Sistemasi: Jurnal Sistem Informasi ISSN:2302-8149
Volume 15, Nomor 8, 2026: 3000-3013 e-ISSN:2540-9719

Analisis dan Pengelompokan Produksi Tanaman Perkebunan di Indonesia menggunakan Metode *K-Means Clustering* dan *Principal Component Analysis*

Analysis and Clustering of Plantation Crop Production in Indonesia using K-Means Clustering and Principal Component Analysis Methods

[1]Afredo Grasia Valdano, [2]Styawati*
[1,2]Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia
[1,2]Jl. Zainal Abidin Pagaralam No.9-11 Labuhan Ratu, Bandar Lampung, Indonesia
*e-mail: alfredo_grasia_valdano@teknokrat.ac.id, Styawati@teknokrat.ac.id

(***received*:** 5 June 2026, ***revised*:** 28 July 2026, ***accepted*:** 29 July 2026)

Abstrak

Indonesia memiliki potensi besar pada sektor perkebunan dengan tingkat produksi yang bervariasi di setiap provinsi. Perbedaan tersebut dipengaruhi oleh kondisi geografis, luas lahan, dan komoditas unggulan masing-masing wilayah. Tujuan studi ini adalah mengidentifikasi pola produksi tanaman perkebunan di Indonesia menggunakan metode *K-Means Clustering* serta memvisualisasikan hasil pengelompokan dengan *Principal Component Analysis (PCA)*. Publikasi resmi BPS tahun 2024 yang memuat data produksi kelapa sawit, kelapa, karet, kopi, kakao, teh, dan tebu pada 38 provinsi di *preprocessing* penanganan pencilan *(outlier)* menggunakan metode pembatasan nilai *(capping)* berbasis *Interquartile Range (IQR)*, transformasi logaritmik, normalisasi menggunakan *StandardScaler*, reduksi dimensi menggunakan PCA, dan proses *clustering* menggunakan algoritma *K-Means*. Penetapan jumlah *cluster* optimal dilakukan menggunakan *Elbow Method*, *Silhouette Score*, dan *Davies-Bouldin Index (DBI)*. Temuan studi menunjukkan bahwa jumlah *cluster* terbaik adalah K = 5 dengan nilai *Silhouette Score* sebesar 0,5883 dan DBI sebesar 0,4744. Hasil pengelompokan membagi provinsi di Indonesia kedalam lima kategori, yaitu sangat tinggi, tinggi, sedang, rendah, dan sangat rendah. Provinsi di Sumatera dan Kalimantan mendominasi kategori sangat tinggi dan tinggi, beberapa provinsi berada pada kategori sedang, sedangkan sebagian besar wilayah di Indonesia bagian timur berada pada kategori rendah dan sangat rendah. Selain itu, kombinasi *K-Means Cluster*ing dan PCA terbukti efektif untuk mengkategorikan data serta mengidentifikasi pola produksi tanaman perkebunan antarprovinsi di Indonesia. Hasil penelitian ini dapat menjadi bahan pendukung bagi pemerintah dan pemangku kepentingan dalam penyusunan kebijakan pengembangan sektor perkebunan dan penentuan prioritas wilayah berdasarkan karakteristik produksi masing-masing provinsi.

Kata kunci: *data mining*, *K-Means Clustering*, pengelompokan provinsi, *principal component analysis (PCA)*, produksi tanaman perkebunan, visualisasi *cluster*.

Abstract

Indonesia has substantial potential in the plantation sector, with production levels varying considerably across provinces. These differences are influenced by geographical conditions, land availability, and the leading commodities of each region. This study aims to identify production patterns of plantation crops in Indonesia using K-Means Clustering and to visualize the resulting clusters using Principal Component Analysis (PCA). Official 2024 BPS data covering the production of oil palm, coconut, rubber, coffee, cocoa, tea, and sugarcane across 38 provinces were analyzed. The preprocessing procedure included outlier treatment using Interquartile Range (IQR)-based capping, logarithmic transformation, normalization using StandardScaler, dimensionality reduction using PCA, and clustering using the K-Means algorithm. The optimal number of clusters was determined using the Elbow Method, Silhouette Score, and Davies-Bouldin Index (DBI). The results indicate that K = 5 was the optimal clustering solution, achieving a Silhouette Score of 0.5883 and a

 ISSN:2302-8149
e-ISSN:2540-9719

DBI of 0.4744. The clustering results classified Indonesian provinces into five production categories: very high, high, moderate, low, and very low. Provinces in Sumatra and Kalimantan dominated the very-high and high production categories, while several provinces fell into the moderate category. In contrast, most provinces in eastern Indonesia were classified into the low and very-low production categories. Furthermore, the combination of K-Means Clustering and PCA proved effective for categorizing provincial production data and identifying spatial patterns in plantation crop production across Indonesia. The findings can support government agencies and other stakeholders in formulating plantation-sector development policies and establishing regional priorities based on the production characteristics of individual provinces.
***Keywords**: data mining, K-Means clustering, principal component analysis (PCA), plantation crop production, provincial clustering.*

1 Pendahuluan

Indonesia adalah negara agraris dengan potensi yang sangat besar untuk industri Perkebunan [1]. Berbagai komoditas perkebunan seperti kelapa sawit, kelapa, karet, kopi, kakao, teh, dan tebu menjadi sumber utama pendapatan daerah serta berkontribusi terhadap pertumbuhan ekonomi nasional [2]. Produksi tanaman perkebunan pada setiap provinsi memiliki karakteristik yang berbeda-beda karena dipengaruhi oleh kondisi geografis, iklim, luas lahan, serta tingkat produktivitas komoditas [3]. Perbedaan tersebut menyebabkan terjadinya ketimpangan tingkat produksi antarwilayah sehingga diperlukan suatu analisis untuk mengetahui pola persebaran produksi perkebunan di Indonesia [4].

Perkembangan teknologi informasi, khususnya pada bidang *data mining* dan *machine learning*, memungkinkan proses analisis data dengan efektif dan efisien [5]. Saat mengkategorikan data, metode pengelompokan *K-Means* sering digunakan [6]. Dengan menggabungkan data dengan fitur yang sebanding, teknik ini menciptakan sejumlah *cluster*, yang masing-masing mencerminkan tingkat kesamaan tertentu antara item yang sedang dipelajari [7]. *K-Means* memiliki keunggulan dalam proses komputasi yang relatif cepat, mudah diimplementasikan, serta mampu menghasilkan *cluster* yang baik pada data numerik [8]. Metode *K-Means* digunakan dalam penelitian ini untuk mengklasifikasikan provinsi berdasarkan tingkat hasil tanaman perkebunan guna menemukan daerah dengan tingkat produksi tinggi dan rendah [9].

Pendekatan *K-Means* telah digunakan di sektor pertanian dan perkebunan dalam sejumlah penelitian sebelumnya untuk mengkategorikan wilayah berdasarkan kesamaan hasil produksi [10], [11]. Namun, kedua penelitian tersebut hanya menerapkan algoritma *K-Means* dan belum mengintegrasikan *Principal Component Analysis (PCA)* sebagai teknik reduksi dimensi untuk mendukung visualisasi hasil *cluster*, sehingga interpretasi hubungan antarobjek masih terbatas [12].

