PENGELOMPOKKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR KRIMINALITAS NASIONAL MENGGUNAKAN ALGORITMA K-MEANS DENGAN METODE DAVIES BOULDIN INDEX

Taufik Tri Wijanarko¹, Hardian Oktavianto², Reni Umilasari³

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember Email : taufiktriwijanarko87@gmail.com

ABSTRAK

Kriminalitas adalah perbuatan yang merugikan korban, juga masyarakat karena fenomena ini menghilangkan rasa aman pada kehidupan sosial. Tingkat resiko terkena tindak kejahatan (*crime rate*) pada tahun 2018 sebesar 113.000 penduduk. Melihat resiko terkena tindak kejahatan masih tinggi perlu dibuat sebuah pengelompokkan data kepolisian daerah provinsi di Indonesia agar dapat mengetahui tingkat terendah dan tingkat tertinggi pada Indikator Kriminalitas Nasional dengan mengukur menggunakan jumlah angka kejahatan (crime total), jumlah kejahatan yang telah diselesaikan (crime cleared), dan risiko penduduk terkena kejahatan (crime rate). Implementasi data mining menggunakan algoritma K-Means untuk mengelompokkan kepolisian daerah provinsi di Indonesia berdasarkan data Indikator Kriminalitas Nasional pada tahun 2018 melalui situs resmi Badan Pusat Statistik Indonesia, diperoleh cluster optimumnya dengan hasil 3 cluster berdasarkan indeks nilai *Davies Bouldin* sebesar 0,468 dari serangkaian pengujian dengan skenario 2 cluster sampai dengan 5 cluster. Sedangkan jumlah anggota pada masing masing cluster vaitu cluster 1 terdapat 2 provinsi, cluster 2 terdapat 28 provinsi, dan *cluster* 3 terdapat 4 provinsi.

Kata Kunci: Kriminalitas, K-Means, Davies Bouldin

ABSTRACT

Criminality is an act that is detrimental to the victim, as well as society, because this phenomenon eliminates the sense of security in social life. The level of risk of being exposed to crime (crime rate) in 2018 was 113,000 people. Seeing that the risk of being exposed to crime is still high, it is necessary to make a grouping of provincial police data in Indonesia in order to find out the lowest and highest levels of the National Crime Indicator by measuring using the number of crimes (total crime), the number of crimes that have been resolved (crime cleared), and the risk of the population being exposed to crime (crime rate). The implementation of data mining uses the K-Means algorithm to group provincial police in Indonesia based on National Crime Indicator data in 2018 through the official website of the Indonesian Central Statistics Agency, the optimum cluster is obtained with 3 clusters based on the Davies Bouldin value index of 0.468 from a series of test scenarios 2 clusters up to 5 clusters. While the number of members in each cluster, namely cluster 1 has 2 provinces, cluster 2 has 28 provinces, and cluster 3 has 4 provinces

Keywords: Criminality, K-Means, Davies Bouldin

1. PENDAHULUAN

Masalah kriminalitas atau kejahatan adalah masalah yang sering terjadi di negara-negara berkembang seperti di Indonesia. Berita-berita kriminal akan menjadi salah satu hal yang menarik perhatian masyarakat karena objek dalam setiap berita bisa terjadi kapanpun dan dimanapun. Istilah kriminal dalam Kamus Besar Bahasa Indonesia berarti "berkaitan dengan kejahatan (pelanggaran hukum) yang dapat dihukum menurut undang-undang pindana". Sebagai komoditas, peristiwa kriminalitas juga tentu menjadi berita yang bisa disajikan media untuk mempertahankan minat khalayaknya. Namun dengan melihat sudut pandang pemberitaan yang berbeda, maka berita-berita kriminalitas akan berdampak luas pada masyarakat.

Masyarakat secara rasional akan memberikan reaksi terhadap sanksi yang berat dengan menghindari atau mencegah untuk melakukan perbuatan kriminal atau melanggar hukum (Atmasasmita & Wibowo, 2016). Seiring dengan itu salah satu kewajiban pemerintah dan negara Indonesia adalah memberikan rasa aman pada seluruh rakyatnya. Kewajiban ini secara jelas juga tercantum dalam Pasal 30 ayat (4), amandemen kedua UUD 1945 yang antara lain menyebutkan bahwa Kepolisian Negara Republik Indonesia (POLRI) adalah alat negara yang menjaga keamanan dan ketertiban masyarakat serta bertugas melindungi, mengayomi, dan melayani masyarakat serta menegakkan hukum.

