

**Jutisi:** Jurnal Ilmiah Teknik Informatika dan Sistem Informasi  
<https://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/index>  
 Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru  
 Loktabat – Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com  
 e-ISSN: 2685-0893

## Implementasi Algoritma *K-Means Clustering* untuk Mengelompokkan Wilayah Produksi Kopi di Jawa Barat

DOI: <http://dx.doi.org/10.35889/jutisi.v15i2.3622>

Creative Commons License 4.0 (CC BY – NC)



**Jingga Pratama<sup>1\*</sup>, Baenil Huda<sup>2</sup>, Shofa Shofiah Hilabi<sup>3</sup>, Elvina Novalia<sup>4</sup>**  
 Sistem Informasi, Universitas Buana Perjuangan Karawang, Jawa Barat, Indonesia  
 \*e-mail *Corresponding Author*: [si22.jinggapratama@mhs.ubpkarawang.ac.id](mailto:si22.jinggapratama@mhs.ubpkarawang.ac.id)

### Abstract

*Regional disparities in Arabica coffee output across West Java Province underscore the necessity of evidence-based analytical approaches to systematically map production distribution. This study applies K-means clustering to segment Arabica coffee-producing districts and cities based on their production volumes. The analytical pipeline encompasses data preprocessing, cluster modeling, performance evaluation, and result visualization. Three distinct production tiers emerged from the analysis: low, moderate, and high. A silhouette score of 0.883 confirmed excellent cluster cohesion and separation quality. The novelty of this research lies in its territory-level, data-driven segmentation approach paired with interpretable visualization to inform regional policy. The findings reveal pronounced inter-regional production disparities and offer an empirical foundation for crafting targeted, data-informed, and sustainable development strategies to improve the equity and competitiveness of Arabica coffee production at the regional level.*

**Keywords:** *K-Means clustering; Arabica coffee; Coffee production; Regional clustering; Data analysis*

### Abstrak

Kesenjangan tingkat produksi kopi Arabika antar daerah di Provinsi Jawa Barat mengindikasikan perlunya pendekatan analitis berbasis data untuk memetakan pola sebaran produksi secara sistematis. Penelitian ini menerapkan metode *K-means clustering* guna mengelompokkan daerah penghasil kopi Arabika berdasarkan volume produksinya. Rangkaian analisis mencakup pra-pemrosesan data, pemodelan *cluster*, evaluasi performa, dan visualisasi hasil. Tiga kelompok wilayah berhasil terbentuk, yakni kategori produksi rendah, menengah, dan tinggi. Nilai *silhouette score* sebesar 0,883 mengonfirmasi kualitas pengelompokan yang sangat baik dengan separasi antar *cluster* yang tegas. Kebaruan penelitian terletak pada penerapan segmentasi berbasis wilayah yang bersifat *data-driven* serta penyajian visual yang informatif sebagai fondasi pengambilan kebijakan. Temuan ini mengungkap adanya ketimpangan produksi yang nyata antar wilayah sekaligus menyediakan landasan empiris untuk merancang strategi pengembangan kopi Arabika yang lebih tepat sasaran, berbasis data, dan berkelanjutan ditingkat daerah.

**Kata kunci:** *K-Means Clustering; Kopi Arabika; Produksi kopi; Pengelompokan wilayah; Analisis data*

### 1. Pendahuluan

Sebagai salah satu produsen kopi terkemuka di dunia berada di belakang Kolombia, Brasil, dan Vietnam, Indonesia menjadikan kopi Arabika sebagai komoditas unggulan dengan nilai ekonomi yang signifikan [1]. Jawa Barat termasuk wilayah dengan kondisi geografis yang mendukung budidaya kopi Arabika, di mana karakteristik produk dipengaruhi oleh faktor lingkungan dan kemampuan sumber daya manusia [2]. Salah satu varietas kopi bernilai ekonomi tinggi adalah kopi Arabika yang dikenal memiliki cita rasa kompleks, aroma khas, serta tingkat keasaman yang lebih ringan dan halus dibandingkan kopi Robusta [3]. Tingginya permintaan pasar terhadap kopi Arabika menjadikan komoditas ini memiliki prospek strategis dalam meningkatkan kesejahteraan petani, mendorong pertumbuhan ekonomi daerah, serta

memperkuat daya saing sektor perkebunan. Oleh karena itu, pengembangan produksi kopi Arabika di Jawa Barat memerlukan dukungan perencanaan yang terarah dan berbasis data agar potensi masing-masing wilayah dapat dimanfaatkan secara optimal.