Untuk memperoleh hasil analisis yang lebih informatif, penelitian ini mengintegrasikan *K-Means Cluster*ing dengan *Principal Component Analysis (PCA)*. Penerapan PCA bertujuan mengurangi kompleksitas data melalui reduksi dimensi sehingga data dapat divisualisasikan tanpa mengurangi informasi utama yang terkandung di dalamnya [12], [13]. Oleh karena itu, penelitian ini bertujuan menganalisis karakteristik produksi tanaman perkebunan pada berbagai provinsi di Indonesia menggunakan *K-Means Clustering* serta memvisualisasikan hasil pengelompokan dengan PCA.

Data pada studi ini merupakan data produksi tanaman perkebunan pada 38 provinsi di Indonesia tahun 2024 yang meliputi produksi kelapa sawit, kelapa, karet, kopi, kakao, teh, dan tebu [2]. Penelitian ini menerapkan tahapan *preprocessing*, reduksi dimensi menggunakan *Principal Component Analysis (PCA)*, serta pengelompokan menggunakan algoritma *K-Means*. Jumlah *cluster* optimal ditentukan menggunakan *Elbow Method*, *Silhouette Score*, dan *Davies–Bouldin Index (DBI)* [14], [15].

Penelitian ini diharapkan dapat memberikan informasi mengenai pola distribusi produksi tanaman perkebunan di Indonesia sehingga dapat menjadi bahan pendukung dalam penyusunan kebijakan pengembangan sektor perkebunan berbasis data serta menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan analisis dan pengelompokan data produksi pertanian maupun perkebunan.

2 Tinjauan Literatur

Indonesia adalah negara agraris yang industri perkebunannya memainkan peran penting dalam perekonomian negara [16]. Komoditas perkebunan menjadi sumber pendapatan daerah serta memiliki

 ISSN:2302-8149
e-ISSN:2540-9719

peranan dalam kegiatan ekspor nasional. Perbedaan kondisi geografis, luas lahan, tingkat produktivitas, dan karakteristik wilayah menyebabkan setiap provinsi memiliki tingkat produksi perkebunan yang berbeda-beda [17]. Untuk meningkatkan pengambilan keputusan di sektor perkebunan, keadaan ini menuntut penerapan teknik analisis data yang dapat mengenali pola kesamaan produksi di berbagai Lokasi [18].

Perkembangan teknologi informasi dan *data mining* memungkinkan proses analisis data dilakukan secara lebih efektif, khususnya dalam pengelompokan data berskala besar. Pendekatan pengelompokan data populer yang disebut *K-Means Cluster*ing menciptakan *cluster* berdasarkan seberapa mirip properti suatu objek satu sama lain, sehingga memudahkan untuk menemukan pola dalam data [19]. Metode ini memiliki proses komputasi yang relatif cepat, mudah diimplementasikan, dan efektif digunakan pada data numerik. Penelitian Rochmawati membuktikan algoritma *K-Means* dapat memberikan pengelompokan data penjualan terbaik dengan metode *Knowledge Discovery in Database (KDD)*. Berdasarkan hasil penelitian, penerapan *K-Means* terbukti mampu menghasilkan pengelompokan data yang representatif sesuai dengan kesamaan atribut atau karakteristik tertentu yang dimiliki oleh setiap data.

Penerapan metode *K-Means* juga telah digunakan pada berbagai bidang penelitian lain, seperti pengelompokan data kesehatan, pendidikan, hingga sektor pertanian [20]. Penelitian sebelumnya menerapkan metode *K-Means* dan *Hierarchical Clustering* untuk menyatukan data pengangguran sesuai dengan wilayah [21]. Temuan studi membuktilkan jika metode *clustering* mampu membantu proses identifikasi kelompok data yang memiliki kemiripan tertentu sehingga mempermudah proses analisis data regional [22]. Selain itu, penelitian Dewi dan Pakereng menerapkan kombinasi *PCA* dan *K-Means* pada data tingkat pendidikan penduduk [23], [24]. Temuan penelitian membuktikan jika metode PCA efektif dalam mereduksi dimensi data sehingga proses visualisasi *cluster* menjadi lebih mudah dipahami, tanpa mengurangi informasi esensial yang merepresentasikan data [25].

Untuk mempermudah proses analisis, penelitian ini memanfaatkan PCA sebagai teknik penyederhanaan data dengan merangkum berbagai variabel ke dalam sejumlah komponen utama yang tetap dan merepresentasikan karakteristik penting data [26]. Penelitian Rosyada dan Utari menunjukkan bahwa PCA mampu meningkatkan efisiensi proses *cluster*ing dengan mengurangi jumlah variabel tanpa menghilangkan karakteristik utama data. Penggunaan PCA juga membantu visualisasi hasil *clustering* sehingga pola persebaran data dapat diamati dengan lebih jelas. Metode yang efisien untuk menganalisis data numerik dengan banyak variabel adalah dengan menggabungkan PCA dengan *K-Means* [27].

Beberapa penelitian sebelumnya pada sektor pertanian dan perkebunan umumnya hanya berfokus pada proses *clustering* tanpa melakukan visualisasi pola persebaran data secara optimal. Selain itu, sebagian penelitian belum menerapkan *preprocessing* data secara menyeluruh, seperti penanganan *outlier*, transformasi logaritmik, dan normalisasi data sebelum proses *clustering* dilakukan [28]. Padahal, distribusi data produksi perkebunan cenderung memiliki rentang nilai yang besar sehingga berpotensi memengaruhi kualitas hasil *clustering* apabila tidak dilakukan *preprocessing* yang tepat [29].

Berdasarkan analisis literatur tersebut, penelitian ini mengombinasikan metode *K-Means Clustering* dan PCA untuk menganalisis serta mengelompokkan produksi tanaman perkebunan di Indonesia berdasarkan data produksi pada 38 provinsi [30]. Penelitian ini tidak hanya melakukan proses *clustering*, tetapi juga menerapkan *preprocessing* data berupa penanganan *missing value*, *capping outlier* berbasis *Interquartile Range (IQR)*, transformasi logaritmik, dan normalisasi data menggunakan *StandardScaler* sebelum proses *cluster*ing dilakukan. Selain itu, penelitian ini memanfaatkan *PCA* untuk membantu visualisasi persebaran *cluster* dalam bentuk *scatter plot* sehingga pola kemiripan produksi antarprovinsi dapat diinterpretasikan secara lebih informatif. Oleh karena itu, dibandingkan studi terdahulu, diharapkan penelitian ini dapat memberikan temuan pengelompokan yang lebih ideal dan informasi yang lebih mudah dipahami [31].

