

## Pengelompokan Daerah Rawan Kriminalitas di Sulawesi Selatan Menggunakan Metode K-means Clustering

Irwan<sup>1,a)</sup>, Wahidah Sanusi<sup>1, b)</sup>, dan Febriyanto Saman<sup>1, c)</sup>

<sup>1</sup>Jurusan Matematika FMIPA Universitas Negeri Makassar

<sup>a)</sup> irwan@unm.ac.id

<sup>b)</sup> wahidahsanusi@unm.ac.id

<sup>c)</sup> febriyantosaman@gmail.com

**Abstrak.** Penelitian ini merupakan penelitian terapan yang menekankan cara melaksanakan analisis cluster secara matematis, mengetahui bagaimana aplikasi k-means clustering, dan ciri dari setiap kelompok daerah rawan kriminalitas. Adapun data simulasi yang digunakan pada penelitian ini adalah data yang diperoleh dari Badan Pusat Statistika (BPS) Propinsi Sulawesi Selatan. Data tersebut selanjutnya dianalisis dengan metode K-means clustering. Hasil penelitian menunjukkan bahwa terdapat empat ciri dari tiap kelompok daerah rawan kriminalitas di Sulawesi Selatan. Kelompok 1 masuk kategori daerah yang cukup aman kriminalitas, Kelompok 2 masuk kategori daerah yang rawan kriminalitas, kelompok 3 masuk ketegori daerah yang aman kriminalitas, dan kelompok 4 masuk kategori daerah yang cukup rawan kriminalitas.

**Kata Kunci:** Analisis Cluster, K-means Clustering, Kriminalitas.

**Abstract.** This research is an applied research that emphasizes how to carry out cluster analysis mathematically, knowing how to apply k-means clustering, and the characteristics of each group of crime-prone areas. The simulation data used in this study is data obtained from the Central Statistics Agency (BPS) of South Sulawesi Province. The data was then analyzed by the K-means clustering method. The results of the study show that there are four characteristics of each group of crime-prone areas in South Sulawesi. Group 1 is categorized as a crime-safe area, Group 2 is categorized as a crime-prone area, group 3 is categorized as a crime-safe area, and group 4 is categorized as an area that is quite prone to crime.

**Keywords:** Cluster Analysis, K-means Clustering, Crime.

## PENDAHULUAN

Clustering dalam data mining adalah pengelompokkan sejumlah data atau objek ke dalam cluster (group) sehingga setiap dalam cluster tersebut akan berisi data yang semirip mungkin dan berbeda dengan objek dalam cluster yang lainnya. Sampai saat ini, para ilmuwan masih terus melakukan berbagai usaha untuk melakukan perbaikan model cluster dan menghitung jumlah cluster yang optimal sehingga dapat dihasilkan cluster yang paling baik. ada dua metode yang kita kenal, yaitu *hierarchical clustering* dan *partitioning*. Metode *hierarchical clustering* sendiri terdiri dari *complete linkage clustering*, *single linkage clustering*, *average linkage clustering* dan *centroid linkage clustering*. Sedangkan metode *partitioning* sendiri terdiri dari *K-means* dan *fuzzy K-means* (Alfina, Santosa dan Barakbah, 2012; Suryadi, Supriatna, 2019; Sianturi, Yuni, 2020; Prima, D., T. dan Widodo, E. 2017).

Metode *K-means* merupakan metode *clustering* yang paling sederhana dan umum. Hal ini dikarenakan *K-means* mempunyai kemampuan mengelompokkan data dalam jumlah yang

cukup besar dengan waktu komputasi yang relatif cepat dan efisien. Namun, *K-means* mempunyai kelemahan yang diakibatkan oleh penentuan pusat awal *cluster* (Andayani, S. 2007).

Hasil *cluster* yang terbentuk dari metode *K-means* ini sangatlah tergantung pada inisiasi nilai pusat awal *cluster* yang diberikan. Hal ini menyebabkan hasil *clusternya* berupa solusi yang sifatnya local optimal (Zahrotun, 2015; Awalia, Resky. 2018).

Algoritma K-means terdapat kelebihan yaitu dinilai cukup efisien, yang ditunjukkan dengan kompleksitasnya  $O(tkn)$ , dengan catatan  $n$  adalah banyaknya obyek data,  $k$  adalah jumlah cluster yang dibentuk, dan  $t$  banyaknya iterasi. Biasanya, nilai  $k$  dan  $t$  jauh lebih kecil dari pada nilai  $n$ . selain itu, dalam iterasinya, algoritma ini akan berhenti dalam kondisi optimum local (Adhe et al, 2020; Goreti, & Yuki Novia. 2016).

