

Pemetaan Tingkat Kriminalitas Kota Pagar Alam Dengan Memanfaat Metode *Machine Learning*

Sasmita¹, Fitria Rahmadayanti², Rika Rahayu³

^{1,2,3} Teknik Informatika, Institut Teknologi Pagar Alam, Pagar Alam, Sumatera Selatan

Email: ¹samitha661@gmail.com, ²ria.ria.rr71@gmail.com, ³rikarupiah@gmail.com

Email Penulis Korespondensi: rikarupiah@gmail.com

Article History:

Received Dec 10th, 2023

Revised Dec 27th, 2023

Accepted Jan 15th, 2024

Abstrak

Tujuan dari penelitian ini adalah menghasilkan pemetaan tingkat kriminalitas kota Pagar Alam dengan memanfaatkan metode *machine learning*. Penelitian ini dilatar belakangi data laporan polisi kurang lengkapnya data dalam berita acara pemeriksaan yang digunakan, berdampak pada informasi terkait lokasi kejadian yang ada pada data laporan Satreskrim sehingga kurangnya keakuratan informasi. Metode pengembangan sistem menggunakan *CRISP-DM* yang terdiri dari 6 tahap yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation* dan *Deployment*. Pada penelitian ini algoritma yang digunakan yaitu *K-Means Clustering*, data diolah menjadi 3 *Cluster* yaitu *Cluster* tingkat kriminalitas tidak rawan (*C0*), *Cluster* rawan (*C1*), *Cluster* sangat rawan (*C2*). Hasil diperoleh tahun 2020 untuk *cluster_0* yaitu sebanyak 28 kelurahan, tahun 2021 sebanyak 24 kelurahan dan tahun 2020 sebanyak 20 kelurahan. Untuk *cluster_1* pada tahun 2020 yaitu 6 kelurahan, tahun 2021 sebanyak 10 kelurahan dan tahun 2022 sebanyak 11 kelurahan. Dan *cluster_2* pada tahun 2020 dan 2021 yaitu hanya 1 kelurahan dan tahun 2022 sebanyak 4 kelurahan. Metode pengujian menggunakan *Elbow*. Hasil dari pengujian metode *Elbow* untuk menghitung hasil *SSE* terbentuk 3 *cluster* ($K=3$) dengan nilai 3244.766. Hasil pengujian dengan jumlah *cluster* 3 dapat dikatakan *valid* atau sesuai dengan hasil *clustering k-means* pada *Rapid Miner*, hasil dari klasterisasi berupa pemetaan menggunakan *ArcGIS*.

Kata Kunci : Kriminalitas, *CRIPS-DM*, *K-Means*, *Clustering*, *Elbow Method*, Pemetaan

Abstract

The aim of this research is to produce a mapping of the crime rate of Pagar Alam city by utilizing machine learning methods. This research was based on the incomplete data in the police report data in the investigation report used, which had an impact on the information related to the location of the incident in the Satreskrim report data, resulting in less accurate information. The system development method uses *CRISP-DM* which consists of 6 stages, namely *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation* and *Deployment*. In this research, the algorithm used is *K-Means Clustering*, the data is processed into 3 clusters, namely the not-prone crime level cluster (*C0*), the vulnerable cluster (*C1*), the very vulnerable cluster (*C2*). The results obtained in 2020 for *cluster_0* were 28 subdistricts, in 2021 there were 24 subdistricts and in 2020 there were 20 subdistricts. For *cluster_1* in 2020 there are 6 sub-districts, in 2021 there are 10 sub-districts and in 2022 there are 11 sub-districts. And *cluster_2* in 2020 and 2021 is only 1 sub-district and in 2022 there will be 4 sub-districts. The test method uses *Elbow*. The results of testing the *Elbow* method for calculating *SSE* results formed 3 clusters ($K=3$) with a value of 3244,766. The test results with a cluster number of 3 can be said to be valid or in accordance with the results of *k-means clustering* on *Rapid Miner*, the results of clustering are in the form of mapping using *ArcGIS*.

Keyword : Crime, *CRIPS-DM*, *K-Means*, *Clustering*, *Elbow Method*, Mapping

1. PENDAHULUAN

Perkembangan teknologi dan informasi saat ini mengalami kemajuan yang sangat pesat. Hal ini dipicu oleh kemajuan zaman dan pola pikir manusia yang semakin cerdas. Teknologi adalah kumpulan pengetahuan, keterampilan, dan alat yang digunakan untuk menciptakan, mengembangkan, dan memperbaiki barang dan layanan yang dapat memenuhi kebutuhan manusia [1]. Teknologi yang semakin maju berdampak pada perkembangan masyarakat, hal ini dapat

menimbulkan persaingan perekonomian dan banyaknya angka pengangguran menyebabkan munculnya kejahatan termasuk kriminalitas. Kriminalitas adalah mencakup berbagai tindakan dan perilaku yang menyebabkan kerugian baik secara ekonomis maupun psikologis, yang bertentangan dengan hukum yang berlaku di Indonesia, serta norma-norma sosial dan agama [2]. Oleh sebab itu peneliti memanfaatkan metode *Machine learning* yang bekerja apabila tersedia data sebagai input untuk dilakukan analisis terhadap kumpulan data besar (big data) sehingga menemukan pola tertentu.

Machine learning adalah cabang ilmu bagian dari kecerdasan buatan (artificial intelligence), dengan pemrograman untuk memungkinkan komputer menjadi cerdas berperilaku seperti manusia, dan dapat meningkatkan pemahamannya melalui pengalaman secara otomatis [3] sehingga *Machine learning* sangat membutuhkan data untuk dapat dipelajari atau yang biasa disebut data training [4]. Menurut Larose dalam [5] data mining merupakan sebuah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning yang mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakrit dari berbagai database yang besar dan data mining sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan field [6]. Salah satu teknik data mining untuk mengelompokkan wilayah kriminalitas adalah menggunakan Clustering.

Clustering merupakan metode pengelompokan data yang tidak bergantung pada kelas-kelas tertentu, di mana data dikelompokkan ke dalam kelompok objek yang sama. Dalam suatu cluster, terdapat sejumlah catatan yang memiliki kesamaan satu sama lain dan bersifat berbeda dari catatan dalam cluster lainnya.[7]. Salah satu metode untuk melakukan *cluster* adalah *K-Means*. *K-means* merupakan metode pengelompokan data yang berupaya mempartisi data ke dalam kelompok, sehingga data dengan kesamaan dikelompokkan dalam satu cluster, sementara data dengan karakteristik yang berbeda dikelompokkan ke dalam cluster yang berbeda pula [8] yang mana proses pengklasteran algoritma ini dengan cara mempartisi atau mengelompokkan dataset menjadi beberapa klaster [9].

