Idealis: Indonesia Journal Information System
Volume 9, Nomor 2, Juli 2026
ISSN 2684-7280 (*online*)
Halaman 440-449
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

Application of K-Means Clustering to Group Deforestation Areas in Indonesia

Nasrul Mahruf Aznawi[1*], Abdul Halim Hasugian[2]

[1,2]Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
Email: [1*]nasrulmahrufaznawi@gmail.com, [2]abdulhalimhasugian@uinsu.ac.id
(*: *corresponding author*)

Abstract - *Deforestation remains one of the major environmental challenges in Indonesia because the rate of tree cover loss varies considerably among regions, making it difficult for stakeholders to identify priority areas for forest monitoring and management using conventional descriptive analysis. This study aims to identify spatial patterns of deforestation by clustering Indonesian districts/cities based on multi-year Tree Cover Loss and to evaluate the effectiveness of the K-Means clustering algorithm for supporting data-driven environmental analysis. The dataset was obtained from Global Forest Watch (GFW) and consists of Tree Cover Loss data for 402 districts/cities in Indonesia during 2023–2025, represented by three numerical attributes measured in hectares. The research methodology includes data cleaning, attribute selection, Min-Max normalization, determination of the optimal number of clusters using the Elbow Method, K-Means clustering, and cluster evaluation using the Davies–Bouldin Index (DBI). Experimental results show that the optimal number of clusters is three, producing 333 districts (82.84%) in the low-loss cluster, 61 districts (15.17%) in the moderate-loss cluster, and 8 districts (1.99%) in the high-loss cluster, with a DBI value of 0.6599, indicating good clustering quality. The findings reveal that tree cover loss is unevenly distributed across Indonesia and provide a data-driven regional categorization that can support priority setting for forest monitoring and conservation. The scientific contribution of this study lies in utilizing multi-year Tree Cover Loss data (2023–2025) at the district/city level across Indonesia to characterize regional deforestation patterns using K-Means clustering.*

Keywords: *Clustering, Deforestation, K-Means, Regions, Data Mining.*

Abstrak - Deforestasi masih menjadi salah satu permasalahan lingkungan utama di Indonesia karena tingkat kehilangan tutupan pohon berbeda secara signifikan antarwilayah sehingga menyulitkan penentuan prioritas pengawasan dan pengelolaan hutan apabila hanya menggunakan analisis deskriptif. Penelitian ini bertujuan mengidentifikasi pola spasial deforestasi melalui pengelompokan kabupaten/kota di Indonesia berdasarkan hilangnya tutupan pohon multi tahun serta mengevaluasi efektivitas algoritma K-Means sebagai pendekatan analisis lingkungan berbasis data. *Dataset* berasal dari Global Forest Watch (GFW) yang terdiri atas data *Tree Cover Loss* (Hilangnya tutupan pohon) sebanyak 402 kabupaten/kota di Indonesia selama periode 2023–2025 dengan tiga atribut numerik berupa luas kehilangan tutupan pohon (hektar). Tahapan penelitian meliputi *data cleaning*, seleksi atribut, normalisasi Min-Max, penentuan jumlah *cluster* optimal menggunakan Elbow Method, proses *clustering* menggunakan K-Means, serta evaluasi menggunakan Davies–Bouldin Index (DBI). Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah tiga *cluster*, terdiri atas 333 kabupaten/kota (82,84%) pada *cluster* kehilangan rendah, 61 kabupaten/kota (15,17%) pada *cluster* kehilangan sedang, dan 8 kabupaten/kota (1,99%) pada *cluster* kehilangan tinggi dengan nilai DBI sebesar 0,6599 yang menunjukkan kualitas pengelompokan yang baik. Hasil penelitian mengungkap bahwa pola kehilangan tutupan pohon di Indonesia tidak tersebar secara merata serta menghasilkan kategorisasi wilayah berbasis data yang dapat mendukung penentuan prioritas pengawasan dan konservasi hutan. Kontribusi ilmiah penelitian ini terletak pada pemanfaatan data *Tree Cover Loss* multi tahun (2023–2025) pada tingkat kabupaten/kota di seluruh Indonesia untuk mengidentifikasi karakteristik pola deforestasi menggunakan algoritma K-Means.

Kata Kunci: *Cluster*ing, Deforestasi, K-Means, Wilayah, Data Mining.

1. PENDAHULUAN

Deforestasi merupakan proses perubahan kawasan hutan menjadi lahan non-hutan[1] yang disebabkan oleh aktivitas manusia maupun faktor alam. Fenomena ini menjadi salah satu permasalahan lingkungan global karena berpengaruh terhadap keseimbangan ekosistem, keanekaragaman hayati, perubahan iklim, serta keberlanjutan kehidupan manusia[2]. Hilangnya tutupan hutan juga menurunkan kesanggupan lingkungan dalam menyerap karbon sehingga mempercepat peningkatan emisi gas rumah kaca dan memperbesar risiko terjadinya perubahan iklim. Selain itu, deforestasi dapat memicu berbagai bencana hidrometeorologi, seperti banjir dan tanah longsor, akibat berkurangnya kemampuan tanah dalam menyerap air dan menahan erosi[3].

Indonesia adalah satu dari negara yang memiliki kawasan hutan tropis terluas nomor tiga di dunia[4] sehingga memiliki peranan penting dalam menjaga keseimbangan ekosistem global. Namun, Indonesia masih menghadapi kehilangan tutupan hutan yang terjadi di berbagai wilayah setiap tahunnya. Berdasar pada laporan dari Kementerian Kehutanan, luas deforestasi Indonesia pada tahun 2024 mencapai sekitar 175,4 ribu hektar. Besarnya angka tersebut menunjukkan bahwa deforestasi masih menjadi persoalan yang memerlukan perhatian serius. Meskipun begitu, tingkat kehilangan tutupan hutan tidak terjadi secara merata pada setiap wilayah. Beberapa kabupaten menunjukkan kehilangan tutupan hutan yang sangat tinggi, sedangkan wilayah lainnya

Idealis: Indonesia Journal Information System
Volume 9, Nomor 2, Juli 2026
ISSN 2684-7280 (*online*)
Halaman 440-449
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

memiliki tingkat kehilangan yang relatif rendah[5]. Perbedaan tersebut menyebabkan perlunya pengelompokan wilayah agar pola deforestasi di Indonesia dapat dipahami secara lebih sistematis.

