PENERAPAN DATA MINING *CLUSTERING* MENGGUNAKAN METODE *K-MEANS* PADA DATA TINDAK KRIMINALITAS DI POLRES KABUPATEN KUNINGAN

Stedila[1], Rini Astuti[2], Fadhil Muhammad Basysyar[3]
[1] Teknik Informatika, STMIK IKMI CIREBON
[2] Sistem Informasi, STMIK LIKMI BANDUNG
[3] Sistem Informasi, STMIK IKMI CIREBON
Jl. Perjuangan No. 10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat, 415125
stedilaa@gmail.com

ABSTRAK

Dengan meningkatnya tingkat tindak kriminalitas di Kabupaten Kuningan tentunya data yang akan disimpan akan terus bertambah. Menurut Polres Kabupaten Kuningan jumlah kasus tindak kejahatan yang terjadi di Kabupaten Kuningan pada 3 Tahun terakhir mencapai 800 kasus. Salah satu upaya untuk menanggulangi kejahatan dengan mengetahui pola dan karakteristik dari tindak kejahatan yang terjadi, dengan mengetahui pola dan karakteristik dari tindak kejahatan maka dapat diidentifikasi fakto-faktor yang menyebabkan terjadinya tindak kejahatan. Oleh karena itu, untuk mengelompokan data tindak kejahatan yang terjadi di masyarakat akan memanfaatkan Dara Mining menggunakan metode K-Means Clustering dengan Tools RapidMiner versi 10.3. Nilai *Davies Bouldin Index* yang dihasilkan dari algoritma *k-means* ini sebesar 3.323. Pengelompokan data tindak kejahatan ini dapat digunakan oleh Polres Kabupaten Kuningan untuk menentukan pola dan karakteristik dari tindakan kejahatan yang terjadi.

Kata kunci : *Data Mining, Algoritma K-Means Clustering, Kriminalitas, Tools Rapid miner 10.3*

1. PENDAHULUAN

Tindak kriminalitas merupakan perbuatan yang merugikan secara ekonomis dan psikologis dengan melanggar hukum yang berlaku dalam Negara Republik Indonesia serta norma-norma sosial dan agama [1]. Biasanya tindak kriminalitas juga terjadi dikarenakan beberapa faktor kondisi misalnya kemiskinan yang makin meningkat, kondisi lingkungan yang mendukung individu melakukan kejahatan serta kebencian terhadap seseorang [2]. Saat ini kriminalitas menjadi isu tren pada saat ini dikarenakan banyaknya jenis tindak kejahatan yang beraneka ragam, hal ini membuat fenomena kriminalitas muncul di tengah masyarakat [3]. Dengan bertambahnya data kriminalitas, maka perlu dirancang suatu sistem yang dapat mengolah data tindak kriminalitas. Oleh karena itu, salah satu metode yang digunakan dalam pengolahan data yaitu metode Algoritma K-Means Clustering guna membantu pihak kepolisan di Kabupaten Kuningan mendapatkan gambaran untuk mengelompokan data tindak kejahatan berdasarkan pola dan karakteristiknya dari tindak kejahatan yang terjadi.

Dalam era informasi modern, penanganan tindak kriminalitas menjadi kompleks dan menantang terutama di Polres Kabupaten Kuningan. Pertumbuhan jumlah data terkait Dari data Polres Kabupaten Kuningan yang meningkat setiap tahunnya menjadi isu yang memerlukan perhatian serius dari pihak kepolisian, terutama di tingkat Kabupaten Kuningan. Penerapan teknologi dengan menggunakan metode data mining clustering khususnya menggunakan metode K-Means sebagai suatu potensial untuk mengatasi permasalahan dalam pengolahan data skala besar [4]. Dengan volume data yang terus meningkat terkait dengan kriminalitas, perlu adanya pendekatan inovatif dalam mengelola dan menganalisis data kejahatan untuk mendukung proses pengambilan keputusan di Polres Kabupaten Kuningan. Dengan menerapkan metode K-Means dalam data mining Clustering, diharapkan dapat mengidentifikasi pola-pola yang muncul dalam data kejahatan [2]. Meskipun demikian, dalam upaya penerapan teknologi ini terdapat tantangan yang perlu diatasi seperti parameter yang tepat, interpretasi hasil Clustering, dan pengelolaan data besar. [5].

Penelitian tentang penggunaan penggunaan data mining clustering menggunakan metode K-Means pada tindak kriminalitas telah menjadi topik yang menarik di bidang analisis keamanan dan prediksi kejahatan, penelitian sebelumnya dengan judul "Algoritma *K-Means Clustering* dalam mengklasifikasi data daerah rawan tindak kriminalitas Polres Kepulauan Mentawai" menerapkan algoritma *K-Means* memfokuskan pada pengumpulan dan analisis data kriminilitas yang relevan. Data ini mungkin meliputi jenis kejahatan, lokasi, waktu kejadian dan lainnya dengan pemodelan pola kriminalitas oleh peneliti [6]. Pada penelitian yang dilakukan oleh [7] dengan judul "Analisa pemetaan kriminalitas Kabupaten Bangkalan menggunakan metode *K-Means* dan *K-Means ++"* dengan menggunakan 2 metode yaitu Algoritma *K-means* dan *K-Means ++* cara kerja *K-Means Clustering* itu untuk melakukan partisi set data kedalam Cluster yang telah diimplementasikan dari awal sedangkan *K-Means ++* tujuan utama dari algoritma ini mengambil titik-titik data sebagai pusat awal yang berjarak satu sama lain.