3 Metode Penelitian

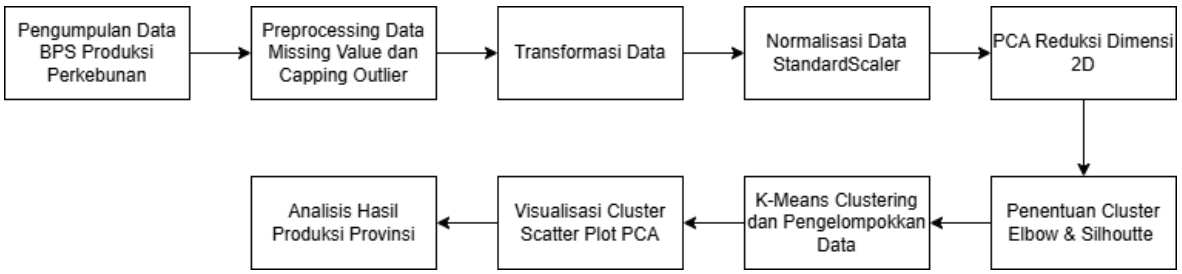
Penelitian ini menganalisis dan mengkategorikan hasil produksi tanaman perkebunan di Indonesia menggunakan metodologi kuantitatif dan metode *K-Means Clustering* dan *Principal Component Analysis (PCA)* [14]. Sumber utama untuk analisis tren produksi tanaman perkebunan di Indonesia pada tahun 2024 dalam studi ini adalah data sekunder dari publikasi resmi BPS. Variabel

 ISSN:2302-8149
e-ISSN:2540-9719

yang digunakan meliputi produksi kelapa sawit, kelapa, karet, kopi, kakao, teh, dan tebu. Data terdiri atas 38 provinsi dengan tujuh variabel numerik yang merepresentasikan jumlah produksi masing-masing komoditas [32].

3.1 Tahapan Penelitian

Tahapan pada penyelidikan ini yaitu pengumpulan data, *preprocessing* data, transformasi data, reduksi dimensi menggunakan PCA, *clustering* menggunakan *K-Means*, evaluasi *cluster*, dan visualisasi hasil penelitian (Gambar 1) [14].



Gambar 1 Alur tahapan penelitian

3.2 Pengumpulan Data

Studi ini memakai data sekunder dari publikasi Badan Pusat Statistik (BPS) *https://www.bps.go.id/id/statistics-table/2/MjU2NiMy/produksi-tanaman-perkebunan-menurut-provinsi-dan-jenis-tanaman--ribu-ton-.html* Laporan ini mencakup informasi tentang produksi tanaman perkebunan di Indonesia berdasarkan provinsi dan jenis tanaman pada tahun 2024. Data tersebut diakses dan dikumpulkan pada tanggal 5 April 2026. Berikut Tabel 1 menunjukkan variabel pada studi ini.

Tabel 1 Variabel yang digunakan

Variabel	Keterangan
Kelapa Sawit	Produksi Kelapa Sawit
Kelapa	Produksi Kelapa
Karet	Produksi Karet
Kopi	Produksi Kopi
Kakao	Produksi Kakao
Teh	Produksi Teh
Tebu	Produksi Tebu

3.3 *Preprocessing* Data

Langkah-langkah *preprocessing* dilakasanakan guna memastikan kualitas data sebelum proses pengelompokan. Langkah-langkah ini yaitu penanganan nilai yang *missing value*, *outlier*, transformasi data, dan normalisasi data [33].

a. Penanganan *Missing value*

Nilai kosong pada data diisi menggunakan nilai rata-rata *(mean)* dari masing-masing variabel numerik agar tidak memengaruhi proses analisis *cluster*.

b. Penanganan *Outlier*

Data produksi perkebunan memiliki rentang nilai yang cukup besar sehingga diperlukan penanganan *outlier* menggunakan metode *capping* berbasis *Interquartile Range (IQR)*[12]. Membatasi nilai-nilai yang berada di luar batas bawah dan atas distribusi data adalah cara pendekatan ini diimplementasikan.
Rumus *IQR* yang digunakan ditunjukkan pada Persamaan (1).

$$IQR = Q_3 - Q_1 \quad (1)$$

Batas bawah dan batas atas ditetapkan menggunakan Persamaan (2) dan Persamaan (3).

 ISSN:2302-8149
e-ISSN:2540-9719

$$Lower\ Bound = Q_1 - 1.5(IQR) \quad (2)$$
$$Upper\ Bound = Q_3 + 1.5(IQR) \quad (3)$$

Nilai yang berada di luar batas tersebut akan disesuaikan ke nilai batas minimum atau maksimum tanpa menghapus data asli.

c. Transformasi Data

Karena distribusi data produksi memiliki skala yang tidak merata, dilakukan transformasi logaritmik menggunakan fungsi *log10* untuk mengurangi *skewness* pada data[33].
Transformasi logaritmik dilakukan menggunakan Persamaan (4).

$$X' = \log_{10}(X + 1) \quad (4)$$

d. Normalisasi Data

Untuk mencegah variabel tertentu mendominasi proses pengelompokan, data kemudian dinormalisasi dengan *StandardScaler* sehingga setiap variabel memiliki skala yang konsisten[32].
Persamaan normalisasi ditunjukkan pada Persamaan (5).

$$Z = \frac{X - \mu}{\sigma} \quad (5)$$

Keterangan:
X= nilai data
μ= rata-rata data
σ= standar deviasi

3.4 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) berfungsi untuk membagi dimensi data menjadi dua komponen utama tanpa mengorbankan informasi penting[14]. Penggunaan PCA bertujuan untuk mempermudah visualisasi *cluster* dan meningkatkan efisiensi proses analisis. Pada penelitian ini, PCA direduksi menjadi dua dimensi utama, yakni *Principal Component 1 (PC1*) dan *Principal Component 2 (PC2)*.
Secara umum, PCA dapat dinyatakan menggunakan Persamaan (6).

$$Z_i = a_1X_1 + a_2X_2 + \cdots + a_nX_n \quad (6)$$

Keterangan:
Z_i= principal component
a_n= bobot eigenvector
X_n= variabel data

3.5 *K-Means Clustering*

Pengelompokan *K-Means* adalah teknik utama yang digunakan dalam penelitian ini[14]. Pendekatan ini menggunakan jarak terdekat ke centroid *cluster* untuk mengelompokkan data. *Elbow Method* dan penilaian *Silhouette Score* digunakan untuk mengidentifikasi jumlah *cluster* [34]. Berdasarkan hasil evaluasi, total *cluster* terbaik dipilih untuk memperoleh hasil maksimal.
Tahapan algoritma *K-Means* terdiri atas [35]:

1. Hitung jumlah *cluster* (k).
2. Pilih *centroid* awal secara acak.
3. Tentukan seberapa jauh setiap titik data dari *centroid*.
4. Urutkan data berdasarkan *centroid* terdekat.
5. Revisi *centroid*.
6. Lanjutkan proses ini hingga *centroid* stabil.

Perhitungan jarak menggunakan *Euclidean Distance* yang dibuktikan pada Persamaan (7) [36]:

 ISSN:2302-8149
e-ISSN:2540-9719

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (7)$$

Fungsi objektif *K-Means* untuk meminimalkan WCSS ditunjukkan pada Persamaan (8) [37]:

$$WCSS = \sum_{i=1}^{k} \sum_{x_j \in C_i} || x_j - c_i ||^2 \quad (8)$$

Keterangan:
C_i= *cluster* ke-i
c_i= centroid *cluster*
x_j= data ke-j

3.6 Evaluasi *Cluster*

Untuk mengukur tingkat kualitas pengelompokan data, penelitian ini menggunakan indikator *Silhouette Score* dan *Davies-Bouldin Index (DBI***)**.[38]. Dengan memeriksa kedekatan antar anggota dalam *cluster* yang sama dan jarak antara mereka dengan *cluster* lain, nilai *Silhouette Score* digunakan untuk menilai kualitas pengelompokan data.[39]. Semakin tinggi nilai *Silhouette Score*, maka kualitas *cluster* semakin baik[40].
Persamaan *Silhouette Score* ditunjukkan pada Persamaan (9).