Pengelompokan wilayah berdasarkan kehilangan tutupan hutan dapat membantu pemerintah maupun pemegang kepentingan dalam menentukan prioritas pengawasan, konservasi, serta penyusunan kebijakan pengelolaan hutan. Dengan mengetahui wilayah-wilayah yang memiliki deforestasi serupa, upaya pencegahan dapat dilakukan secara lebih terarah sesuai kondisi masing-masing daerah. Pendekatan menggunakan data juga memberikan informasi yang lebih terarah dibandingkan penilaian yang hanya didasarkan pada pengamatan deskriptif.

Diantara pendekatan yang layak dipakai guna menjalankan pengelompokan data adalah *data mining*. *Data mining* adalah elemen proses *Knowledge Discovery in Databases* (KDD) yang berfokus mengekstraksi pola atau informasi yang bermanfaat dari kumpulan data berukuran besar[6]. Salah satu teknik dalam *data mining* adalah *clustering*, yaitu proses pengelompokan data tanpa label berdasarkan tingkat kemiripan karakteristik antar objek[7]. Di antara berbagai algoritma *clustering*, K-Means ialah satu dari metode yang sering digunakan karena mempunyai proses komputasi yang mudah, mampu menangani data numerik dalam skala besar, serta menghasilkan pengelompokan berdasarkan kedekatan setiap data terhadap pusat *cluster* (*centroid*)[8].

Berbagai riset lampau telah mengaplikasikan algoritma K-Means pada beragam permasalahan pengelompokan wilayah, di antaranya sebagai berikut. Pada penelitian sebelumnya [9] membahas terkait *cluster*ing pulau yang menghasilkan kelapa sawit berdasar pada luas areal, produksi, dan tenaga kerja dengan memakai K-Means berhasil menunjukkan pemisahan wilayah, bahwa diperoleh jumlah klaster yang optimal sebanyak k=2. Penelitian [10] dalam mengelompokkan wilayah provinsi di Indonesia berdasar pada karakteristik kemiskinan serta juga pembangunan menggunakan K-Means juga berhasil mengelompokkan provinsi kedalam 4 klaster yang mempunyai karakteristik sosial dan ekonomi yang berbeda. Penelitian [11] menyampaikan bahwa algoritma K-Means juga efektif dipakai untuk pengelompokan data wilayah dengan tingkat kerawanan kriminalitas. Hasil riset-riset itu membuktikan bahwasanya K-Means sanggup menghasilkan kelompok data yang memiliki karakteristik serupa.

Berbagai riset terdahulu telah memperlihatkan bahwa algoritma K-Means berhasil guna dipakai untuk pengelompokan data wilayah pada berbagai bidang, seperti pengelompokan daerah penghasil kelapa sawit, wilayah berdasar pada indikator sosial ekonomi, maupun tingkat kriminalitas. Namun, sebagian besar penelitian tersebut menggunakan objek penelitian yang berbeda dengan fokus pada indikator sosial, ekonomi, pertanian, atau kriminalitas, sehingga belum mampu menggambarkan karakteristik spasial deforestasi berdasarkan data kehilangan tutupan hutan. Di sisi lain, penelitian mengenai deforestasi umumnya lebih menitikberatkan pada identifikasi faktor penyebab, analisis perubahan tutupan lahan, maupun pemetaan menggunakan citra satelit, sementara pendekatan *data mining* terutama *clustering* untuk mengelompokkan kabupaten/kota yang berdasar pada deforestasi masih sangat terbatas. Selain itu, riset terdahulu umumnya menggunakan data pada satu periode pengamatan atau cakupan wilayah tertentu sehingga belum memberikan gambaran pola kehilangan tutupan hutan secara nasional menggunakan data multi tahun.

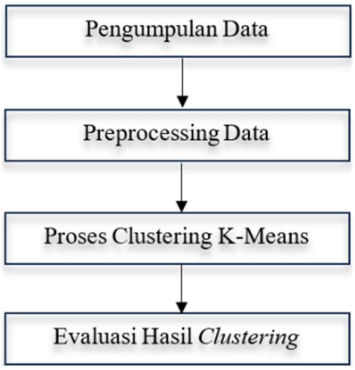
Berdasarkan kondisi tersebut terdapat kesenjangan penelitian, yaitu belum adanya penelitian yang mengelompokkan seluruh kabupaten/kota di Indonesia berdasarkan deforestasi tahun 2023–2025 memakai algoritma K-Means juga menilai kualitas *cluster* dengan memakai kombinasi *Elbow Method* dan Davies-Bouldin Index. Kebaruan penelitian ini terletak pada penggunaan data deforestasi tiga tahun terakhir (2023–2025) pada tingkat kabupaten/kota di seluruh Indonesia, yang belum banyak dikaji pada penelitian-penelitian pengelompokan wilayah sebelumnya. Berdasarkan permasalahan yang tertera, riset ini mengaplikasikan algoritma K-Means *Clustering* guna mengelompokkan kabupaten/kota di Indonesia berdasar pada data *tree cover loss* tahun 2023–2025 yang diperoleh dari Global Forest Watch [12]. Sebelum proses *clustering* dijalankan, data dinormalisasi memakai metode *Min-Max Scaling* untuk menyamakan rentang nilai antar variabel. Selanjutnya, jumlah *cluster* optimal ditentukan menggunakan *Elbow Method*, sementara kualitas hasil pengelompokan dievaluasi dengan Davies–Bouldin Index (DBI). Hasil riset diinginkan dapat memberikan informasi mengenai tingkat kehilangan tutupan hutan pada setiap kelompok wilayah sehingga dapat menjadi salah satu referensi dalam mendukung penentuan prioritas pengawasan dan pengelolaan hutan di Indonesia.

2. METODE PENELITIAN

2.1. Tahapan Penelitian

Tahapan penelitian yaitu alur langkah penelitian yang dijalankan bermulai dari langkah awal hingga diperoleh hasil akhir. Tahapan dilaksanakan dalam penelitian ini diawali dari pengumpulan data, *preprocessing* data, penentuan jumlah *cluster*, proses *cluster*ing memakai algoritma K-Means, evaluasi hasil *cluster*ing, hingga analisis karakteristik hasil pengelompokkan. Dalam penelitian ini tahap penelitian sepenuhnya terdapat pada Gambar 1.