Penelitian ini menggunakan metode *k-means clustering* untuk kasus kriminalitas di Sulawesi Selatan tahun 2020. Analisis cluster ini akan menghasilkan pengelompokan daerah rawan kriminalitas yang terjadi di Sulawesi Selatan sehingga warga Sulawesi Selatan bisa melakukan antisipasi dengan mencegah tindakan kejahatan tersebut agar intensitasnya semakin berkurang. Penelitian ini diharapkan dapat menambah pengetahuan sekaligus informasai kepada masyarakat umum mengenai kondisi kasus kriminalitas di Sulawesi Selatan tahun 2020.

## METODE PENELITIAN

Pada penelitian ini data yang digunakan adalah data kriminalitas yang terjadi pada tahun 2020 di Provinsi Sulawesi Selatan yang diperoleh dari Badan Pusat Statistika Provinsi Sulawesi Selatan. Variabel yang digunakan adalah jenis-jenis kejahatan, yaitu:

- X1 = jumlah kasus kejahatan terhadap nyawa
- X2 = jumlah kasus kejahatan terhadap fisik atau badan
- X3 = jumlah kasus kejahatan terhadap kesusilaan
- X4 = jumlah kasus kejahatan terhadap kemerdekaan
- X5 = jumlah kasus kejahatan terhadap hak milik atau barang
- X6 = jumlah kasus kejahatan terkait narkoba
- X7 = jumlah kasus kejahatan terkait penipuan, penggelapan, dan korupsi
- X8 = jumlah kasus kejahatan terhadap ketertiban umum .

Adapun teknik analisis data dalam penelitian ini yaitu:

1. Penelitian ini menggunakan aplikasi dari analisis *cluster* yang merupakan bagian dari analisis multivariat dengan metode interdependensi.
2. Data mentah berupa data sekunder yang diambil dari Badan Pusat Statistik Provinsi Sulawesi Selatan. Pada kolom akan diisi dengan sejumlah  $n$  objek, sedangkan untuk baris pertama akan diisi dengan sejumlah  $p$  variabel.
3. Memastikan bahwa asumsi sebagai syarat pengaplikasian analisis *cluster* terpenuhi, yakni data representative serta tidak adanya multikolinearitas antar variabel.
  - a. Representatif (keterwakilan data). Data yang dipakai adalah data populasi dari data sekunder sensus yang berarti bahwa keterwakilan data jelas terpenuhi.

- b. Multikolinieritas (korelasi tinggi antar variabel). Gangguan multikolinieritas terjadi jika dalam sebuah model terdapat korelasi yang tinggi antara dua atau lebih variabel bebas. Gangguan ini tidak akan muncul, sebab variabel-variabel dalam yang digunakan merupakan variabel yang dijamin kelayakannya oleh Badan Pusat Statistik Provinsi Sulawesi Selatan dan telah dipakai oleh banyak penelitian lain.
- Selanjutnya menentukan ukuran ketidaksamaan atau jarak antar dua objek ke-I dengan objek ke-j, disimbolkan dengan  $d_{ij}$  menggunakan perhitungan jarak kuadrat *Euclidean*:
  - Membuat *cluster* atau kelompok sebanyak 4 *cluster*. Pengelompokan data dilakukan secara non hirarki yakni terlebih dahulu tidak ditentukan berapa banyak kelompok yang nantinya akan terbentuk. Metode yang digunakan untuk proses *clustering* secara non hirarki yaitu k-means.

$$d_{ih}^2 = \sum_{j=i}^p (x_{ij} - x_{hj})^2 \quad (1)$$

Dimana:

$d_{ih}$  = jarak kuadrat *Euclidean* antar obyek ke-I dengan obyek ke-h.

$P$  = jumlah variabel *cluster*.

$x_{ij}$  = nilai atau data dari obyek ke-i pada variabel ke-j.

$x_{hj}$  = nilai atau data dari obyek ke-h pada variabel ke-j (Everitt, 1993; Nuningsih, 2010).