Sedangkan dari penelitian [8]. Penerapan metode *data mining* pada kasus kriminalitas, melakukan evaluasi terhadap keberhasilan algoritma *K-Means* dalam mengidentifikasi *Cluster* rawan tindak kriminalitas, evaluasi ini melibatkan metrik atau validasi yang relvan dalam mengevaluasi seberapa baik *Cluster* mewakili pola kriminalitas yang sebenarnya.

Dari tujuan penelitian ini dalam mengidentifikasi pola kriminalitas di Kabupaten Kuningan dengan data yang ada di Polres Kabupaten Kuningan serta menerapkan Algoritma *K-Means Clustering* dalam mengelompokan area yang memiliki karakteristik serupa dari segi tingginya kriminalitas. Dalam menentukan wilayah rawan dan memberikan analisis prediktif penelitian ini juga untuk mengidentifikasi wilayah yang rawan terhadap kriminalitas, hal ini membantu pihak Kepolisian Kabupaten Kuningan untuk lebih fokus dalam alokasi sumber daya dan perencanaan strategi keamanan di masing-masing wilayah polsek yang ada di Kabupaten Kuningan serta memberkan pengetahuan dasar dalam mengetahui pola analisis bagi anggota kepolisian yang bertanggungjawab dalam memegang masalah kriminalitas. Selain itu juga dalam signifikasi penelitian dapat meningkatkan efektifitas penegakan hukum dengan memahami pola kriminalitas, penegak hukum dapat lebih efektif dalam merespon situasi kejahatan dan dapat menghasilkan penanggulangan kejahatan, penangkapan pelaku, serta pengurangan kejahatan.

Dalam serangkaian metode pendekatan untuk penelitian ini dibutuhkan pengumpulan data, penting untuk mengumpulkan data Kriminalitas ini termasuk jenis kriminalitas, wilayah (Polsek) kriminalitas, dan waktu (tahun) terjadinya kriminalitas. Selain data kriminalitas informasi tambahan seperti data *demografi* dan *geografis* dari wilayah yang diteliti diambil dari Polres Kabupaten Kuningan, diperlukannya data-data tersebut guna melengkapi pemahaman yang akan menjadi fokus penelitian. Jika data sudah terkumpul pemprosesan selanjutnya dalam persiapan data, dalam proses tersebut langkah pertama adalah pembersihan dan pemrosesan data hal ini meliputi pemilihan fitur, normalisasi data, dan pengelompokan [9]. Proses ekstraksi fitur yang tepat dari data merupakan langkah yang krusial, ini melibatkan identifikasi variabel yang paling berpengaruh dala pola kriminalitas untuk digunakan dalam analisis clustering. Setelah selesai beranjak kedalam penerapan algoritma *K-Means*, algoritma *K-Means* membutuhkan jumlah *Cluster* sebagai paramete input selain itu juga memerlukan inisialisasi titik-titik pusat *Cluster* [10]. Dalam proses terakhir evaluasi hasil dan interpretasi, metrik evaluasi dilakukan untuk memastikan bahwa *Cluster* yang dihasilkan menggambarkan pola kriminalitas dengan baik, hasil dari algoritma *K-Means* harus diinterpretasikan ini melibatkan pemahaman terhadap setiap *Cluster*, mengidentifikasi wilayah atau waktu yang terkonsentrasi dengan jenis kejahatan tertentu dan mengevaluasi apakah klasterisasi sesuai dengan situasi kejahatan sebenarnya.

2. TINJAUAN PUSTAKA

Clustering merupakan salah satu teknik data science yang bertujuan untuk mengklasifikasikan data ke dalam kelompok berdasarkan persamaan atau perbedaan. Biasanya, data dari satu kelompok memiliki karakteristik yang serupa tetapi berbeda dengan data dari kelompok lain. Selain itu, ada juga algoritma pengelompokan yang berbeda untuk mencapai tujuan yang berbeda. Pemilihan algoritma yang tepat bergantung pada jenis data dan tujuan analisis.

K-Means Metode clustering K-Means bekerja dengan mencari pusat atau centroid cluster yang mewakili kelompok. Proses ini melibatkan iterasi antara penghitungan jarak antara masing-masing objek data dengan pusat cluster yang ada dan memperbarui lokasi pusat cluster berdasarkan rata-rata objek data yang terdapat dalam cluster. Iterasi ini dilakukan berulang kali hingga diperoleh pusat cluster yang stabil atau terkonvergensi.