Idealis: Indonesia Journal Information System
Volume 9, Nomor 2, Juli 2026
ISSN 2684-7280 (*online*)
Halaman 440-449
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index



Gambar 1. Tahapan Penelitian

2.2. Dataset Penelitian

Dataset yang dipakai dalam riset ini yakni data sekunder yang didapat dari Global Forest Watch (GFW)[12]. *Dataset* ini bersifat publik dan dapat diakses oleh pihak manapun. *Dataset* berisi informasi *Tree Cover Loss* yang merupakan luas kehilangan tutupan pohon (*tree canopy*) yang direkam oleh Global Forest Watch menggunakan citra satelit Landsat melalui algoritma deteksi perubahan tutupan lahan. Nilai *Tree Cover Loss* dinyatakan dalam satuan hektar dan menunjukkan luas area yang mengalami kehilangan tutupan pohon pada masing-masing tahun pengamatan. Data yang digunakan merupakan data dalam tiga tahun terakhir yaitu tahun 2023, 2024, dan 2025 dengan jumlah sebanyak 402 kabupaten/kota. Setiap data terdiri atas tiga atribut dengan nilai numerik yang melambangkan luas kehilangan tutupan pohon dalam satuan hektar pada masing-masing tahun pengamatan. Ringkasan dari *dataset* disajikan pada Tabel 1.

Tabel 1. Karakteristik *Dataset*

Komponen	Keterangan
Sumber Data	Global Forest Watch
Jumlah Data	402 Kabupaten/Kota
Variabel	*Tree Cover Loss* 2023 *Tree Cover Loss* 2024 *Tree Cover Loss* 2025
Tipe Data	Numerik
Satuan	Hektar

2.3. *Preprocessing* Data

Preprocessing dilaksanakan untuk menghasilkan *dataset* yang siap digunakan pada proses *cluster*ing. Tahapan ini bermaksud meningkatkan kualitas data sehingga hasil pengelompokan menjadi lebih representatif. Proses *preprocessing* terdiri atas *data cleaning*, seleksi atribut, dan normalisasi data[13].

2.4. *Data Cleaning* dan Seleksi Atribut

Tahap ini dijalankan pemeriksaan pada *dataset* untuk menjamin tidak terdapat data yang duplikat maupun atribut yang tidak digunakan dalam penelitian. Selanjutnya dilakukan seleksi atribut sehingga hanya tiga variabel numerik yang digunakan, yaitu *Tree Cover Loss* 2023, *Tree Cover Loss* 2024, *dan Tree Cover Loss* 2025. Ketiga variabel tersebut dipilih karena secara langsung merepresentasikan tingkat kehilangan tutupan pohon pada masing-masing wilayah.

2.5. Normalisasi Data

Normalisasi dilakukan dengan memakai metode *Min-Max Scaling*[14] hingga seluruh atribut mempunyai jarak nilai antara 0 hingga 1. Normalisasi bertujuan mencegah atribut yang memiliki nilai lebih besar mendominasi proses pembentukan *cluster*. Persamaan normalisasi *Min-Max* ditunjukkan pada persamaan (1).

$$X' \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (1)$$

Idealis: Indonesia Journal Information System
Volume 9, Nomor 2, Juli 2026
ISSN 2684-7280 (*online*)
Halaman 440-449
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

2.6. Penentuan Jumlah *Cluster*

Sebelum proses *cluster*ing dijalankan, ditentukan lebih dahulu jumlah *cluster* (k) yang optimal memakai *Elbow Method*[15]. Metode tersebut bekerja dengan menghitungkan nilai *Within Cluster Sum of Squares* (WCSS)[16] pada sejumlah alternatif jumlah *cluster*. Nilai k dipilih berdasarkan titik siku (*elbow point*) yang menunjukkan bahwasanya kenaikan jumlah *cluster* berikutnya tak lagi memberi hasil penurunan nilai WCSS yang signifikan. Tahap ini bermaksud memperolehkan jumlah *cluster* yang sanggup merepresentasikan karakteristik data secara optimal.

2.7. Proses *Cluster*ing Menggunakan Algoritma K-Means

Langkah *cluster*ing dibuat dengan memakai algoritma K-Means. Setelah jumlah *cluster* optimum diperoleh, algoritma diawali dengan menetapkan *centroid* awal secara *random*. Berikutnya dihitung jarak dari tiap data atas seluruh *centroid* memakai *Euclidean Distance*[17]. Setiap data lalu diletakkan pada *cluster* yang mempunyai jarak paling dekat terhadap *centroid*. Setelah seluruh data memperoleh *cluster* sementara, *centroid* baru dihitungkan menurut rata-rata anggota pada tiap *cluster*. Proses tersebut diterapkan dengan berulang sampai tiada pergantian keanggotaan *cluster* atau sudah mendapati keadaan konvergen. Persamaan *Euclidean Distance* dinyatakan oleh Persamaan (2).

$$d_{ij} = \sqrt{\sum_{k=1}^{n}(v_{ik} - v_{jk})^2} \qquad (2)$$

2.8. Evaluasi

Kualitas hasil *cluster*ing dievaluasi dengan memakai Davies-Bouldin Index (DBI)[18]. Metode ini dipakai untuk menilai kualitas *cluster* berdasar pada tingkat kedekatan anggota didalam satu *cluster* (*cohesion*) dan tingkat keterpisahan antar *cluster* (*separation*). Semakin kecil nilai DBI memperlihatkan bahwa kualitas *cluster* yang semakin bagus dikarena memiliki tingkat kesamaan yang tinggi di dalam *cluster* dan variasi yang jelas antar*cluster*.

2.9. Desain Eksperimen

Eksperimen pada penelitian ini dibuat secara bertahap supaya proses pembentukan *cluster* dapat diamati dengan sistematis. Tahapan eksperimen dimulai dari *preprocessing dataset*, normalisasi memakai *Min-Max Scaling*, penetapan jumlah *cluster* optimal memakai *Elbow Method*, demonstrasi mekanisme algoritma K-Means melalui perhitungan manual terhadap sebagian data sampel, kemudian implementasi algoritma pada seluruh *dataset* dengan memakai Google Colaboratory dengan bahasa pemrograman Python dan pustaka *scikit-learn*. Pada tahap ini juga, jumlah cluster tidak ditetapkan secara langsung, melainkan ditentukan melalui pengujian beberapa alternatif nilai k pada rentang 2 sampai 10 memakai Elbow Method. Nilai k yang dipilih adalah nilai yang menunjukkan titik elbow pada kurva WCSS. Hasil *cluster*ing selanjutnya dievaluasi dengan Davies-Bouldin Index (DBI) dan dianalisis berdasarkan karakteristik masing-masing *cluster*. Skenario eksperimen secara ringkas ditampilkan pada Tabel 2.