Berdasarkan hasil wawancara dan observasi kepada pihak Kepolisian Resor (Polres) kota Pagar Alam saat ini dalam data laporan polisi kurang lengkapnya data dalam berita acara pemeriksaan (BAP) yang digunakan, berdampak pada informasi terkait lokasi kejadian (TKP) yang ada pada data laporan Satreskrim, pada penelitian ini pemetaan akan memuat semua komponen penting berupa alamat lengkap kejadian, kedepannya pihak kepolisian akan menghimpun data selengkap mungkin untuk lokasi yang belum diketahui, untuk lokasi yang telah diketahui akan dilanjutkan dengan penerapannya di dalam sebuah pemetaan. Dari kurangnya kelengkapan data tersebut pihak kepolisian Resor (Polres) sulit untuk menentukan wilayah mana saja yang memiliki tingkat kerawanan kriminalitas. Sehingga mereka membutuhkan sebuah pemetaan terkait data kriminalitas yang terjadi di kota Pagar Alam. Dengan adanya penelitian dapat memaksimalkan pihak Kepolisian Resor (Polres) kota Pagar Alam untuk mengatasi kekurangan tersebut, peneliti memutuskan untuk melakukan pemetaan pada data tersebut. Pemetaan ini dilakukan dengan tujuan menyusun informasi secara visual, memungkinkan pemahaman yang lebih baik terhadap lokasi kejadian. Tindakan ini diambil untuk memperbaiki ketidaklengkapan data BAP dan memastikan bahwa data terkait TKP menjadi lebih lengkap dan mudah dipahami. Dari proses pengumpulan data tersebut maka dibutuhkan sebuah data mining dengan metode *Clustering* menggunakan Algoritma *K-Means* untuk mengetahui tindak kriminalitas yang ada di Pagar Alam, dan dapat dijadikan sebagai pedoman bagi pihak kepolisian untuk melakukan kebijakan selanjutnya.

Menurut penelitian yang dilakukan oleh [10], hasil dari penerapan algoritma *K-Means* menunjukkan bahwa terdapat 6 anggota dalam cluster 1 yang memiliki tingkat kriminalitas tinggi, dan 28 anggota dalam cluster 2 yang memiliki tingkat kriminalitas rendah. Di sisi lain, ketika menggunakan perhitungan manual algoritma *K-Medoids*, ditemukan bahwa cluster 1 memiliki 7 anggota dengan tingkat kriminalitas tinggi, sementara cluster 2 memiliki 27 anggota dengan tingkat kriminalitas rendah. Hal yang menarik dari penelitian ini adalah penggunaan metode *K-Means* dan *K-Medoids* untuk mengelompokkan tindak kriminalitas, khususnya dalam pemetaan data kriminalitas di Pagar Alam.

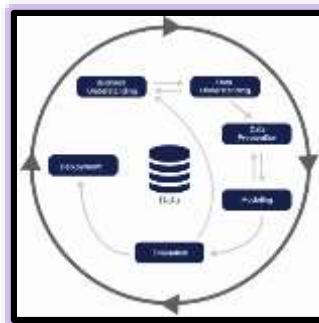
Berdasarkan penelitian dari [11] dengan menerapkan Algoritma *Fp-Growth*, pencarian frequent itemset dapat dilakukan dengan lebih mudah karena memanfaatkan struktur data *Fp-Tree*. Algoritma ini juga bermanfaat dalam mengidentifikasi tingkat kejahatan yang sering terjadi, khususnya pencurian yang memiliki persentase tinggi. Informasi ini dapat memberikan manfaat bagi pihak berwenang dalam penanganan kejahatan di wilayah tersebut. Keterkaitan dengan penelitian ini adalah penggunaan metode *Fp-Growth* untuk memudahkan pencarian frequent itemset dengan *Fp-Tree*, sementara penelitian selanjutnya akan menerapkan algoritma *K-Means* dalam analisis kriminalitas di kota Pagar Alam. Menurut penelitian yang dilakukan oleh [12], penerapan algoritma *k-means* telah berhasil dalam mengidentifikasi pengelompokan daerah rawan kriminalitas di Kabupaten Karawang selama periode 2019 hingga 2020, berdasarkan lokasi, jumlah kejadian, dan jenis kasus. Evaluasi terhadap pemetaan tingkat kriminalitas tersebut menggunakan algoritma *k-means*, dengan menggunakan *silhouette coefficient* untuk mengukur kualitas klaster. Hasil evaluasi menunjukkan bahwa baik pada clustering tahun 2019 maupun 2020, klaster masuk dalam kategori struktur medium, yang diinterpretasikan sebagai penempatan klaster yang wajar. Hubungan dengan penelitian ini adalah bahwa penelitian tersebut menggunakan metode menggunakan *silhouette coefficient* untuk menguji kualitas sebuah klaster dan penelitian ini menggunakan metode *elbow* untuk mengevaluasi hasil dari pemodelan yang diterapkan.

Dari ketiga penelitian terdahulu dapat disimpulkan peneliti, bahwa sistem klasterisasi menggunakan Algoritma *K-Means* dari hasil masing-masing metode yang di pakai menghasilkan kelas-kelas dan memiliki tingkat akurasi yang cukup baik dalam pemetaan wilayah, dan akan dilakukan hasil dari pengujian *rapidminer* akan dibandingkan dengan hasil pengujian yang ada di *google colab* dengan menggunakan metode *elbow*. Dengan beberapa penelitian di atas maka akan

menjadi referensi bagi penulis yang akan dirancang dan dibangun secara terkomputerisasi dalam pemetaan tingkat kriminalitas berdasarkan wilayah yang sering terjadi di kota Pagar Alam.

2. METODOLOGI PENELITIAN

Metode penelitian adalah rangkaian aktivitas untuk menemukan kebenaran dalam suatu studi penelitian. Proses ini dimulai dengan pembentukan rumusan masalah berdasarkan pemikiran, yang kemudian menghasilkan hipotesis awal. Dengan dukungan literatur dan wawasan dari penelitian sebelumnya, penelitian dapat diolah, dianalisis, dan pada akhirnya membentuk kesimpulan [13]. Pada Penelitian ini dilakukan dengan metodologi *CRIPS-DM* dengan menerapkan metode *clustering* menggunakan algoritma *k-means*. *CRISP-DM* merupakan metode yang bisa digunakan dalam mengatasi strategi pemecahan masalah umum dan menjadi metodologi yang memberikan standar baku dalam kegiatan data mining. Terdapat enam tahap dalam metode *CRISP-DM*, yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan implementasi [14]. Pada tahun 1996, *CRISP-DM*, sebuah metodologi untuk proyek data mining, dikembangkan oleh praktisi dan perusahaan yang berpengalaman di bidang tersebut, termasuk *Daimler Chrysler*, *SpSS Inc.* (yang pada saat itu dikenal sebagai *Integral Solution Limited*), dan *NCR Corporation*. [15].



