



Clustering Menggunakan Algoritma K-Means untuk Mengelompokan Wilayah Rawan Kejahatan di Wilayah Kabupaten Solok

Rahmat Hidayat

Universitas Putra Bangsa

Ayat150190@gmail.com

ARTICLE INFO

Article History:

Received: August 13th 2022

Accepted: August 15th 2022

Published: August 24th 2022

Keywords:

Data Mining, Clustering,

K-Means

ABSTRACT

Kabupaten solok merupakan salah satu wilayah yang terdapat pada provinsi sumatera barat. Kabupaten solok memiliki 14 kecamatan dan 74 nagari/desa. Berdasarkan data yang terdapat pada lumbung data Badan Pusat Statistik Kabupaten Solok indeks kemiskinan yang terdata pada tahun 2021 adalah 30,36 ribu jiwa atau 8,01% dari jumlah penduduk kabupaten solok. Tentu tingkat kemiskinan ini mempengaruhi banyak tindak pidana yang terjadi pada wilayah ini. Tercatat selama tahun 2021 berdasarkan data perkara yang dilimpahkan oleh Kejaksaan Solok terdapat 202 dakwaan yang telah dilimpah kepada pengadilan Negeri Koto Baru. Dari permasalahan tersebut muncullah gagasan untuk menggali informasi dari data tindak pidana terjadi pada tahun 2021 untuk mengelompokan wilayah yang rawan tindak kejahatan dengan menggunakan metode algoritma K-Means. Diharapkan hasil dari penelitian ini dengan mengelompokan wilayah yang rawan tindak kejahatan, dapat memberikan pengetahuan baru atau informasi baru bagi warga yang tinggal disatu wilayah di kabupaten solok agar lebih meningkat kewaspadaan agar terhindar dari tindak kejahatan sehingga dapat menurunnya tingkat kriminalitas diwilayah kabupaten solok.

Pendahuluan

Kriminalitas merupakan segala macam bentuk Tindakan dan perbuatan yang merugikan secara ekonomis dan psikologis yang melanggar hukum yang berlaku dalam Negara Indonesia serta norma-norma social dan agama (Dewi *et al.*, 2019). Ada banyak tindak kejahatan yang terjadi yang dianggap kriminalitas seperti pencurian, penganiayaan, Perampokan, Pembunuhan, Pencabulan, Narkotika dll. Banyaknya tindak kejahatan yang terjadi tidak hanya menjadi tugas pemerintah akan tetapi adalah tugas kita bersama bagaimana tindak kejahatan yang terjadi menurun disetiap tahunnya. Pemberian informasi yang tepat serta memberikan pemahaman kepada penduduk tentu juga akan berperan untuk menurunkan tindak kejahatan yang terjadi. Sehingga orang-orang bisa waspada akan tindak kejahatan agar tidak menjadi korban dari tindak kejahatan.

Analisa dengan metode k-means merupakan metode clustering yang yang paling rendah dan sederhana, hal ini dikarenakan k-means mempunyai kemampuan menglompokan data dalam jumlah yang cukup besar dengan waktu komputasi yang relative cepat dan efisien (Dewi *et al.*, 2019). Pada penelitian yang dilakukan oleh Tampubolon *et al.* (2020) dengan menggunakan metode cluster analysis, yaitu metode Kernel Density, Metode K-Means dan metode K-Medoids. Metode Kernel Density digunakan untuk menganalisis titik kejadian curanmor untuk menentukan tingkat kerawanan. Dan metode k-means dan metode K-Medoids digunakan untuk menentukan tingkat kewawanan dan jumlah kejadian yang terjadi.