Penelitian oleh [9] Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining *Algoritma K-Means Clustering*, mengggunakan metode algoritma *K-means* dengan menggunakan *software rapidminer*, yang bertujuan pengelompokkan untuk menentukan tingkat daerah rawan. Dengan dibuatnya sistem ini diharapkan dapat membantu pihak kepolisian dalam menentukan daerah rawan tindak kriminal. Dan dari hasil penelitian menyatakan kelompok daerah rawan tindak kriminal, yakni POLRESTA MEDAN dan POLRES LABUHAN BATU, Hasil implementasi dari penelitian ini Pada Kepolisian Daerah Sumatera Utara (POLDASU) terdapat 28 Polres dan semua data kriminal tiap-tiap Polres akan dijadikan sampel untuk penerapan Data Mining. Hasil tersebut di atas dibandingkan dengan hasil iterasi ke-3 sebelumnya dan hasilnya adalah sama.Maka oleh karena itu, pencarian iterasi dihentikan dan hasil telah ditemukan dimana hasil dari pada pengelompokan dapat dilihat pada tabel 5, yaitu C3 adalah kelompok daerah rawan tindak kriminal, yakni POLRESTA MEDAN dan POLRES LABUHAN BATU.

Penelitian oleh [1] Algoritma K-Means dalam Mengelompokkan Daerah Rawan Kriminalitas Pencurian Sepeda Motor Di Kabupaten Asahan, membahas mengenai sulitnya Polres Asahan mengelompokkan daerah yang rawan tindak kriminalitas dalam kategori paling rawan, cukup rawan dan tidak rawan, menggunakan Algoritma *K-Means Clustering* sebagai salah satu metode data clustering non-hirarki mempartisi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok., yang bertujuan untuk menemukan *Cluster* yang berkualitas dalam waktu yang layak. *Clustering*

dalam data mining berguna untuk menemukan pola distribusi di dalam sebuah data set yang berguna untuk proses analisa data. Analisis clustering dapat diterapkan secara luas pada gambar pemrosesan, pengenalan model, pengambilan dokumen, diagnosis medis, analisis web, Dari proses *K-Means Clustering* yang telah dilakukan, pengelompokan daerah rawan tindak kriminalitas pencurian sepeda motor di Asahan terdiri dari cluster 1 (paling rawan) yaitu 1 desa, cluster 2 (cukup rawan) yaitu 2 desa, dan cluster 3 (tidak rawan) yaitu 10 Kecamatan, Berdasarkan hasil iterasi 1 dan 2 diatas, maka dapat ditemukan hasil sebagai berikut :

a. Untuk *Cluster* 1 (C1) /Paling Rawan: adalah 1 Cluster 1: Kecamatan Kisaran Timur
b. Untuk *Cluster* 2 (C2) /Cukup Rawan adalah 2,7 Cluster 2: Kecamatan Kisaran Barat dan Kecamatan Air Baru
c. Untuk *Cluster* 3 (C3)/Tidak Rawan:adalah 3,4,5,6,8,9,10 *Cluster* 3: Kecamatan Meranti, Kecamatan Rawang panca Arga, Kecamatan Aek Songosongan, Kecamatan Simpang Empat, Kecamatan Air Joman, Kecamatan Simpang Empat, Kecamatan Air Joman, Kecmatan BP Mandoge dan Kecamatan Setia Janji.

Penelitian oleh [10] Analisis Pemetaan Tingkat Kriminalitas Di Kabupaten Karawang Menggunakan Algoritma *K-Means*, Penelitian ini menggunakan metode clustering dengan algoritma *k-means* dan dilakukan pemetaan daerah rawan kriminalitas menggunakan QGIS. Tujuan belum dapat ? , Hasil pengelompokan daerah rawan kriminalitas di Karawang pada 2019 didapatkan *cluster* tidak rawan sebanyak 23 kecamatan, *cluster* rawan sebanyak 3 kecamatan dan *cluster* sangat banyak sebanyak 4 kecamatan. Sedangkan pada 2020 didapatkan *cluster* tidak rawan sebanyak 22 kecamatan, cluster rawan sebanyak 4 kecamatan, dan *cluster* sangat rawan sebanyak 4 kecamatan. Hasil evaluasi *clustering* menggunakan *silhouette coefficient* pada tahun 2019 yaitu sebesar 0,52 dan 0,54 pada tahun 2020, keduanya masuk dalam kategori *medium strucutre* dengan interpretasi penempatan klaster yang wajar.