Gambar 1. Model *Crisp-DM*

1. Pemahaman Bisnis (*Business Understanding*)

Yaitu dilakukan analisis masalah mengacu pada data kriminalitas. Pada langkah ini, pemahaman tentang kepentingan menggunakan data kriminalitas diperlukan agar dapat mengurangi tingkat kejahatan.

2. Pemahaman Data (*Data Understanding*)

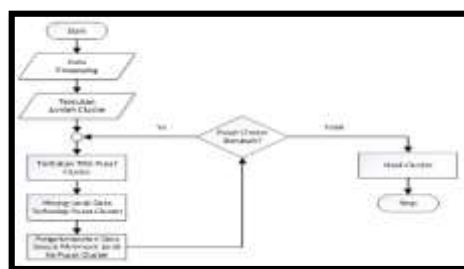
Peneliti melakukan pemahaman terhadap kebutuhan data terkait dengan kasus Kriminalitas di kota Pagar Alam. Data diambil dari bagian Kepolisian Resor (Polres) kota Pagar Alam berupa file *microsoft word* selama 3 tahun dari 2020 sampai tahun 2022.

3. Persiapan Data (*Data Preparation*)

Setelah melakukan fase pemahaman dari data kehadiran pegawai, maka pada fase pengolahan data ini membuat dataset final yang akan diterapkan ke dalam pemodelan data. Dalam tahapan ini yaitu membangun dataset akhir dari berupa data mentah. Ada beberapa hal yang dilakukan antara lain pembersihan data, pembersihan *record*, kemudian dilakukan seleksi data, record dan atribut.

4. Pemodelan (*Modelling*)

Adapun pemodelan yang dilakukan pada penelitian ini adalah metode clustering menggunakan algoritma *k-means*. Berikut langkah-langkah algoritma *K-Means* yang digunakan dalam pemodelan ini :



Gambar 2. Langkah-langkah algoritma *k-means*

a. Tentukan Jumlah *Cluster*

Berdasarkan *dataset* yang disiapkan, digunakan *cluster* yang dibentuk berdasarkan 3 kelompok kasus kriminalitas yakni kelompok Sangat Rawan, Rawan, Cukup Rawan dan Tidak Rawan pada kota Pagar Alam.

C0 : Tidak Rawan

C1 : Rawan

C2 : Sangat Rawan

b. Tentukan *centroid* secara acak

Karena ada 3 cluster yang diharapkan, diperlukan 43 centroid sebagai pusat cluster, yang ditentukan secara acak.

c. Hitung jarak data ke centroid

Untuk menghitung jarak ke centroid digunakan rumus *Euclidean Distance* [16] :

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

Keterangan :

- d(x,y) = Jarak antara data x ke data
- x_i = Data *testing* ke-i
- y_i = Data *training* ke-i

Dalam perhitungan ini, langkah awalnya adalah menetapkan nilai pusat cluster awal untuk setiap variabel pada setiap cluster. Pada iterasi pertama, nilai pusat cluster awal diberikan secara acak. Pada iterasi berikutnya, nilai pusat cluster awal (dari iterasi ke-1 hingga mencapai iterasi normal atau maksimal) dihitung dengan mengambil rata-rata nilai data pada setiap cluster. Jika nilai pusat cluster awal yang baru sama dengan nilai pusat cluster awal yang lama, proses iterasi akan berlanjut hingga nilai tersebut sama atau mencapai batas iterasi maksimal yang telah ditetapkan sebelumnya. Namun, jika nilai pusat cluster awal yang baru sama dengan nilai pusat cluster awal yang lama, proses pengelompokan berhenti.

d. Perbarui titik *Centroid*

Perhitungan titik *centroid* baru dengan berdasarkan pengelompokan yang pertama untuk melakukan iterasi selanjutnya.

e. Lakukan langkah 3 hingga 5 secara berulang hingga nilai titik centroid tidak mengalami perubahan. Perhitungan dilakukan seterusnya sampai dengan jumlah data yang telah ditentukan. Setelah menghitung jarak data dari pusat, langkah selanjutnya adalah mengelompokkan data berdasarkan jarak terdekat. Kemudian Anda dapat melakukan lebih banyak pengulangan untuk membandingkan hasilnya. Jika centroid berubah, ulangi langkah tersebut sampai centroid tidak berubah.

5. *Evaluation*

Pada tahap ini lebih difokuskan pada model yang dihasilkan sudah sesuai standar *K-Means Clustering* dan tidak ada yang dilewatkan saat melakukan tahap awal hingga tahap pemodelan selesai.

6. *Deployment*

Setelah proses pemodelan dengan menerapkan teknik data mining clustering, selanjutnya adalah tahapan penyusunan laporan dan proses pelaporan hasil kepada pihak kepolisian yang diharapkan kepolisian dapat menentukan keputusan yang tepat dalam mengatasi dan melakukan pencegahan kriminalitas yang terjadi [17].

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Dari hasil penelitian ini, diperoleh pengelompokan data kriminalitas menggunakan model algoritma k-means, yang dapat dijadikan sebagai rekomendasi untuk clustering yang ideal untuk melihat wilayah mana saja yang memiliki tingkat kerawanan kriminalitas selama 3 tahun yaitu dari tahun 2020 sampai dengan 2022 dengan jumlah 15 jenis kasus kriminalitas. Peneliti menggunakan algoritma *k-means*, peneliti menggunakan bantuan aplikasi *Rapid Miner* terlebih dahulu untuk menghasilkan *cluster*. Data total kriminalitas yaitu 15 jenis kasus kriminalitas yang akan di *cluster* dan total *record* dari 470 data setelah diproses *cleaning* menjadi 35 data kriminalitas per kelurahan kota Pagar Alam. Setelah itu peneliti melakukan proses analisa kriminalitas dengan melihat rata-rata kasus kriminalitas di *excel* yang kemudian akan menjadi *datasheet* yang diolah ke dalam *Rapid Miner*. Kemudian hasil dari tahapan pengujian dengan metode *Elbow* pada aplikasi *Google Colab* dengan bahasa pemrograman *Python* untuk menghitung hasil *Sum of square error (SSE)*.