Rasa aman merupakan salah satu kebutuhan hidup masyarakat yang harus terpenuhi. Indeks keamanan sebagai salah satu alat pengukuran tingkat keamanan dan digunakan untuk mengukur perubahan tingkat keamanan, misalnya jumlah angka kejahatan (crime total), jumlah orang yang berisiko terkena tindak kejahatan (crime rate) setiap 100.000 penduduk. Semakin tinggi angka kriminalitas menunjukkan semakin banyak tindak kejahatan pada masyarakat yang merupakan indikasi bahwa kondisi suatu wilayah menjadi semakin tidak aman.

Selama periode tahun 2016 sampai 2018, jumlah kejadian kejahatan atau tindak kriminalitas di Indonesia cenderung fluktuatif. Data POLRI menunjukkan bahwa jumlah kejadian kejahatan (crime total) pada tahun 2016 sebanyak 357.197 kejadian, menurun menjadi sebanyak 336.652 kejadian pada tahun 2017 dan menurun pada tahun 2018 menjadi 294.281 kejadian. Dari data tersebut crime rate (risiko penduduk terkena kejahatan) yang merupakan angka yang dapat menunjukkan tingkat kerawanan kejahatan pada suatu daerah provinsi tertentu dalam waktu tertentu. Semakin tinggi angka crime rate maka tingkat kerawanan akan kejahatan suatu daerah semakin tinggi pula, dan sebaliknya. Tingkat kriminalitas dapat menjadi sebuah gambaran yang jelas dalam menganalisis kondisi keamanan dari sebuah wilayah

2. PENELITIAN TERKAIT

A. Kepolisian Daerah Provinsi di Indonesia

Kepolisian Negara Republik Indonesia (POLRI) adalah Kepolisian Nasional di Indonesia, yang bertanggung jawab langsung di bawah Presiden. POLRI juga mempunyai moto Rastra Swakotama yang artinya Abdi Utama bagi Nusa Bangsa. Berdasarkan dengan kehidupan bernegara POLRI merupakan alat yang berperan dalam memelihara keamanan dan ketertiban masyarakat, menegakan hukum, serta memberikan perlindungan, pengayoman, dan pelayanan pada masyarakat dalam rangka terpeliharanya keamanan dalam negeri, agar dalam melaksanakan fungsinya dan perannya di seluruh wilayah Indonesia dibagi dalam daerah hukum menurut kepentingan pelaksaan tugas Kepolisian Negara Republik Indonesia, maka dari itu

ditentukan dalam peraturan pemerintah. (Budi Rizki Husin, studi lembaga penegak hukum, 2014. Op.cit, hlm 15).

Di Indonesia terdapat 34 Provinsi dan setiap provinsi memiliki POLDA. Provinsi Papua Barat ada di wilayah hukum POLDA Papua namun sejak tanggal 19 Desember 2014, Papua Barat resmi memiliki POLDA sendiri. Provinsi 7 Sulawesi Barat juga sudah memiliki wilayah hukum sendiri berpisah dari POLDA Sulawesi Selatan sejak Juli 2016. Terakhir, provinsi Kalimantan Utara resmi memiliki wilayah hukum sendiri setelah berpisah dari POLDA Kalimantan Timur pada awal tahun 2018 ini

B. Definisi Kejahatan

Kejahatan menurut Kamus Bahasa Indonesia adalah perilaku yang bertentangan dengan nilai-nilai dan norma-norma yang berlaku yang telah disahkan oleh hukum tertulis (hukum pidana). Pengertian sederhananya yaitu suatu tindakan kejahatan merupakan segala tindakan atau sesuatu yang dilakukan individu, kelompok, ataupun komunitas yang melanggar hukum, sehingga mengganggu keseimbangan atau stabilitas sosial dalam masyarakat.

Kejahatan bukan merupakan peristiwa hereditas (bawaan sejak lahir, warisan), juga bukan merupakan warisan biologis. Tindak kejahatan bisa dilakukan siapapun, baik wanita maupun pria, dengan tingkat pendidikan yang berbeda. Tindak kejahatan bisa dilakukan secara sadar yaitu difikirkan, direncanakan dan diarahkan pada maksud tertentu secara sadar benar. Kejahatn merupakan suatu konsepsi yang bersifat abstrak, dimana kejahatan tidak dapat diraba dan dilihat kecuali akibatnya saja (Kartono, 2005).