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (9)$$

Keterangan:
$a(i)$= rata-rata jarak data dalam *cluster* yang sama
$b(i)$= rata-rata jarak data ke *cluster* terdekat

4 Hasil dan Pembahasan

4.1 Hasil *Preprocessing* Data

Sebelum menggunakan metode pengelompokan *(clustering)*, langkah pertama dalam penelitian ini adalah mempersiapkan data untuk meningkatkan kualitasnya.

Tabel 2 Tahapan dan hasil *preprocessing* data

Keterangan	Hasil
Jumlah Provinsi	38
Jumlah Variabel	7
Missing value	Ditangani dengan Mean
Outlier	Ditangani dengan Metode Capping IQR
Transformasi Data	Logaritmik
Normalisasi	StandardScaler

Berdasarkan hasil *preprocessing* (Tabel 2), seluruh data berhasil diproses tanpa menghapus observasi sehingga jumlah data tetap sebanyak 38 baris dan tujuh variabel numerik. Penanganan *missing value* dilakukan menggunakan rata-rata *(mean)* masing-masing variabel numerik. Selanjutnya, *outlier* ditangani memakai metode *capping* berbasis *Interquartile Range (IQR)* agar nilai ekstrem tidak mendominasi proses *clustering.* Setelah itu, data ditransformasikan menggunakan logaritma basis 10 untuk mengurangi *skewness* dan dilanjutkan dengan normalisasi menggunakan *StandardScaler* sehingga seluruh variabel berada pada skala yang seimbang.

 ISSN:2302-8149
e-ISSN:2540-9719

4.2 Hasil Reduksi Dimensi Menggunakan PCA

PCA digunakan untuk mereduksi dimensi data dengan mengubah variabel asli menjadi beberapa komponen utama yang memiliki informasi terbesar. Berdasarkan hasil analisis PCA, diperoleh tujuh komponen utama dari tujuh variabel produksi tanaman perkebunan yang digunakan dalam penelitian. Namun, hanya tiga komponen utama pertama yaitu PC1, PC2, dan PC3 yang dipilih karena mampu mempertahankan sebagian besar informasi data sebesar 80,97%, sedangkan PC1 dan PC2 digunakan untuk kebutuhan visualisasi dua dimensi. Penggunaan PCA bertujuan untuk menyederhanakan data multidimensi sehingga lebih mudah divisualisasikan dan dianalisis. Proses reduksi dimensi dilakukan dengan mentransformasikan variabel asli ke dalam komponen utama menggunakan kombinasi linier dari setiap variabel sebagaimana ditunjukkan pada Persamaan (6).

$$Z_i = a_1X_1 + a_2X_2 + \cdots + a_nX_n \quad (6)$$

Keterangan:

Z_i= principal component
a_n= bobot eigenvector
X_n= variabel data

Tabel 3 Hasil *explained variance principal component analysis (PCA)*

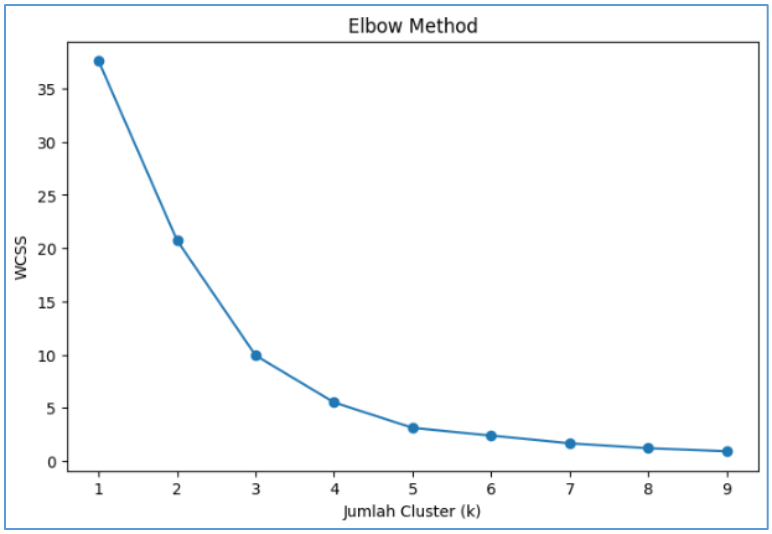
Komponen	*Explained Variance* (%)	*Cumulative Explained Variance* (%)
PC1	41,54	41,54
PC2	24,51	66,06
PC3	14,91	80,97

Berdasarkan hasil analisis Principal Component Analysis (PCA) (Tabel 3), diperoleh tujuh komponen utama berdasarkan tujuh variabel produksi tanaman perkebunan yang digunakan. Komponen utama pertama (PC1) mampu menjelaskan variasi data sebesar 41,54%, sedangkan komponen utama kedua (PC2) menjelaskan sebesar 24,51%. Secara kumulatif, kedua komponen tersebut mampu mempertahankan informasi data sebesar 66,06%. Namun, berdasarkan *threshold cumulative explained variance* sebesar 80%, dipilih tiga komponen utama, yaitu PC1, PC2, dan PC3, karena ketiganya mampu menjelaskan 80,97% variasi data asli variasi data asli. Oleh karena itu, ketiga komponen utama tersebut dianggap mampu merepresentasikan sebagian besar informasi penting dari data produksi tanaman perkebunan. Meskipun pemilihan komponen utama didasarkan pada *Threshold Cumulative Explained Variance* sebesar ≥80% sehingga dipilih tiga komponen utama (PC1, PC2, dan PC3), visualisasi hasil *cluster*ing tetap menggunakan PC1 dan PC2 karena visualisasi *scatter plot* PCA direpresentasikan dalam dua dimensi, sedangkan kedua komponen tersebut memiliki kontribusi variasi terbesar. Visualisasi hasil *clustering* menggunakan PC1 dan PC2 sehingga pola persebaran data dan pemisahan antarcluster dapat diamati secara lebih jelas melalui *scatter plot* dua dimensi.

4.3 Penentuan Jumlah *Cluster*

Penetapan total *cluster* dilakukan menggunakan *Elbow Method* dan evaluasi *Silhouette Score*. Elbow Method digunakan untuk melihat perubahan nilai *WCSS* terhadap jumlah *cluster*. Dari grafik *Elbow Method*, terjadi penurunan nilai *WCSS* yang cukup signifikan pada beberapa jumlah *cluster* awal, kemudian mulai melandai setelah mencapai jumlah *cluster* tertentu. Kondisi ini menunjukkan bahwa jumlah *cluster* ideal berada pada titik siku, yang menandakan kompromi optimal antara jumah *cluster* dan varians data *cluster*.

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 2 Elbow method penentuan jumlah *cluster*

Nilai K = 5 dipilih sebagai jumlah *cluster* ideal dalam penelitian ini karena grafik pada Gambar 2 menunjukkan bahwa penurunan nilai WCSS mulai mengalami perlambatan setelah jumlah *cluster* mencapai lima kelompok.