- Selanjutnya data dianalisis dengan menggunakan program SPSS dengan membentuk 4, 5, dan 6 *cluster*.
- Metode *k-means* dimulai dengan menentukan seberapa banyak kelompok yang akan dibentuk, kemudian operasi diteruskan sampai semua objek menempati satu *cluster*.
- Profiling *cluster* adalah mendeskripsikan karakter tiap *cluster* yang telah terbentuk dengan membandingkan skor rata-rata dari profil *cluster*. Deskripsi nama dari tiap *cluster* yang terbentuk betul-betul akan mewakili ciri dari *clusternya*.
- Validasi *cluster*. Setelah *cluster* terbentuk, Langkah selanjutnya adalah menguji validasi *cluster* menggunakan uji parsial F.  
yakni menganalisis *cluster* pada sampel terpisah, membandingkan solusi dan mencocokkan hasilnya.

Hipotesis:

$H_0$  = variabel i bukan variabel pembeda dalam pengclusteran.

$H_1$  = variabel i merupakan pembeda dalam pengclusteran.

Taraf Signifikansi  $\alpha$

Statistik uji:

$$F = \frac{\text{means.squar.cluster}}{\text{means.square.error}} \quad (2)$$

Kriteria Uji:

$$\text{Tolak } H_0 \text{ Jika } F > F_{\alpha, k-1, n-k}$$

- Menentukan *Cluster* Terbaik.

Tahap evaluasi dapat dilakukan dengan melakukan analisis *cluster* dengan ukuran jarak atau metode *cluster* yang berbeda kemudian dibandingkan hasilnya. Pemilihan *cluster* yang menghasilkan kualitas pengelompokan terbaik dilakukan dengan memperhatikan nilai rasio rata-rata simpangan baku dalam *cluster* terhadap simpangan baku antar *cluster*. Rata rata simpangan baku dalam *cluster* ( $s_w$ ) dinyatakan dengan:

$$(s_w) = \sum_{k=1}^c s_k \quad (3)$$

Simpangan baku antar *cluster* ( $s_b$ ) dinyatakan sebagai:

$$(s_b) = \left[ \frac{1}{c-1} \sum_{k=1}^c (\bar{x}_k - \bar{x}_j)^2 \right]^{\frac{1}{2}} \quad (4)$$

Dimana  $c$  adalah jumlah *cluster*,  $s_k$  merupakan simpangan baku didalam *cluster* ke- $k$ .  $\bar{x}_k$  sebagai rata rata *cluster* ke- $k$  dan  $\bar{x}_j$  adalah rata-rata dari semua *cluster*. Semakin kecil nilai ( $s_w$ ) dan semakin besar nilai ( $s_b$ ) maka kelompok tersebut memiliki kinerja yang baik, artinya memiliki homogenitas yang tinggi. Kelompok yang dipilih adalah yang memberikan nilai rasio  $s_w/s_b$  terkecil (Goreti, 2016).

## HASIL DAN PEMBAHASAN

Berdasarkan metode penelitian dan data yang diperoleh, maka yang harus dilakukan adalah pengklasteran data kriminalitas yang ada di 24 Kabupaten/Kota di Provinsi Sulawesi selatan Berdasarkan jenis kasus kejahatan dengan 4 clustar, 5 cluster dan 6 cluster. Langkah awal mengelompokkan menggunakan metode *K-Means* adalah menentukan jumlah  $k$  cluster yang akan dibentuk. Pada tahap clustering menggunakan metode *K-Means* diperlukan penentuan *centroid* awal. Dalam menentukan nilai *centroid* awal dipilih secara random dari objek data yang dipakai.

### Hasil Pengelompokan dengan 4 kelompok.

Hasil simulasi dengan 4 kelompok dapat dilihat pada tabel 1

**TABEL 1.** Hasil Rata-Rata 4 Kelompok

	Cluster			
	1	2	3	4
kejahatan terhadap nyawa	3	8	1	4
kejahatan terhadap fisik dan atau badan	4	23	2	14
kejahatan terhadap kesusilaan	1	4	1	2
kejahatan terhadap kemerdekaan	0	1	0	0
kejahatan terhadap hak milik atau barang	24	114	20	63
kejahatan terhadap narkoba	2	14	1	5
kejahatan terhadap penipuan, penggelapan dan korupsi	6	26	1	15
kejahatan terhadap ketertiban umum	7	26	47	17
<b>Jumlah</b>	<b>47</b>	<b>216</b>	<b>73</b>	<b>120</b>

Tabel 2 menunjukan banyak anggota pada masing-masing kelompok.