C. Data Mining

Data *mining* adalah suatu proses pencarian pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data 14 mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat bergantung pada tujuan dan proses KDD secara keseluruhan.

D. Clustering

Prasetyo Eko (2013) mengatakan bahwa *Clustering* adalah teknik menemukan sekelompok data dari pemecahan atau pemisahan sekumpulan data menurut karakteristik tertentu yang telah ditentukan. Dalam pengelompokkan tersebut nilai labelnya belum diketahui sehingga diharapkan setelah melakukan 17 pengelompokkan data dapat diketahui label dari data tersebut. Metode *clustering* juga sering disebut dengan tahap awal sebelum melakukan metode lain seperti klasifikasi.

E. K-Means

Algoritma K-Means adalah salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster* atau kelompok. Metode ini mempartisi ke dalam cluster atau kelompok sehingga data yang memiliki karakteristik yang sama (*High intra class similarity*) dikelompokkan ke dalam satu *cluster* yang sama dan yang memiliki karakteristik yang berbeda (*Law inter class similarity*) dikelompokkan pada kelompok yang lain.

Proses *clustering* dimulai dengan mengidentifikasi data yang akan dikelompokkan, Xij(i=1, ..., n; j=1, ..., m) dengan n adalah jumlah data yang akan dicluster dan m adalah jumlah variabel. Pada awal iterasi, pusat setiap cluster ditetapkan secara bebas (sembarang), Ckj(k=1, ..., k; j=1, ..., m). Kemudian

dihitung jarak antara setiap data dengan setiap pusat cluster. Untuk melakukan penghitungan jarak data ke-i (Xi) pada pusat cluster ke-k (Ck) diberi nama (dik), dapat digunakan formula Euclidean, seperti pada persamaan (1), yaitu:

$$d_{ik} = \sqrt{\sum_{j=1}^{m} (x_{ij} - c_{ij})^2}$$
 (1)

Suatu data akan menjadi anggota *cluster* ke-k apabila jarak data tersebut ke pusat *cluster* ke-k bernilai paling kecil jika dibandingkan dengan jarak ke pusat cluster lainnya. Hal ini dapat dihitung dengan menggunakan persamaan (2) Selanjutnya, kelompokkan data-data yang menjadi anggota pada setiap *cluster*

$$Min \sum_{j=1}^{k} d_{ik} = \sqrt{\sum_{j=1}^{m} x_{ij} - c_{ij}}$$
 (2)

Nilai pusat cluster yang baru dapat dihitung dengan cara mencari nilai ratarata dari data-data yang menjadi anggota pada cluster tersebut, dengan menggunakan rumus pada persamaan (3):

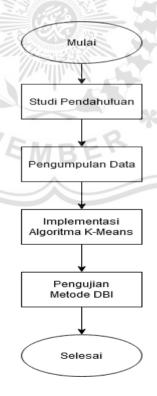
$$c_{ij} = \frac{\sum_{i=1}^{p} x_{ij}}{p}$$
Dimana $X_{ij} \in cluster \text{ ke-}k$ (3)

p = banyaknya anggota cluster ke k

F. Metode Davies Bouldin Index

Davies Bouldin Index (DBI) merupakan cara validasi cluster yang ditemukan oleh D.L.Davies. DBI adalah fungsi rasio dari jumlah distribusi didalam cluster untuk pemisahan antar cluster. Adapun langkah-langkah untuk menghitung Davies Bouldin Index (Sujacka, 2019).

Tahapan Penelitian



Gambar 1. Diagram Alur Penelitian

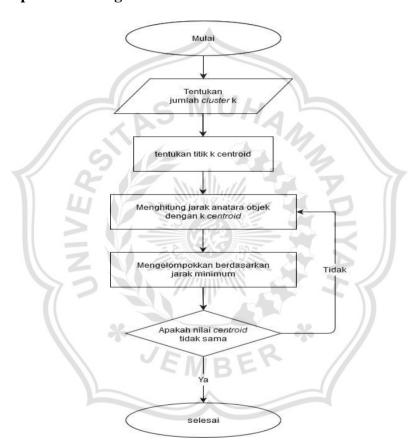
H. Studi Pendahuluan

Tahapan awal dari penelitian ini adalah dengan mencari dan mempelajari masalah yang akan diteliti, kemudian menentukan ruang lingkup masalah, latar belakang, dan mempelajari beberapa literatur yang berkaitan dengan permasalahan dan bagaimana mencari solusi dari masalah tersebut.

