

Penerapan Data Mining Clustering Algoritma K-Means Untuk Menganalisa Pola Kejadian Tindak Kejahatan (Studi Kasus Polrestabes Semarang)

Ematitania Kharisma¹, Artef Jananto²

^{1,2}Fakultas Teknologi Informasi dan Industri, Universitas Stikubank, Semarang, Indonesia

E-mail: ematitania28@gmail.com¹, ajananto09@edu.unisbank.ac.id²

Abstract

Crime is a complex social problem that affects the security of the community, especially in the city of Semarang. Therefore, in an effort to deal with the increasing crime, the use of information technology and data analysis becomes very relevant. This research aims to implement data mining algorithm K-Means Clustering in analyzing crime patterns in Semarang City. This research method involves the use of historical crime data from January to December 2022 with a total data of 305 incidents in Semarang City. The K-Means Clustering algorithm in data mining was chosen because of its ability to group data based on similar characteristics effectively. The data was analyzed using RapidMiner software, which facilitated the clustering of crime patterns into seven clusters cluster 1 with 60, cluster 2 with 31, cluster 3 with 36, cluster 4 with 35, cluster 5 with 51, cluster 6 with 55, and cluster 7 with 37. These findings provide a strong basis for the police to design more targeted and efficient crime handling strategies. The implementation of the K-Means Clustering algorithm in this study proved effective in identifying crime patterns and providing useful insights for security policy decision-making. This research also opens up opportunities for the development of more sophisticated information systems in city security management in the future.

Keywords: Crime, Data Mining, K-Means, Clustering, RapidMiner

Abstrak

Tindak Kejahatan merupakan permasalahan sosial yang kompleks dan berdampak kepada keamanan Masyarakat, terutama pada kota Semarang. Penelitian ini bertujuan untuk mengimplementasikan data mining algoritma K-Means Clustering dalam menganalisis pola tindak kriminal di Kota Semarang. Metode penelitian ini melibatkan penggunaan data historis tindak kejahatan dari bulan januari sampai desember 2022 dengan jumlah data 305 kejadian di Kota Semarang. Algoritma K-Means Clustering dalam data mining dipilih karena kemampuannya dalam mengelompokkan data berdasarkan kesamaan karakteristik secara efektif. Data dianalisis menggunakan perangkat lunak RapidMiner, yang memfasilitasi pengelompokan pola tindak kejahatan menjadi tujuh cluster cluster 1 sebanyak 60, cluster 2 sebanyak 31, cluster 3 sebanyak 36, cluster 4 sebanyak 35, cluster 5 sebanyak 51, cluster 6 sebanyak 55, dan cluster 7 sebanyak 37. Temuan ini memberikan dasar yang kuat bagi pihak kepolisian untuk merancang strategi penanganan tindak kejahatan yang lebih tepat sasaran dan efisien. Implementasi algoritma K-Means Clustering dalam penelitian ini terbukti efektif dalam mengidentifikasi pola tindak kriminal dan memberikan wawasan yang berguna bagi pengambilan keputusan terkait kebijakan keamanan. Penelitian ini juga membuka peluang untuk pengembangan sistem informasi yang lebih canggih dalam manajemen keamanan kota di masa depan.

Kata kunci: Tindak Kejahatan, Data Mining, K-Means, Clustering, RapidMiner

1. Pendahuluan

Tindak kejahatan merupakan suatu tindakan anti sosial yang melanggar hukum, berkaitan dengan merampas hak milik orang lain dan dapat menimbulkan suatu kerugian, ketidaknyamanan, dan ketidakpatuhan dalam bermasyarakat [1]. Tindak kejahatan adalah salah satu permasalahan utama yang dihadapi oleh berbagai kota, termasuk Semarang. POLRESTABES (Kepolisian Resort Besar Kota) Semarang sebagai lembaga penegak hukum memiliki tanggung jawab besar untuk menjaga keamanan dan ketertiban masyarakat. Untuk Mencapai tujuan ini, diperlukan strategi yang efektif dalam menganalisis dan mengidentifikasi pola-pola tindak kejahatan agar tindakan preventif dan responsive dapat dilakukan dengan tepat.