Penelitian oleh [11] Sistem clustering tindak kejahatan pencurian di wilayah jawa barat menggunakan Algoritma *k-means* , Metodologi yang digunakan adalah algoritma *K-Means Clustering*, yang bertujuan untuk sebagai salah satu usaha untuk membantu pihak kepolisian dalam mengambil keputusan apakah suatu daerah memerlukan pengawasan ekstra atau tidak, dan hasil telah ditemukan dari penelitian ini dibentuk 3 *cluster*, dimana cluster 1 masuk kedalam kategori tingkat pencurian tinggi, *cluster* 2 masuk kedalam kategori tingkat pencurian sedang dan *cluster* 3 masuk kedalam tingkat pencurian rendah. *Iterasi clustering* data pencurian terjadi sebanyak 6 kali iterasi dengan menggunakan WCV dan BCV sebagai acuan iterasi. Terdapat 8 wilayah dengan tingkat pencurian tinggi, 10 wilayah dengan tingkat pencurian sedang dan 9 wilayah dengan tingkat pencurian rendah. Ditemukan pusat cluster dengan Cluster 1 = 297,65 ; 22,375, Cluster 2 = 169,200 ; 11 dan Cluster 3 = 35,222 ; 6,111.

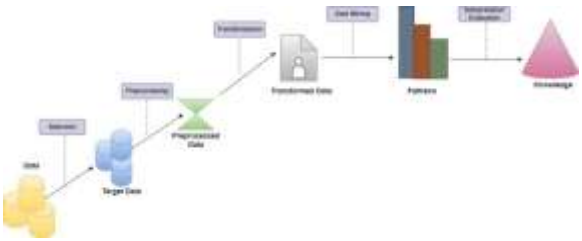
2.1. Algoritma K-Means

Deuclidean (x,y) = $\sqrt{\sum}$(xini=1−yi)2
Dimana:x dan y : representasi nilai atribut dari dua record .

3. METODE PENELITIAN

3.1. Metode Penelitian

Knowledge Discovery in Database (KDD) adalah sebuah proses untuk mengekplorasi dan menganalisa data dalam jumlah yang besar untuk menentukan pola atau informasi yang menarik. Proses KDD merupakan proses Iteratif dan berulang yang terdiri dari beberapa tahapan, yaitu :



Gambar 1. Proses KDD

Dari Langkah-langkah yang terdapat pada gambar 1 sebagai berikur :

a. Data, pada tahapan ini, data diambil dari sumber-sumber yang relevan, seperti database, file text, atau lainnya
b. Target Data, Dalam tahapan ini, varibel atau atribut yang akan dianalisis secara khusus untuk memecahkan masalah yang ingi diselesaikan diidentifikasi, misalnya dalam konteks klasifikasi
c. *Preprocessed* Data, Pada tahap ini, data dinormalisasi, data yang hilang diisi, outlier diidentifikasi dan data yang tidak relevan dihapus
d. *Transformed* Data, Dalam tahapan ini data disesuaikan dengan kebutuhan analisis, misalnya konversi format, ekstraksi fitur, atau pengubahan skala data
e. *Patterns*, Melalui teknik analisis data seperti Clustering atau Regresi, pola-pola dan hubungan yang signifikan diidentifikasi dalam data
f. *Knowledge*, Dari pola-pola yang ditemukan pengetahuan baru diekstraksi yang dapat membantu dalam pemecahan masalah, pengambilan keputusan atau prediksi di masa depan

3.2. Sumber Data

Sumber data yang diambil dalam penelitian ini yaitu data Tindak Kriminalitas yang di dapatkan dari

Polres Kabupaten Kuningan. Data Kriminalitas diambil dari tahun 2021 hingga 2023 (selama 3 tahun) menampilkan 11 variabel jenis tindak kriminalitas diantaranya curanmor, curas, KDRT, pemerasan, penipuan, penganiayaan, pemerkosaan, NAPZA, curat, pembunuhan, dan pencurian biasa, dengan jumlah total keseluruhan hasil rekap data kasus tindak kriminalitas sebanyak 1.133 kasus selama 3 tahun.

3.3. Data Training dan Data Testing

Data pengujian harus berasal dari data terpisah dari data training, hal ini dilakukan untuk memastikan model clustering bekerja dengan baik pada data baru. Dalam penelitian ini data testing terdiri dari 11 data Polsek kecamatan Kabupaten Kuningan, Data kecamatan dipilih secara acak dari daftar kecamatan yang ada di Kabupaten Kuningan. Jumlah kejadian kriminalitas yang dijadikan atribut pada data pengujian adalah jumlah kejadian kriminalitas yang terjadi di setiap kecamatan pada 3 tahun terakhir. Data training digunakan untuk melatih model clustering, data training harus cukup besar untuk membuat model yang akurat. Pada penelitian ini data latih terdiri dari 23 data kecamatan Kabupaten Kuningan. Data kecamatan dipilih secara acak dari daftar kecamatan yang ada di Kabupaten Kuningan. Jumlah kejadian kriminalitas yang dijadikan atribut pada data training merupakan jumlah kejadian kriminalitas yang terjadi pada masing-masing kecamatan pada 3 tahun. Atribut Data yang digunakan dalam penelitian ini adalah jumlah kejadian kejahatan per kecamatan. Atribut ini dipilih karena merupakan metrik yang paling relevan untuk mengukur tingkat kejahatan di suatu daerah, jumlah perkara pidana per kecamatan dihitung berdasarkan jenis kejahatan yang terjadi pada masing-masing kecamatan. Kejahatan yang dihitung antara lain pencurian, perampokan, penyerangan, dan kejahatan narkoba.