Tabel 2. Skenario Eksperimen

Tahapan	Metode/Parameter
Dataset	Global Forest Watch
Jumlah Data	402 Kabupaten/Kota
Variabel	Tree Cover Loss 2023 – 2025
Preprocessing	Cleaning, seleksi atribut, Min-Max Normalization
Penentuan jumlah *cluster*	*Elbow Method*
Algoritma	K-Means
Implementasi	Google Colaboratory
Evaluasi	Davies-Bouldin Index

2.10. Lingkungan Implementasi Eksperimen

Seluruh langkah eksperimen pada riset kali ini dilakukan memakai Google Colaboratory dengan bahasa pemrograman Python. *Platform* tersebut digunakan sebab menyiapkan lingkungan komputasi berbasis *cloud* yang mendukung analisis data dan implementasi algoritma *machine learning* tanpa memerlukan instalasi perangkat lunak tambahan. Implementasi penelitian memanfaatkan beberapa pustaka (*library*) Python untuk

Idealis: Indonesia Journal Information System
Volume 9, Nomor 2, Juli 2026
ISSN 2684-7280 (*online*)
Halaman 440-449
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

mendukung setiap tahapan eksperimen, mulai dari pembacaan *dataset*, *preprocessing*, proses *cluster*ing, visualisasi hasil, hingga evaluasi kualitas *cluster*. *Library* yang dipakai disajikan pada Tabel 3.

Tabel 3. *Library* yang digunakan

Library/Modul	Fungsi
Pandas	Membaca, mengelola, dan memanipulasi dataset
NumPy	Operasi numerik dan pengolahan array
Matplotlib.pyplot	Membuat grafik Elbow, scatter plot, dan visualisasi *cluster*
Sklearn.preprocessing.MinMaxScaler	Normalisasi data memakai metode Min-Max Scaling
Sklearn.*cluster*.KMeans	Menerapkan algoritma K-Means *Cluster*ing
Sklearn.metrics.davies_bouldin_score	Menghitung nilai Davies–Bouldin Index sebagai evaluasi kualitas *cluster*

Library pandas digunakan untuk membaca *dataset* berformat Microsoft Excel serta melakukan pengelolaan data sebelum proses analisis. *Library* NumPy dimanfaatkan dalam proses komputasi numerik, sedangkan Matplotlib digunakan untuk menghasilkan visualisasi berupa grafik *Elbow Method*, sebaran data, hasil *cluster*ing, dan posisi *centroid*. Proses normalisasi dilakukan menggunakan kelas *MinMaxScaler* dari modul *sklearn.preprocessing*, sedangkan proses *cluster*ing menggunakan kelas KMeans dari modul *sklearn.cluster*. Kualitas hasil *cluster*ing dievaluasi memakai fungsi *davies_bouldin_score* dari modul *sklearn.metrics*.

3. HASIL DAN PEMBAHASAN

3.1. Karakteristik *Dataset*

Dataset memuat tiga atribut numerik, ketiga atribut dari *dataset* tersebut dipergunakan sebagai fondasi dalam langkah pengelompokan wilayah dengan algoritma K-Means *Clustering*. Karakteristik *dataset* penelitian ditampilkan pada Tabel 4.

Tabel 4. Data Penelitian

Provinsi	Kabupaten	Tc_loss_ha_2023	Tc_loss_ha_2024	Tc_loss_ha_2025
Aceh	Aceh Barat	601	719	1041
Aceh	Aceh Barat Daya	489	520	448
Aceh	Banda Aceh	0	0	0
Aceh	Aceh Tamiang	70	94	128
Aceh	Aceh Selatan	1995	1783	2315
...	...	...	...	...
Sumatera Utara	Toba Samosir	77	222	360

Berdasarkan Tabel 4 dapat dilihat bahwa rentang nilai kehilangan tutupan pohon disetiap wilayah sangat bervariasi. Sebagian wilayah memiliki nilai kehilangan tutupan pohon mendekati nol bahkan nol, sedangkan beberapa wilayah lainnya memiliki nilai yang jauh lebih tinggi. Perbedaan rentang nilai tersebut dapat memengaruhi proses *clustering* sehingga diperlukan tahapan *preprocessing* yang berupa normalisasi data sebelum dilakukan proses pengelompokan.

3.2. Hasil *Preprocessing*

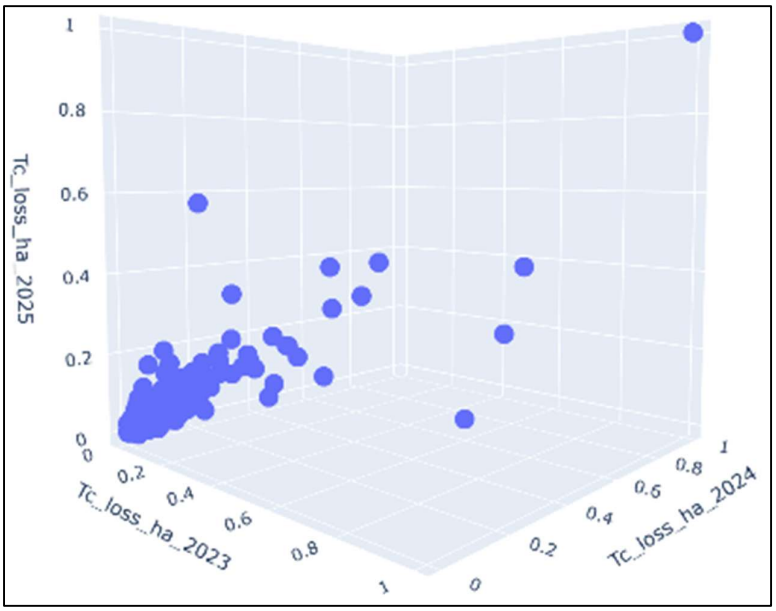
Sebelum dilaksanakan langkah *clustering*, *dataset* lebih dahulu melalui tahap *preprocessing* yang bertujuan menghasilkan data yang siap dipakai dalam proses analisis. Tahapan *preprocessing* meliputi pemeriksaan data, seleksi atribut, serta normalisasi memakai metode Min-Max *Scaling*. Pada penelitian ini digunakan tiga atribut numerik. Seluruh atribut kemudian dinormalisasi sehingga mempunyai rentang nilai antara 0 sampai 1.