Seiring dengan perkembangan teknologi informasi, pemanfaatan data besar (big data) dan Teknik-teknik data mining menjadi semakin relevan dalam membantu mengatasi permasalahan kriminalitas. Salah satu teknik data mining yang efektif dalam mengelompokkan data berdasarkan kemiripannya adalah algoritma *K-Means clustering*. Algoritma ini memungkinkan pengelompokan data tindak kejahatan berdasarkan karakteristik tertentu, sehingga dapat mengidentifikasi kejahatan yang sering terjadi dengan pola yang mirip.

Penelitian ini menerapkan algoritma *k-means clustering* dalam menganalisa data tindak kejahatan yang terjadi di wilayah hukum POLRESTABES (Kepolisian Resort Kota Besar) Semarang. Dengan menggunakan data historis yang dimiliki oleh POLRESTABES Semarang, diharapkan dapat ditemukan pola-pola tindak kejahatan yang terjadi. Informasi ini sangat berharga bagi pihak kepolisian dalam merancang strategi keamanan yang lebih tepat sasaran.

Metode *K-Means clustering* dipilih karena kemampuannya dalam mengelompokkan data efisien dan efektif. Objek-objek atau data yang dikelompokkan ke dalam suatu grup memiliki ciri-ciri yang sama berdasarkan kriteria tertentu, salah satu aktifitas analisis data adalah klasifikasi atau pengelompokan data ke dalam beberapa kategori atau cluster. Untuk mengelompokkan terhadap data pengelompokan suku cadang dana sensoris computer menggunakan teknik *clustering*. Dengan menggunakan metode ini data-data yang didapat bisa dikelompokkan ke dalam beberapa cluster, berdasarkan kemiripan data-data tersebut, sehingga data-data yang dimiliki karakteristik yang sama dikelompokkan dalam satu *cluster* yang lain [2].

Penelitian ini memiliki tujuan yang ingin dicapai dalam penelitian ini yaitu *pertama*, mengimplementasikan Data Mining Algoritma *K-Means Clustering* untuk mengekstrak ilmu pengetahuan, informasi penting dan menarik dari database. *Kedua*, mengelompokkan data kejadian tindak kejahatan untuk mendapatkan pola kejadian tindak kejahatan dengan menggunakan algoritma K-Means. *Ketiga*, menggunakan alat bantu (*Tools*) *RapidMiner* untuk melakukan proses *clustering*.

Berdasarkan hasil penelitian yang dilakukan oleh Penelitian oleh [3] dengan judul Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering. Dalam penelitian tersebut metode yang digunakan yaitu algoritma *k-means clustering* dengan *software rapid miner* dengan tujuan untuk menentukan daerah rawan kriminalitas. Penulis menyebutkan bahwa pembagian daerah rawan Poldasu belum efektif dan efisien, dikarenakan belum diadakannya secara manual. Penelitian ini juga membagikan kategori berdasarkan data-data JTPJPTP, kemudian akan dikategorikan menjadi daerah rawan, sedang, tidak rawan, dengan nilai-nilai yang didapat pada setiap *cluster*.

Berdasarkan hasil penelitian oleh [4] dengan judul Penerapan Data Mining Menggunakan Metode K-Means Clustering Pada Data Tindak Kriminalitas Di POLRES Kabupaten Kuningan. Dalam penelitian ini penulis mengidentifikasi pola kriminalitas di Kabupaten Kuningan dengan data yang ada di Polres Kabupaten Kuningan serta menerapkan algoritma *K-Means Clustering* dalam mengelompokkan area yang memiliki