3.4. Teknik Pengumpulan Data

Teknik pengumpulan data yang dilakukan adalah teknik pengumpulan data kuantitatif, pengumpulan data tindak kriminalitas berkoordinasi secara langsung kepada pihak Polres Kabupaten Kuningan bagian bidang Reskrim. Pada pengumpulan data awal menggunakan microsoft excel untuk rekap data yang masuk dalam tindak kriminalitas.

3.5. Teknik Analisis Data

Teknik analisis data merupakan langkah yang paling penting dalam suatu penelitian karena berfungsi untuk menghasilkan kesimpulaaan tentang hasil penelitian. Berikut adalah teknik analisis data yang digunakan pada penelitian ini, antara lain :

a. Data Collection
   Data yang digunakan dalam penelitian ini dikumpulkan dari catatan kriminalitas Polres Kabupaten Kuningan. Data tersebut mencakup informasi mengenai jenis kejahatan, jumlah kejahatan, lokasi kejahatan, dan atribut lainnya. Periode pengumpulan data mencakup 1 bulan dengan tujuan untuk mencakup variasi musiman dantren temporal.
b. Preprocessing Data
   Ini mencakup pengelompokan dan normalisasi data untuk memastikam konsistensi dan akurasi analisis, selain itu penanganan missing data untuk memastikan keberlanjutan proses analisis
c. Implementasi Algoritma *K-Means*
   Algoritma *K-Means Clustering* diterapkan pada data kriminalitas menggunakan jumlah *Cluster* (C) yang telah ditentukan. Proses ini melibatkan literasi untuk mengelompokan data dalam *cluster-cluster* yang memiliki kesamaan berdasarkan metrik jarak yang dipilih.
d. Evaluasi Hasil *Clustering*
   Hasil *Clustering* dievaluasi untuk memastikan interpretabilitas dan keberlanjutan model, serta visualisasi hasil *clustering* menggunakan peta atau grafik.
e. Validasi dan Sensitivitas analisis
   Validasi hasil *Clustering* dilakukan dengan mempertimbangkan interpretasi domain ahli dan membandingkan temuan dengan situasi lapangan yang sebenarnya.

4. HASIL DAN PEMBAHASAN

Hasil penelitian ini yang dilakukan pada pembehasan ini menjelaskan tentang proses klasifikasi atau pengelompokan dataset kriminalitas menggunakan Algoritma K-Means , dalam pengelompokan ini dilakukan melalui proses pengujian marchine learning yaitu software Rapidminer 10.3

4.1. Data

Dalam dataset yang digunakan di penelitian ini sebanyak 800 data dan terdiri dari 7 atribut dari tahun 2021 sampai dengan tahun 2023, dataset didapatkan dari Polres Kabupaten Kuningan yang beralamat di Jalan RE Martadinata No.526, Ancaran Kecamatan Kuningan Kabupaten Kuningan. Dari hasil penelusuran data dalam bentuk soft file Microsoft Excel seperti di tabel 4.3 berikut : (Dataset disajikan sebanyak 10 record, dari dataset rindak kriminalitas sebanyak 800 record)

Tabel 1. Dataset Kriminalitas

No.	Inisial Pelaku	Inisial Korban	Jenis Tindak Kriminal	Barang Bukti	TKP	Kasus Ditangani	Waktu
1.	Dalam Lidik	HI	Curat	Sepeda Motor	Sawah Pangsor	Polsek Ciniru	12 Febuari 2021
2.	NS	ADR	Penipuan	Lembar Kwitansi	Ciwaru	Polsek Ciwaru	3 Januari 2021
3.	Dalam Lidik	Pelapor	Curat	Kartu Keluarga	Puhun	Polsek Ciawigebang	06 Juni 2021
4.	Dalam Lidik	Pelapor	Curat	ATM	Pangbuyutan	Polsek Lebakwangi	13 Mei 2021
5.	Dalam Lidik	AMA	Curat	Ijazah	Gunung Keling	Polsek Cigugur	10 Agustus 2021
6.	Dalam Lidik	JS	Curat	Kartu Keluarga	Cilame	Polsek Cigugur	09 Januari 2021
7.	Dalam Lidik	SN	Curat	Ktp	Gibug	Polsek Cigugur	09 Januari 2021
8.	Dalam Lidik	DDH	Curat	Paspor	Windu Herang	Polsek Cigugur	13 Januari 2021
9.	Dalam Lidik	YA	Penipuan	Kartu Yp3js	Cipari	Polsek Cigugur	21 Januari 2021
10.	Dalam Lidik		Curat	Sepeda Motor	Ciwaru	Polsek Cigugur	27 Febuari 2021
..	…….	……	……	…….	…….	…….	…….
..	…….	……	……..	…….	…….	…….	…….
..	…….	……	……..	…….	…….	…….	…….
800.	HS	CP	KDRT	Ktp	Rumah	Polsek Cilimus	18 Desember 2023

4.2. Data Selection

Data yang dipilih untuk penelitian ini adalah data tintak kriminalitas tahun 2021, 2022 dan 2023. Data ini terdiri dari 800 record dan terdiri dari tujuh atribut: nomor, Inisial Pelaku, Inisial Korban, Jenis Tindak Kriminalitas, Barang Bukti, TKP, Kasus Ditangani, dan Waktu .