Tabel 4 Hasil evaluasi jumlah *cluster*

Jumlah *Cluster*	Silhouette Score	Davies-Bouldin Index
2	0.4290	1.0241
3	0.5251	0.6153
4	0.5521	0.5642
5	0.5883	0.4744
6	0.5612	0.5076
7	0.5628	0.4962
8	0.5578	0.5281
9	0.5692	0.4972

Selain *Elbow Method*, evaluasi menggunakan *Silhouette Score* dan *Davies–Bouldin Index (DBI)* (Tabel 4) menunjukkan bahwa konfigurasi K = 5 memberikan hasil terbaik dengan nilai Silhouette Score sebesar 0,5883 dan DBI sebesar 0,4744. Oleh karena itu, jumlah *cluster* yang digunakan pada tahap *cluster*ing adalah lima.

4.4 Hasil *Clustering K-Means*

Setelah diperoleh jumlah *cluster* optimal sebanyak lima (K=5) berdasarkan hasil evaluasi pada Subbab 4.3, proses *clustering* menggunakan algoritma *K-Means* menghasilkan pengelompokan 38 provinsi ke dalam lima kategori berdasarkan karakteristik produksi tanaman perkebunan.

 ISSN:2302-8149
e-ISSN:2540-9719

Tabel 5 Representasi provinsi pada masing-masing kategori *cluster*

Provinsi	Kategori
Riau	Sangat Tinggi
Aceh	Tinggi
Lampung	Sedang
Bali	Rendah
DKI Jakarta	Sangat Rendah

Berdasarkan hasil pengujian pada Tabel 5, jumlah *cluster* sebanyak lima dipilih sebagai konfigurasi optimal karena menghasilkan nilai *Silhouette Score* tertinggi, yaitu 0,5883, serta nilai *DBI* sebesar 0,4744. Hasil tersebut menunjukkan bahwa konfigurasi *cluster* memiliki kualitas pengelompokan yang baik dengan tingkat homogenitas dalam *cluster* dan pemisahan antarkelompok yang memadai. Nilai DBI sebesar 0,4744 yang mendekati nol semakin memperkuat bahwa kualitas *cluster* yang dihasilkan tergolong baik. Hasil *cluster*ing mengelompokkan 38 provinsi ke dalam lima kategori, yaitu sangat tinggi, tinggi, sedang, rendah, dan sangat rendah, sebagaimana ditunjukkan pada Tabel 5. Karakteristik masing-masing *cluster* selanjutnya dianalisis berdasarkan rata-rata produksi setiap komoditas yang disajikan pada Tabel 6.

Tabel 6 Rata-rata produksi tiap *cluster*

Kategori	Kelapa Sawit	Kelapa	Karet	Kopi	Kakao	Teh	Tebu	Total Produksi
Sangat Tinggi	4595.55	88.62	109.16	1.03	0.74	0.00	0.00	4795.09
Tinggi	2500.48	71.84	218.18	79.93	19.11	4.08	23.15	2916.78
Sedang	156.62	128.71	24.69	43.64	44.40	15.70	370.85	784.61
Rendah	50.75	95.89	0.08	5.91	21.89	0.02	10.15	184.69
Sangat Rendah	141.29	9.03	1.92	0.67	1.28	0.00	0.00	154.18

Hasil rata-rata produksi pada setiap *cluster* menunjukkan bahwa label *cluster* tidak hanya menggambarkan tinggi rendahnya total produksi, tetapi juga mencerminkan kemiripan pola produksi beberapa komoditas perkebunan secara bersamaan. *Cluster* sangat tinggi didominasi oleh provinsi dengan produksi kelapa sawit yang besar dan kontribusi perkebunan yang dominan secara nasional, seperti Riau, Kepulauan Bangka Belitung, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, dan Kalimantan Timur.

Hasil pengelompokan menunjukkan bahwa Aceh, Sumatera Utara, Sumatera Barat, Jambi, Sumatera Selatan, dan Bengkulu berada pada kategori *cluster* tinggi. Kelompok ini memiliki kesamaan berupa tingginya kontribusi beberapa komoditas perkebunan utama, terutama kelapa sawit, kopi, dan karet.

Sementara itu, Lampung, Jawa Barat, Jawa Tengah, Jawa Timur, Sulawesi Tengah, dan Sulawesi Selatan termasuk dalam kategori *cluster* sedang dengan tingkat produksi yang cukup stabil. Walaupun memberikan kontribusi yang signifikan terhadap sektor perkebunan, tingkat produksinya masih lebih rendah dibandingkan wilayah utama di Sumatera dan Kalimantan.

Di sisi lain, *cluster* rendah dan sangat rendah didominasi oleh daerah perkotaan serta provinsi dengan keterbatasan lahan produktif untuk perkebunan. Kondisi tersebut menyebabkan produksi komoditas perkebunan di wilayah seperti DKI Jakarta, Banten, Papua Barat, Papua Tengah, Papua Selatan, dan Papua Pegunungan relatif lebih kecil dibandingkan daerah yang menjadi pusat produksi perkebunan nasional.

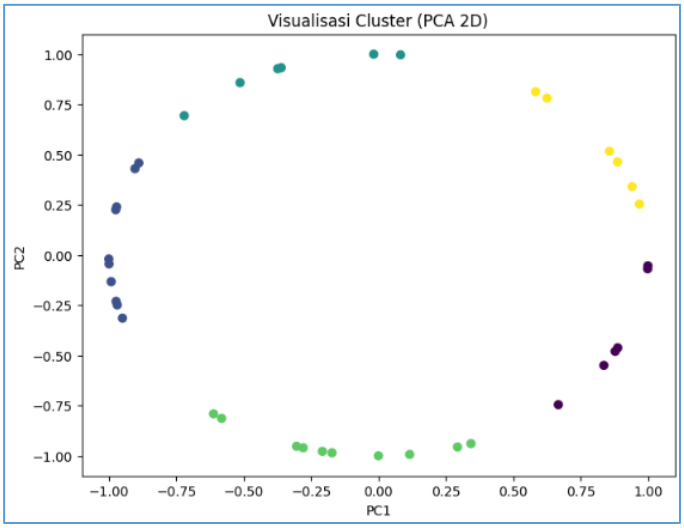
Secara umum, hasil pengelompokan menunjukkan bahwa sebagian besar provinsi di Pulau Sumatera dan Kalimantan berada pada kategori produksi tinggi hingga sangat tinggi. Kondisi ini menunjukkan bahwa Sumatera dan Kalimantan masih menjadi pusat utama produksi tanaman

 ISSN:2302-8149
e-ISSN:2540-9719

perkebunan di Indonesia. Sebaliknya, beberapa wilayah di Indonesia bagian timur cenderung berada pada kategori rendah hingga sangat rendah karena luas area produksi dan infrastruktur perkebunan yang masih terbatas.