3.2 Pembahasan

Metodologi CRISP-DM terdiri dari serangkaian tahapan sebagai berikut:

1. Pemahaman Bisnis

Proses pengelolaan data kriminalitas untuk menentukan jenis kriminalitas apa saja yang sering terjadi masih dilakukan secara manual, yaitu melihat dan menghitung di *Excel*. Dari proses pengelompokan tersebut belum adanya pengelolaan lebih lanjut. Maka dari itu peneliti akan melakukan pemetaan tingkat kriminalitas di kota Pagar Alam untuk mengetahui wilayah mana saja yang sering terjadi kriminalitas, sehingga dinilai perlu bagi pihak kepolisian untuk meningkatkan keamanan masyarakat kota Pagar Alam, dan untuk menentukan serta membuat kebijakan baru terhadap pihak kepolisian dalam menjaga keamanan dan ketertiban masyarakat.

2. Pemahaman Data

Pada fase pemahaman data ini, data didapat dari pihak polres kota Pagar Alam yaitu data kriminalitas selama 3 tahun pada tahun 2020 sampai dengan tahun 2022 dengan jumlah 470 *reccord* dengan 8 atribut yaitu no, laporan polisi, korban, pelaku, uraian kejadian, pasal, keterangan da penyidik. Kategori data yang diterima dalam bentuk *word* dan data yang didapat perlu di *cleaning* dan akan dilakukan pemilihan data dengan atributnya.

3. Pengolahan Data

Pengolahan data langsung diimplementasikan pada *Rapid Miner* dengan tiga tahapan yang ditetapkan pada pengolahan data yaitu :

a. Data Selection

Pada tahapan ini diperoleh 8 atribut yaitu No, laporan Polisi, Korban, Pelaku, Uraian Kejadian, Pasal, Ket dan Penyidik. Dilakukan seleksi menggunakan *excel* dengan cara difilter. Setelah dilakukan filter, data yang peneliti gunakan ada 18 atribut pada tahap *selection data* diantaranya Kecamatan, Kelurahan, Jenis Kriminalitas yaitu KDRT, Penipuan, Cabul, Pencurian, Curat, Penganiayaan, Curas, Kekerasan Anak dibawah Umur, Pengeroyokan, Persetubuhan, Penganiayaan Anak, Penyebaran Foto UU ITE, Penggelapan, persetubuhan, kekerasan dan total dengan total jumlah kriminalitas 470 *record* setelah di *selection* menjadi 35 data kriminalitas berdasarkan kelurahan. Dan hasil *record* data kriminalitas dari 470 data setelah di *cleaning* menjadi 35 data yang akan dijadikan *datashee* dan peneliti melakukan proses *clustering* berdasarkan tahun. Dengan total data kriminalitas pada tahun 2020 sampai dengan tahun 2022 berjumlah 35 data kriminaitas berdasarkan kelurahan. Selanjutnya peneliti melakukan input data dan analisa data kriminalitas dari setiap laporan kejadian menggunakan *excel*. Dengan melihat rata-rata data kriminalitas yang akan di jadikan *datasheet* pertahun. Kemudian *datasheet* dari hasil tersebut dapat dilihat pada tabel dibawah ini.

Tabel 1. Data Selection Tahun 2020

Kecamatan	Kelurahan	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14	K15	Total
1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
...
5	35	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	2

b. Data processing/cleaning

Proses *processing* merupakan proses yang mencakup *cleaning*, dan *transformasi* data. Pada tahap data *processing*, peneliti sudah memastikan tidak ada lagi *missing value* atau data yang kosong. Pada tahun 2020 data yang di *cleaning* sebanyak 37 data, tahun 2021 sebanyak 43 data dan tahun 2022 sebanyak 24 data yang di *cleaning*. Sehingga pembersihan data ini juga akan mempengaruhi hasil dari proses pengolahan data yang dihasilkan dalam pemrosesannya.

c. Data Transformation

Pada fase transformasi data, data yang diproses menunjukkan atribut yang dipilih dan diproses di *Rapid Miner*. Adapun transformasi data sample kriminalitas yang dilakukan selama tahun 2020 sampai 2022 sehingga menghasilkan data sebagai berikut:

Tabel 2. Transformation Kriminalitas Kecamatan

Kecamatan	Keterangan
Dempo Selatan	1
Dempo Tengah	2
Dempo Utara	3
Pagar Alam Selatan	4
Pagar Alam Utara	5

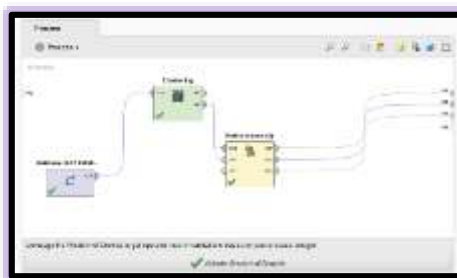
Tabel 3. Tabel Transformasi kriminalitas kelurahan

Kelurahan	Keterangan	Kelurahan	Keterangan
Penjalang	1	Muara Siban	13
Lubuk Buntak	2	Jangkar Mas	14
Perahu Dipo	3	Burung Dinang	15
Kance Diwe	4	Bumi Agung	16

Kelurahan	Keterangan	Kelurahan	Keterangan
Atung Bungsu	5	Gunung Agung	17
Pelang Kenidai	6	Tanjung Agung	18
Padang Temu	7	Ulu Lurah	19
Karang Dalo	8	Tumbak Ulas	20
Jokoh	9	Tebat Giri Indah	21
Candi Jaya	10	Nendagung	22
Rebah Tinggi	11	Besemah Serasan	23
Pagar Wangi	12	Sidorejo	24
Gunung Dempo	25	Selibar	31
Beringn Jaya	26	Curup Jare	32
Pagar Alam	27	Bangun Rejo	33
Bangun Jaya	28	Dempo Makmur	34
Alun Dua	29	Sukorejo	35
Kuripan Babas	30		

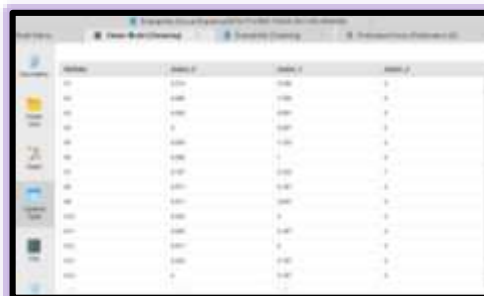
4. Modelling

Datasheet kriminalitas pada tahun 2020 sampai tahun 2022 sebanyak 35 data kriminalitas berdasarkan kelurahan yang kemudian di proses menjadi *datasheet* pertahun dan selanjutnya dilakukan proses *modeling* menggunakan algoritma *k-means clustering*. Untuk proses *modelling* dapat dilihat pada gambar dibawah ini.