Pada penelitian ini data tindak pidana yang digunakan adalah data yang didapatkan dari perkara pidana yang dilimpahkan oleh kejaksaan Negeri Solok ke Pengadilan Negeri Koto Baru. Data yang digunakan adalah data tindak pidana yang terjadi dari januari sampai dengan dengan desember 2021 di wilayah kabupaten solok. Data tindak pidana dikelompokan menjadi tiap kecamatan dan jorong/desa untuk dijadikan bahan untuk data mining. Dalam melakukan clustering, data yang diperoleh akan dihitung berdasarkan jumlah kriminalitas disatu kecamatan dan jorong/desa serta jenis tindak pidana yang terjadi diwilayah tersebut. Penulis menggunakan data tindak pidana yang terjadi pada tahun 2021 untuk diolah dengan data mining dengan tujuan mengelompokan data kejahatan untuk menentukan cluster tingkat kejahatan tingkat tinggi dan cluster kejahatan tingkat rendah.

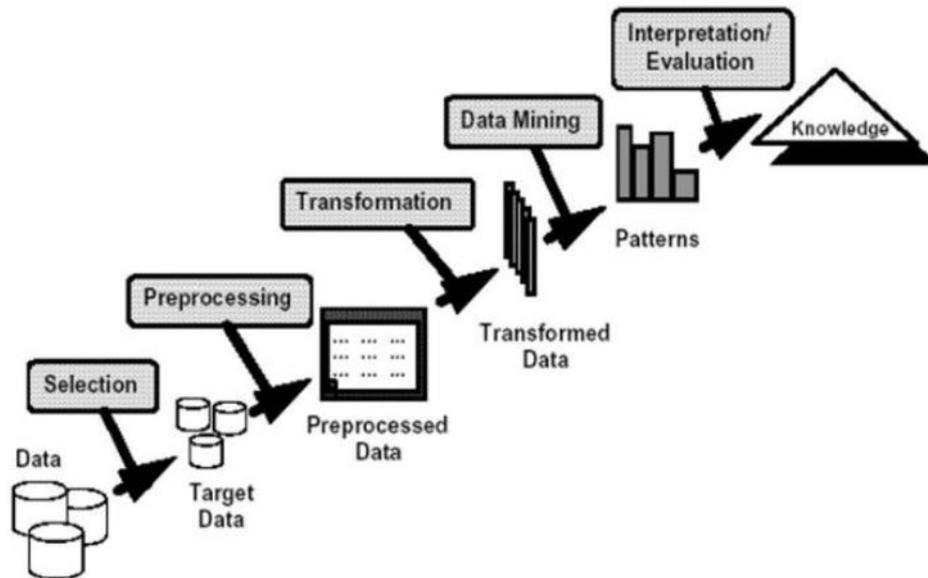
Metode Penelitian

Data Mining

Data mining adalah Analisa terhadap data untuk menemukan hubungan yang jelas serta menyimpulkan yang belum diketahui sebelumnya dengna cara tekini dipahami dan berguna bagi pemilik data tersebut (Darmi & Setiawan, 2016). Sedangkan menurut Wijanarko *et al.* (2021) data mining adalah suatu peroses pencarian pola atau informasi menarik dalam data terpilih dengan menggunakan Teknik atau metode tertentu.

Secara garis besar, data mining dapat dikelompokan menjadi 2 kategori utama (Darmi & Setiawan, 2016), yaitu: Deskriptive mining, yaitu proses menemukan karakteristik penting dari data dalam satu basis data. Yang termasuk kedalam deskriktive mining adalah *clustering*, dan *asosiation sequential mining*.

Predictive, yaitu proses untuk menemukan pola dari data dengan menggunakan beberapa variable lain dimasa depan. Salah satu Teknik yang terdapat dalam predictive mining adalah kalsifikasi. Secara sederhananya data mining biasa dikatakan sebagai proses menambang pengetahuan baru dari kumpulan data yang ada. Adapun istilah lain dari data mining adalah *Knowledge Discovery in Database* (KDD). Walaupun data mining sendiri merupakan bagian dari tahapan proses KDD seperti yang dapat dilihat pada gambar 1.