Tabel 2. Data Colection

No	Nama Atribut	Type
1	No	Integer
2	Inisial Pelaku	Nominal
3	Inisial Korban	Nominal
4	Jenis Tindakan Kejahatan	Nominal
5	Bukti	Nominal
6	TKP	Nominal
7	Kasus Ditangani	Nominal
8.	Waktu	Integer

4.3. Data Transformasi

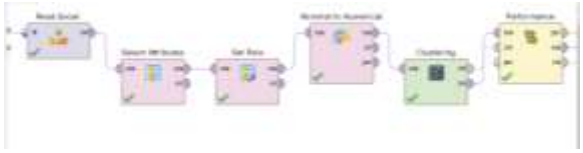
Gambar 1. Data Sebelum Transformasi

Konversi data dilakukan untuk mengubah data bertipe nominal menjadi data numerik dengan menggunakan operator nominal ke numerik.

Gambar 2. Data Sesudah transformasi

4.4. Data Mining

Data mining melibatkan penggunaan algoritma K-Means dan pengukuran besar kecilnya nilai Davies-Bouldin yang diperoleh dari algoritma K-Means menggunakan operator Cluster Distance Performance.



Gambar 3. Proses Pengelolaan Data Mining

Dengan menggunakan operator algoritma K-Means dan pengukuran ukuran Davies-Bouldin yang diperoleh dari algoritma K-Means dengan menggunakan operator Cluster Distance Performance maka dihasilkan parameter-parameter.

Tabel 3. Parameter

Parameter	Keterangan
K min	3
K max	10
Measure type	*NumericalMeasurement*
Numerical measure	*EunclideanDistance*
Clustering algorithm	*Kmeans*

4.5. Interpretasi Data

Hasil data mining yang diperoleh dengan menggunakan algoritma K-Means adalah sebagai berikut :

a) Hasil pengelompokan menggunakan k-means diperoleh terbaik sebanyak 9 cluster model, yaitu cluster 0 sebayak 141 items, cluster 1 sebanyak 128 item, cluster 2 106 items, cluster 3 110 items, cluster 5 151 items, cluster 6 49 items, cluster 7 45 items, dan cluster 8 2 items

Cluster Model

```
Cluster 0: 141 items
Cluster 1: 128 items
Cluster 2: 106 items
Cluster 3: 110 items
Cluster 4: 151 items
Cluster 5: 49 items
Cluster 6: 64 items
Cluster 7: 45 items
Cluster 8: 2 items
Total number of items: 706
```

Gambar 4. Cluster Model

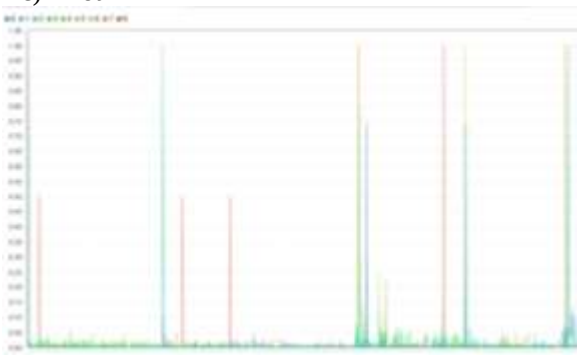
b) Nilai Davies Bouldin Indek Yang dihasilkan dari algoritma k-means sebesar -0.004 dengan menghitung jarak antara cluster diperoleh bahwa cluster 1 merupakan kelompok tertinggi

