Catch ex As Exception
End Try
End Sub

Private Sub Form_Proses_Load(sender As Object, e As EventArgs) Handles MyBase.Load
    Call koneksi()
    lblpesan_proses.Visible = False
    lblpesan_simpan.Visible = False
    Call tampil()
    Call lebar_grid()
End Sub

Sub cluster()
    lblpesan_proses.Visible = True
    lblpesan_proses.Refresh()
    lblpesan_simpan.Refresh()

    txtIterasi.Text = Val(txtIterasi.Text) + 1
    txtIterasi.Refresh()
    dgv3.Columns.Clear()
    dgv4.Columns.Clear()
    dgv5.Columns.Clear()
    dgv6.Columns.Clear()

    dgv3.ColumnCount = 7
    dgv3.Columns(0).HeaderText = "Lokasi"
    dgv3.Columns(1).HeaderText = "C1"
    dgv3.Columns(2).HeaderText = "C2"
    dgv3.Columns(3).HeaderText = "C3"
    dgv3.Columns(4).HeaderText = "JCT"
    dgv3.Columns(5).HeaderText = "JCT^2"
    dgv3.Columns(6).HeaderText = "Cluster"
    dgv3.RowCount = dgv1.RowCount

    For a As Integer = 0 To dgv1.RowCount - 1
        dgv3.Rows(a).Cells(0).Value = dgv1.Rows(a).Cells(0).Value
        dgv3.Rows(a).Cells(1).Value =
 $\text{Math.Sqrt}((\text{dgv1.Rows}(a).\text{Cells}(1).\text{Value} - \text{dgv2.Rows}(0).\text{Cells}(1).\text{Value})^2 + (\text{dgv1.Rows}(a).\text{Cells}(2).\text{Value} - \text{dgv2.Rows}(0).\text{Cells}(2).\text{Value})^2)$ 
        dgv3.Rows(a).Cells(2).Value =
 $\text{Math.Sqrt}((\text{dgv1.Rows}(a).\text{Cells}(1).\text{Value} - \text{dgv2.Rows}(1).\text{Cells}(1).\text{Value})^2 + (\text{dgv1.Rows}(a).\text{Cells}(2).\text{Value} - \text{dgv2.Rows}(1).\text{Cells}(2).\text{Value})^2)$ 
        dgv3.Rows(a).Cells(3).Value =
 $\text{Math.Sqrt}((\text{dgv1.Rows}(a).\text{Cells}(1).\text{Value} - \text{dgv2.Rows}(2).\text{Cells}(1).\text{Value})^2 + (\text{dgv1.Rows}(a).\text{Cells}(2).\text{Value} - \text{dgv2.Rows}(2).\text{Cells}(2).\text{Value})^2)$ 
    Next

    For a1 As Integer = 0 To dgv3.RowCount - 1
        Dim min As Double = 0
        min = Math.Min(dgv3.Rows(a1).Cells(1).Value,
 $\text{Math.Min}(\text{dgv3.Rows}(a1).\text{Cells}(2).\text{Value}, \text{dgv3.Rows}(a1).\text{Cells}(3).\text{Value}))$ 
        dgv3.Rows(a1).Cells(4).Value = min
        dgv3.Rows(a1).Cells(5).Value = min ^ 2
        If min = dgv3.Rows(a1).Cells(1).Value Then

```

```

        dgv3.Rows(a1).Cells(6).Value = "C1"
    ElseIf min = dgv3.Rows(a1).Cells(2).Value Then
        dgv3.Rows(a1).Cells(6).Value = "C2"
    Else
        dgv3.Rows(a1).Cells(6).Value = "C3"
    End If
Next

Dim SSW As Double
SSW = 0
For t As Integer = 0 To dgv3.RowCount - 1
    SSW = SSW + (dgv3.Rows(t).Cells(5).Value)
Next
txtWCV.Text = SSW

'menghitung Nilai SSB
Dim jlh_c1 As Integer = dgv2.RowCount - 1
Dim SSB As Double
SSB = 0

Dim m1m2 As Double = Math.Sqrt((dgv2.Rows(0).Cells(1).Value -
dgv2.Rows(1).Cells(1).Value) ^ 2 + (dgv2.Rows(0).Cells(2).Value -
dgv2.Rows(1).Cells(2).Value) ^ 2)
Dim m1m3 As Double = Math.Sqrt((dgv2.Rows(0).Cells(1).Value -
dgv2.Rows(2).Cells(1).Value) ^ 2 + (dgv2.Rows(0).Cells(2).Value -
dgv2.Rows(2).Cells(2).Value) ^ 2)
Dim m2m3 As Double = Math.Sqrt((dgv2.Rows(1).Cells(1).Value -
dgv2.Rows(2).Cells(1).Value) ^ 2 + (dgv2.Rows(1).Cells(2).Value -
dgv2.Rows(2).Cells(2).Value) ^ 2)
SSB = m1m2 + m1m3 + m2m3

txtBCV.Text = SSB
Dim rasio As Double = SSB / SSW
txtrasio.Text = rasio

Call hasil_pengelompokan()