4.5 Visualisasi *Cluster* PCA

Visualisasi *cluster* melalui *scatter plot* PCA dua dimensi (Gambar 3) memperlihatkan adanya



Gambar 3 Visualisasi *cluster* menggunakan PCA 2D

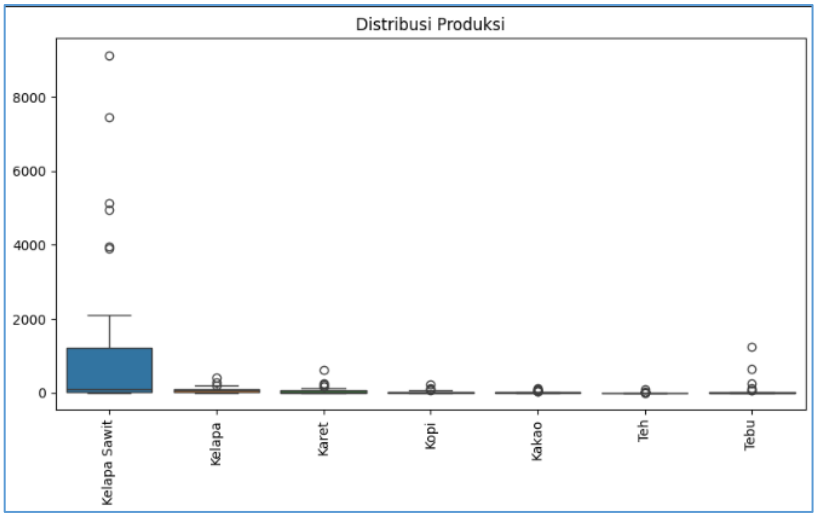
perbedaan pola persebaran pada setiap kelompok data. Meskipun berdasarkan *Cumulative Explained Variance* tiga komponen utama (PC1, PC2, dan PC3) dipilih untuk mempertahankan informasi sebesar 80,97%, visualisasi dilakukan menggunakan PC1 dan PC2 karena kedua komponen tersebut memiliki kontribusi variasi terbesar. Kombinasi PC1 dan PC2 menghasilkan representasi visual sebesar 66,06% dari total variasi data, sehingga cukup memberikan gambaran awal mengenai pola hubungan antarcluster yang terbentuk. Visualisasi dua dimensi tersebut tidak menggambarkan seluruh informasi data, namun mampu memberikan gambaran umum mengenai pola kedekatan dan perbedaan karakteristik antarcluster berdasarkan dua komponen utama dengan kontribusi terbesar.

Terutama pada *cluster* dengan tingkat produksi sangat tinggi dan sangat rendah, terlihat adanya kecenderungan pemisahan yang lebih jelas dibandingkan dengan kelompok *cluster* lainnya. Hal ini menunjukkan bahwa provinsi dengan karakteristik produksi yang berbeda memiliki posisi yang berbeda dalam ruang PCA. Kondisi tersebut mengindikasikan bahwa penerapan PCA membantu hasil *clustering* menjadi lebih mudah dipahami melalui visualisasi data.

4.6 Distribusi Produksi Perkebunan

Hasil visualisasi distribusi data menggunakan *boxplot* (Gambar 4) menunjukkan bahwa variabel kelapa sawit memiliki rentang produksi paling besar dibandingkan variabel lainnya. Hal ini mengindikasikan bahwa kelapa sawit merupakan komoditas perkebunan dominan di Indonesia. Selain itu, beberapa variabel lain seperti karet dan tebu juga memiliki *outlier* dengan nilai produksi cukup tinggi pada beberapa provinsi tertentu.

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 4 Distribusi produksi tanaman perkebunan

Distribusi data yang tidak merata menjadi salah satu alasan penting dilakukannya transformasi logaritmik dan normalisasi sebelum proses *clustering* dilakukan. Dengan *preprocessing* tersebut, hasil *clustering* menjadi lebih stabil dan mampu menggambarkan pola data secara lebih baik.

4.7 *Insight* Hasil Penelitian

Dari hasil *clustering* dan visualisasi data, penelitian ini menghasilkan beberapa insight penting, yaitu:

1. Provinsi di Pulau Sumatera dan Kalimantan mendominasi kategori produksi perkebunan tinggi hingga sangat tinggi, terutama pada komoditas kelapa sawit dan karet.
2. Wilayah Indonesia bagian timur cenderung berada pada kategori produksi rendah karena keterbatasan luas lahan perkebunan dan tingkat produksi komoditas yang lebih kecil.
3. Penggunaan PCA berhasil membantu visualisasi pola kemiripan produksi antarprovinsi sehingga hubungan antarcluster dapat diamati dengan lebih jelas.
4. Penerapan *K-Means Clustering* berhasil mengidentifikasi kelompok provinsi yang memiliki karakteristik produksi perkebunan serupa, sehingga hasil pengelompokan dapat dimanfaatkan sebagai bahan pendukung dalam perumusan kebijakan pada sektor perkebunan nasional.
5. PCA mampu mengurangi kompleksitas data dengan mempertahankan 80,97% informasi asli melalui tiga komponen utama, sehingga proses interpretasi pola *cluster* menjadi lebih mudah dilakukan.

Secara keseluruhan, kombinasi metode *K-Means Clustering* dan PCA mampu menghasilkan pengelompokan data yang baik serta memberikan gambaran mengenai pola produksi tanaman perkebunan di Indonesia berdasarkan karakteristik masing-masing provinsi.

5 Kesimpulan

Penelitian ini menerapkan metode *K-Means Cluster*ing dan *Principal Component Analysis (PCA)* untuk menganalisis dan mengelompokkan produksi tanaman perkebunan pada 38 provinsi di Indonesia tahun 2024. Hasil evaluasi membuktikan jika konfigurasi terbaik didapat pada lima *cluster* (K = 5), dengan nilai *Silhouette Score* sebesar 0,5883 dan nilai DBI 0,4744. Temuan ini membuktikan jika model pengelompokan memiliki tingkat kesamaan data yang tinggi di setiap kelompok dan dapat memberikan pemisahan *cluster* yang cukup baik. Temuan penelitian menunjukkan bahwa hasil pengelompokan membagi 38 provinsi di Indonesia ke dalam lima kategori, yaitu kategori produksi sangat tinggi, tinggi, sedang, rendah, dan sangat rendah. Sebagian besar provinsi di Pulau Sumatera dan Kalimantan berada pada kategori produksi tinggi dan sangat tinggi, sedangkan sebagian besar

 ISSN:2302-8149
e-ISSN:2540-9719

wilayah Indonesia bagian timur berada pada kategori rendah dan sangat rendah. Pengelompokan tersebut mampu memberikan gambaran mengenai pola kemiripan produksi tanaman perkebunan antarprovinsi di Indonesia. Pemanfaatan metode PCA mempermudah visualisasi hasil pengelompokan melalui grafik *scatter plot*. Berdasarkan hasil analisis PCA, tiga komponen utama (PC1, PC2, dan PC3) mampu mempertahankan 80,97% informasi data, sedangkan PC1 dan PC2 digunakan sebagai representasi visual dua dimensi karena secara kumulatif menjelaskan 66,06% variasi data. Secara keseluruhan, kombinasi metode *K-Means Clustering* dan PCA mampu menghasilkan pengelompokan provinsi berdasarkan karakteristik produksi tanaman perkebunan serta memberikan gambaran mengenai pola distribusi produksi perkebunan di Indonesia. Temuan penelitian ini dapat menjadi bahan pertimbangan bagi pemerintah dan pemangku kepentingan dalam penyusunan kebijakan pengembangan sektor perkebunan, khususnya untuk penentuan wilayah sentra produksi dan pemerataan pengembangan komoditas perkebunan antarprovinsi. Penelitian ini masih memiliki keterbatasan karena hanya menggunakan data produksi tanaman perkebunan tahun 2024 yang bersumber dari data sekunder Badan Pusat Statistik (BPS). Selain itu, penelitian ini hanya memanfaatkan tujuh variabel produksi tanpa mempertimbangkan faktor lain, seperti luas lahan, produktivitas, kondisi iklim, dan aspek sosial ekonomi yang juga dapat memengaruhi karakteristik produksi tanaman perkebunan di setiap provinsi. Penelitian selanjutnya disarankan menggunakan data produksi tanaman perkebunan dalam rentang waktu beberapa tahun sehingga dapat dianalisis perubahan pola produksi secara temporal. Selain itu, penelitian berikutnya dapat membandingkan performa *K-Means* dengan metode *cluster*ing lain seperti *Agglomerative Clustering*, DBSCAN, atau *Fuzzy C-Means*, serta mengintegrasikan variabel pendukung seperti luas lahan, curah hujan, produktivitas, dan kondisi iklim agar hasil pengelompokan menjadi lebih komprehensif.