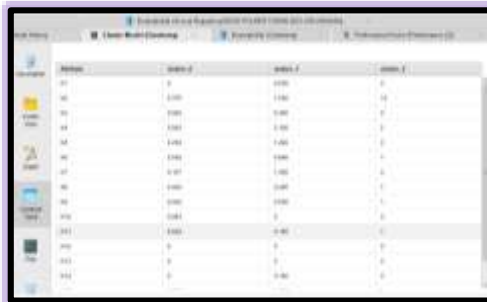
Gambar 2. Model Algoritma K-Means

- a. Tentukan Jumlah *Cluster*
Setelah menyiapkan dataset maka langkah selanjutnya adalah menentukan jumlah cluster disini peneliti menggunakan 3 cluster. Untuk menentukan jumlah cluster dilakukan beberapa percobaan dilakukan, yang dapat dilihat pada fokus terkecil.
- b. Tentukan Titik *Centroid*
Pertama klasterisasi kriminalitas untuk tahun 2020 Setelah dilakukan beberapa percobaan, percobaan dengan 3 cluster yang paling tepat. Dengan pengukuran *performance vector* rata-rata dalam *centroid distance* yang didapat dengan nilai 3.744, kemudian rata-rata *cluster_0* dengan nilai 3.620, rata-rata *cluster_1* dengan nilai 9.611, dan rata-rata dalam *centroid cluster_2* dengan nilai 0.000 dan nilai *Device Bouldi Index* adalah 0.593.



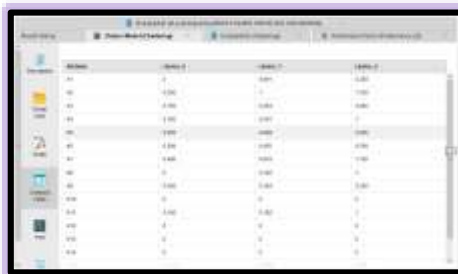
Gambar 3. Titik *centroid* kriminalitas tahun 2020

Percobaan dengan jumlah *cluster* 3. Percobaan 3 *cluster* itu C0, C1, dan C2 pada tahun 2021. Dengan pengukuran *performance vector* rata-rata dalam *centroid distance* yang didapat dengan nilai 5.561, kemudian rata-rata *cluster_0* dengan nilai 3.535, rata-rata *cluster_1* dengan nilai 10.980, dan rata-rata dalam *centroid cluster_2* dengan nilai 0.000 dan nilai *Device Bouldi Index* adalah 0.571.



Gambar 4. Titik *centroid* kriminalitas tahun 2021

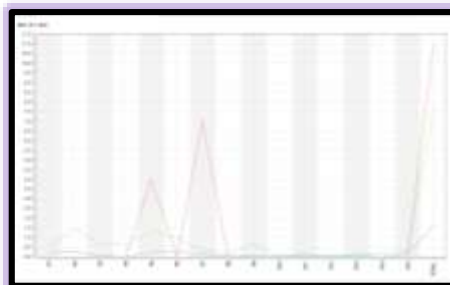
Selanjutnya percobaan kedua untuk kriminalitas tahun 2022 dengan jumlah *cluster* 3. Percobaan 3 *cluster* itu C0, C1, dan C2 Dengan pengukuran *performance vector* rata-rata dalam *centroid distance* yang didapat dengan nilai 6.811, kemudian rata-rata *cluster_0* dengan 2.580, rata-rata *cluster_1* dengan nilai 6.843, dan rata-rata dalam *centroid cluster_2* dengan nilai 27.875 dan nilai *Device Bouldi Index* adalah 0.893.



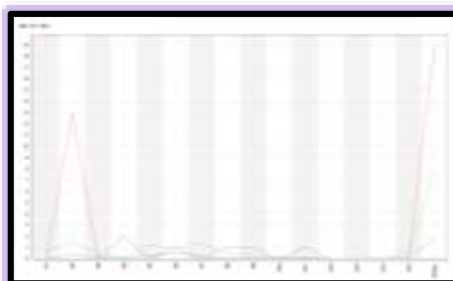
Gambar 5. Titik *centroid* kriminalitas tahun 2022

c. Hitung jarak dengan *Euclidean Distance*

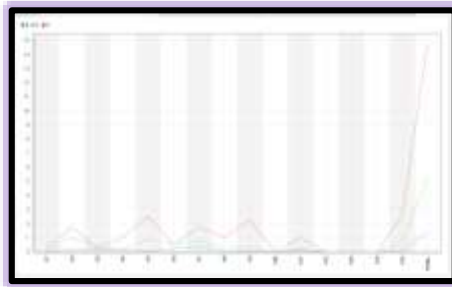
Setelah mengimpor dataset dan algoritme *k-means* ke dalam *Rapid Miner*, proses data dengan mengklik *Run* untuk menampilkan jarak ke pusat data.



Gambar 6. Jarak data ke centroid Tahun 2020



Gambar 7. Jarak data ke centroid Tahun 2021



Gambar 8. Jarak data ke centroid Tahun 2022

d. Perbarui nilai titik *centorid*

Tahap 1 nilai titik *centroid* untuk kriminalitas tahun 2020 pada iterasi 1 dengan nilai 5.402, pada iterasi 2 dengan nilai 2.964, dan iterasi 3 dengan nilai 12.374. Kemudian titik *centorid* pada tahun 2021 pada iterasi 1 nilai titik *centroid* adalah 5.561, pada iterasi 2 dengan nilai 4.162 dan iterasi 3 dengan nilai 9.678. Untuk tahun 2022 nilai iterasi 1 dengan nilai 9.678, iterasi 2 dengan nilai 6.811 dan iterasi 3 dengan nilai 4.128.

e. Lakukan langkah 3 hingga 5 secara berulang hingga nilai titik *centroid* tidak mengalami perubahan.

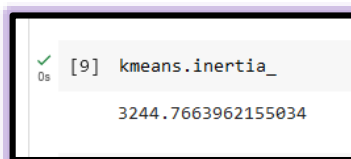
Jika titik pusat *centorid* berubah maka iterasi diulangi, tetapi jika tidak berubah maka iterasi diberikan dan diperoleh hasil dari masing-masing kelompok.