I. Pengumpulan Data

Pada pengumpulan data peneliti ini menggunakan data Statistik Kriminal di Indonesia yang dimana terdapat 34 kepolisian daerah provinsi di Indonesia dengan 3 atribut data Indikator Kriminalitas Nasional 3 atribut yang berupa data pada tahun 2018.

J. Implementasi Algoritma K-Means



Gambar 2. Flowchart Algoritma K-Means

3. HASIL DAN PEMBAHASAN

Pada bab ini membahas tentang hasil dan pembahasan dari pengujian data yang diperoleh dari Bada Pusat Statistik (BPS) di Indonesia pada tahun 2019. Data yang digunakan adalah data Indikator Kriminalitas Nasional yang meliputi 34 kepolisian daerah provinsi di Indonesia dengan 3 variabel yaitu, jumlah kejahatan yang dilaporkan (*crime total*), jumlah kejahatan yang telah diselesaikan (*crime cleared*), dan risiko penduduk terkena kejahatan (*crime rate*). **Pengujian Kasus Tertinggi dan Terendahnya.**

A. Skenario dan Hasil Pengujian

Pengujian terhadap K-Means dengan mencari nilai optimum pada uji validitas menggunakan Davies Bouldin Index dan sekenari nilai k=2 sampai dengan k=5Berikut nilai akurasi, presisi, dan recall pengujian k-fold.Berikut adalah hasil dari metode DBI:

1. Untuk k = 2 diperoleh nilai SSW sebagai berikut:

 $SSW_1 = 9150,732$

 $SSW_2 = 3481,618$

Sedangkan nilai SSB untuk kluster (d_{ii}) adalah:

 $d_{C1,C2} = 25578,655$

Jika dibuat dalam bentuk matriks, maka hasil nilai dari menghitung jumlah rasionya adalah:

$$R = \begin{bmatrix} 0 & 0,494 \\ 0.494 & 0 \end{bmatrix}$$

 $R = \begin{bmatrix} 0 & 0,494 \\ 0,494 & 0 \end{bmatrix}$ Sehingga, hasil DBI untuk k = 2 diperoleh sebagai berikut:

DBI =
$$\frac{1}{3} \sum_{i=1}^{2} Max(R) = \frac{1}{2} (0,494 + 0,494) = 0,494$$

Untuk k = 3 diperoleh nilai SSW sebagai berikut:

$$SSW_1 = 4158,548$$

 $SSW_3 = 4814,39644$

 $SSW_2 = 3242,717$

Sedangkan nilai SSB untuk kluster (d_{ij}) adalah:

$$d_{C1,C2} = 36085,885$$
 $d_{C2,C3} =$

$$d_{C2,C3} = 16776,2001$$

 $d_{C1,C3} = 20215,55122$

Jika dibuat dalam bentuk matriks, maka hasil nilai dari menghitung jumlah rasionya adalah:

$$R = \begin{bmatrix} 0 & 0,201 & 0,44386345 \\ 0,201 & 0 & 0,48027047 \\ 0,4438634 & 0,480270468 & 0 \end{bmatrix}$$

Sehingga, hasil DBI untuk k = 2 diperoleh sebagai berikut:

DBI =
$$\frac{1}{3} \sum_{i=1}^{2} Max(R) = \frac{1}{3} (0,444 + 0,480 + 0,480270468) = 0,468$$

Untuk k = 4 diperoleh nilai SSW sebagai berikut:

$$SSW_1 = 4814,375$$

$$SSW_3 = 4158,547882$$

$$SSW_2 = 2034,739$$

$$SSW_4 = 1200,188664$$

Sedangkan nilai SSB untuk kluster (d_{ij}) adalah:

$$d_{C1,C2} = 13853,796$$

$$d_{C2,C3} = 33915,80484$$

$$d_{C1,C3} = 20216,08283$$

$$d_{C2,C4} = 6300,801887$$

$$d_{C1,C4} = 20150,07821$$
 d_{C3}

$$d_{C3,C4} = 40148,86532$$

Jika dibuat dalam bentuk matriks, maka hasil nilai dari menghitung jumlah rasionya adalah:

$$R = \begin{bmatrix} 0 & 0,494 & 0,44385072 & 0,298488356 \\ 0,494 & 0 & 0,18260768 & 0,523415163 \\ 0,44385072 & 0,18260768 & 0 & 0,133471681 \\ 0,298488356 & 0,513415163 & 0,133471681 & 0 \end{bmatrix}$$

