PENGELOMPOKAN PENYEBARAN VIRUS COVID-19 MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING YANG BERADA DI WILAYAH JAWA BARAT

Syifa Restillah, Ade Irma Purnamasari
Program Studi Manajemen Informatika D3 Sekolah Tinggi Manajemen Informatika dan Komputer IKMI Cirebon, Jalan Perjuangan No 10 B Kota Cirebon Indonesia
Syifarestillah123@gmail.com

ABSTRAK

Indonesia telah terkena penyakit kategori baru yang dikenal sebagai *COVID-19*, yang telah menyebar ke seluruh dunia. China mengalami pendeteksian *COVID-19* pertamanya pada akhir tahun 2019. Pasar basah Wuhan di China adalah lokasi klaster pertama di mana virus tersebut ditemukan. Pada 2 Maret 2020, virus tersebut ditemukan di kota Depok, Jawa Barat, Indonesia. virus penyebab *COVID-19* dapat mengganggu sistem pernafasan dan menimbulkan berbagai gejala mulai dari gejala mirip *flu* hingga infeksi paru-paru seperti *pneumonia*. Hingga 27 Desember 2022, jumlah total orang yang terinfeksi *COVID-19* mencapai 1.231.878 di Provinsi Jawa Barat, menjadikan Jawa Barat sebagai Provinsi dengan tingkat infeksi *COVID-19* tertinggi. Untuk membuat strategi menghentikan penyebaran *COVID-19*, maka penting untuk mengelompokkan Kota atau Kabupaten di Provinsi Jawa Barat. Informasi ini merupakan salah satu titik data yang tersedia di *website Open Data* Jabar. Pendekatan *K-Means clustering* yang membagi data menjadi banyak *cluster* berdasarkan kesamaan data dan diimplementasikan menggunakan Rapidminer. Rapidminer digunakan untuk mengelompokkan penyebaran *COVID-19*. Berdasarkan temun kajian, terdapat dua *cluster*, dengan *cluster* 0 yang penyebaran *C0VID-19* terbanyak tersebar di 23 Kabupaten atau kota. 4 Kabupaten atau Kota diperoleh untuk *cluster* 1, *cluster* dengan sebaran *COVID-19* paling sedikit. Pemerintah Provinsi Jawa Barat diharapkan dapat mengambil manfaat dari temuan studi tersebut untuk menangani kasus *COVID-19.*

Kata kunci: *Covid-19, Cluster, K-Means*

1. PENDAHULUAN

Virus Covid-19 awalnya teridentifikasi pada akhir Desember 2019 di Wuhan, China. Awal Maret 2020, wabah virus ini meluas ke Indonesia. Virus covid-19, mengganggu sistem pernapasan dan dapat menyebabkan infeksi paru-paru seperti pneumonia atau gejala mirip flu lainnya [1]. Per 27 Desember 2022, jumlah penduduk Indonesia memiliki total penyebaran Covid-19 sebanyak 6.717.395 juta, Provinsi Jawa Barat memiliki angka tertinggi sebesar 1.231.878 juta. Provinsi Jawa Barat saat ini menempati posisi teratas penyebaran Covid-19 di Indonesia. Untuk itu, pemerintah daerah harus menyusun rencana penanganan wabah Covid-19 di Provinsi Jawa Barat yang sudah meluas menjadi 27 kota dan kabupaten. Untuk menerapkan pendekatan ini, diperlukan klaster atau pengelompokan data wabah Covid-19 di Provinsi Jawa Barat menggunakan Algoritma K-Means.

K-Means *clustering* adalah teknik analisis data atau data mining yang menggunakan sistem partisi untuk mengkategorikan data. Data yang memiliki sifat yang sama atau sebanding dikelompokkan bersama ke dalam *cluster* menggunakan teknik data mining *clustering* [2]. Hasil klasterisasi ini diharapkan dapat mengungkap sebaran kasus covid-19 di setiap wilayah Provinsi Jawa Barat yang dibagi berdasarkan wilayah kabupaten atau kota.