Meskipun data produksi kopi Arabika tersedia melalui berbagai sumber resmi, pemanfaatannya masih cenderung bersifat deskriptif tanpa analisis lanjutan yang mendukung pengambilan kebijakan berbasis data [4]. Padahal, optimalisasi pengolahan data produksi diperlukan untuk meningkatkan daya saing dan keberlanjutan sektor kopi [5]. Perbedaan tingkat produksi antarwilayah seharusnya dapat dimanfaatkan untuk mengidentifikasi daerah dengan kategori produksi rendah, sedang, maupun tinggi sehingga kebijakan pengembangan dapat disusun secara lebih tepat sasaran. Apabila data hanya disajikan dalam bentuk statistik agregat tanpa penggalian pola yang mendalam, maka nilai informasi yang tersimpan di dalamnya belum dimanfaatkan secara optimal. Kondisi ini menandai adanya kesenjangan antara ketersediaan data dan kegunaannya secara analitis, sehingga diperlukan pendekatan komputasional yang mampu mengubah data mentah menjadi informasi strategis yang dapat ditindaklanjuti.

Guna mendukung kebutuhan analisis yang lebih mendalam, pendekatan *clustering* dalam *data mining* dapat didayagunakan untuk mengelompokkan wilayah berdasarkan kesamaan pola dan karakteristik produksinya. *Data mining* pada dasarnya merupakan suatu proses ekstraksi pengetahuan dari kumpulan data berskala besar yang bertujuan mengungkap pola tersembunyi, keterkaitan antar variabel, serta informasi laten yang bernilai bagi proses pengambilan keputusan [6]. Di antara berbagai algoritma yang tersedia, *k-means* menonjol sebagai metode yang mudah diimplementasikan, hemat komputasi, dan terbukti andal dalam mengelompokkan data berdasarkan kedekatan nilai antar observasi [7]. Secara prinsip, teknik ini berupaya membentuk kelompok yang seragam secara internal namun berbeda secara eksternal, sehingga setiap *cluster* memiliki profil karakteristik yang khas dan dapat dibedakan secara jelas dari *cluster* lainnya [8]. Hasil clusterisasi dapat memberikan informasi terstruktur, seperti kategori produksi tinggi, sedang, dan rendah, yang bermanfaat dalam perencanaan kebijakan.

Sejumlah penelitian sebelumnya telah mengaplikasikan *k-means* untuk menjawab permasalahan pengelompokan di sektor pertanian. Santoso dkk. [9] mengklusterisasi petani agroforestri menggunakan variabel luas lahan, pengalaman, dan pendapatan, menghasilkan tiga kelompok dengan karakteristik ekonomi yang berbeda signifikan. Putra [10] menerapkan *k-means* untuk mengklasifikasikan negara tujuan ekspor kopi Indonesia berdasarkan volume dan nilai ekspor guna mendukung strategi pemasaran. Wijayanto dkk. [11] mengelompokkan wilayah di Jawa Tengah berdasarkan produktivitas padi per hektar ke dalam tiga kategori sebagai dasar distribusi sumber daya pertanian. Febriani dkk. [12] serta Riono dkk. [13] mengaplikasikan pendekatan serupa masing-masing pada data produksi beras dan pemetaan produktivitas padi di Pulau Jawa; keduanya mengonfirmasi bahwa *k-means* efektif dalam mengidentifikasi pola yang tersimpan dalam data kuantitatif pertanian.