**TABEL 2.** Banyak Anggota ke-4 Cluster

Cluster	Banyak Anggota	Anggota
1	4	Gowa, Bone, Wajo, dan Kota Makassar.
2	9	Bulukumba, Bantaeng, Jeneponto, Takalar, Maros, Sindenreng Rappang, Pinrang, Luwu, Luwu Timur.
3	10	Selayar, Sinjai, Pangkep, Barru, Soppeng, Enrekang, Tana Toraja, Luwu Utara, Pare-Pare, dan Palopo.

Dari tabel 2 dilihat bahwa sampel terbanyak ada di cluster 3 dengan 10 anggota diantaranya yaitu Selayar, Sinjai, Pangkep, Barru, Soppeng, Enrekang, Tana Toraja, Luwu Utara, Pare-Pare, dan Palopo, cluster 1 dengan 4 anggota yaitu Gowa, Bone, Wajo, dan Kota Makassar cluster 2 dengan 9 anggota. Bulukumba, Bantaeng, Jeneponto, Takalar, Maros, Sindenreng Rappang, Pinrang, Luwu, Luwu Timur, dan cluster 4 dengan 4 anggota yaitu Toraja Utara.

**Penentuan Cluster Terbaik**

Untuk melihat kinerja ketiga *cluster* tersebut digunakan kriteria nilai simpangan baku, yaitu simpangan baku dalam *cluster* ( $s_w$ ) dan simpangan baku antar *cluster* ( $s_b$ ). Nilai rata-rasio  $s_w/s_b$  yang terkecil menunjukkan ketepatan pengelompokan yang paling baik. Hasil perhitungan adalah sebagai berikut:

Rata-rata simpangan baku dalam *cluster* ( $s_w$ ) dinyatakan dengan:

$$(s_w) = \sum_{k=1}^c s_k \tag{5}$$

Simpangan baku antar *cluster* ( $s_b$ ) dinyatakan sebagai:

$$(s_b) = \left[ \frac{1}{c-1} \sum_{k=1}^c (\bar{x}_k - \bar{x}_j)^2 \right]^{\frac{1}{2}} \tag{6}$$

**TABEL 3.** Ratio Sw/Sb untuk Metode Terbaik

Jumlah Cluster	Nilai Rataan Rasio
4	0,193
5	0,204
6	0,233

Tabel 3 menunjukkan nilai rata-rasio simpangan baku dalam *cluster* ( $s_w$ ) dan antar *cluster* ( $s_b$ ) dengan pengelompokan empat *cluster* memiliki nilai yang lebih kecil yaitu sebesar 0,193, sedangkan pengelompokan lima *cluster* memiliki nilai yaitu 0,204 dan pengelompokan enam *cluster* memiliki nilai yaitu sebesar 0,233. Dengan demikian, untuk kasus pengklasifikasian daerah rawan kriminalitas di Sulawesi Selatan tahun 2020, pengelompokan empat *cluster* lebih baik dalam melakukan klasifikasi dibanding lima *cluster* dan enam *cluster*.

**Ciri-Ciri Hasil Pengelompokan 4 Cluster**

Dari keempat kelompok yang terbentuk kita dapat mengklasifikasikan sebagai berikut:

a. Kelompok I

Anggota dari kelompok ini adalah Gowa, Bone, Wajo, dan Kota Makassar.

Kasus kejahatan yang terjadi pada kelompok ini dapat dikatakan tinggi (Tabel 4). Hal ini dikarenakan tujuh dari de/lapan jenis kejahatan yang memiliki Jumlah rata-rata kejahatan tertinggi seperti kejahatan terhadap nyawa, kejahatan terhadap fisik atau badan, kejahatan terhadap kesusilaan, kejahatan terhadap kemerdekaan, kejahatan terhadap hak milik atau barang, kejahatan terhadap narkoba, dan kejahatan terhadap penipuan, penggelapan dan korupsi, berada di kelompok ini dan di atas rata-rata kelompok lain, hanya kenjataan terhadap ketertiban umum yang berada di bawah rata-rata kelompok III. sehingga masuk

ke dalam kategori daerah yang rawan kriminalitas.

a. Kelompok II

Anggota dari kelompok ini adalah Bulukumba, Bantaeng, Jeneponto, Takalar, Maros, Sindenreng Rappang, Pinrang, Luwu, Luwu Timur.