Tabel 5. Normalisasi Data

Provinsi	Kabupaten	Tc_loss_ha_2023	Tc_loss_ha_2024	Tc_loss_ha_2025
Aceh	Aceh Barat	0,039209	0,049463	0,061492
Aceh	Aceh Barat Daya	0,031902	0,035773	0,026463
Aceh	Banda Aceh	0,000000	0,000000	0,000000
Aceh	Aceh Tamiang	0,004567	0,006467	0,007561
Aceh	Aceh Selatan	0,130154	0,122661	0,136748
…	…	…	…	…
Sumatera Utara	Toba Samosir	0,005023	0,015272	0,021265

Idealis: Indonesia Journal Information System
Volume 9, Nomor 2, Juli 2026
ISSN 2684-7280 (*online*)
Halaman 440-449
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

Hasil normalisasi pada Tabel 5 terlihat bahwa keseluruhan atribut sudah berada pada rentang nilai yang sama sehingga tiap atribut mempunyai peran yang seimbang dalam proses pembentukan *cluster*. Dengan demikian, proses *clustering* tidak dipengaruhi oleh perbedaan skala nilai antar atribut.



Gambar 2. Sebaran data penelitian

Sebaran data hasil normalisasi pada Gambar 2 divisualisasikan menggunakan *scatter plot* untuk melihat distribusi data sebelum dilakukan proses *clustering*. Berdasarkan Gambar 2 terlihat bahwa data masih tersebar tanpa adanya batas kelompok yang jelas sehingga diperlukan algoritma K-Means untuk mengidentifikasi kelompok data yang punya karakteristik serupa.

3.3. Penentuan Jumlah Cluster

Sebelum algoritma K-Means diterapkan, lebih dulu dilakukannya penetapan jumlah *cluster* ideal memakai *Elbow Method*. Penetapan jumlah *cluster* dijalankan dengan menghitungkan nilai *Within Cluster Sum of Squares* (WCSS) pada rentang nilai k = 1 hingga k = 10. Nilai WCSS menunjukkan jumlah kuadrat jarak setiap data atas *centroid* pada tiap-tiap *cluster* sehingga makin kecil nilai WCSS menunjukkan semakin kompak anggota di dalam *cluster*. Nilai WCSS dihitung menggunakan Persamaan (3).

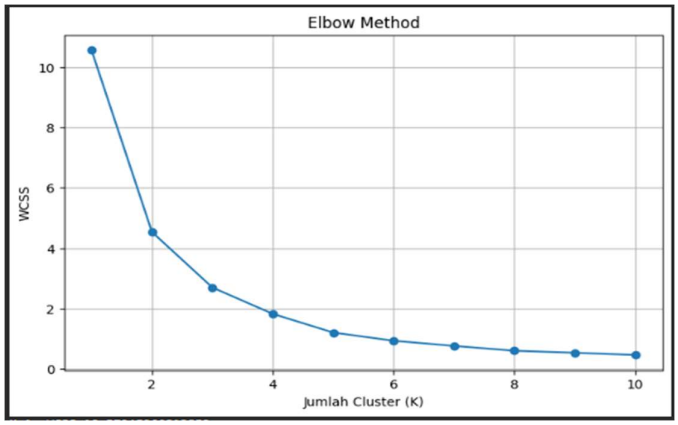
$$WCSS = \sum_{i=1}^{k} \sum_{x \in C_i} ||x - c_i||^2 \qquad (3)$$

Berdasarkan Tabel 6 dapat dilihat bahwa telah terjadi pengurangan nilai WCSS yang lumayan besar dari k=1 ke k=2, yaitu sebesar 6,0370. Selanjutnya dari k=2 ke k=3 nilai WCSS kembali menurun sebesar 1,8430. Setelah jumlah *cluster* mencapai k=3, penurunan WCSS mulai melandai. Misalnya dari k=3 ke k=4 hanya menurun sebesar 0,8750, kemudian semakin kecil pada nilai k berikutnya.

Tabel 6. Tabel Nilai WCSS

K	WCSS	Penurunan	Persentase
1	10,5725	-	-
2	4,5354	6,0370	57,10%
3	2,6924	1,8430	40,64%
4	1,8174	0,8750	32,50%
5	1,1969	0,6205	34,14%
6	0,9248	0,2722	22,74%
7	0,7524	0,1723	18,63%
8	0,5932	0,1592	21,16%
9	0,5252	0,0680	11,46%
10	0,4534	0,0719	13,69%

Idealis: Indonesia Journal Information System
Volume 9, Nomor 2, Juli 2026
ISSN 2684-7280 (*online*)
Halaman 440-449
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index



Gambar 3. *Elbow Method*

Berdasarkan Gambar 3 dapat dilihat bahwa turunnya nilai WCSS berlangsung sangat tajam pada perubahan jumlah *cluster* dari k=1 hingga k=3. Setelah nilai k=3, penurunan WCSS mulai melandai sehingga membentuk titik siku (*elbow point*). Penurunan WCSS dari k=1 ke k=2 mencapai 57,10%, sedangkan dari k=2 ke k=3 masih sebesar 40,64%. Setelah k=3, penurunan WCSS menurun secara signifikan menjadi hanya 32,50% pada k=4 dan terus mengecil pada jumlah *cluster* berikutnya. Kondisi ini membuktikan bahwa peningkatan *cluster* setelah k=3 hanya memberikan peningkatan yang relatif kecil sehingga k=3 dipilih sebagai jumlah *cluster* optimal. Berdasarkan hasil tersebut, jumlah *cluster* yang dipakai pada riset ini ditetapkan sebanyak tiga *cluster* (k=3) sebab dinilai telah sanggup menggambarkan struktur data dengan baik tanpa menambah kompleksitas model.

3.4. Demonstrasi Perhitungan Manual K-Means

Untuk memudahkan pemahaman terhadap mekanisme algoritma K-Means, dilakukan demonstrasi perhitungan manual memakai 15 data sampel yang dipilih dari keseluruhan *dataset* penelitian. Demonstrasi ini hanya bertujuan menjelaskan tahapan kerja algoritma, sedangkan proses *clustering* utama tetap dilakukan terhadap seluruh *dataset* penelitian yang terdiri atas 402 kabupaten/kota. Tahap pertama dilakukan dengan menentukan *centroid* awal secara acak sebagaimana diperlihatkan oleh Tabel 7.