Meskipun demikian, masih terdapat celah (*gap*) yang belum dijawab oleh literatur yang ada. Hingga saat ini, belum ditemukan penelitian yang secara spesifik mengimplementasikan *k-means clustering* untuk memetakan kapasitas produksi kopi Arabika pada level kabupaten/kota di Provinsi Jawa Barat, sekaligus melengkapinya dengan evaluasi kualitas *cluster* yang terukur menggunakan *silhouette score*. Lalu pada penelitian kali ini mengusulkan pendekatan *data-driven* yang menggabungkan *K-Means*, *Elbow Method*, dan *silhouette score* secara terintegrasi. Pendekatan ini dinilai efektif karena menghasilkan segmentasi wilayah yang berbasis data aktual sekaligus dapat diverifikasi secara kuantitatif. Kebaruan penelitian terletak pada integrasi ketiga komponen tersebut untuk memetakan distribusi produksi kopi Arabika di Jawa Barat pada level kabupaten/kota, yang sebelumnya belum pernah dilakukan. Hasil penelitian diharapkan menjadi dasar penentuan prioritas pembinaan, alokasi bantuan, serta perumusan strategi pengembangan kopi Arabika yang lebih tepat sasaran, berkelanjutan, dan berdaya saing.

## 2. Metodologi

Penelitian ini mengadopsi pendekatan kuantitatif berbasis *data mining* dengan rancangan eksperimen komputasional untuk mengkaji pola distribusi produksi kopi Arabika di Provinsi Jawa Barat menggunakan algoritma *k-means clustering*. Penerapan *data mining* dalam penelitian ini difokuskan pada pengelompokan data produksi kopi guna menghasilkan segmentasi wilayah yang dapat dijadikan acuan dalam perumusan kebijakan yang lebih terarah

[14]. Data mining efektif dalam mengolah data pengguna untuk menghasilkan rekomendasi perbaikan sistem dan pengelompokan [15]. Teknik *clustering* banyak diterapkan di berbagai bidang, termasuk pertanian dan keuangan, karena mampu mengelompokkan objek atau wilayah berdasarkan kesamaan atribut tertentu [16]. Pendekatan ini bertujuan untuk mengelompokkan wilayah berdasarkan kesamaan nilai produksi sehingga dapat diidentifikasi pola distribusi produksi antar wilayah. Proses analisis dilakukan secara sistematis melalui tahapan persiapan data, pemodelan *clustering*, serta evaluasi kualitas cluster menggunakan metrik validasi internal.

### 2.1 Dataset dan Objek Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder produksi kopi Arabika perkebunan rakyat yang bersumber dari portal *Open Data* resmi Pemerintah Provinsi Jawa Barat. Cakupan data meliputi periode 2013–2024 dengan rentang wilayah sebanyak 27 kabupaten/kota di Provinsi Jawa Barat, menghasilkan total 324 rekaman observasi.

Unit analisis pada penelitian ini adalah kabupaten/kota per tahun, sehingga setiap observasi merepresentasikan jumlah produksi kopi Arabika pada suatu wilayah dalam satu periode tertentu. Variabel utama yang digunakan dalam proses *clustering* adalah jumlah produksi kopi Arabika yang dinyatakan dalam satuan ton. Seluruh data digunakan sebagai populasi penelitian tanpa proses *sampling*, karena algoritma *k-means* termasuk ke dalam metode *unsupervised learning* yang tidak memerlukan pemisahan data *training* dan *testing*.

### 2.2 Persiapan Data dan Parameter Model

Sebelum proses *clustering* dijalankan, data dipersiapkan secara menyeluruh guna memastikan kualitas, konsistensi, dan kesesuaian strukturnya. Tahap ini mencakup identifikasi nilai hilang (*missing value*), deteksi duplikasi data, koreksi kesalahan pencatatan, serta penyesuaian format. Setelah pembersihan data selesai, dilakukan normalisasi menggunakan metode *Min-Max Scaling* untuk memetakan nilai data ke dalam rentang 0 hingga 1. Normalisasi ini diperlukan karena algoritma *k-means* mengandalkan perhitungan jarak *Euclidean* yang sensitif terhadap perbedaan skala antar variabel.