Tabel. 1 Jumlah yang terkonfirmasi Covid-19 Provinsi Jawa Barat

No	Kab/kota	Konf.sembuh	Konf.meninggal	Konf.aktif
1.	Kabupaten Bandung	616	13	99
2.	Kabupaten Bandung Barat	177	6	97
3.	Kabupaten Bekasi	1566	46	1024
4.	Kabupaten Bogor	855	15	858
5.	Kabupaten Ciamis	33	2	52
….	….	….	….	…
836.	Kota Sukabumi	258	5	124
837.	Kota Tasikmalaya	61	5	95

Tujuan dalam penelitian ini untuk mengetahui kabupaten atau kota mana saja yang termasuk dalam *cluster* dengan tingkat sebaran kasus Covid-19 tertinggi dan terendah di wilayah kabupaten atau kota Provinsi Jawa Barat [3]. Algoritma K-Means mengelompokkan data berdasarkan kriteria seperti jumlah pasien sembuh, jumlah pasien yang aktif dan jumlah pasien meninggal dunia.

2. TINJAUAN PUSTAKA

2.1. Covid-19

Virus SARS-CoV-2 yang mengakibatkan covid-19, mengganggu sistem pernapasan dan dapat menyebabkan infeksi paru-paru seperti pneumonia atau gejala mirip flu lainnya. Virus ini berpotensi menyebar dengan cepat dan mungkin menjangkau hampir semua negara, termasuk Indonesia [1].

2.1.1. Knowledge Discovery in Database (KDD)

KDD menghasilkan pengetahuan yang sebelumnya belum ditemukan dan berpotensi bermanfaat. Salah satu langkah dalam rangkaian prosedur iteratif KDD adalah data mining [4].

2.1.2. Data Mining

Data Mining merupakan suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari data tersebut [5].

2.1.3. Rapidminer

RapidMiner adalah platform perangkat lunak data ilmu pengetahuan yang dikembangkan oleh perusahaan dengan nama yang sama, yang menyediakan lingkungan terpadu untuk pembelajaran *machine learning* [6].

2.1.4. K-Means

algoritma K-Means adalah pembentukan partisi cluster, yang kemudian diperbaiki secara iteratif hingga tidak ada perubahan yang terlihat pada partisi cluster. Data dibagi menjadi beberapa kelompok dengan menggunakan algoritma K-Means sehingga data dengan ciri yang mirip ditempatkan pada kelompok yang sama dan data dengan ciri yang berbeda ditempatkan pada kelompok yang berbeda [7].

2.1.5. Clustering

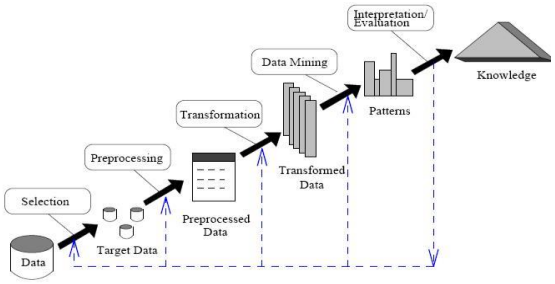
clustering adalah proses mengatur titik data menjadi dua atau lebih kelompok sehingga titik data yang termasuk dalam kelompok yang sama lebih mirip satu sama lain dari pada titik data dalam kelompok yang berbeda [8].

2.1.6. Davies Bouldin Index

Untuk menghitung jumlah cluster yang paling efektif, gunakan Indeks Davies Bouldin Index. David L. Davies dan Donald W. Bouldin menemukan Davies Bouldin Index (DBI) pada tahun 1979. Dalam metode clustering dimana keterpaduan didefinisikan sebagai penjumlahan kedekatan data dengan titik pusat *cluster* dari *cluster* yang diikuti, salah satu metode yang digunakan untuk menilai validitas jumlah cluster yang paling efektif adalah Davies-Bouldin Indeks [9].