Dim ttK1 As Double = 0
Dim ttK2 As Double = 0
Dim ttK3 As Double = 0

Dim jlh As Integer = dgv4.RowCount - 1
For i4 As Integer = 0 To dgv4.RowCount - 1
    ttK1 = ttK1 + dgv4.Rows(i4).Cells(1).Value / jlh
    ttK2 = ttK2 + dgv4.Rows(i4).Cells(2).Value / jlh
Next

dgv2.Rows(0).Cells(1).Value = Replace(ttK1, "NaN", 0)
dgv2.Rows(0).Cells(2).Value = Replace(ttK2, "NaN", 0)

ttK1 = 0
ttK2 = 0

jlh = dgv5.RowCount - 1
For i5 As Integer = 0 To dgv5.RowCount - 1
    ttK1 = ttK1 + dgv5.Rows(i5).Cells(1).Value / jlh
    ttK2 = ttK2 + dgv5.Rows(i5).Cells(2).Value / jlh
Next
dgv2.Rows(1).Cells(1).Value = Replace(ttK1, "NaN", 0)

```



```

dgv2.Rows(1).Cells(2).Value = Replace(ttK2, "NaN", 0)

ttK1 = 0
ttK2 = 0
ttK3 = 0
j1h = dgv6.RowCount - 1
For i6 As Integer = 0 To dgv6.RowCount - 1
    ttK1 = ttK1 + dgv6.Rows(i6).Cells(1).Value / j1h
    ttK2 = ttK2 + dgv6.Rows(i6).Cells(2).Value / j1h
Next

dgv2.Rows(2).Cells(1).Value = Replace(ttK1, "NaN", 0)
dgv2.Rows(2).Cells(2).Value = Replace(ttK2, "NaN", 0)

If txtIterasi.Text <= 1 Then
    'btnlanjutproses.Enabled = True
    btntampilkan.Enabled = False
    rasio_lama = rasio
    dgv2.Refresh()
    dgv3.Refresh()
    dgv4.Refresh()
    dgv5.Refresh()
    dgv6.Refresh()
    dgv7.Refresh()
    Call cluster()
Else
    If rasio_lama = rasio Then
        txtIterasi.Refresh()
        dgv2.Refresh()
        dgv3.Refresh()
        dgv4.Refresh()
        dgv5.Refresh()
        dgv6.Refresh()
        dgv7.Refresh()

        MsgBox("Rasio Sama Pada Iterasi Ke-" & Val(txtIterasi.Text)
- 1 & " dan Ke-" & txtIterasi.Text & ", Maka Iterasi Berakhir",
MsgBoxStyle.Information)
        rasio_lama = rasio
        btnlanjutproses.Enabled = False
        btntampilkan.Enabled = True

        lblpesan_proses.Visible = False
        lblpesan_simpan.Visible = True
        lblpesan_proses.Refresh()
        lblpesan_simpan.Refresh()

        Call koneksi()
        str = "Delete * from Tb_Pengelompokan"
        cmd = New OleDbCommand(str, con)
        cmd.ExecuteNonQuery()
        con.Close()

        For ii As Integer = 0 To dgv7.RowCount - 1
            Call koneksi()
            str = "Insert into Tb_Pengelompokan
(Kode_kasus,Cluster, Keterangan) values ('" & dgv7.Rows(ii).Cells(0).Value
& "', '" & dgv7.Rows(ii).Cells(1).Value & "', '" &
dgv7.Rows(ii).Cells(2).Value & "')"
            cmd = New OleDbCommand(str, con)

```

```

        cmd.ExecuteNonQuery()
        con.Close()
    Next
    lblpesan_proses.Visible = False
    lblpesan_simpan.Visible = False
    lblpesan_proses.Refresh()
    lblpesan_simpan.Refresh()

Else
    MsgBox("Rasio Belum Sama, Lanjutkan Proses",
MsgBoxStyle.Exclamation)
    rasio_lama = rasio
    'btnlanjutproses.Enabled = True
    btntampilkan.Enabled = False
    dgv2.Refresh()
    dgv3.Refresh()
    dgv4.Refresh()
    dgv5.Refresh()
    dgv6.Refresh()
    dgv7.Refresh()
    Call cluster()
End If
End If
End Sub