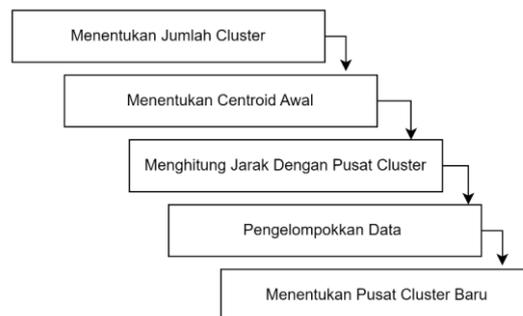
karakteristik serupa dari segi tingginya kriminalitas. Atribut dengan variabilitas tinggi membantu algoritma *K-Means* membedakan *cluster*, sedangkan atribut dengan distribusi berbeda membantu algoritma *K-Means* menemukan *cluster* terbaik. Berdasarkan hasil penelitian pengelompokan kriminalitas menggunakan algoritma *k-Means*, maka hasilnya 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.

Berdasarkan hasil penelitian oleh [5] dengan judul Clustering Menggunakan Algoritma K-Means Untuk Mengelompokkan Wilayah Rawan Kejahatan Di Wilayah Kabupaten Solok. Dalam penelitian ini data tindak pidana yang dijadikan sebagai sumber data untuk proses clustering menggunakan algoritma *K-Means*. Penulis melakukan pengujian dengan aplikasi *RapidMiner* sebagai hasil perbandingan antara pencarian dengan manual dengan menggunakan aplikasi.

Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi nyata dalam upaya peningkatan keamanan di wilayah Semarang. Penerapan data mining *clustering* menggunakan algoritma *k-means* tidak hanya akan membantu dalam identifikasi pola-pola tindak kejahatan, tetapi juga membuka peluang untuk pengembangan sistem informasi yang lebih canggih dalam manajemen keamanan kota.

2. Metodologi Penelitian

Penelitian ini menggunakan metode kuantitatif dengan pendekatan data mining untuk menganalisis data tindak kejahatan menggunakan algoritma *k-means clustering*. Peneliti memperoleh data melalui proses observasi dan wawancara kepada instansi. Data yang digunakan oleh penelitian ini adalah data tindak kejahatan Kota Semarang Tahun 2022 dengan jumlah data 305. Sebelum data diproses menggunakan metode *k-means clustering*, peneliti sudah melakukan proses kdd pada data mining agar data yang diteliti dapat digunakan pada metode *k-means clustering*.



Gambar 1. Proses Metode K-Means Clustering

Tahapan proses perhitungan data Tindak Kejahatan dengan algoritma *k-mean clustering* sebagai berikut:

2.1. Menentukan Jumlah Cluster

Untuk menemukan berapa banyak cluster yang akan digunakan, peneliti menggunakan software *RapidMiner* untuk menentukan *average* jarak terkecil tiap *cluster* dengan jumlah *cluster*.

2.2. Menentukan Centroid Awal

Untuk menentukan *centroid* awal dilakukan secara random / acak dari objek-objek yang tersedia sebanyak *k cluster*, kemudian untuk menghitung centroid cluster ke-1

berikutnya, digunakan rumus sebagai berikut [6]:

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

Di mana \Leftrightarrow v : centroid pada cluster.
 x_i : objek ke- i.
 n : banyaknya objek.

2.3. Menghitung Jarak Dengan Pusat Cluster

Untuk menghitung jarak antara objek dengan centroid menggunakan *Euclidian Distance* :

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

Di mana \Leftrightarrow x_i : objek x ke-i.
 y_i : daya y ke-i.
 n : banyaknya objek.

2.4. Pengelompokkan Data

Data yang dihitung dikelompokkan dari jarak terdekat objek kedalam masing – masing *cluster*. Ini pertanda bahwa data masuk kedalam kelompok yang sama atau data memiliki karakteristik yang sama.

2.5. Menentukan Pusat Cluster Baru

Menentukan pusat *cluster* baru dengan menghitung rata – rata anggota *cluster* dan pusat cluster. Jika hasil yang di peroleh tidak sama, maka dilanjutkan dengan iterasi berulang hingga menemukan pusat *cluster* sebelumnya dengan pusat *cluster* setelahnya sama.