3. METODE PERANCANGAN

3.1 Tahapan perancangan



Gambar 1. Knowledge Discovery in Data base

KDD atau *Knowledge Discovery in Databases*, digunakan dalam penelitian ini. *Knowledge Discovery in Database* memiliki tahapan dan proses sebagai berikut.

a. Data Selection
   Proses pemilihan data merupakan tahap KDD pertama dalam penelitian ini, dimana diputuskan jenis dan kategori data mana yang akan digunakan dan dibutuhkan untuk penelitian dari sekumpulan data yang ada. karena tidak semua data akan digunakan dalam perhitungan
b. *Preprocessing* Data
   Proses *cleaning* data dilakukan selama tahap *preprocessing* ini. Untuk memastikan bahwa data yang diproses benar-benar data yang relevan, *cleaning* data melibatkan penghapusan atribut yang tidak akan digunakan dan bidang yang tidak bernilai (null) dan tidak diperlukan dalam proses perhitungan setelahnya.
c. Transformasi Data
   Setelah tahap *preprocessing* selesai, lanjut ke prosedur transformasi data. Untuk mempermudah proses penambangan data, transformasi data melibatkan penempatan beberapa subdataset ke dalam format yang sama. Untuk membuat data lebih mudah disajikan dan diproses, beberapa kode juga diubah nama atributnya.
d. Data Mining
   Data mining adalah suatu metodologi atau proses untuk mengidentifikasi pola atau mengekstraksi informasi yang khas dan menarik dari dalam kumpulan data yang dipilih dengan menggunakan metodologi atau algoritma tertentu. Regresi, klasifikasi, dan pengelompokan adalah tiga teknik yang paling sering digunakan dalam penambangan data.
e. *Interpretation* atau Evaluasi
   Tahap ini melibatkan penerjemahan dan analisis pola atau informasi spesifik yang muncul dari perhitungan data mining. Dengan menentukan apakah pola, informasi, atau pengetahuan yang diperoleh konsisten dengan fakta atau hipotesis sebelumnya atau bahkan bertentangan, prosedur peninjauan ini dilakukan. Data yang terkumpul selanjutnya ditampilkan melalui visualisasi [10].

4. HASIL DAN PEMBAHASAN

4.1. Hasil Tujuan 1

4.1.1. Data Selection

Operator paling mendasar yang digunakan sebelum memulai proses adalah *Read Excel*. Data dapat dimuat menggunakan operator ini dari spreadsheet Microsoft Excel

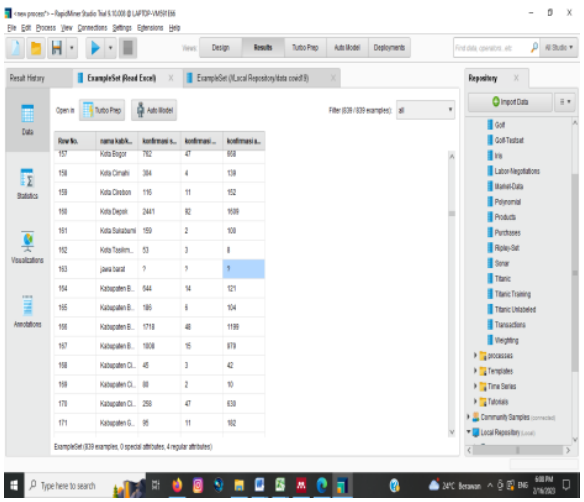
Gambar 2. Operator Read Excel

4.1.2. Preprocessing data

Preprocessing data ini menggunakan operator *Replace Missing Value* yang berfungsi untuk menyaring data yang noise atau data yang *double.*

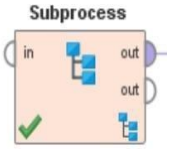
Gambar 3. Operator Replace Missing Value

Tampilan gambar ini merupakan tampilan dari hasil *Replace Missing Value yang* sudah dilakukan.



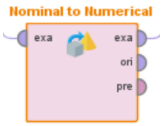
Gambar 4. Hasil dari replace missing value

4.1.3. Transformasi Data

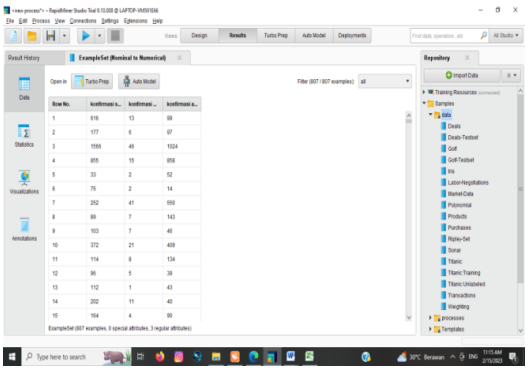


Gambar 5. Operator Subprocces

Operator *Subprocess* yang berfungsi untuk *cleansing* data. Didalam operator *Subprocess* terdapat 2 tahapan, tahap pertama menggunakan operator *Nominal to Numerical* yang berfungsi untuk mengubah semua tipe atribut yang bersifat *non-numeric* menjadi tipe *numeric*, juga digunakan untuk mengatur semua nilai atribut ke nilai *numeric*. Data yang akan di numerical yaitu nama kabupaten atau kota.