Sub hasil_pengelompokan()
    dgv4.Columns.Clear()
    dgv4.Rows.Clear()
    Dim jctn As Double = 1000
    Try
        dgv4.ColumnCount = dgv1.ColumnCount
        dgv5.ColumnCount = dgv1.ColumnCount
        dgv6.ColumnCount = dgv1.ColumnCount
    Catch ex As Exception

    End Try

    For iii As Integer = 0 To dgv1.ColumnCount - 1
        dgv4.Columns(iii).Name = dgv1.Columns(iii).Name
        dgv5.Columns(iii).Name = dgv1.Columns(iii).Name
        dgv6.Columns(iii).Name = dgv1.Columns(iii).Name
    Next

    For i As Integer = 0 To dgv3.RowCount - 1
        'Dim data As Double = 100
        If dgv3.Rows(i).Cells(6).Value = "C1" Then
            For iii As Integer = 0 To dgv1.RowCount - 1
                If dgv3.Rows(i).Cells(0).Value =
dgv1.Rows(iii).Cells(0).Value Then
                    dgv4.Rows.Add(dgv3.Rows(i).Cells(0).Value,
dgv1.Rows(iii).Cells(1).Value, dgv1.Rows(iii).Cells(2).Value)
                End If
            Next
        ElseIf dgv3.Rows(i).Cells(6).Value = "C2" Then
            For iii As Integer = 0 To dgv1.RowCount - 1
                If dgv3.Rows(i).Cells(0).Value =
dgv1.Rows(iii).Cells(0).Value Then
                    dgv5.Rows.Add(dgv3.Rows(i).Cells(0).Value,
dgv1.Rows(iii).Cells(1).Value, dgv1.Rows(iii).Cells(2).Value)
                End If
            Next
        End If
    Next
End Sub

```

```

        Next
        ElseIf dgv3.Rows(i).Cells(6).Value = "C3" Then
            For iii As Integer = 0 To dgv1.RowCount - 1
                If dgv3.Rows(i).Cells(0).Value =
dgv1.Rows(iii).Cells(0).Value Then
                    dgv6.Rows.Add(dgv3.Rows(i).Cells(0).Value,
dgv1.Rows(iii).Cells(1).Value, dgv1.Rows(iii).Cells(2).Value)
                End If
            Next
        End If
    Next
    ' Exit Sub
    Call hasil_k_means()
End Sub

Sub hasil_k_means()
    dgv7.Columns.Clear()
    dgv7.ColumnCount = 3
    dgv7.Columns(0).Name = "Kode Kasus"
    dgv7.Columns(1).Name = "Cluster"
    dgv7.Columns(2).Name = "Keterangan"
    dgv7.RowCount = dgv3.RowCount

    For i As Integer = 0 To dgv3.RowCount - 1
        dgv7.Rows(i).Cells(0).Value = dgv3.Rows(i).Cells(0).Value
        dgv7.Rows(i).Cells(1).Value = dgv3.Rows(i).Cells(6).Value
        If dgv3.Rows(i).Cells(6).Value = "C1" Then
            dgv7.Rows(i).Cells(2).Value = "Tinggi"
            dgv7.Rows(i).DefaultCellStyle.BackColor = Color.Red
            dgv7.Rows(i).DefaultCellStyle.ForeColor = Color.White
        ElseIf dgv3.Rows(i).Cells(6).Value = "C2" Then
            dgv7.Rows(i).Cells(2).Value = "Sedang"
            dgv7.Rows(i).DefaultCellStyle.BackColor = Color.Yellow
            dgv7.Rows(i).DefaultCellStyle.ForeColor = Color.Green
        Else
            dgv7.Rows(i).Cells(2).Value = "Rendah"
            dgv7.Rows(i).DefaultCellStyle.BackColor = Color.Green
            dgv7.Rows(i).DefaultCellStyle.ForeColor = Color.Yellow
        End If
    Next
End Sub

Private Sub btnlanjutproses_Click(sender As Object, e As EventArgs)
Handles btnlanjutproses.Click
    Call cluster()
    Call lebar_grid()
End Sub

Private Sub btntampilkan_Click(sender As Object, e As EventArgs)
Handles btntampilkan.Click
    If btnlanjutproses.Enabled = False Then
        Form_Laporan.Close()
        Form_Laporan.ShowDialog()
    End If
End Sub

Private Sub btnReset_Click(sender As Object, e As EventArgs) Handles
btnReset.Click
    btnlanjutproses.Enabled = True

```