Tabel 7. *Centroid* awal

Cluster	X1	X2	X3
C1	0,039209	0,049463	0,061492
C2	0,049843	0,091635	0,169473
C3	0,000000	0,000000	0,000000

Label C1, C2, dan C3 pada Tabel 7 hanya digunakan sebagai penanda sementara untuk keperluan demonstrasi perhitungan manual sehingga penomorannya dimulai dari satu, dan tidak selalu berkorespondensi langsung dengan penomoran *cluster* hasil implementasi algoritma pada seluruh *dataset* yang menggunakan indeks berbasis nol (0, 1, 2) sebagaimana dihasilkan oleh pustaka *scikit-learn*. Setelah ditentukan *centroid* awal, langkah selanjutnya perhitungan iterasi 1 dengan menghitung jarak *Euclidean Distance* antar setiap data dengan nilai *centroid* pada Tabel 6 menggunakan persamaan *Euclidean Distance*.

Perhitungan jarak *centroid* pada data pertama terhadap C1

$$C1 = \sqrt{(0{,}039209 - 0{,}039209)^2 + (0{,}049463 - 0{,}049463)^2 + (0{,}061492 - 0{,}061492)^2} = 0$$

Perhitungan jarak *centroid* pada data pertama terhadap C2

$$C2 = \sqrt{(0{,}039209 - 0{,}049843)^2 + (0{,}049463 - 0{,}091635)^2 + (0{,}061492 - 0{,}169473)^2} = 0{,}116414$$

Perhitungan jarak *centroid* pada data pertama terhadap C3

$$C3 = \sqrt{(0{,}039209 - 0)^2 + (0{,}049463 - 0)^2 + (0{,}061492 - 0)^2} = 0{,}088124$$

Perhitungan jarak dikerjakan terhadap seluruh data sampel terhadap tiap *centroid*. Hasil perhitungan diatas, diperoleh jarak antara data pertama terhadap masing-masing *centroid*. Dapat dilihat bahwa nilai terkecil ditunjukkan oleh jarak data pertama terhadap C1 yang mana artinya data pertama lebih dekat jaraknya terhadap

Idealis: Indonesia Journal Information System
Volume 9, Nomor 2, Juli 2026
ISSN 2684-7280 (*online*)
Halaman 440-449
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

C1 atau *cluster* satu. Hal tersebut juga menunjukkan bahwa diantara ketiga *centroid* yang ada, data pertama mempunyai tingkat kesamaan yang lebih tinggi dengan C1 sehingga data pertama merupakan anggota *cluster* satu. Langkah berikutnya yakni menentukan *centroid* baru melalui cara hitungkan rata-rata dari setiap data yang ada dalam *cluster* di iterasi pertama. Kemudian setelah didapatkan *centroid* baru, lanjut menghitung jarak euclidean dan *cluster* ditetapkan kembali sama seperti pada iterasi pertama. Iterasi selesai dilakukan ketika sudah tidak terjadi perubahan *cluster* atau sudah menggapai kondisi konvergen.

3.5. Implementasi Algoritma pada Seluruh *Dataset*

Setelah mekanisme algoritma dijelaskan melalui demonstrasi perhitungan manual, algoritma K-Means selanjutnya diterapkan pada seluruh *dataset* penelitian sebanyak 402 kabupaten/kota memakai Google Colaboratory dengan bahasa pemrograman Python dan pustaka *scikit-learn. Dataset* yang dipakai ialah hasil *preprocessing* dan normalisasi menggunakan *Min-Max Scaling*, sedangkan jumlah *cluster* ditetapkan sebanyak tiga *cluster* berdasarkan hasil *Elbow Method*. Proses *clustering* dilakukan secara iteratif hingga *centroid* mencapai kondisi stabil dan tidak terjadi lagi perubahan keanggotaan *cluster*.

3.6. Hasil *Clustering*

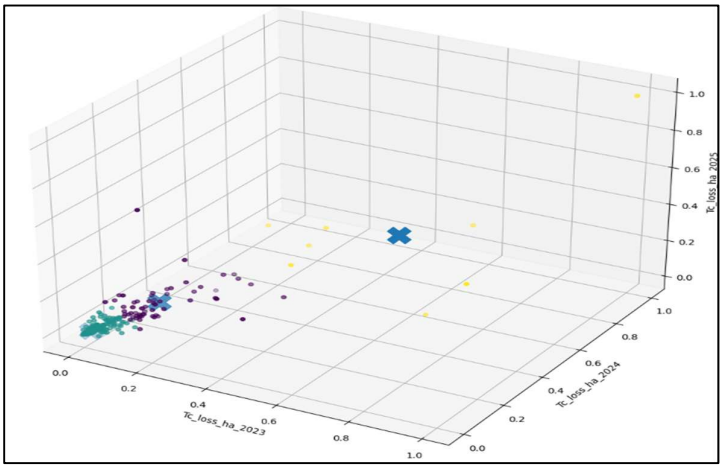
Hasil *clustering* pada Tabel 8 memperlihatkan bahwa seluruh data sukses dikelompokkan ke dalam tiga *cluster* berdasarkan karakteristik kehilangan tutupan pohon. Distribusi anggota pada tiap-tiap *cluster* menyatakan adanya perbedaan tingkat deforestasi antar wilayah.

Tabel 8. Persentase Distribusi Wilayah

Cluster	Jumlah Wilayah	Persentase
0	61	15,17%
1	333	82,84%
2	8	1,99%

Berdasarkan Tabel 8 *cluster* 1 merupakan kelompok terbesar yang terdiri atas 333 wilayah dengan kehilangan tutupan pohon relatif rendah, beberapa wilayah yang masuk kedalam *cluster* 1 yakni Aceh Barat, Aceh Besar, Tapanuli Tengah, Toba Samosir dan lainnya. *Cluster* 0 terdiri atas 61 wilayah yang menunjukkan tingkat kehilangan tutupan pohon sedang, beberapa wilayahnya meliputi Aceh Utara, Bengkulu Utara, Musi Banyuasin, Solok Selatan, Mandailing Natal dan lainnya, Sedangkan *cluster* 2 hanya terdiri atas 8 wilayah tetapi memiliki tingkat kehilangan tutupan pohon paling tinggi sehingga terpisah cukup jauh dari kelompok lainnya, wilayah *cluster* 2 meliputi Kapuas Hulu, Kayong Utara, Ketapang, Kapuas, Katingan, Berau, Kutai Barat dan Morowali.