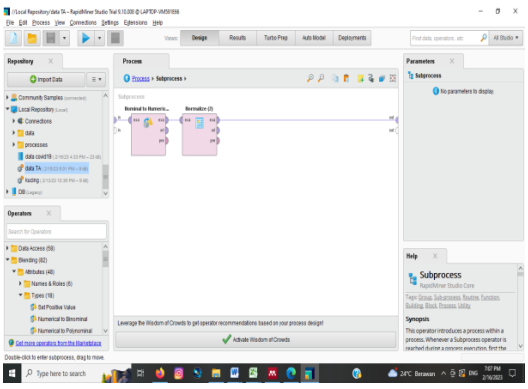


Gambar 6. Operator Nominal to Numerical

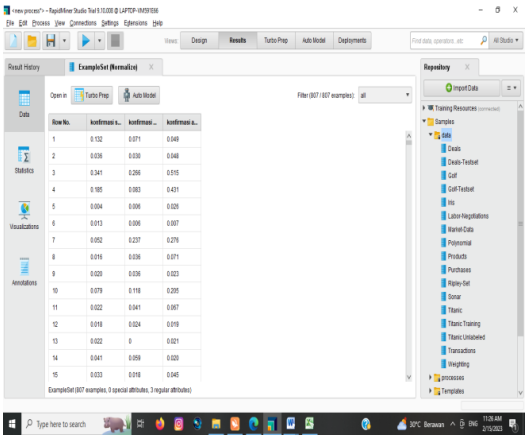


Gambar 7. Hasil Proses Nominal to Numerical

Tahap yang kedua adalah menggunakan operator *Normalize* untuk melakukan normalisasi. Operator ini berfungsi untuk memperkecil jarak atribut yang ingin kita pakai. Operator ini dilakukan besarnya jarak antar selisih.



Gambar 8. Operator Normalize

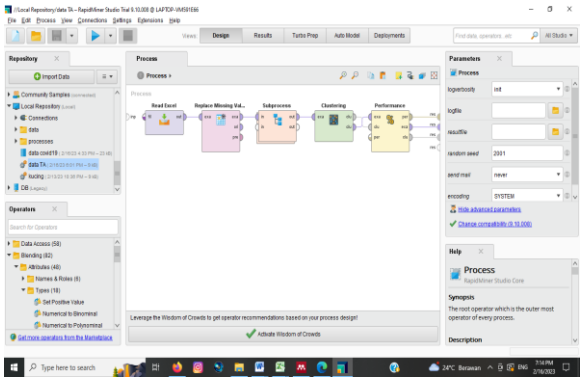


Gambar 9. Hasil Proses Normalize

Hasil dari operator *Normalize* menggunakan *method* yang digunakan yaitu *range transformation* karena dapat merubah jarak yang kita inginkan. Oleh karena itu penelitian ini mengubah jarak dengan range minimal 0.0 dan maksimal 1.0.

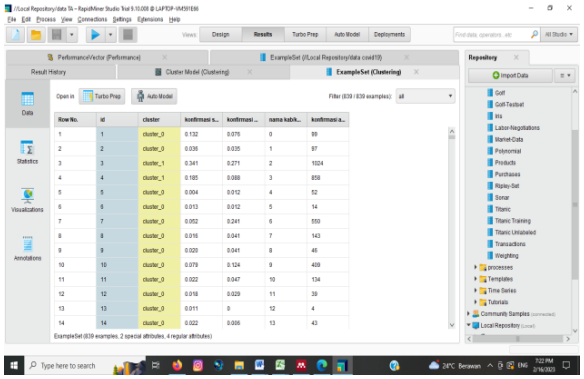
4.1.4. Data Mining

Pada tahap ini adalah metode yang digunakan dalam proses data mining ini adalah pengelompokan dengan menggunakan algoritma K-Means.