3. Hasil dan Pembahasan

Data yang digunakan oleh penelitian ini adalah data tindak kejahatan Kota Semarang Tahun 2022 dengan jumlah data 305. Sebelum data diproses menggunakan metode *k-means clustering*, peneliti sudah melakukan proses *KDD* pada data mining agar data yang diteliti dapat digunakan pada metode *k-means clustering*.

Tabel 1. Sampel Data Tindak Kejahatan

No	Hari Kejadian	Bulan												Jml
		1	2	3	4	5	6	7	8	9	10	11	12	
1.	SENIN	4	2	0	4	2	3	2	9	7	2	1	0	36
2.	SELASA	4	0	4	1	2	5	6	3	3	0	5	1	34
3.	RABU	3	2	2	5	1	5	3	3	5	3	7	3	42
4.	KAMIS	1	2	5	1	4	5	5	2	7	5	3	3	43
5.	JUM'AT	3	2	4	6	4	6	3	3	6	5	4	4	50
6.	SABTU	4	4	5	5	4	6	6	7	9	5	3	4	62
7.	MINGGU	2	0	3	2	2	6	2	3	10	2	4	2	38
	JUMLAH	21	12	23	24	19	36	27	30	47	22	27	17	305

3.1. Analisa Data

Sebelum data dilakukan proses metode *k-means clustering*, akan dilakukan tranformasi data. Proses ini dilakukan agar data dapat diolah dengan *algoritma k-means clustering*. Dari data tindak kejahatan ada beberapa data yang ditranformasikan diantaranya :

a) Bulan

Tabel 2. Tranformasi Atribut Bulan

Bulan	Kode
Januari	1
Februari	2
Maret	3

Bulan	Kode
April	4
Mei	5
Juni	6
Juli	7
Agustus	8
September	9
Oktober	10

b) Hari

Tabel 3. Tranformasi Atribut Hari

Hari	Kode
Senin	1
Selasa	2
Rabu	3
Kamis	4
Jumat	5
Sabtu	6
Minggu	7

c) Jam

Tabel 4. Tranformasi Atribut Jam

Jam	Kode
00.01 - 03.00 WIB	1
03.01 - 06.00 WIB	2
06.01 - 09.00 WIB	3
09.01 - 12.00 WIB	4
12.01 - 15.00 WIB	5
15.01 - 18.00 WIB	6
18.01 - 21.00 WIB	7
21.01 - 00.00 WIB	8

d) Jenis Kejadian

Tabel 5. Tranformasi Atribut Jenis Kejadian

Jenis Kejadian	Kode
Curat	1
Curas	2
Curanmor	3
Rampas	4

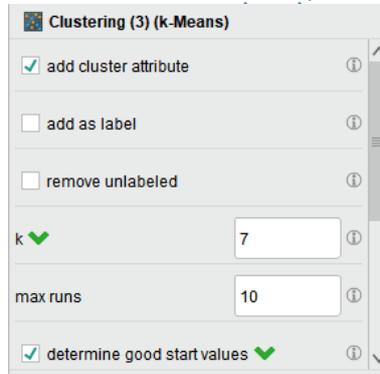
Dari hasil tranformasi tersebut, berikut hasil data yang telah di tranformasikan:

Tabel 6. Data Setelah di Konversi

No.	Bulan	Hari	Jam	Jenis Kejadian
1	1	1	1	2
2	1	1	1	2
3	1	1	1	4
4	1	1	2	3
5	1	1	2	3
...
301	12	6	8	1
302	12	7	2	3
303	12	7	2	3
304	12	7	8	1
305	12	7	5	4

3.2. Menentukan nilai K

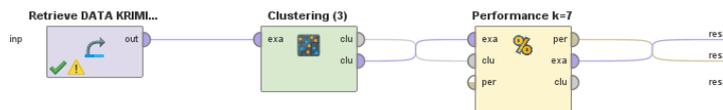
Dalam algoritma *k-means clustering* menentukan nilai k berdasarkan kebutuhan yang nantinya akan dijadikan acuan untuk menentukan hasil dalam *cluster*. Dalam penelitian ini, jumlah k yang digunakan adalah k=7.