Implementasi model dilakukan menggunakan bahasa pemrograman Python di lingkungan *Google Colab* dengan dukungan pustaka *pandas*, *numpy*, *scikit-learn*, dan *matplotlib*. Parameter yang ditetapkan meliputi jumlah *cluster* sebanyak tiga ( $k=3$ ), inisialisasi sentroid menggunakan metode *k-means++*, batas iterasi maksimum sebesar 300, nilai *n\_init* mengikuti pengaturan bawaan sistem, serta pengukuran jarak berbasis *Euclidean Distance*. Konfigurasi parameter ini dirancang untuk memastikan stabilitas proses *clustering* dan optimalitas struktur kelompok yang dihasilkan.

### 2.3 Prosedur Clustering

Algoritma *k-means* diimplementasikan untuk mengelompokkan data produksi kopi Arabika berdasarkan tingkat kemiripan nilai antar observasi. Prinsip kerja algoritma ini bertumpu pada minimalisasi variasi di dalam setiap kelompok (*within-cluster variance*) dengan cara mengalokasikan tiap titik data ke sentroid terdekatnya. Fungsi objektif yang dioptimalkan dalam proses ini dinyatakan sebagai berikut:

$$J = \sum_{i=1}^k \sum_{x \in C_i} |x - \mu_i|^2 \quad (1)$$

dengan  $k$  menyatakan jumlah *cluster*,  $C_i$  adalah himpunan data pada *cluster* ke- $i$ ,  $x$  merupakan data ke- $x$ , dan  $\mu_i$  adalah centroid pada *cluster* ke- $i$ .

Penentuan jumlah *cluster* yang paling representatif dilakukan melalui *elbow method* dengan menganalisis nilai *Within-Cluster Sum of Squares* (WCSS). Metode ini bekerja dengan mengidentifikasi titik patahan (*elbow point*) pada kurva penurunan WCSS sebagai indikator jumlah *cluster* optimal. Dari hasil pengujian, penurunan WCSS yang paling tajam terjadi pada  $k=3$ , setelah itu kurva mulai melandai secara konsisten. Berdasarkan temuan ini, penelitian menetapkan tiga *cluster* sebagai jumlah optimal untuk mengkategorikan wilayah produksi kopi Arabika ke dalam tiga segmen: produksi rendah, menengah, dan tinggi.

## 2.4 Evaluasi Kinerja Model

Tahap evaluasi dilaksanakan untuk mengukur kualitas pengelompokan yang dihasilkan oleh model. Metrik validasi internal yang digunakan adalah *silhouette score*, yakni suatu ukuran yang mencerminkan seberapa baik suatu titik data berada dalam *cluster*-nya sendiri dibandingkan dengan *cluster* lain yang berdekatan [17]. Nilai *silhouette* untuk setiap observasi dihitung menggunakan persamaan berikut:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

di mana  $a(i)$  merupakan rata-rata jarak data ke- $i$  terhadap seluruh anggota dalam *cluster* yang sama, sementara  $b(i)$  adalah rata-rata jarak data ke- $i$  terhadap *cluster* lain yang paling dekat.

Adapun nilai *silhouette score* keseluruhan dihitung sebagai rata-rata seluruh nilai *silhouette* individu, yang dirumuskan sebagai:

$$S = \frac{1}{n} \sum_{i=1}^n s(i) \quad (3)$$

Rentang nilai *silhouette score* berada antara  $-1$  hingga  $1$ . Nilai yang semakin mendekati  $1$  menandakan kohesi internal *cluster* yang semakin kuat disertai separasi antar kelompok yang semakin tegas.

## 2.5 Luaran Analisis

Proses *clustering* menghasilkan sejumlah luaran utama yang menjadi bahan analisis dalam penelitian ini. Luaran tersebut mencakup pembentukan tiga segmen wilayah produksi kopi Arabika, nilai sentroid masing-masing *cluster* sebagai representasi rata-rata produksi, label *cluster* untuk setiap observasi, serta nilai *silhouette score* sebagai tolok ukur performa model. Seluruh hasil tersebut digunakan untuk mengidentifikasi pola distribusi produksi kopi Arabika antarwilayah di Provinsi Jawa Barat dan memberikan dasar analisis terhadap ketimpangan tingkat produksi antar daerah.