Gambar 10. Hasil Proses Clustering

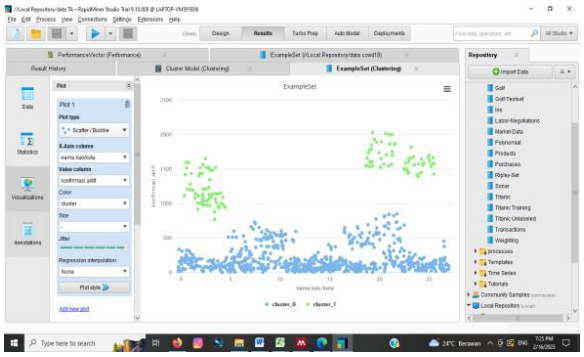
Tahap selanjutnya memasukan operator *clustering* K-Means, dan operator *cluster distance performance.*



Gambar 11. Hasil operator Clustering menggunakan algoritma K-Means

4.2. Hasil Tujuan 2

4.2.1. Jumlah Cluster optimal



Gambar 12. Hasil Visualizations Cluster

Cluster_0 yang berwarna biru merupakan cluster dengan anggota terbanyak yaitu berjumlah 23 anggota, Cluster_1 yang berwarna hijau berjumlah 4 anggota. Cluster ini memuat Kabupaten atau kota yang angka penyebaran Covid-19 nya tertinggi di Jawa Barat, yaitu Kabupaten Bandung.

Tabel 2. Hasil pengelompokan cluster 0

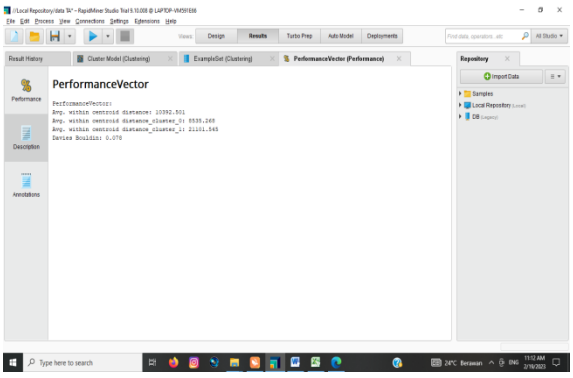
No	Kab/kota	sembuh	meninggal	aktif	cluster
1	Kab Bandung	616	13	99	cluster_0
2	Kab Bandung Barat	177	6	97	cluster_0
3	KabCiamis	33	2	52	cluster_0
4	KabCianjur	75	2	14	cluster_0
…	…	…	…	…	…
714	Kota Sukabumi	258	5	124	cluster_0
715	Kota Tasikmalaya	61	5	95	cluster_0

Tabel 3. Hasil pengelompokan cluster 1

No	Kab/kota	sembuh	meninggal	aktif	cluster
1	Kabupaten Bekasi	1566	46	1024	cluster_1
2	Kabupaten Bogor	855	15	858	cluster_1
3	Kota Bekasi	2245	49	1841	cluster_1
4	Kota Depok	2254	82	1578	cluster_1
…	…	…	…	…	…
123	Kota Bekasi	4563	114	1949	cluster_1
124	Kota Depok	4246	170	1477	cluster_1

4.2.2. Performa Cluster

Untuk menghasilkan nilai cluster (k) yang paling *optimum* menggunakan operator *Performance* untuk mengetahui hasil *Davies Bouldin Index*. Nilai *cluster* (k) yang paling *optimum* adalah yang nilai *Davies Bouldin Index* nya mendekati 0, dan sudah dilakukan 10 kali percobaan untuk mengetahui hasil *Davies Bouldin Index* yang mendekati 0.



Gambar 13. Hasil Performance

Untuk mengetahui hasil *Davies Bouldin Index yang* paling *optimum* dilakukan beberapa kali percobaan, melakukan 10 kali percobaan untuk

mengetahui nilai *Davies Bouldin Index* yang paling *optimum* yaitu, 0.078 pada *cluster* k=2.