Referensi

[1] S. Fevriera and F. Safara Devi, "Analisis Produksi Kelapa Sawit Indonesia: Pendekatan Mikro dan Makro Ekonomi," *Transformatif*, Vol. 12, No. 1, pp. 1–16, May 2023.

[2] R. Situmeang, S. Marta Lina, N. Rodiyah Hisan, and S. Vaulina, "Peran Subsektor Perkebunan terhadap Pertumbuhan Ekonomi di Provinsi Riau: *The Role of Plantation Subsector to Economic Growth in Riau Province*," *Jurnal Agroteknologi Agribisnis dan Akuakultur*, Vol. 4, No. 1, pp. 51–60, Jan. 2024.

[3] A. Said, A. Akhmad, I. Sribianti, M. Natsir, and M. Maulina, "Analisis Pengaruh Produksi dan Luas Lahan Kelapa Sawit terhadap PDRB Sektor Pertanian: Pendekatan Regresi Linier Berganda menggunakan Data Sekunder 2013-2022," *Jurnal Ilmu Manajemen Sosial Humaniora (JIMSH)*, Vol. 6, No. 1, pp. 46–56, Jul. 2024, DOI: 10.51454/jimsh.v6i1.632.

[4] N. P. Daulay, A. Alyani Putri, O. N. Ramadani, E. V. Agata, and M. Hamid, "Analisis Perkembangan Produksi Kelapa Sawit berdasarkan Luas Lahan dan Jumlah Pabrik Kelapa Sawit (Studi Pada Kabupaten Rokan Hulu)," *Jurnal Pengabdian Masyarakat dan Riset Pendidikan*, Vol. 4, No. 2, pp. 11335–11342, 2025, DOI: 10.31004/jerkin.v4i2.3935.

[5] I. Akbar, I. S. Samad, R. Rahmat, and S. Rosmiana, "*Data Mining Analysis of K-Means Algorithm and Decision Tree for Early Detection of Students at Risk of Dropping Out*," *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, Vol. 7, No. 2, pp. 148–162, May 2025, DOI: 10.20895/inista.v7i2.1630.

[6] M. Rochmawati, G. W. C. Bagaskara, I. A. Adha, Y. Umaidah, and A. Voutama, "Implementasi Algoritma *K-Means* dalam Klasterisasi Penjualan pada Sebuah Perusahaan menggunakan Metodologi KDD," *SISTEMASI: Jurnal Sistem Informasi*, Vol. 13, No. 1, pp. 54–62, Jan. 2024, [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[7] A. Mafera, H. Susilo, and I. N. Sulrieni, "*Application of Data Mining With K-Means Clustering Algorithm for Hypertension Disease Classification at Puskesmas Lubuk Buaya Padang in 2024*," *Journal of Medical Records and Information Technology (JOMRIT)*, Vol. 2, No. 2, pp. 1–5, Dec. 2024, [Online]. Available: https://jurnal.syedzasaintika.ac.id/index.php/jomrit

[8] I. Firman Ashari, R. Banjarnahor, D. R. Farida, S. P. Aisyah, A. P. Dewi, and N. Humaya, "*Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index*

for Grouping IMDB Movies," *Journal of Applied Informatics and Computing (JAIC)*, Vol. 6, No. 1, pp. 7–15, 2022, [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[9] A. Alif *et al.*, "Penerapan Algoritma *K-means Clustering* dan *Hierarchical Clu*stering dalam Mengelompokkan Data Pengangguran di Karawang," *Algoritma*, Vol. 21, No. 2, pp. 209–220, Nov. 2024, DOI: 10.33364/algoritma/v.21-2.2155.

[10] R. Maulidina and S. Y. Riska, "*Application of the K-Means Algorithm for Clustering Plantation Crop Production in Indonesia*," *SMATIKA JURNAL*, Vol. 13, No. 02, pp. 339–349, Dec. 2023, DOI: 10.32664/smatika.v13i02.991.

[11] A. A. Simangunsong, I. Gunawan, Z. M. Nasution, and G. Artikel, "Pengelompokkan Hasil Produksi Tanaman Perkebunan berdasarkan Provinsi menggunakan Metode *K-Means Clustering Production of Plantation Crops by Province using the K-Means Method Article Info ABSTRAK*," *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, Vol. 1, No. 4, pp. 2828–9099, 2022, DOI: 10.55123/jomlai.v1i4.1661.

[12] I. A. Rosyada and D. T. Utari, "Penerapan *Principal Component Analysis* untuk Reduksi Variabel pada Algoritma *K-Means Clustering*," *Jambura J. Probab. Stat*, Vol. 5, No. 1, pp. 6–13, 2024, DOI: 10.34312/jjps.v5i1.18733.

[13] R. Ishak, "Optimalisasi Seleksi Atribut *K-Means* Menggunakan *Correlation Matrix* pada *Clustering* Penyakit Pasien *Optimization of K-Means Attribute Selection using Correlation Matrix in Patient Disease Clustering*," *Jambura Journal of Electrical and Electronics Engineering*, Vol. 7, No. 2, pp. 141–148, Jul. 2025.

[14] S. Dewi and M. A. I. Pakereng, "Implementasi *Principal Component Analysis* pada *K-Means* untuk Klasterisasi Tingkat Pendidikan Penduduk Kabupaten Semarang," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, Vol. 8, No. 4, pp. 1186–1195, Dec. 2023, DOI: 10.29100/jipi.v8i4.4101.

[15] D. A. Awaliyah, Budi Prasetiyo, R. Muzayanah, and A. D. Lestari, "*Optimizing Customer Segmentation in Online Retail Transactions through the Implementation of the K-Means Clustering Algorithm*," *Scientific Journal of Informatics*, Vol. 11, No. 2, pp. 539–548, Jun. 2024, DOI: 10.15294/sji.v11i2.6137.

[16] R. Situmeang, S. M. Lina, N. R. Hisan, and S. Vaulina, "Peran Subsektor Perkebunan terhadap Pertumbuhan Ekonomi di Provinsi Riau: *The Role of Plantation Subsector to Economic Growth in Riau Province*," *Jurnal Agroteknologi Agribisnis dan Akuakultur*, Vol. 4, No. 1, pp. 51–60, Jan. 2024.

[17] P. Villalobos Perna, M. Di Febbraro, M. L. Carranza, F. Marzialetti, and M. Innangi, "*Remote Sensing and Invasive Plants in Coastal Ecosystems: What We Know so Far and Future Prospects*," *Land (Basel).*, Vol. 12, No. 2, p. 341, Feb. 2023, DOI: 10.3390/land12020341.