Gambar 1. Knowledge Discovery in Database (KDD)

Clustering

Clustering atau klusterisasi adalah metode pengelompokan data, merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan cluster (Mayona *et al.* 2022). Adapun tujuan dari clustering data dapat dibedakan mejadi dua (Suryadi & Supriatna, 2019), yaitu:

1. Pengelompokan untuk pemahaman terbentuk harus menangkap struktur alami data yang bertujuan hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan inti seperti peringkasan atau summarization (rata-rata standar deviasi), pemberian label kelas pada setiap kelompok bertujuan untuk digunakan sebagai data latih klasifikasi.
2. Pengelompokan untuk penggunaan bertujuan untuk mencari prototipe kelompok yang paling representatif terhadap data, memberikan abstraksi dari setiap objek data dalam kelompok dimana sebuah data terletak di dalamnya.

Algoritma K-Means

Algoritma K-Means merupakan algoritma non hirarki yang berasal dari metode clustering yang paling sederhana dan mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang cepat dan efisien metode K-Means ini mempartisi data kedalam kelompok (Mayona *et al.*, 2022). K-Means merupakan salah satu algoritma Clustering yang masuk dalam kelompok Unsupervised Learning yang digunakan untuk membagi data menjadi beberapa kelompok dengan system partisi. Algoritma ini menerima masukan berupa data tanpa label kelas. Pada setiap cluster terdapat titik pusat (centroid) yang mempresentasikan cluster tersebut. Secara sederhana algoritma K-Means dapat dijelaskan sebagai algoritma data mining yang digunakan untuk menyelesaikan

masalah pengelompokan (Clustering). Langkah-langkah algoritma K-Means dapat dijelaskan sebagai berikut (Tampubolon *et al.*, 2020):

1. Tentukan jumlah cluster(k) pada data set.
2. Tentukan titik pusat (Centroid) secara random

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

Keterangan:

- V_{ij} = centroid rata-rata cluster ke-I untuk variable ke-j
- N_i = Jumlah anggota cluster ke-i
- I,k = indeks dari cluster
- J = indeks dari variable
- X_{kj} = nilai data ke-k variable ke-j untuk cluster tersebut

3. Untuk menghitung jarak terdekat dengan centroid, dapat menggunakan rumus dibawah ini:

$$d = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Keterangan:

- d = Euclidean Distance
 - i = Banyak objek
 - x,y = Titik koordinat objek
 - s,t = Titik koordinat centroid
4. Golongkan objek menurut jarak ke centroid terdekat.

Lakukan Kembali Langkah ke-3 dan Langkah ke -4 dan iterasi mencapai centroid bernilai optimal.

Hasil dan Pembahasan

Dalam penelitian ini data tindak pidana yang dijadikan sebagai sumber data untuk data mining nantinya adalah data tindak pidana yang terjadi pada tahun 2021 dari bulan januari sampai dengan bulan desember. Data tindak pidana merupakan data tindak pidana yang terjadi di wilayah kabupaten seelok yang dikelompokkan berdasarkan kecamatan. Sedangkan untuk tindak pidana agar lebih memudahkan dalam clustering dilakukan penyederhanaan seperti pada table 1 berikut:

Tabel 1. Jenis Tindak Pidana

| No | Tindak Pidana |
|----|-------------------------|
| 1 | Lingkungan Hidup (LH) |
| 2 | Narkotika (NKT) |
| 3 | Pembunuhan (PBN) |
| 4 | Pencurian (PCR) |
| 5 | Penganiayaan (PGN) |
| 6 | Penipuan (PNP) |
| 7 | Perjudian (PJD) |
| 8 | Perlindungan Anak (PLA) |