Gambar 2. Menentukan Jumlah Cluster

3.3. Centroid Data

Menentukan *centroid* data pada algoritma *k-means clustering* merupakan prasyarat. Dalam proses ini pada *RapidMiner* ini merupakan proses design yang nantinya akan di jalankan dan tampil hasilnya.



Gambar 3. Pengolahan K-Means Clustering pada *RapidMiner*

3.4. Pengelompokkan Data

Setelah *design* di *run*, maka akan tampil data yang telah dikelompokkan berdasarkan *cluster* yang telah ditentukan di awal. Hasil pengolahan data tindak kejahatan menggunakan algoritma *K-Means Clustering* pada software *RapidMiner* dapat dilihat pada Gambar 4.

Row No.	id	cluster	BULAN	HARI	JAM	JENIS KEJA...
1	1	cluster_2	1	1	1	2
2	2	cluster_2	1	1	1	2
3	3	cluster_2	1	1	1	4
4	4	cluster_2	1	1	2	3
5	5	cluster_2	1	1	2	3
6	6	cluster_2	1	1	2	3
7	7	cluster_2	1	2	2	3
8	8	cluster_2	1	2	3	1
9	9	cluster_2	1	2	2	3
10	10	cluster_2	1	2	3	1
11	11	cluster_2	1	3	1	1
12	12	cluster_2	1	3	1	1
13	13	cluster_2	1	3	1	1
14	14	cluster_6	1	4	4	1

Gambar 4. Data View

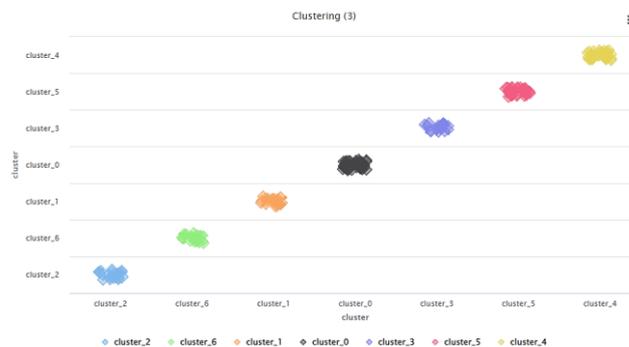
Dari Gambar 4, *RapidMiner* juga dapat menampilkan hasil jumlah tiap *cluster* pada data Tindak Kejahatan. Pada Gambar 5, jumlah tiap *cluster* yaitu *cluster 0* sebanyak 60, *cluster 1* sebanyak 31, *cluster 2* sebanyak 36, *cluster 3* sebanyak 35, *cluster 4* sebanyak 51, *cluster 5* sebanyak 55, dan *cluster 6* sebanyak 37.

Cluster Model

Cluster 0: 60 items
 Cluster 1: 31 items
 Cluster 2: 36 items
 Cluster 3: 35 items
 Cluster 4: 51 items
 Cluster 5: 55 items
 Cluster 6: 37 items
 Total number of items: 305

Gambar 5. Hasil Clustering

Berdasarkan Gambar 5 hasil *clustering* yang tampil merupakan *database* yang telah diolah secara keseluruhan lengkap dengan *cluster* dan jumlah tiap *cluster*.



Gambar 6. Plot View hasil clustering

Plot view merupakan halaman di *result perspective* untuk menampilkan *centroid* data sebagai plot dalam *software RapidMiner*.