Sehingga, hasil DBI untuk k = 2 diperoleh sebagai berikut:

DBI =
$$\frac{1}{3} \sum_{i=1}^{2} Max(R) = \frac{1}{4} (0,494 + 0,513 + 0,44385072 + 0,513415163) = 0,491$$

Untuk k = 5 diperoleh nilai SSW sebagai berikut:

 $SSW_1 = 1200,185366$ $SSW_4 = 1693,816276$

 $SSW_2 = 41548,547854$ $SSW_5 = 3106,830298$

 $SSW_3 = 2722,006987$

Sedangkan nilai SSB untuk kluster (d_{ii}) adalah:

```
d_{C1,C2} = 40148,1723
                          d_{C2,C4} = 34250,56426
d_{C1,C3} = 14234,25956
                          d_{C2,C5} = 16348,74507
                          d_{C3,C4} = 8296,016974
d_{C1,C4} = 5944,0303335
                           d_{C3,C5} = 10493,7923
d_{C1,C5} = 24618,82728
d_{C2,C3} = 25988,00506
                           d_{C4,C5} = 18683,68
Jika dibuat dalam bentuk matriks, maka hasil nilai dari menghitung
jumlah rasionya adalah:
                    0,13347390
                                  0,275545934
                                                  0,486875315
                                                                0,174948043
     0,13347390
                         0
                                   0,264758869
                                                  0,17086913
                                                                 0,444399746
R = 0.191229264
                   0,264758869
                                                  0,532282332
                                                                0,555455751
                    0,17086913
     0,486875315
                                   0,532282332
                                                       0
                                                                 0,259643256
    L0,174948043 0,444399746 0,555455751 0,256943256
Sehingga, hasil DBI untuk k = 2 diperoleh sebagai berikut:
DBI = \frac{1}{3} \sum_{i=1}^{2} Max(R) = \frac{1}{2} {0,486875315 + 0,444399746 + 0,555455751 + 0,515} + 0,515455751 = 0,515
```

Tabel 4. 1 Hasil Nilai Proses Metode DBI

Cluster	Nilai DBI	Anggota Cluster
2	0,494	Cluster 1: 5 items dan Cluster 2: 29 items
3	0,468	Cluster 1: 2 items, Cluster 2: 28 items, dan Cluster 3: 4 items
4	0.491	Cluster 1: 4 items, Cluster 2: 15 items, Cluster 3: 2 items, dan Cluster 4: 13 items
5	0,515	Cluster 1: 13 items, Cluster 2: 2 items, Cluster 3: 3 items, Cluster 4: 14 items, dan Cluster 5: 2 items

Hasil segmentasi yang terbentuk akan dievaluasi menggunakan *metode Davies Bouldin Index* (DBI) yang dimana metode ini merupakan metode validasi *cluster* dari hasil *clustering*. Pendekatan pengukuran DBI yaitu memaksimalkan jarak inter *cluster* serta meminimalkan jarak intra *cluster*. Dari perhitungan tersebut telah diperoleh hasil pada **Tabel 4.2** dalam pengujian skenario *cluster* 2 sampai dengan *cluster* 5 yang menerangkan bahwa nilai DBI terdapat pada 3 *cluster* dengan nilai yaitu 0,468 berdasarkan nilai indeks terkecil.

4. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil dan pembahasan yang telah diuraikan, maka dapat disimpulkan bahwa:

- 1. Berdasarkan algoritma *K-Means* untuk proses *clustering* data Indikator Kriminalitas Nasional di Indonesia dengan 34 provinsi dan 3 variabel yaitu *crime total, crime cleared,* dan *crime rate.* Hasil dari pengelompokkan diperoleh *cluster* optimum dengan hasil 3 *cluster* berdasarkan indeks nilai *Davies Bouldin* terkecil yaitu sebesar 0,468 dengan skenario *cluster* 2 sampai dengan 5 *cluster*.
- 2. Hasil pengelompokkan berdasarkan Indikator Kriminalitas Nasional menurut provinsai dengan algoritma K-Means didapatkan untuk cluster tertinggi (C1) terdapat 2 Kepolisian daerah provinsi diantaranya Sumater Utara dan Metro Jaya. Pada *cluster* terendah (C2) terdapat 28 kepolisan daerah provinsi diantaranya Aceh, Sumatera Barat, Riau dan 25 provinsi lainnya. Sedangkan pada *cluster* rendah (C3) terdapat 4 kepolisan daerah provinsi diantaranya Sumatera Selatan, Jawa Barat, Jawa Timur, dan Sulawesi Selatan.
- 3. Dalam perankingan pada suatu daerah atau suatu wilayah menggunakan nilai rata rata setiap *cluster*. Maka hasil rata-rata karakteristik kasus kejahatan Crime Total, Crime Cleared, dan Crime Rate. Pada cluster 1 menempati posisi tertinggi dengan jumlah kasus, 33.789 kasus, 26.510 kasus, dan 184 kasus dibandingkan dengan cluster 2 dan cluster 3. Sedangkan pada cluster 2 menempati posisi terendah dengan jumlah kasus, 6.675 kasus, 3.629 kasus, dan 221,5 kasus dibandingkan pada cluster 3 dengan jumlah kasus, 19.927 kasus, 12.805 kasus, dan 148,5 kasus. Sehingga apabila kepala Kepolisian setiap daerah provinsi akan mengalokasikan tenaga keamanan, petugas Kepolisian dapat melihat daftar kasus terbanyak pada hasil rata-rata karakteristik kasus kejahatan.

B. Saran

Pada penelitian ini masih sangat jauh dari kata sempurna, disarankan bagi penelitian selanjutnya agar :

- 1. Dalam penelitian ini yaitu sebagai penerapan dari algoritma *K-Means*, agar lebih bermanfaat dengan harapan dapat menghasilkan penelitian yang lebih baik lagi maka disarankan membangun Sistem Informasi sebagai media implementasi algoritma *K-Means*.
- 2. Diharapkan adanya penelitian dan kajian lebih lanjut mengenai pengembangan aplikasi Sistem Informasi yang menggunakan algoritma *K-Means*. Serta melakukan uji validitas dapat menggunkan teknik yang berbeda seperti *elbow*, *silhoute coefisien* dan dapat meningkatkan jumlah atribut dan *variable* pada data yang digunakan terhadap kinerja program sebagai bahan masukan untuk keperluan pengembangan program selanjutnya.

5. DAFTAR PUSTAKA

- Atmasasmita, R., & Wibowo, K. (2016). *Analisis Ekonomi Tentang Hukum Pidana*. Jakarta: Prenada Media Group.
- Davies, G. M., Hollin, C. R., & Bull, R. (2008). Forensic Psychology Paperback. Fayyad, U. (1996). Advances in Knowledge Discovery and Data mining. MIT Press. Kartono, K. (2005). Patologi Sosial. Jakarta: Raja Grafindo Persada. Kumar, W. d. (2009). The Top Ten Algorithms in Data Mining. CRC Press.
- Larose, D. T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. John Willey & Sons, Inc.
- Maulida, L. (2018). PENERAPAN DATAMINING DALAM MENGELOMPOKKANKUNJUNGAN WISATAWAN KE OBJEK WISATA UNGGULAN DIPROV. DKI JAKARTA DENGAN K-MEANS.
 - JISKa (Jurnal Informatika Sunan Kalijaga), Vol. 2, No. 3, Program Studi Manajemen Informatika, AMIK BSI TangerangKomplek BSD Sektor XIV- C11 Jl Letnan Sutopo Banten, 167–174.
- Ong, J. O. (2013). Algoritma Clustering Untuk Menentukan Strategi Marketing Pada President University.
- Ong, J. O. (2013). Implementasi algoritma k-means clustering untuk menentukan strategi marketing president university
- Pranoto, S. W. (2008: 39). Bandit berdasi: korupsi berjamaah: merangkai hasil kejahatan pasca-reformasi. Yogyakarta: Kanisius.
- Rinawati, Sihombing, E. G., Dew, L. S., & Arisawati, E. (2020). Analisis Algoritma Datamining pada Kasus Daerah Pelaku Kejahatan Pencurian Berdasarkan Provinsi. *Jurnal Sains Komputer & Informatika (J-SAKTI) Volume 4Nomor 1Maret2020, pp. 77-87ISSN:2548-9771/EISSN:2549-7200*
- Sujacka. (2019). Peningkatan Akurasi Algoritma K-Means dengan Clustering Purity Sebagai Titik Pusat Cluster Awal (Centroid). Sumatera Utara: Universitas Sumatera Utara.