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -0.004
Avg. within centroid distance_cluster_0: -0.004
Avg. within centroid distance_cluster_1: -0.003
Avg. within centroid distance_cluster_2: -0.004
Avg. within centroid distance_cluster_3: -0.004
Avg. within centroid distance_cluster_4: -0.005
Avg. within centroid distance_cluster_5: -0.002
Avg. within centroid distance_cluster_6: -0.003
Avg. within centroid distance_cluster_7: -0.004
Avg. within centroid distance_cluster_8: -0.001
Davies Bouldin: -0.004
```

Gambar 5. Performance Vector

c) Plot

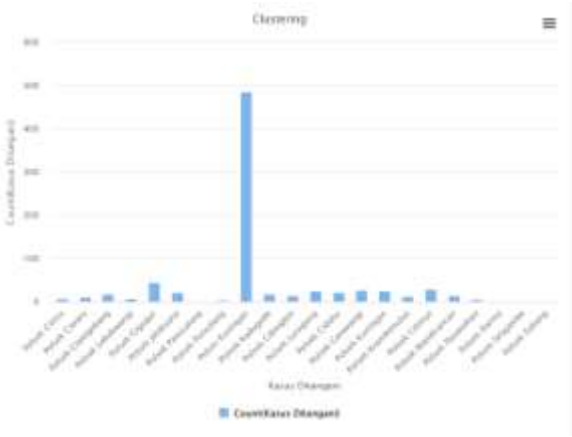


Gambar 6. Plot

Setiap bar pada plot akan mewakili kontribusi persentase dari setiap cluster terhadap total kriminalitas

d) Visualizasions

Visualisasi data membantu kita lebih memahami data, mengidentifikasi pola dan tren, serta membuat keputusan yang lebih tepat. Oleh karena itu, visualisasi data merupakan langkah penting dalam proses penambangan data.



Gambar 7. Visualization Kasus Ditangani

e) Nominal Values

Di RapidMiner, nilai nominal diwakili oleh tipe data nominal. Tipe data nominal memiliki beberapa karakteristik, sebagai berikut :

Index	Nominal value	Absolute count	Fraction
1	Polres Kuningan	485	0.609
2	Polsek Cigugur	45	0.057
3	Polsek Cilimus	28	0.035
4	Polsek Garawangi	26	0.033
5	Polsek Kuningan	25	0.031
6	Polsek Luragung	25	0.031
7	Polsek Jalaksana	22	0.028
8	Polsek Cidahu	21	0.026
9	Polsek Kadugede	18	0.023
10	Polsek Ciawigebang	17	0.021
11	Polsek Mandirancan	15	0.019
12	Polsek Cibingbin	14	0.018
13	Polsek KramMulya	13	0.016
14	Polsek Ciwaru	11	0.014
15	Polsek Ciniru	8	0.010
16	Polsek Lebakwangi	7	0.009
17	Polsek Pasawahan	6	0.008
18	Polsek Pancalang	4	0.005
19	Polsek Darma	2	0.003
20	Polsek Subang	2	0.003
21	Polsek Panrcalang	1	0.001
22	Polsek Selajamba	1	0.001

Gambar 8. Hasil Nominal Values

Proses pengelompokan *K-Means* menggunakan beberapa operator untuk menganalisis data :

a. Muat Excel dan muat data yang akan dianalisis.
b. Pilih Atribut dan pilih atribut yang diperlukan. Yaitu kode, nama produk, persediaan awal, dan sisa persediaan.
c. Perubahan dari nominal ke numerik. Yaitu mengubah atribut Nominal Type menjadi Numeric.
d. Clustering digunakan untuk mengelompokkan sesuatu menjadi beberapa cluster.
e. *Operator Cluster Performance Distance* digunakan untuk mengevaluasi kinerja

pengelompokan berbasis *centroid* untuk menghasilkan model *cluster* pusat. Model *cluster* centroid berisi informasi tentang pembentukan *cluster* yang dilakukan. Jarak kinerja *cluster* Operator mendukung dua ukuran kinerja: jarak *cluster* rata-rata dan indeks *Davies-Bouldin*. Indikator kinerja ini dijelaskan berdasarkan parameter. Untuk mencari nilai k yang optimal, harus terlebih dahulu bereksperimen dengan nilai k tersebut untuk mencari nilai indeks *Davies-Bouldin* yang terkecil. Parameter berikut diuji untuk menentukan nilai *Davies-Bouldin*:

Tabel 4. Parameter

Parameters	Value
K/Cluster	*3 sampai 9*
Max Runs	*10*
Measure Types	*NumericalMeasures*
Numerical Measure	*EuclideanDistance*
Main Criterion	*Davies Bouldin*

Dalam percobaan yang sudah dilakukan menggunakan k=3 sampai k=9, berikut hasil yang diperoleh :

Tabel 5. Hasil Cluster dan DBi

Cluster	Davies Bouldin
K0	-0.004
K1	-0.003
K2	-0.004
K3	-0.004
K4	-0.005
K5	-0.002
K6	-0.003
K7	-0.004
K8	-0.001

Hasil DBi = -0,001 dari nilai k optimal k=8 menunjukkan bahwa akurasi model k-means yang dihasilkan baik, untuk nilai -0.001. Oleh karena itu, nilai DBi -0.001 dapat dikategorikan sebagai nilai yang sangat baik. Nilai tersebut menunjukkan bahwa model k-means yang digunakan untuk mengklasifikasikan data dari Polsek Salajambe memiliki akurasi yang sangat tinggi. Nilai DBi yang negatif menunjukkan bahwa model dapat memprediksi kelas data dengan lebih akurat dibandingkan hanya menebak secara acak. Nilai DBi yang lebih kecil menunjukkan model yang lebih akurat. Nilai k=8 menunjukkan bahwa model K-means yang dihasilkan menggunakan delapan cluster. Jumlah cluster yang optimal dapat ditentukan dengan menggunakan berbagai metode, misalnya metode siku. Metode ini membandingkan nilai DBi untuk nilai k yang berbeda. Nilai k yang menyebabkan perubahan besar pada nilai DBi menunjukkan bahwa nilai k tersebut merupakan nilai k optimal.