Tabel 4. Hasil Davies Bouldin Index

No	*Cluster*	*Davies Bouldin Index*
1	K2	0.078
2	K3	0.091
3	K4	0.091
4	K5	0.119
5	K6	0.114
6	K7	0.128
7	K8	0.118
8	K9	0.120
9	K10	0.123

Berdasarkan Tabel diatas menunjukan hasil *cluster* yang optimal melakukan 10 kali percobaan yaitu k=2 dengan nilai *Davies Bouldin Index* nya 0.078

5. KESIMPULAN DAN SARAN

Kesimpulan pada penelitian ini, berdasarkan data yang sudah diolah kemudian menggunakan *Rapidminer* dengan menggunakan metode algoritma K-Means. Berdasarkan hasil diatas menjelaskan bahwa kesimpulan sebagai berikut: 1. Dengan menerapkan algoritma K-Means pada pengelompokan covid-19 berdasarkan wilayah kabupaten atau kota didapatkan 2 cluster yaitu cluster tertinggi dan terendah. 2. Hasil pengelompokan penyebaran Covid-19 berdasarkan kabupaten atau kota menggunakan algoritma K-Means yaitu *cluster_0* merupakan cluster dengan anggota terbanyak yaitu berjumlah 23 anggota dan *cluster_1* berjumlah 4 anggota, *cluster* ini memuat kabupaten yang penyebaran Covid-19 tertingi di Jawa Barat, yaitu Kabupaten Bandung.

Data yang disampaikan pada situs Open Data Jabar agar lebih *up to date* dan *real time.* Untuk mencegah penyebaran covid-19 perlu diadakannya penyuluhan terhadap masyarakat tentang bahayanya covid-19.

DAFTAR PUSTAKA

[1] N. Mirantika, A. Tsamaratul'Ain, and F. Diviana Agnia, "Volume 15 Nomor 2 , Juli 2021 PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN PENYEBARAN COVID-19 JURNAL NUANSA INFORMATIKA Volume 15 Nomor 2 , Juli 2021," *J. Nuansa Inform.*, vol. 15, pp. 92–98, 2021.

[2] M. W. Goni, D. Gustian, and F. Sembiring, "Implementasi K-Means Dalam Pengelompokan Penyebaran COVID-19 di Jawa Barat," *Progresif J. Ilm. Komput.*, vol. 17, no. 2, p. 107, 2021, doi: 10.35889/progresif.v17i2.648.

[3] D. N. P. Sari and Y. L. Sukestiyarno, "Analisis Cluster dengan Metode K-Means pada Persebaran Kasus Covid-19 Berdasarkan Provinsi di Indonesia," *Prism. Pros. Semin. Nas. Mat.*, vol. 4, pp. 602–610, 2021, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/

[4] M. R. Muttaqin and M. Defriani, "Algoritma K-Means untuk Pengelompokan Topik Skripsi Mahasiswa," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 121–129, 2020, doi: 10.33096/ilkom.v12i2.542.121-129.

[5] R. Ordila, R. Wahyuni, Y. Irawan, and M. Yulia Sari, "PENERAPAN DATA MINING UNTUK PENGELOMPOKAN DATA REKAM MEDIS PASIEN BERDASARKAN JENIS PENYAKIT DENGAN ALGORITMA CLUSTERING (Studi Kasus : Poli Klinik PT.Inecda)," *J. Ilmu Komput.*, vol. 9, no. 2, pp. 148–153, 2020, doi: 10.33060/jik/2020/vol9.iss2.181.

[6] R. Nofitri and N. Irawati, "ANALISIS DATA HASIL KEUNTUNGAN MENGGUNAKAN PENDAHULUAN Penerapan teknologi informasi saat ini berkembang begitu pesat . Salah satunya penerapan teknologi yang dapat diterapkan didunia industri yaitu untuk evaluasi terhadap kinerja perusahaan . Evaluasi me," vol. V, no. 2, pp. 199–204, 2019.

[7] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustring dalam Penetuan Siswa Kelas Unggulan," vol. 15, no. 2, pp. 25–36, 2020.

[8] V. Herlinda and D. Darwis, "Analisis Clustering Untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means," *Darwis, Dartono*, vol. 2, no. 2, pp. 94–99, 2021, [Online]. Available: http://jim.teknokrat.ac.id/index.php/JTSI

[9] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *J. Sains dan Manaj.*, vol. 9, no. 1, pp. 95–100, 2021, [Online]. Available: www.bps.go.id

[10] A. Supriyadi, A. Triayudi, and I. D. Sholihati, "Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 6, no. 2, pp. 229–240, 2021, doi: 10.29100/jipi.v6i2.2008.