```
txtIterasi.Text = ""
Try
    dgv2.Rows.Clear()

    dgv3.Columns.Clear()
    dgv4.Columns.Clear()
    dgv5.Columns.Clear()
    dgv6.Columns.Clear()
    dgv7.Columns.Clear()
    lblpesan_proses.Visible = False
    lblpesan_simpan.Visible = False

    Call tampil()
Catch ex As Exception

End Try
End Sub

Private Sub dgv1_CellContentClick(sender As Object, e As
DataGridViewCellEventArgs) Handles dgv1.CellContentClick

End Sub

Private Sub dgv2_CellContentClick(sender As Object, e As
DataGridViewCellEventArgs) Handles dgv2.CellContentClick




End Sub
End Class
```



2. Sample Data

Bulan	Tempat	JENIS KEJAHATAN	CT	CC
Februari	POLRESTABES MEDAN.xlsx	PENGGELAPAN	99	58
September	POLRES LAB.BATU.xlsx	PENCURIAN RINGAN	98	54
Februari	POLRESTABES MEDAN.xlsx	LAIN-LAIN (AGAR DIISI TINDAK PIDANANYA)	92	98
Juni	POLRES LAB. BATU.xlsx	PENCURIAN RINGAN	90	47
September	POLRES PS.xlsx	PENIPUAN / PERBUATAN CURANG	9	4
Oktober	POLRES LAB. BATU.xlsx	PEMERASAN DAN PENGANCAMAN	9	6
November	POLRES PSP.xlsx	PENCURIAN DENGAN PEMBERATAN	9	
Mei	POLRES TOBA.xlsx	PENGANIAYAAN BIASA	9	12
Mei	POLRES TAPSEL.xlsx	PENCABULAN	9	4
Mei	POLRES T.BALAI.xlsx	PENCURIAN DENGAN PEMBERATAN	9	7
Mei	POLRES SIBOLGA.xlsx	PENCURIAN BIASA	9	5
Mei	POLRES SERGAI.xlsx	PENGANIAYAAN BERAT	9	7
Mei	POLRES PALAS.xlsx	PENGANIAYAAN	9	5
Mei	POLRES BELAWAN.xlsx	PENGHAPUSAN KEKERASAN DALAM RUMAH TANGGA (PKDRT)	9	8
Juli	POLRES ASAHAN.xls	PERADILAN ANAK	9	0

3. Surat Keterangan Pembimbing

	UNIVERSITAS MEDAN AREA FAKULTAS TEKNIK	
<small>Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 ☎ (061) 7366878, 7360168, 7364348, 7366781, Fax (061) 7366998 Medan 20223 Kampus II : Jalan Selabud Nomor 79 / Jalan Sei Serayu Nomor 70 A ☎ (061) 8225602, Fax (061) 8226331 Medan 20122 Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id</small>		
Nomor	: 312/FT.6/01.10/VII/2024	5 Juli 2024
Lamp	:	
H a l	: Perpanjang SK Pembimbing Tugas Akhir	
Yth. Pembimbing Tugas Akhir Nanda Novita, S.Kom, M.Kom di Tempat		
Dengan hormat, Sehubungan telah berakhirnya waktu masa berlaku SK pembimbing nomor 30/FT.6/01.10/I/2024 tertanggal 10 Januari 2024 maka perlu diterbitkan kembali SK Pembimbing Skripsi baru atas nama mahasiswa berikut		
N a m a	: Evimai Indri Sitorus	
N P M	: 208160018	
Jurusan	: Teknik Informatika	
Oleh karena itu kami mengharapkan kesediaan saudara :		
Nanda Novita, S.Kom, M.Kom		(Sebagai Pembimbing)
Adapun Tugas Akhir Skripsi berjudul :		
"Implementasi Data Mining dalam Pengelompokan Data Kriminalitas Umum di Provinsi Sumatera Utara menggunakan Algoritma K-Means Clustering"		
SK Pembimbing ini berlaku selama enam bulan dihitung sejak SK ini diterbitkan. Jika proses pembimbing melebihi batas waktu yang telah ditetapkan, SK ini dapat ditinjau ulang.		
Demikian kami sampaikan, atas kesediaan saudara diucapkan terima kasih.		
Dekan,		
 Dr. Eng. Supriatno, ST, MT		
		

4. Surat Pengantar Riset



UNIVERSITAS MEDAN AREA FAKULTAS TEKNIK

Kampus I : Jalan Kolan Nomor 1 Medan Estate/Jalan PBSI Nomor 1 (061) 7366878, 7360168, 7364348, 7366781, Fax.(061) 7366998 Medan 20223
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A, (061) 8225602, Fax (061) 8226331 Medan 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : 244 /FT.6/01.10/V/2024

17 Mei 2024

Lamp : -