3.7. Analisis Karakteristik *Cluster*



Gambar 4. Visualisasi *cluster* dan *centroid*

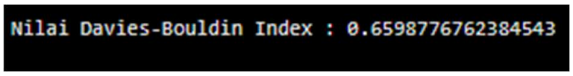
Visualisasi *centroid* dari Gambar 4 menunjukkan adanya perbedaan karakteristik antar *cluster*. Posisi *centroid* yang semakin jauh dari titik nol menunjukkan semakin besarnya nilai kehilangan tutupan pohon. *Cluster* 1 memiliki *centroid* yang paling dekat dengan titik nol sehingga merepresentasikan wilayah dengan

Idealis: Indonesia Journal Information System
Volume 9, Nomor 2, Juli 2026
ISSN 2684-7280 (*online*)
Halaman 440-449
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

tingkat kehilangan tutupan pohon rendah. Sebaliknya, *Cluster* 2 memiliki *centroid* yang paling jauh sehingga mencerminkan daerah dengan tingkat kehilangan tutupan pohon tinggi.

Hasil *clustering* memperlihatkan bahwa sebagian besar kabupaten/kota di Indonesia berada pada kelompok dengan tingkat kehilangan tutupan pohon rendah, sedangkan hanya sebagian kecil wilayah termasuk dalam kelompok kehilangan tinggi. Informasi tersebut dapat dimanfaatkan pemerintah sebagai dasar penyusunan prioritas pengawasan hutan secara lebih efisien. Wilayah pada *cluster* tinggi dapat menjadi prioritas utama dalam kegiatan *monitoring*, rehabilitasi hutan, maupun penegakan hukum terhadap aktivitas yang berpotensi menyebabkan deforestasi. Sementara itu, wilayah pada *cluster* sedang dapat difokuskan pada upaya pencegahan agar tidak mengalami peningkatan kehilangan tutupan hutan, sedangkan wilayah pada *cluster* rendah dapat dipertahankan melalui program konservasi yang berkelanjutan. Penelitian ini juga memberikan keterkaitan metodologis bahwa kombinasi *Min-Max Scaling, Elbow Method*, K-Means, dan Davies-Bouldin Index mampu menghasilkan pengelompokan wilayah pada data *Tree Cover Loss* berskala nasional.

3.8. Evaluasi Hasil *Clustering*



Gambar 5. Nilai DBI

Kualitas hasil *clustering* yang merupakan tingkat kekompakan anggota di dalam suatu *cluster* (*cohesion*) dan tingkat keterpisahan antar anggota *cluster* (*separation*) dievaluasi dengan memakai Davies-Bouldin Index (DBI). Berdasar pada hasil pengujian didapat nilai DBI sebesar 0,6599 pada Gambar 5. Nilai DBI yang berada di bawah angka 1 menandakan bahwasanya hasil *clustering* memiliki kualitas yang baik. Hal itu menandakan bahwa anggota pada setiap *cluster* mempunyai tingkatan kesamaan yang tinggi (*cohesion*), sementara jarak antar *cluster* relatif cukup jauh (*separation*). Dengan begitu, proses pengelompokan memakai algoritma K-Means dapat membentuk *cluster* yang cukup mewakili terhadap karakteristik kehilangan tutupan pohon pada data penelitian.

Secara menyeluruh, rangkaian eksperimen mulai dari *preprocessing* data, penetapan jumlah *cluster* dengan *Elbow Method*, penerapan algoritma K-Means, hingga evaluasi memakai DBI membuktikan bahwa metode yang digunakan mampu memperoleh pengelompokan wilayah yang stabil dan dapat digunakan sebagai dasar analisis tingkat deforestasi di Indonesia.

4. KESIMPULAN

Riset ini sukses mengaplikasikan algoritma K-Means *Clustering* untuk mengelompokkan 402 kabupaten/kota di Indonesia berdasar pada data *Tree Cover Loss* tahun 2023–2025 yang diperoleh dari Global Forest Watch. Sebelum langkah *clustering* dijalankan, *dataset* awalnya melewati tahap *preprocessing* berbentuk seleksi atribut dan normalisasi dengan memakai *Min-Max Scaling*, kemudian jumlah *cluster* optimal ditentukan memakai *Elbow Method*, sehingga diperoleh nilai k = 3 sebagai jumlah *cluster* yang paling representatif.

Hasil implementasi algoritma K-Means memperlihatkan bahwa seluruh wilayah sukses dikelompokkan ke dalam tiga *cluster* dengan kehilangan tutupan pohon yang berbeda. *Cluster* 1 ialah kelompok terbesar yang beranggotakan 333 wilayah dengan tingkat kehilangan tutupan pohon relatif rendah. *Cluster* 0 terdiri atas 61 wilayah dengan kehilangan tutupan pohon sedang, sedangkan *Cluster* 2 beranggotakan 8 wilayah yang menunjukkan tingkat kehilangan tutupan pohon paling tinggi. Perbedaan tersebut memperlihatkan dimana tingkat hilangnya tutupan pohon di Indonesia tidak tersebar secara merata, sehingga setiap kelompok wilayah memerlukan perhatian dan strategi pengelolaan yang berbeda.

Evaluasi memakai Davies–Bouldin Index (DBI) memperoleh nilai 0,6599, yang mengisyaratkan bahwa hasil *clustering* mempunyai kualitas yang baik dikarenakan mampu membentuk kelompok dengan tingkat kesamaan yang tinggi di dalam *cluster* (*cohesion*) dan keterpisahan yang cukup jelas antarcluster (*separation*). Hasil ini menandakan penggunaan gabungan *Min-Max Scaling*, *Elbow Method*, dan K-Means *Clustering* berhasil mendapatkan pengelompokan yang representatif terhadap kehilangan tutupan pohon pada data penelitian.

Penelitian ini memberikan gambaran mengenai pola kehilangan tutupan pohon pada tingkat kabupaten/kota di Indonesia serta menunjukkan bahwa pendekatan *clustering* dapat digunakan sebagai salah satu metode pendukung dalam mengidentifikasi wilayah yang memiliki deforestasi serupa. Informasi yang dihasilkan diharapkan dapat menjadi referensi awal dalam mendukung proses pemantauan, penyusunan prioritas pengawasan, dan perencanaan pengelolaan hutan berbasis data. Kontribusi utama riset ini berada pada penerapan K-Means *Clustering* pada data *Tree Cover Loss* multi tahun (2023–2025) yang mencakup 402 kabupaten/kota di Indonesia, cakupan yang belum banyak dikaji pada penelitian pengelompokan wilayah sebelumnya. Penelitian

ini juga memiliki keterbatasan, yaitu belum dilakukannya evaluasi komparatif dengan algoritma *clustering* lain serta belum melibatkan variabel penyebab deforestasi di luar *Tree Cover Loss*.