5. **Evaluation**

Setelah hasil *cluster* didapatkan dari data kriminalitas selama 3 tahun yang sudah dimodelkan sesuai dengan standar *K-Means Clustering*, langkah selanjutnya akan dilakukan tahap *evaluation* menggunakan *Elbow Method* yang sudah diprogram kedalam *Python* melalui *Google Colaboration*. Sedangkan metode *elbow* adalah metode untuk menentukan jumlah *cluster* yang tepat atau terbaik dengan cara membandingkan persentase hasil antara jumlah *cluster* pembentuk siku pada suatu titik [18]. Langkah selanjutnya akan dilakukan tahap *evaluation* menggunakan *Elbow Method*. Dengan menghitung hasil dari *Sum of Square Error (SSE)* : Persamaan (6) Menghitung *SSE* :

$$SSE = \sum_{k=1}^k \sum_{xi} [Xi - Ck]^2 \dots\dots\dots(2)$$

Beberapa tahapan metode *elbow* yaitu inialisasi nilai awal k, naikan nilai k, kemudian hitung hasil *sum of square error (SSE)*, melihat hasil *SSE*, didapatkan nilai *cluster* yang berbentuk siku. Kemudian diperoleh hasil perhitungan *SSE* dari *Python* dibawah ini :



Gambar 9. Hasil Perhitungan SSE

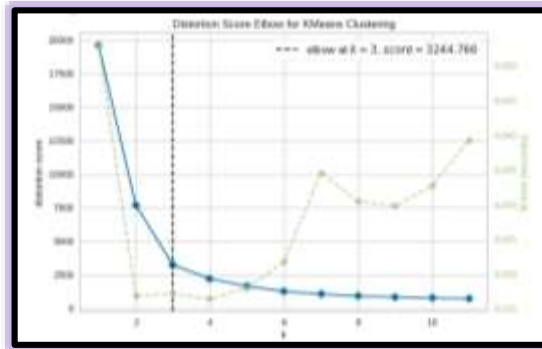
a. **Menampilkan Hasil Pengujian *Elbow Method***

Berikut *coding* untuk menampilkan hasil dari pengujian *elbow method* untuk menentukan *cluster* yang tepat.

```
< [6] from yellowbrick.cluster import KElbowVisualizer
model = KMeans()
visualizer = KElbowVisualizer(model, k=(1,12)).fit(df)
visualizer.show()
```

Gambar 10. Coding Program *Elbow Method*

Kemudian didapatkan hasil *cluster* yang berbentuk siku, dari data kriminalitas selama 3 tahun sebanyak 361 *record* maka didapatkan hasil *cluster* yang tepat berjumlah 3 *cluster* (K=3) dengan nilai *sum of square error* yaitu 3244.766 Hasil *Elbow* dapat dilihat pada gambar dibawah ini :

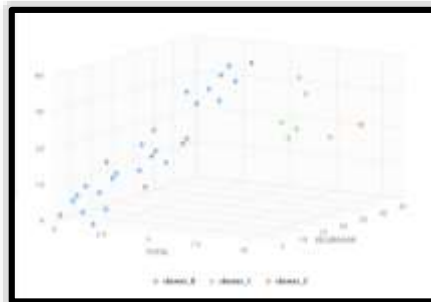


Gambar 11. Hasil Pengujian *Elbow Method*

Berdasarkan hasil pengujian tersebut dari hasil *clustering k-means* pada *Rapid Miner* dengan jumlah *cluster* 3 dapat dikatakan *valid* atau sesuai dengan hasil *clustering k-means* pada *python* menggunakan *elbow method*.

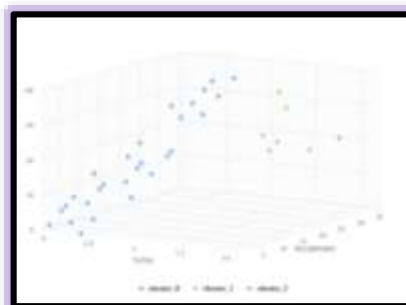
6. Deployment

Pada tahap ini *deployment* merupakan tahapan terakhir berupa pengetahuan atau informasi yaitu mengenai pemetaan tingkat kriminalitas kota Pagar Alam selama 3 tahun sehingga dapat diketahui ada 3 *cluster* yaitu *C0* keterangan tidak rawan, *C1* yaitu rawan dan *C2* yaitu sangat rawan. Dari data kriminalitas yang diolah dengan *Rapid Miner* yang digunakan pada sistem menggunakan metode *Clustering K-Means* maka diperoleh data kriminalitas ditahun 2020 yaitu *Cluster_0* memiliki tingkat kriminalitas Tidak Rawan dengan Jumlah 28 kelurahan *Cluster_1* memiliki tingkat kriminalitas Rawan dengan 6 kelurahan, *Cluster_2* memiliki tingkat kriminalitas dengan jumlah 1 kelurahan Sangat Rawan.



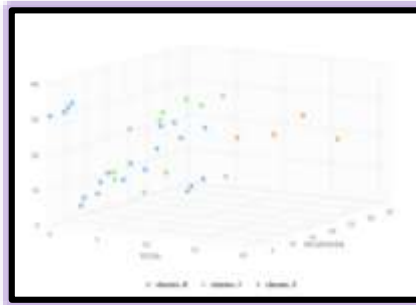
Gambar 12. Grafik kriminalitas Tahun 2020

Gambar diatas merupakan hasil *clustering* kriminalitas kota pagar alam pada tahun 2020 dengan *cluster_0* yang terdiri dari 28 kelurahan dengan data kriminalitas ditahun 2020 yaitu *Cluster_0* tingkat kriminalitas Tidak Rawan berjumlah 28 kelurahan, *Cluster_1* tingkat kriminalitas Rawan terdiri dari 6 kelurahan, *Cluster_2* tingkat kriminalitas berjumlah 1 kelurahan Sangat Rawan.



Gambar 13. Grafik kriminalitas Tahun 2021

Selanjutnya ditahun 2021 diperoleh *Cluster_0* sebanyak 24 kelurahan dengan tingkat kriminalitas Tidak Rawan, *Cluster_1* sebanyak 10 kelurahan dengan tingkat kriminalitas Rawan dan *Cluster_2* sebanyak 1 kelurahan dengan tingkat kriminalitas Sangat Rawan.



Gambar 14. Grafik kriminalitas Tahun 2022

Pada tahun 2022 diperoleh cluster_0 sebanyak 20 kelurahan dengan tingkat kriminalitas tidak rawan, cluster_1 sebanyak 11 kelurahan dengan tingkat kriminalitas rawan, dan cluster_2 sebanyak 4 kelurahan dengan keterangan tingkat kriminalitas Sangat Rawan.