5. KESIMPULAN DAN SARAN

Atribut dengan variabilitas tinggi membantu algoritma *K-Means* membedakan *cluster*, sedangkan atribut dengan distribusi berbeda membantu algoritma *K-Means* menemukan *cluster* terbaik.Berdasarkan hasil penelitian pengelompokkan kriminalitas menggunakan algoritma *k-Means*, maka hasilnya dapat disimpulkan sebagai berikut : Dapat menentukan penilaian yang objektif dan tepat sasaran dalam menentukan jenis kriminalitas dengan menerapkan data mining menggunakan algoritma *k-means.* Dapat mengetahui informasi pengelompokan dataset area kasus ditangani yang signifikan Nilai *Davies Bouldin Index* yang dihasilkan dari algoritma *k-means* ini sebesar 3.323. Dapat menganalisa dan pengelompokan dataset kriminalitas menggunakan algoritma k-means. Nilai K optimal yang dihasilkan dari data kejahatan dengan menggunakan algoritma *K-Means* bergantung pada atribut yang digunakan, metode yang digunakan untuk menentukan nilai K, dan tujuan analisis. Secara umum nilai K optimal adalah yang menghasilkan cluster dengan inersia rendah, siluet tinggi, dan tidak ada outlier. Berdasarkan aspek-aspek tersebut, dapat disimpulkan bahwa penerapan data mining dalam situasi kriminal dapat membawa manfaat yang signifikan. Hal ini didukung oleh berbagai penelitian yang menunjukkan bahwa data mining dapat meningkatkan akurasi prediksi kejahatan, efisiensi investigasi kejahatan, dan pemahaman pola dan tren kejahatan.

Saran dari penulis dalam penelitian ini adalah perlunya membandingkan dengan algoritma lain untuk mengelompokan dataset kriminalitas sehingga diharapkan : Dalam pengelompokan yang dilakukan oleh penulis dapat memiliki nilai akurasi yang lebih baik dan akurat. Dengan analisa tindak kriminalitas dengan menggunakan algoritma k-means ini semoga selanjutnya dapat membandingkan dengan algoritma yang lain seperti : Algoritma x-means, algoritma k-means++, algoritma k-mendoids, dan lain-lain.

DAFTAR PUSTAKA

[1] H. L. Mustopa, P. Nopi, and I. Feri, "Algoritma k-means dalam mengelompokkan daerah rawan kriminalitas pencurian sepeda motor di Kabupaten Asahan," *J. Teknol. Komput. Paluta*, vol. 1, no. 2019, pp. 1–8, 2021.

[2] D. Alfania, R. Buaton, S. Ramadani, and S. Kaputama, "Data Mining Pengelompokan Kriminalitas Di Kota Binjai Dengan Menggunakan Metode Clustering," *Agustus*, vol. 6, no. 3, 2022.

[3] S. B. Faradilla, "Komparasi Analisis K-Medoids Clustering Dan Hierarchical Clustering," *(Studi Kasus Data Kriminalitas di Indones. Tahun 2020)*, 2022.

[4] F. Wiza, "Klasterisasi karakteristik kekerasan seksual terhadap anak dengan metode k-means cluster analysis," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 10, no. 1, pp. 44–53, 2019.

[5] N. Septiani, K. Erwansyah, and ..., "Analisis Data Mining Pengelompokkan Kasus Tindak Kejahatan Yang Terjadi Di Kecamatan Medan Polonia Dengan Menggunakan Metode K-

Means Clustering," *J. Cyber Tech*, vol. 3, no. 2, pp. 1–13, 2020.

[6] Y. Aswan, S. Defit, and G. W. Nurcahyo, "Algoritma K-Means Clustering dalam Mengklasifikasi Data Daerah Rawan Tindak Kriminalitas (Polres Kepulauan Mentawai)," *J. Sistim Inf. dan Teknol.*, pp. 245–250, Aug. 2021.

[7] C. A. Sri Fastaf and Y. Yamasari, "Analisa Pemetaan Kriminalitas Kabupaten Bangkalan Menggunakan Metode K-Means dan K-Means++," *J. Informatics Comput. Sci.*, vol. 3, no. 04, pp. 534–546, 2022.

[8] F. Rahmadayanti and R. Rahayu, "PENERAPAN METODE DATA MINING PADA KASUS KRIMINALITAS," vol. 15, no. 1, pp. 52–61, 2023.

[9] L. Suriani, "Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering," *J. Sist. Komput. dan Inform.*, vol. 1, no. 2, p. 151, 2020.

[10] R. N. Fahmi, M. Jajuli, and N. Sulistiyowati, "Analisis Pemetaan Tingkat Kriminalitas di Kabupaten Karawang menggunakan Algoritma K-Means," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 4, no. 1, pp. 67–79, 2021.