Tabel 2. Data Tindak Pidana Kabupaten Solok

| No | Wilayah | Tindak Pidana | | | | | | | | Total |
|----|--------------------|---------------|-----|-----|-----|-----|-----|-----|-----|-------|
| | | LH | NKT | PBN | PCR | PGN | PNP | PJD | PLA | |
| 1 | Bukit Sundi | | 3 | | 2 | 1 | 2 | | 1 | 9 |
| 2 | Gunung Talang | | 4 | | 2 | 2 | 2 | 2 | | 12 |
| 3 | Hiliran Gumanti | | 1 | | | 1 | | | | 2 |
| 4 | Jujuhan | | 1 | | | | | | | 1 |
| 5 | KPGD | 5 | 1 | | 3 | 3 | 2 | | 1 | 15 |
| 6 | Kubung | 1 | 16 | | 10 | 2 | 1 | 1 | 4 | 35 |
| 7 | Lembah Gumanti | 1 | | | | 3 | | | | 4 |
| 8 | Lembang Jaya | | 1 | | 1 | 1 | 1 | | | 4 |
| 19 | Lubuk Sikarah | 1 | | | | | | | | 1 |
| 10 | Pauh Duo | | | | 6 | 1 | | | | 7 |
| 11 | Payung Sekaki | | | | 1 | 4 | | | | 5 |
| 12 | Sangir | 4 | 4 | 2 | 22 | | 4 | | 1 | 37 |
| 13 | Sangir Batang Hari | | | | 2 | 1 | 1 | | | 4 |
| 14 | Sungai Pagu | | 2 | 1 | 2 | | | | | 5 |

Wilayah yang memiliki jumlah total tindak pidana yang hamper sama akan berada pada kelompok wilayah yang sama. Dan sebaliknya wilayah yang memiliki jumlah tindak pidana yang berbeda akan akan berada pada kelompok wilayah yang berbeda pula.

Berdasarkan data pada table 2 di atas maka kita dapat masuk ketahap Analisa, diantaranya:

Jumlah Cluster = 2

Jumlah Data = 14

Jumlah Atribut = 2 (Cluster kejahatan tingkat tinggi dan cluster kejahatan tingkat sedang).

Centroid Data

Dalam menerapkan algoritma K-means akan menghasilkan centroid atau titik tengah.cluster dibagi menjadi 2 bagian, maka titik tengah atau centroid terdapat 2 titik diantaranya, (C1) cluster tinggi dan (C2) cluster rendah. Pada titik ini dilakukan dengan nilai tinggi (maximum), dan nilai rendah (minimum). Nilai titik tersebut dapat dilihat pada table berikut:

Tabel 3. Centroid Data Awal

| Attribut | C1 | C2 |
|----------|----|----|
| Nilai | 37 | 1 |

Clustering Data

Dari centroid tersebut, maka didapat 2 cluster dengan mengambil jarak terdekat dari setiap data yang diolah. Dengan didapatkan pengelompokkan pada iterasi 1 untuk 2 cluster tersebut. Cluster tinggi (C1) diantaranya Kubung dan Sangir. Dan cluster rendah (C2) diantaranya Bukit Sundi, Gunung Talang, Hiliran Gumanti, Jujuhan, KPGD, Lembah Gumanti, Lembang Jaya, Lubuk Sikarah, Pauh Duo, Payung Sekaki, Sangir Batang Hari dan Sungai Pagu. Pada proses tersebut dapat dilihat pada table 4 berikut :

Tabel 4. Penghitungan Cluster Keanggotaan

| Wilayah | Total | C1 | C2 | Keanggotaan |
|--------------------|-------|----|----|-------------|
| Bukit Sundi | 9 | 28 | 8 | C2 |
| Gunung Talang | 12 | 25 | 11 | C2 |
| Hiliran Gumanti | 2 | 35 | 1 | C2 |
| Jujuhan | 1 | 36 | 0 | C2 |
| KPGD | 15 | 22 | 14 | C2 |
| Kubung | 35 | 2 | 34 | C1 |
| Lembah Gumanti | 4 | 33 | 3 | C2 |
| Lembang Jaya | 4 | 33 | 3 | C2 |
| Lubuk Sikarah | 1 | 36 | 0 | C2 |
| Pauh Duo | 7 | 30 | 6 | C2 |
| Payung Sekaki | 5 | 32 | 4 | C2 |
| Sangir | 37 | 0 | 36 | C1 |
| Sangir Batang Hari | 4 | 33 | 3 | C2 |
| Sungai Pagu | 5 | 32 | 4 | C2 |