Berdasarkan perhitungan *Rapid Miner*, didapatkan sebuah model yang akan diaplikasikan dalam Python. Pola yang digunakan untuk mengelompokkan atau mengklaster data berdasarkan hasil perhitungan jarak adalah:

- Jika $C0 < C1$ dan $C0 < C2$ maka *Cluster_0* dikategorikan sebagai Tidak Rawan
- Jika $C1 < C0$ dan $C1 < C2$ maka *Cluster_1* dikategorikan sebagai Rawan
- Jika $C1 < C0$ dan $C2 < C1$ maka *Cluster_2* dikategorikan sebagai Sangat Rawan

Hasil yang didapatkan dari pengklasteran divisualisasikan dalam bentuk peta dimana daerah dengan tingkat kriminalitas masuk dalam *cluster* (C0) tidak rawan kriminalitas akan diberi warna hijau, wilayah yang masuk dalam *cluster* rawan kriminalitas (C1) akan di beri warna kuning dan wilayah yang masuk kedalam *cluster* sangat rawan kriminalitas (C2) akan diberi warna merah. Akan dilakukan pembagian pemetaan wilayah kriminalitas berdasarkan kelurahan. Sehingga dengan pemetaan perkelurahan ini akan mengetahui titik-titik lokasi yang sering terjadi kriminalitas di kota Pagar Alam.



Gambar 15.Peta Kriminalitas Tahun 2020

Pada tahun 2020 dengan *cluster_0* dengan keterangan berwarna hijau yang terdiri dari 28 kelurahan dengan kategori tidak rawan yaitu kelurahan penjalang, lubuk buntak, perahu dipo, kance diwe, atung bungsu, pelang kenidai, padang temu, karang dalo, jokoh, candi jaya, rebah tinggi, pagar wangi, muara siban, jangkar mas, burung dinang, bumi agung, agung lawangan, tanjung agung (payang), ulu lurah, tumbak ulas, beringin jaya, pagar alam, bangun jaya, alun dua, selibar, curup jare, dempo makmur dan sukorejo. Kemudian pada *cluster_1* dengan keterangan berwarna kuning yaitu terdiri dari 7 kelurahan dengan kategori rawan yaitu kelurahan tebat giri indah, nendagung, besemah serasan, sidorejo, kuripan babas dan bangun rejo. Selanjutnya pada *cluster_2* dengan keterangan berwarna merah yaitu terdiri dari 1 kelurahan dengan kategori Sangat Rawan yaitu kelurahan nendagung.



Gambar 16. Peta kriminalitas Tahun 2021

Pada tahun 2021 dengan *cluster_0* dengan keterangan berwarna hijau yang terdiri dari 24 kelurahan dengan kategori tidak rawan yaitu kelurahan penjalang, lubuk buntak, perahu dipo, kance diwe, atung bungsu, pelang kenidai, padang temu, karang dalo, jokoh, candi jaya, rebah tinggi, pagar wangi, muara siban, jangkar mas, burung dinang, beringin jaya, pagar alam, bangun jaya, kuripan babas, selibar, curup jare, dempo makmur, agung lawangan dan ulu lurah. Kemudian pada *cluster_1* dengan keterangan berwarna kuning yaitu terdiri dari 10 kelurahan dengan kategori rawan yaitu kelurahan alun dua, bangun rejo, sukorejo, bumi agung, tanjung agung (payang), tumbak ulas, tebat giri indah, besemah serasan, sidorejo dan gunung dempo. Selanjutnya pada *cluster_2* dengan keterangan berwarna merah yaitu terdiri dari 1 kelurahan dengan kategori Sangat Rawan yaitu kelurahan nendagung.



Gambar 17. Peta Kriminalitas Tahun 2022

Pada tahun 2022 dengan *cluster_0* dengan keterangan berwarna hijau yang terdiri dari 20 kelurahan dengan kategori tidak rawan yaitu kelurahan selibar, curup jare, dempo makmur, pelang kenidai, jokoh, candi jaya, muara siban, jangkar mas, burung dinang, agung lawangan, tanjung agung (payang), tumbak ulas, beringin jaya, bangun jaya, alun dua, kuripan babas, penjalang, lubuk buntak, perahu dipo dan kance diwe. Kemudian pada *cluster_1* dengan keterangan berwarna kuning yaitu terdiri dari 11 kelurahan dengan kategori rawan yaitu kelurahan sukorejo, pagar wangi, atung bungsu, padang temu, karang dalo, rebah tinggi, bumi agung, ulu lurah, sidorejo, gunung dempo dan pagar alam. Selanjutnya pada *cluster_2* dengan keterangan berwarna merah yaitu terdiri dari 4 kelurahan dengan kategori Sangat Rawan yaitu kelurahan tebat giri indah, besemah serasan, nendagung dan bangun rejo.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, diperoleh hasil bahwa penelitian ini menghasilkan pemetaan tingkat kriminalitas kota Pagalaran. Hasil dari proses clustering data dalam mengetahui pemetaan tingkat kriminalitas selama tiga tahun melalui aplikasi Rapid Miner sama dengan hasil yang diterapkan di *python* dengan jumlah cluster yang terdiri dari 3 cluster yakni jenis kriminalitas yang terjadi selama 3 tahun dengan total 35 data per kelurahan. Diketahui *cluster_0* yaitu Tidak Rawan pada tahun 2020 ke 2021 yaitu terjadi peningkatan kriminalitas sebanyak yaitu 28 kelurahan menjadi 24 kelurahan. Kemudian pada tahun 2021 ke tahun 2022 terjadi peningkatan kriminalitas yaitu dari 24 kelurahan menjadi 20 kelurahan. Untuk *cluster_1* dengan kriminalitas Rawan pada tahun 2020 ke 2021 terjadi peningkatan yaitu dari 6 kelurahan menjadi 10 kelurahan, kemudian pada tahun 2021 ke 2022 terjadi peningkatan sebanyak 11 kelurahan untuk kriminalitas dengan tingkat Rawan. Dan *cluster_2* yaitu sangat Rawan tidak terjadi peningkatan dari tahun 2020 ke tahun 2021 yaitu hanya 1 kelurahan. Kemudian tahun 2021 ke tahun 2022 sebanyak 4 kelurahan dengan tingkat kriminalitas Sangat Rawan. Kemudian hasil pada tahap pengujian menggunakan *Elbow Method* pada aplikasi *google colab* dengan bahasa pemrograman *python* untuk menghitung hasil dari *Sum of Square Error (SSE)* diperoleh jumlah *cluster* yang tepat yaitu $K=3$ dengan jumlah nilai 3244.766 K adalah *Cluster_0* dengan tingkat kriminalitas Tidak Rawan, *cluster_1* dengan tingkat kriminalitas Rawan, dan *cluster_2* yaitu Sangat Rawan. Sehingga dapat dikatakan *valid* atau sesuai dengan hasil *cluster* pada aplikasi *Rapid Miner*.