Riset yang akan datang dapat menggunakan periode data yang lebih panjang serta menambahkan variabel lain yang berkaitan dengan deforestasi, seperti perubahan penggunaan lahan, kepadatan penduduk, curah hujan, atau aktivitas perkebunan, sehingga setiap *cluster* dapat dianalisis secara lebih mendalam. Selain itu, hasil pengelompokan dapat dibandingkan dengan algoritma *clustering* lain untuk memperoleh perbandingan performa berdasarkan nilai evaluasi yang berbeda.

DAFTAR PUSTAKA

[1] W. O. I. Febryanti, S. Adiningsi, and R. A. Saputra, "Menganalisis Pola Deforestasi Hutan Lindung Di Sulawesi Tenggara Menggunakan Metode K-Means," *JIP (Jurnal Inform. Polinema)*, vol. 10, no. 1, pp. 53–58, 2023, doi: https://doi.org/10.33795/jip.v10i1.1455.

[2] J. Markovic, M. Starke, J. Wilkes-allemann, and O. Wolf, "Drivers of deforestation and forest degradation between 1990 and 2023 - A global meta-analysis," *Environ. Sci. Policy*, vol. 173, no. 6, pp. 1–11, 2025, doi: 10.1016/j.envsci.2025.104242.

[3] F. A. Mahmud and A. Kurniawan, "Vegetasi Di Kota Batu Menggunakan Citra Satelit Multi-Temporal," *J. Penginderaan Jauh Indones.*, vol. 04, no. 01, pp. 1–10, 2025, doi: 10.12962/jpji.v4i1.3209.

[4] P. W. Widodo, "10 Negara Pemilik Hutan Tropis Terluas di Dunia," Internasional Kontan.co.id. [Online]. Available: https://internasional.kontan.co.id/news/10-negara-pemilik-hutan-tropis-terluas-di-dunia

[5] B. Nathania, A. A. Nasution, and B. Asmara, "Ketika 'Benteng Terakhir' Mulai Rapuh: Pergeseran Deforestasi ke Timur Indonesia," WRI Indonesia. [Online]. Available: https://wri-indonesia.org/id/wawasan/ketika-benteng-terakhir-mulai-rapuh-pergeseran-deforestasi-ke-timur-indonesia

[6] P. Mai, S. Tarigan, J. T. Hardinata, H. Qurniawan, M. Safii, and R. Winanjaya, "Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus : Toko Sinar Harahap)," *Just IT J. Sist. Informasi, Teknol. Inf. dan Komput.*, vol. 12, no. 2, pp. 51–61, 2022.

[7] G. Alasi and R. A. Putri, "Penerapan K-Means Clustering untuk Pengelompokan Pasien Rumah Sakit berdasarkan Tingkat Keparahan Penyakit," *Jatilima J. Multimed. Dan Teknol. Inf.*, vol. 07, no. 03, pp. 475–486, 2025.

[8] R. Anggari *et al.*, "Penerapan Algoritma K-Means untuk Pengelompokan Negara di Dunia Berdasarkan Indikator Ekonomi," *JURTI*, vol. 8, no. 2, pp. 195–204, 2024, doi: http://dx.doi.org/10.30872/jurti.v8i2.19745.

[9] G. Oktariani, "Klasterisasi Pulau Penghasil Kelapa Sawit di Indonesia Berdasarkan Luas Areal , Produksi dan Tenaga Kerja Menggunakan Metode K-Means Clustering," *J. Pendidik. Tambusai*, vol. 9, no. 1, pp. 8735–8744, 2025, doi: https://doi.org/10.31004/jptam.v9i1.25967.

[10] E. P. Permatasari, L. A. Iriani, and E. Widodo, "Identification of Indonesian Provinces Based on Socioeconomic Indicators in 2024 Using K-Means," *J. Sintak*, vol. 4, no. 2, pp. 57–70, 2026, doi: https://doi.org/10.62375/jsintak.v4i2.789.

[11] S. Sulistina, P. E. P. Utomo, and B. F. Hutabarat, "Klasterisasi Wilayah Rawan Kriminalitas Di Kota Jambi (2022-2024) Menggunakan Algoritma K-Means," *J. INSTEK Inform. Sains dan Teknol.*, vol. 10, no. 2, pp. 546–560, 2025, doi: https://doi.org/10.24252/instek.v10i2.62051.

[12] G. F. Watch, "Indonesia Dashboard," Global Forest Watch. [Online]. Available: https://www.globalforestwatch.org/dashboards/country/IDN/

[13] A. A. A. Fernandes, M. Koehler, N. Konstantinou, P. Pankin, and N. W. Paton, "Data Preparation : A Technological Perspective and Review," *SN Comput. Sci.*, vol. 4, no. 4, pp. 1–20, 2025, doi: 10.1007/s42979-023-01828-8.

[14] S. Sinsomboonthong, "Performance Comparison of New Adjusted Min-Max with Decimal Scaling and Statistical Column Normalization Methods for Artificial Neural Network Classification," *Int. J. Math. Math. Sci.*, vol. 2022, no. 1, pp. 1–9, 2022, doi: 10.1155/2022/3584406.

[15] N. A. Maori, "Metode Elbow Dalam Optimasi Jumlah Cluster Pada K-Means Clustering," *J. SIMETRIS*, vol. 14, no. 2, pp. 277–287, 2023.

[16] A. Rajsya and A. Rachman, "Rancang Bangun Penerapan Metode Elbow Pada K-Means Untuk Clustering Data Persediaan Barang," *Lit. Inform. Komput.*, vol. 1, no. 4, pp. 395–403, 2024, doi: 10.33096/linier.v1i4.2539.

[17] R. Nooraeni and G. Nurfalah, "Kajian Penerapan Jarak Euclidean , Manhattan , Minkowski , dan Chebyshev pada Algoritma Clustering K-Prototype," *SAKTI – Sains, Apl. Komputasi dan Teknol. Inf.*, vol. 4, no. 2, pp. 72–82, 2022, doi: http://dx.doi.org/10.30872/jsakti.v4i2.9241.

[18] A. Anfossi and D. Chicco, "An easy guide to the Davies-Bouldin index for unsupervised internal clustering evaluation," *Discov. Comput.*, vol. 29, no. 212, pp. 1–24, 2026, doi: https://doi.org/10.1007/s10791-026-10094-0.