Setelah hasil iterasi 1 maka didapatkan hasil sebagai berikut:

Tabel 5. Pengelompokan dari iterasi 1

| Wilayah | Keanggotaan |
|--------------------|-------------|
| Bukit Sundi | C2 |
| Gunung Talang | C2 |
| Hiliran Gumanti | C2 |
| Jujuhan | C2 |
| KPGD | C2 |
| Kubung | C1 |
| Lembah Gumanti | C2 |
| Lembang Jaya | C2 |
| Lubuk Sikarah | C2 |
| Pauh Duo | C2 |
| Payung Sekaki | C2 |
| Sangir | C1 |
| Sangir Batang Hari | C2 |
| Sungai Pagu | C2 |

Berdasarkan iterasi pertama yang terdapat pada table 5 maka juga dapat dihasilkan WCV (Within Cluster Variation) dan juga BVC (Beetwin Cluster Variation).

Tabel 6. Within Cluster Variation

| C1 | C2 | Keanggotaan | Min Jarak | Kuadrat Min Jarak |
|------------|----|-------------|-----------|-------------------|
| 28 | 8 | C2 | 8 | 64 |
| 25 | 11 | C2 | 11 | 121 |
| 35 | 1 | C2 | 1 | 1 |
| 36 | 0 | C2 | 0 | 0 |
| 22 | 14 | C2 | 14 | 196 |
| 2 | 34 | C1 | 2 | 4 |
| 33 | 3 | C2 | 3 | 9 |
| 33 | 3 | C2 | 3 | 9 |
| 36 | 0 | C2 | 0 | 0 |
| 30 | 6 | C2 | 6 | 36 |
| 32 | 4 | C2 | 4 | 16 |
| 0 | 36 | C1 | 0 | 0 |
| 33 | 3 | C2 | 3 | 9 |
| 32 | 4 | C2 | 4 | 16 |
| WCV | | | | 481 |

Tabel 7. Tabel BVC (Beetwin Cluster Variation)

| Jarak Antar Cluster | | |
|---------------------|----|----|
| C1 | C2 | 36 |
| BCV | | 36 |

Selanjutnya untuk menghitung nilai rasio dapat dilakukan setelah nilai WCV dan BCV didapatkan dengan rumus sebagai berikut:

$$Rasio = \frac{BCV}{WCV}$$

Maka rasio adalah nilai BCV=36 dibagi dengan nilai WCV =481 maka didapatkan nilai rasio 0,074844. Setelah rasio iterasi pertama didapatkan barulah dilanjutkan dengan iterasi kedua dengan centroid kedua yang didapatkan dari colom keanggotaan dari table 4.5. setelah didapatkan rasio dari iterasi kedua barulah dilakukan perbandingan. Dan iterasi terhenti pada iterasi kedua karea kondisi yang sudah terpenuhi yaitu rasio iterasi terbaru lebih besar dari rasio iterasi sebelumnya. Maka dari kedua iterasi yang telah dilakukan didapatkan hasil sebagi berikut:

Tabel 8. Pengelompokan Data Iterasi 2

| Wilayah | Keanggotaan |
|-----------------|-------------|
| Bukit Sundi | C2 |
| Gunung Talang | C2 |
| Hiliran Gumanti | C2 |
| Jujuhan | C2 |
| KPGD | C2 |
| Kubung | C1 |
| Lembah Gumanti | C2 |
| Lembang Jaya | C2 |

| | |
|--------------------|----|
| Lubuk Sikarah | C2 |
| Pauh Duo | C2 |
| Payung Sekaki | C2 |
| Sangir | C1 |
| Sangir Batang Hari | C2 |
| Sungai Pagu | C2 |

Dari table 8 di atas dapat dilihat hasil yang didapatkan pada iterasi 2. Cluster tinggi C1 atau daerah yang rawan tindak pidana terdapat pada kecamatan kubung dan kecamatan sangir. Sedangkan untuk cluster rendah (C2) yaitu kecamatan bukit sundi, gunung talang, hiliran gumanti, jujuhan, KPGD, lembah gumanti, lembang jaya, lubuk sikarah, pauh duo, payung sekaki, sangir batang hari dan sungai pagu.