[18] C. C. Mireștean, R. I. Iancu, and D. T. Iancu, "*Radiotherapy and Immunotherapy—A Future Partnership towards a New Standard*," *Applied Sciences*, Vol. 13, No. 9, p. 5643, May 2023, DOI: 10.3390/app13095643.

[19] S. P. Sipayung and P. M. Hasugian, "*Integrating PCA and K-Means for Evidence-based Staple Food Segmentation: An Indonesian Food Policy Approach*," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, Vol. 9, No. 4, pp. 3175–3189, Oct. 2025, DOI: 10.33395/sinkron.v9i4.15343.

[20] A. C. Ehresmann, M. Béjean, and J. P. Vanbremeersch, "*A Mathematical Framework for Enriching Human–Machine Interactions*," *Mach. Learn. Knowl. Extr.*, Vol. 5, No. 2, pp. 597–610, Jun. 2023, DOI: 10.3390/make5020034.

[21] K. H. Izzuddin and A. W. Wijayanto, "Pemodelan *Clustering Ward, K-Means*, Diana, dan PAM dengan PCA untuk Karakterisasi Kemiskinan Indonesia Tahun 2021," *Komputika : Jurnal Sistem Komputer*, Vol. 13, No. 1, pp. 41–53, Apr. 2024, DOI: 10.34010/komputika.v13i1.10803.

[22] H. Lestari Siregar, R. Hidayanthi, and A. Langga Dewa Sakti, "*Implementation of K-Means Clustering on Student Learning Achievements based on Social Economic and Social Related*," *Research and Development in Education*, Vol. 4, No. 2, pp. 1447–1459, Dec. 2024, DOI: 10.22219/raden.v4i2.36742.

[23] E. Arry Kusuma and A. Dharmawati, "Analisis Pemerataan Pendidikan di Indonesia menggunakan Reduksi Dimensi PCA dan Klasterisasi K-Means," *JURNAL FASILKOM*, Vol. 16, No. 1, pp. 105–112, Apr. 2026.

[24] R. Rianti, R. Andarsyah, and R. M. Awangga, "Penerapan PCA dan Algoritma *Clustering* untuk Analisis Mutu Perguruan Tinggi di LLDIKTI Wilayah IV," *NUANSA INFORMATIKA*, Vol. 18, No. 2, pp. 67–77, Jul. 2024, [Online]. Available: https://journal.fkom.uniku.ac.id/ilkom

[25] I. K. O. Jabari, Shofiyah, P. K. S, N. N. Putriwijaya, and N. Yudistira, "*Learning-Augmented K-Means Clustering using Dimensional Reduction*," *ArXiv*, Jan. 2024, DOI: 10.1145/3626641.3627239.

[26] N. Migenda, R. Moller, and W. Schenck, "*Adaptive Dimensionality Reduction for Neural Network-based Online Principal Component Analysis*," *PLoS One*, Vol. 16, No. 3 March, p. e0248896, Mar. 2021, DOI: 10.1371/journal.pone.0248896.

[27] O. Assani-Amate, M. Bakhtyari, É. Roy, and V. Makarenkov, "*Assessing the Impact of Dimensionality Reduction on Clustering Performance -- A Systematic Study*," May 2026, [Online]. Available: http://arxiv.org/abs/2604.22099

[28] P. Chen, L. Wu, and L. Wang, "*AI Fairness in Data Management and Analytics: A Review on Challenges, Methodologies and Applications*," *Applied Sciences*, Vol. 13, No. 18, p. 10258, Sep. 2023, DOI: 10.3390/app131810258.

[29] Y. Fernando, R. Lukman, A. Jayadi, and R. Nuraini, "Sistem Otomatisasi Pemberi Pakan pada Peternakan Bebek dengan Arduino UNO dan Bluetooth menggunakan *SmartPhone*," *TIN: Terapan Informatika Nusantara*, Vol. 3, No. 2, pp. 42–50, Jul. 2022, DOI: 10.47065/tin.v3i2.1789.

[30] J. Arellano-Uson, E. Magaña, D. Morato, and M. Izal, "*Survey on Quality of Experience Evaluation for Cloud-based Interactive Applications*," *Applied Sciences*, Vol. 14, No. 5, p. 1987, Feb. 2024, DOI: 10.3390/app14051987.

[31] S. V. Ludkowski, "*Inverse Spectrum and Structure of Topological Metagroups*," *Mathematics*, Vol. 12, No. 4, p. 511, Feb. 2024, DOI: 10.3390/math12040511.

[32] A. Zaki, Irwan, and I. A. Sembe, "Penerapan *K-Means Clustering* dalam Pengelompokan Data (Studi Kasus Profil Mahasiswa Matematika FMIPA UNM)," *Journal of Mathematics, Computations, and Statistics*, Vol. 5, No. 2, pp. 163–176, Oct. 2022, Accessed: May 16, 2026. [Online]. Available: http://http://www.ojs.unm.ac.id/jmathcos.ojs.unm.ac.id/jmathcos

[33] A. A. A. Daniswara and I. K. D. Nuryana, "Data *Preprocessing* Pola pada Penilaian Mahasiswa Program Profesi Guru," *Journal of Informatics and Computer Science*, Vol. 5, No. 1, pp. 97–100, Mar. 2023.

[34] M. S. Arif, A. Mukheimer, and D. Asif, "*Enhancing the Early Detection of Chronic Kidney Disease: A Robust Machine Learning Model*," *Big Data and Cognitive Computing*, Vol. 7, No. 3, p. 144, Sep. 2023, DOI: 10.3390/bdcc7030144.

[35] L. G. Zamarrenho *et al.*, "*Effects of Three Different Brazilian Green Propolis Extract Formulations on Pro- and Anti-Inflammatory Cytokine Secretion by Macrophages*," *Applied Sciences*, Vol. 13, No. 10, p. 6247, May 2023, DOI: 10.3390/app13106247.

[36] B. Yang, M. H. Arshad, and Q. Zhao, "*Packet-Level and Flow-Level Network Intrusion Detection based on Reinforcement Learning and Adversarial Training*," *Algorithms*, Vol. 15, No. 12, p. 453, Dec. 2022, DOI: 10.3390/a15120453.

[37] M. Almulhem, S. Z. Hassan, A. Al-buainain, M. A. Sohaly, and M. A. E. Abdelrahman, "*Characteristics of Solitary Stochastic Structures for Heisenberg Ferromagnetic Spin Chain Equation*," *Symmetry (Basel).*, Vol. 15, No. 4, p. 927, Apr. 2023, DOI: 10.3390/sym15040927.

[38] Z. Chen, X. Xiong, F. Meng, X. Xiao, and J. Liu, "*Scaling-Invariant Max-Filtering Enhancement Transformers for Efficient Visual Tracking*," *Electronics (Switzerland)*, Vol. 12, No. 18, p. 3905, Sep. 2023, DOI: 10.3390/electronics12183905.

[39] Y. Wang, C. Hu, Z. Li, D. Zheng, F. Cui, and X. Yang, "*Theoretical and Simulation Analysis of Static and Dynamic Properties of MXene-Based Humidity Sensors*," *Applied Sciences*, Vol. 12, No. 16, p. 8254, Aug. 2022, DOI: 10.3390/app12168254.

[40] N. Andriyanov, "*The use of Correlation Features in the Problem of Speech Recognition*," *Algorithms*, Vol. 16, No. 2, p. 90, Feb. 2023, DOI: 10.3390/a16020090.