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4	cluster_5	cluster_6
BULAN	6.683	5.323	2.583	8.257	10.196	10.073	2.541
HARI	4.100	5.419	2.056	1.686	4.686	5.436	5.324
JAM	6.600	1.871	2.361	2.543	6.451	2.382	5.432
JENIS KEJADIAN	2.283	2.194	2.361	2.314	2.196	2.291	2.189

Gambar 7. Hasil Centroid Akhir Clustering

Hasil *clustering* pada Gambar 7 merupakan *centroid* akhir atau kelompok *cluster* pada pengolahan menggunakan *software RapidMiner*. Dari hasil proses clustering maka dapat digambarkan pola dari tiap cluster dengan urutan bulan (bln), hari (hr), jam (jm), dan jenis kejadian (jk). Penelitian ini menggunakan *tools RapidMiner* untuk mengolah 305 dataset dengan 7 cluster dengan pola sebagai berikut :

1. *Cluster 0*, kejadian paling sering pada bln = Juni dan Juli, hr = jumat, jm = 21.01-00.00 WIB, jk = curas dan curanmor.
2. *Cluster 1*, kejadian paling sering pada bln = Mei, hr = jumat, jm = 00.00 - 3.00 WIB, jk = curas.
3. *Cluster 2*, kejadian paling sering pada bln = februari, hr = selasa, jm = 03.01 - 06.00 WIB, jk = curas.
4. *Cluster 3*, kejadian paling sering pada bln = Agustus, hr = senin, jm = 03.01 - 06.00 WIB, jk = curas dan curanmor.
5. *Cluster 4*, kejadian paling sering pada bln = oktober, hr = kamis, jm = 15.01 - 18.00 WIB, jk = curat.

6. *Cluster* 5, kejadian paling sering pada bln = oktober, hr = jumat, jm = 03.01 - 06.00 WIB, jk = curat dan curas.
7. *Cluster* 6, kejadian paling sering pada bln = februari, hr = jumat, jm = 12.01 - 15.00 WIB, jk = curas atau rampas.

4. Kesimpulan

Berdasarkan penelitian data mining *algoritma k-means clustering* pada 305 data, maka didapatkan beberapa kesimpulan: 1) Algoritma *K-Means Clustering* dapat diterapkan pada data tindak kejahatan untuk menemukan pola kejahatan. 2) Berdasarkan hasil *clustering* terbagi menjadi 7 *cluster* yaitu *cluster* 0 sebanyak 60, *cluster* 1 sebanyak 31, *cluster* 2 sebanyak 36, *cluster* 3 sebanyak 35, *cluster* 4 sebanyak 51, *cluster* 5 sebanyak 55, dan *cluster* 6 sebanyak 37.3) Hasil penerapan algoritma *k-means clustering* pada tools *RapidMiner* dapat membantu proses pengolahan dengan mudah.

Daftar Pustaka

- [1] S. Soekanto, Pengantar Penelitian Hukum. Jakarta: Universitas Indonesia Press, 1999.
- [2] R. I. Ndaumanu, Kusriani, and M. R. Arief, "Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode K-Nearest Neighbor," JATISI, vol. 1, no. 1, pp. 1–15, 2014.
- [3] 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, doi: 10.30865/json.v1i2.1955.
- [4] S. Stedila, R. Astuti, and F. M. Basysyar, "Penerapan Data Mining Clustering Menggunakan Metode K-Means Pada Data Tindak Kriminalitas Di Polres Kabupaten Kuningan," JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 2, pp. 1629–1636, 2024, doi: 10.36040/jati.v8i2.8790.
- [5] R. Hidayat, "Clustering Menggunakan Algoritma K-Means untuk Mengelompokan Wilayah Rawan Kejahatan di Wilayah Kabupaten Solok," J. Ilm. Mhs. Manajemen, Bisnis dan Akunt., vol. 4, no. 5, pp. 646–654, 2022, doi: 10.32639/jimmba.v4i5.169.
- [6] F. Safnita, S. Defit, G. W. Nurcahyo, F. I. Komputer, U. Putra, and I. Yptk, "Penerapan Algoritma K-Means Dalam Pengklasteran Hasil Evaluasi Akademik Mahasiswa Abstrak," vol. 5, no. 2, pp. 513–520, 2024.