Penutup

Simpulan

Dari penelitian diatas dapat disimpulkan bahwa, clustering dengan algoritma k-means untuk mengelompokan wilayah rawan tindak kejahatan diwilayah kabupaten solok berdasarkan data dari bulan januari dan desember 2021, maka didapatkan hasil Cluster tinggi (C1) = 2 kecamatan yaitu kecamatan kubung dan kecamatan sangir. Sedangkan untuk klister rendah (C2) = 12 kecamatan, yaitu kecamatan bukit sundi, gunung talang, hiliran gumanti, jujuhan, KPGD, lembah gumanti, lembang jaya, lubuk sikarah, pauh duo, payung sekaki, sangir batang hari dan sungai pagu.

Saran

Adapun saran untuk penelitian selanjutnya yaitu:

1. Diharapkan untuk bisa menggunakan data yang lebih banyak atau data beberapa tahun sebelumnya untuk bahan perbandingan serta meningkatkan akurasi dari metode klustering dengan algoritma k-mans.
2. Menggunakan metode algoritma lain untuk melakukan clustering untuk mengelompokan wilayah yang rawan tindak kejahatan.
3. Dilakukan pengujian dengan aplikasi RapidMiner sebagai hasil perbandingan antara pencarian dengan manual dengan menggunakan aplikasi.

Referensi

- Darmi, Y. D., & Setiawan, A. (2016). Penerapan metode clustering k-means dalam pengelompokan penjualan produk. *Jurnal Media Infotama*, 12(2).
- Dewi, S. M., Windarto, A. P., Damanik, I. S., & Satria, H. (2019). Analisa Metode K-Means pada Pengelompokan Kriminalitas Menurut Wilayah. In *Seminar Nasional Sains dan Teknologi Informasi (SENSASI)* (Vol. 2, No. 1).
- Mayona, Y., Buaton, R., & Simajuntak, M. (2022) Data Mining Clustering Tingkat Kejahatan dengan Metode Algoritma K-Means (Studi Kasus kejahatan Negeri Binjai). *Jurnal Informatika Kaputama (JIK)*, 6(3), 2548-9739.
- Rohmah, A., Sembiring, F., & Erfina, A. (2021). Implementasi Algoritma K-Means Clustering Analysis Untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: SMK Yaspim Gegerbitung). In *Seminar Nasional Sistem Informasi dan Manajemen Informatika Universitas Nusa Putra* (Vol. 1, No. 01, pp. 290-298).

- Suryadi, U. T., & Supriatna, Y. (2019). Sistem clustering tindak kejahatan pencurian di Wilayah Jawa Barat menggunakan algoritma k-means. *Jurnal teknologi informasi dan komunikasi*, 14(1), 15-27.
- Tampubolon, H. D., Gultom, D., Hutabarat, L. Y., Zer, F. I. R., & Hartama, D. (2020). Penerapan Algoritma K-Means Untuk Mengetahui Tingkat Tindak Kejahatan Daerah Pematangsiantar. (*JurTI*) *Jurnal Teknologi Informasi*, 4(1), 146-151.
- Wijanarko, T. T., Oktavianto, H., & Umilasari, R. (2021). Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Kriminalitas Nasional Menggunakan Algoritma K-Means dengan Metode Davies Boulden Index. *Doctoral Dissertation*. Universitas Muhammadiyah Jember