UCAPAN TERIMA KASIH

Terima kasih kepada ibu Sasmita, M.Kom selaku pembimbing I dan juga ibu Fitria Rahmadayanti, M.Kom selaku Pembimbing II yang telah membantu dalam menyelesaikan karya ilmiah ini dengan sebaik-baiknya.

DAFTAR PUSTAKA

- [1] A. Alimuddin, J. N. S. Juntak, R. A. E. Jusnita, I. Murniawaty, and H. Y. Wono, "Teknologi dalam Pendidikan: Membantu Siswa Beradaptasi Dengan Revolusi Industri 4.0," *Menur PJournal Educ.*, vol. 05, no. 04, pp. 11777–11790, 2023, [Online]. Available: <http://jonedu.org/index.php/joe>
- [2] L. Suriani, "Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering," vol. 1, pp. 151–157, 2020, doi: 10.30865/json.v1i2.1955.
- [3] E. Retnoningsih and R. Pramudita, "Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python," *Bina Insa. Ict J.*, vol. 7, no. 2, p. 156, 2020, doi: 10.51211/biict.v7i2.1422.
- [4] Wiki Lofandri, "Analisis Predictive Maintenance Peralatan Lab Berbasis Machine Learning," *J. Sistim Inf. dan Teknol.*, vol. 5, pp. 7–9, 2022, doi: 10.37034/jsisfotek.v5i1.164.
- [5] F. Rahmadayanti, I. Anggraini, and T. Susanti, "Pengklasterisasian Data Penyakit Hipertensi dengan Menggunakan Metode K-Means," *J. Inf. Syst. Res.*, vol. 4, no. 2, pp. 737–741, 2023, doi: 10.47065/josh.v4i2.2905.
- [6] R. I. O. Limabri, F. Putrawansyah, and A. Arif, "Penerapan Data Mining Untuk Mengklasifikasi Nasabah Bank Sumsel Babel Menggunakan Algoritma C4. 5," *Escaf*, pp. 1101–1108, 2023, [Online]. Available: <https://semnas.univbinainsan.ac.id/index.php/escaf/article/view/488>
- [7] E. Luthfi and A. W. Wijayanto, "Analisis perbandingan metode hirearchical , k-means , dan k-medoids clustering dalam pengelompokan indeks pembangunan manusia Indonesia Comparative analysis of hirearchical , k-means , and k-medoids clustering and methods in grouping Indonesia ' s human," *Inovasi*, vol. 17, no. 4, pp. 770–782, 2021, [Online]. Available: <https://journal.feb.unmul.ac.id/index.php/INOVASI/article/download/10106/1437>
- [8] Sasmita and S. Muntari, "Penerapan Algoritma K-Means Clustering Pada Data Keluhan Pelanggan Pt. PIn Persero Kota Pagar Alam," *J. Ilm. Teknosains*, vol. 9, no. 1, pp. 8–12, 2023.
- [9] P. Apriyani, A. R. Dikananda, and I. Ali, "Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi," *Hello World J. Ilmu Komput.*, 2023, [Online]. Available: <https://doi.org/10.56211/helloworld.v2i1.230>
- [10] H. D. Tampubolon, S. Suhada, M. Safii, S. Solikhun, and D. Suhendro, "Penerapan Algoritma K-Means dan K-Medoids Clustering untuk Mengelompokkan Tindak Kriminalitas Berdasarkan Provinsi," *J. Ilmu Komput. dan Teknol.*, vol. 2, no. 2, pp. 6–12, 2021, doi: 10.35960/ikomti.v2i2.703.
- [11] D. Winarti, M. Kom, E. Revita, M. Kom, E. Y. S, and M. Kom, "Penerapan Data Mining untuk Analisa Tingkat Kriminalitas Dengan Algoritma Association Rule Metode FP-Growth," *J. SIMTIKA*, vol. 4, no. 3, pp. 8–22, 2021.
- [12] N. S. Resti Noor Fahmi, Mohamad Jajuli, "ANALISIS PEMETAAN TINGKAT KRIMINALITAS DI KABUPATEN KARAWANG MENGGUNAKAN ALGORITMA K-MEANS," *J. Inf. Technol. Comput. Sci.*, vol. 4, no. 2614–1574, pp. 67–79, 2021.
- [13] S. H. Sahir, *Buku ini di tulis oleh Dosen Universitas Medan Area Hak Cipta di Lindungi oleh Undang-Undang Telah di Deposit ke Repository UMA pada tanggal 27 Januari 2022.* 2022.
- [14] G. Radiana *et al.*, "APLIKASI KAI ACCESS MENGGUNAKAN METODE SUPPORT VECTOR MACHINE," no. April, pp. 1–10, 2023.
- [15] N. N. Hakim, "Implementasi Machine Learning pada Sistem Prediksi Kejadian dan Lokasi Patah Rel Kereta Api di Indonesia," *J. Sist. Cerdas*, vol. 03, no. 01, pp. 25–35, 2020.
- [16] F. Putrawansyah and N. C. Dewi, "Implementasi Algoritma K-Means dengan menggunakan Metode Profile Matching pada Alumni STT Pagar Alam Implementation of the K-Means Algorithm using the Profile Matching Method for Alumni of STT Pagar Alam," *J. Ilm. Bin. STMIK Bina Nusant. Jaya*, vol. 3, no. 2, pp. 2657–2117, 2021.
- [17] F. Rahmadayanti and R. Rahayu, "Penerapan Metode Data Mining Pada Kasus Kriminalitas Indonesia," *J. Teknol. Inf. ...*, vol. 15, no. 1, pp. 52–61, 2023, [Online]. Available: <https://www.jurnal.univbinainsan.ac.id/index.php/jti/article/view/2054%0Ahttps://www.jurnal.univbinainsan.ac.id/index.php/jti/article/download/2054/1046>
- [18] E. Virantika and J. Ipawati, "Evaluasi Hasil Pengujian Tingkat Clusterisasi Penerapan Metode K-Means Dalam Menentukan Tingkat Penyebaran Covid-19 di Indonesia," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 2614–5278, pp. 1657–1666, 2022, doi: 10.30865/mib.v6i3.4325.