Kasus kejahatan yang terjadi pada kelompok ini dapat dikatakan cukup tinggi (Tabel 4). Hal ini dikarenakan hampir semua jenis kasus kejahatan berada di atas rata-rata kelompok I dan III. Seperti kejahatan terhadap nyama, kejahatan terhadap fisik/badan, krjahanterhadap kesusilaan, kejahatan terhadap hak milik/ barang, kejahatan terhadap narkoba, dan kejahatan terhadap penipuan, penggelapan dan korupsi. sehingga masuk ke dalam kategori daerah yang cukup rawan.

b. Kelompok III

Anggota dari kelompok ini adalah Selayar, Sinjai, Pangkep, Barru, Soppeng, Enrekang, Tana Toraja, Luwu Utara, Pare-Pare, dan Palopo.

Kasus kejahatan yang terjadi pada kelompok ini cukup rendah (Tabel 4). Hal ini dikarenakan Jumlah kasus kejahatan terhadap kesusilaan, kejahatan terhadap kemerdekaan, dan kejahatan terhadap ketertiban umum berada di bawah rata-rata kelompok lainnya, bahkan terendah dibandingkan kelompok lain. Namun kejahatan terhadap nyawa, kejahatan terhadap hak milik atau barang, kejahatan terhadap penipuan, penggelapan dan korupsi lebih tinggi dibanding kelompok III, sehingga masuk ke dalam kategori daerah yang cukup aman.

c. Kelompok IV

Anggota dari kelompok ini adalah Toraja Utara

Kasus kejahatan yang terjadi pada kelompok ini rendah (Tabel 4). Hal ini dikarenakan seluruh kasus kejahatan yang terjadi berada di bawah rata-rata kelompok lainnya, bahkan kasus terendah untuk kejahatan terhadap nyawa, fisik atau badan, kesusilaan, kemerdekaan, hak milik atau barang, narkoba, penipuan, penggelapan dan korupsi berada dikelompok ini, hanya kejahatan terhadap ketertiban umum yang berada di atas rata-rata kelompok lain. sehingga masuk ke dalam kategori daerah yang aman.

**Deskripsi Kelompok Terbaik**

**TABEL 4.** deskripsi kelompok terbaik

Cluster	(X1)	(X2)	(X3)	(X4)	(X5)	(X6)	(X7)	(X8)
1	8	23	4	1	114	14	26	26
2	4	14	2	0	63	5	15	17
3	3	4	1	0	24	2	6	7
4	1	2	1	0	20	1	1	47

- a. Pada variabel X1 jumlah kejahatan terhadap nyawa tertinggi berada pada *cluster* satu dengan rata- rata 8 kasus kejahatan dan jumlah kejahatan terendah pada *cluster* empat dengan rata-rata 1 kasus kejahatan.
- b. Pada variabel X2 jumlah kejahatan terhadap fisik atau badan tertinggi berada pada *cluster* satu dengan rata- rata 23 kasus kejahatan dan jumlah kejahatan terendah pada *cluster* empat dengan rata-rata 2 kasus kejahatan.

- c. Pada variabel X3 jumlah kejahatan terhadap kesusilaan tertinggi berada pada *cluster* satu dengan rata-rata 4 kasus kejahatan dan jumlah kejahatan terendah pada *cluster* tiga dan *cluster* empat dengan rata-rata 1 kasus kejahatan.
- d. Pada variabel X4 jumlah kejahatan terhadap kemerdekaan tertinggi berada pada *cluster* satu dengan rata-rata 1 kasus kejahatan dan tidak ada jumlah kejahatan pada *cluster* tiga, *cluster* empat, dan *cluster* dua dengan rata-rata 0 kasus kejahatan.
- e. Pada variabel X5 jumlah kejahatan terhadap hak milik atau barang tertinggi berada pada *cluster* satu dengan rata-rata 144 kasus kejahatan dan jumlah kejahatan terendah pada *cluster* empat dengan rata-rata 20 kasus kejahatan.
- f. Pada variabel X6 jumlah kejahatan terhadap narkoba tertinggi berada pada *cluster* satu dengan rata-rata 14 kasus kejahatan dan jumlah kejahatan terendah pada *cluster* empat dengan rata-rata 1 kasus kejahatan.
- g. Pada variabel X7 jumlah kejahatan terhadap penipuan, penggelapan, dan korupsi tertinggi berada pada *cluster* satu dengan rata-rata 26 kasus kejahatan dan jumlah kejahatan terendah pada *cluster* empat dengan rata-rata 1 kasus kejahatan.
- h. Pada variabel X8 jumlah kejahatan terhadap ketertiban umum tertinggi berada pada *cluster* empat dengan rata-rata 47 kasus kejahatan dan jumlah kejahatan terendah pada *cluster* ketiga dengan rata-rata 7 kasus kejahatan.