Hal : Penelitian Dan Pengambilan Data Tugas Akhir

Yth. DITRESKRIMUM POLDA SUMATERA UTARA
Jln. Sisingamangaraja, KM.10,5
Di
Medan

Dengan hormat,
Kami mohon kesediaan Bapak/Ibu berkenan untuk memberikan izin dan kesempatan kepada mahasiswa kami tersebut dibawah ini :

NO	NAMA	NPM	PRODI
1	Evimai Indri Sitorus	208160018	Teknik Informatika

Untuk melaksanakan Penelitian dan Pengambilan Data Tugas Akhir pada perusahaan/Instansi yang Bapak/Ibu Pimpin.

Perlu kami jelaskan bahwa Pengambilan Data tersebut adalah semata-mata untuk tujuan ilmiah dan Skripsi yang merupakan salah satu syarat bagi mahasiswa tersebut untuk mengikuti ujian sarjana pada Fakultas Teknik Universitas Medan Area dan tidak untuk dipublikasikan. dengan judul penelitian :

Implementasi *Data Mining* dalam Pengelompokan Data Kriminalitas Umum di Provinsi Sumatera Utara menggunakan Algoritma *K-Means Clustering*

Atas perhatian dan kerja sama yang baik diucapkan terima kasih.

Dekan,



Dr. Grg. Supriatno, ST, MT

Tembusan :

1. Ka. BPMPP
2. Mahasiswa
3. File



5. Surat Selesai Riset


KEPOLISIAN NEGARA REPUBLIK INDONESIA
DAERAH SUMATERA UTARA
Jalan Sisingamangaraja Km.10,5 No.60 Medan – 20148

Medan, 26 Juni 2024

Nomor : B/4275 /VI/RES 1.24/2024/Ditreskrim
Klasifikasi: Biasa
Lampiran : 1 (satu) Lembar
Hal : pemberitahuan telah melakukan pengambilan data

Kepada
Yth. DEKAN FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
di
Tempat


1. Rujukan Surat Dekan Fakultas Teknik Universitas Medan Area Nomor : 244/FT.6/01.10/VI/2024, tanggal 17 Mei 2024 Hal Penelitian Dan Pengambilan Data Tugas Akhir.

2. Sehubungan dengan rujukan tersebut diatas, bersama ini diberitahu kepada Dekan bahwa yang tersebut dibawah ini :

Nama : Evimai Indri Sitorus
NPM : 208160018
Jurus / Prodi : Teknik Informatika
Judul : Implementasi Data Mining Dalam Pengelompokan Data Kriminalitas Umum di Provinsi Sumatra Utara Menggunakan Algoritma K-Means Clustering

Bahwa mahasiswa/i tersebut telah selesai melaksanakan penelitian dan pengambilan data pada Direktorat Reserse Kriminal Umum Polda Sumut dengan fokus kajian sebagaimana surat tersebut diatas, selanjutnya seluruh data yang diperoleh dari Ditreskrim Polda Sumut hanya dapat dipergunakan untuk menyusun tugas akhir Mahasiswa sebagai kelengkapan hasil Penelitian dan tidak untuk dipublikasikan.


3. Demikian untuk menjadi maklum.


KEPOLISIAN DAERAH SUMATERA UTARA
DITRESKRIMUM
KEPALA DAERAH SUMATERA UTARA
SUMARYONO, S.H., S.I.K., M.H.
KORPORASI BESAR POLISI NRP 74090546

Tembusan :

1. Kapolda Sumut.
2. Irwasda Polda Sumut.

6. Turnitin

 **Similarity Report ID:** oid:29477:64186982

PAPER NAME
**IMPLEMENTASI DATA MINING DALAM P
ENGELOMPOKAN DATA KRIMINALITAS
UMUM DI PROVINSI SUMATRA UTARA
MENGGU**

AUTHOR
EVIMAI INDRI SITORUS

WORD COUNT
9015 Words

CHARACTER COUNT
56897 Characters

PAGE COUNT
56 Pages

FILE SIZE
1.5MB

SUBMISSION DATE
Aug 9, 2024 11:13 AM GMT+7

REPORT DATE
Aug 9, 2024 11:14 AM GMT+7

● 30% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 24% Internet database
- 6% Publications database
- Crossref database
- Crossref Posted Content database
- 19% Submitted Works database

● Excluded from Similarity Report

- Bibliographic material
- Cited material
- Abstract
- Small Matches (Less than 15 words)

Summary