## KESIMPULAN

Berdasarkan hasil penelitian, maka dapat disimpulkan:

Hasil analisis *cluster* dari 24 kabupaten kota berdasarkan jenis kejahatan (kejahatan terhadap nyawa, kejahatan terhadap fisik/badan, kejahatan terhadap kesusilaan, kejahatan terhadap kemerdekaan, kejahatan terhadap hak milik/barang, kejahatan terhadap narkoba, kejahatan terhadap penipuan, penggelapan, dan korupsi, dan kejahatan terhadap ketertiban umum). Dapat dikelompokkan menjadi 4 kelompok atau *cluster*, yaitu: *cluster* pertama adalah kelompok yang memiliki tingkat kejahatan cukup rendah, *cluster* kedua adalah kelompok yang memiliki tingkat kejahatan tinggi, *cluster* tiga adalah kelompok yang memiliki tingkat kejahatan rendah, dan *cluster* empat adalah kelompok yang memiliki tingkat kejahatan cukup tinggi.

## DAFTAR PUSTAKA

- Adhe, D. *et al.* (2020). Implementasi text mining pengelompokkan dokumen skripsi menggunakan metode *k-means clustering*, *Jurnal FMIPA UNMUL*. 11, pp. 167–174.
- Alfina, T., Santosa, B. & Barakbah, R. (2012). Analisa perbandingan metode *hierarchical clustering*, *k-means* dan gabungan keduanya dalam kluster data ( Studi kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS ). *Jurnal teknik its*, 1.
- Andayani, S. (2007). Pembentukan *cluster* dalam *knowledge discovery in database* dengan algoritma *k-means*. *Jurnal SEMNAS Matematika dan Pendidikan Matematika*. Jurdik Matematika FMIPA UNY Yogyakarta, 2007.
- Awaliah, Resky. (2018). *Analisis clustering untuk mengelompokkan tingkat kesejahteraan kabupaten/kota berdasarkan sosial ekonomi rumah tangga di wilayah Provinsi Sulawesi Selatan*. Skripsi. Fakultas Sains Dan Teknologi, Jurusan Matematika, Universitas Islam Negeri Alauddin, Makassar.

- Dwirohayati, defi. (2020). *Pemetaan daerah rawan kriminalitas menggunakan k-means clustering Polrestabes Bandar Lampung*. Skripsi. Informatic Institute Business Darmajaya, Bandar Lampung.
- Everitt, B.S. (1993). *Cluster Analysis*. Third edition halsted press an imprint of John Wiley and Sons Inc. *Jurnal of American Statistikal*. Ney York.
- Goreti, Maria. (2016). Perbandingan hasil analisis *cluster* menggunakan metode *single linkage* dan metode *k-means*. *Jurnal Eksponensial Volume 7*, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman.
- Goreti, & Yuki Novia. (2016). Perbandingan hasil analisis cluster dengan menggunakan metode *single linkage* dan metode *k-means*, *Jurnal FMIPA UNMUL*. 7, pp. 9–16.
- Nuningsih, Sri, (2010). *K-Means Clustering: Studi kasus pada data pengujian kualitas susu di koperasi peternakan Bandung Selatan*. Skripsi. Fakultas Matematika Dan Ilmu Pengetahuan Alam. UPI. (2013).
- Prima, D., T. & Widodo, E. (2017). Pengelompokan daerah rawan kriminalitas di Indonesia menggunakan analisis *k-means clustering*, *Jurnal Seminal Nasional Integritas Matematika*.1(1), pp. 147–153.
- Sianturi, Yuni Irma, (2020). *Faktor-faktor yang mempengaruhi kriminalitas di Sumatra Utara tahun 2019 Dengan Analisis Jalur*. Skripsi. FMIPA, Statistika, Universitas Sumatra Utara, Medan.
- Suryadi, U. T., Supriatna, Y. & Pendahuluan, I. (2019). Sistem *clustering* tindak kejahatan pencurian di Wilayah Jawa Barat menggunakan algoritma *k-means*. *Jurnal teknologi informasi dan komunikasi*, (April), pp. 15–27.
- Zahrotun, L. (2015). Analisis pengelompokan jumlah penumpang bus trans Jogja menggunakan metode *clustering k-means* dan *agglomerative hierarchical clustering (AHC)*, *Jurnal Ifomatiki*. 9(1), pp. 1039–1047