## 3. Hasil dan Pembahasan

Pada hasil ini disajikan hasil penerapan algoritma *k-means clustering* terhadap data produksi kopi Arabika di Jawa Barat. Proses analisis dimulai dari eksplorasi dataset, penentuan untuk jumlah *cluster* optimal menggunakan *elbow method*, proses pengelompokan wilayah, hingga evaluasi kualitas hasil *clustering* menggunakan *silhouette score*. Hasil yang diperoleh kemudian dianalisis untuk mengidentifikasi pola distribusi produksi kopi Arabika antar wilayah.

### 3.1 Deskripsi Dataset

Data yang digunakan dalam kajian ini mencakup informasi produksi kopi Arabika di Provinsi Jawa Barat yang diperoleh dari portal data resmi pemerintah provinsi setempat. Informasi ini meliputi rentang waktu dari tahun 2013 sampai 2024 dan mencakup 27 kabupaten dan kota, sehingga menghasilkan total sebanyak 324 pengamatan data.

**Tabel 1** Produksi Kopi Jawa Barat

no	nama_kabupaten_kota	produksi_tanaman (ton)	tahun
1	Kabupaten Bogor	101	2013
2	Kabupaten Sukabumi	607	2013
3	Kabupaten Cianjur	78	2013
4	Kabupaten Bandung	6637	2013
5	Kabupaten Garut	1308	2013
320	Kota Bekasi	0	2024
321	Kota Depok	0	2024
322	Kota Cimahi	0	2024
323	Kota Tasikmalaya	0	2024
324	Kota Banjar	0	2024

Sumber: Open Data Pemerintah Provinsi Jawa Barat

Sebagian sampel dataset ditampilkan pada Tabel 1 guna memberikan gambaran awal mengenai struktur data yang digunakan. Tabel tersebut memperlihatkan lima rekaman pertama dan lima rekaman terakhir dari keseluruhan dataset.

Setiap observasi merepresentasikan jumlah produksi kopi Arabika pada suatu wilayah dalam satu periode waktu tertentu. Variabel utama yang digunakan dalam proses *clustering* adalah jumlah produksi kopi Arabika yang dinyatakan dalam satuan ton. Sebelum dilakukan pemodelan, data awalnya dilakukan tahap pembersihan dan normalisasi untuk memastikan konsistensi serta kesesuaian skala data dalam proses analisis *clustering*.

### 3.2 Data Preprocessing

Tahap pra-pemrosesan data dilaksanakan untuk memastikan bahwa dataset yang digunakan berada dalam kondisi bersih, seragam, dan siap untuk dianalisis menggunakan teknik pengelompokan. Proses ini merupakan fondasi penting dalam analisis berbasis data, mengingat kualitas data sangat menentukan akurasi dan keandalan model yang dihasilkan.

Langkah awal yang dilaksanakan adalah pengecekan kelengkapan data untuk memastikan tidak ada nilai yang hilang, data yang terduplikasi, atau kesalahan pencatatan dalam kumpulan data.

```
print("Jumlah data sebelum cleaning:", df.shape[0])
print("Jumlah data setelah cleaning:", df_clean.shape[0])
```

---

```
Jumlah data sebelum cleaning: 324
Jumlah data setelah cleaning: 324
```

**Gambar 1.** Hasil Jumlah Dataset Sebelum Dan Setelah *Cleaning*

Hasil pemeriksaan menunjukkan bahwa dataset produksi kopi Arabika yang digunakan bebas dari nilai kosong dan telah memiliki format yang seragam, sehingga dapat langsung diproses pada tahap berikutnya.

Selanjutnya, normalisasi diterapkan pada variabel produksi kopi Arabika menggunakan metode *Min-Max Normalization*, yang mentransformasikan seluruh nilai data ke dalam rentang 0 hingga 1. Tujuan normalisasi ini adalah menyeragamkan skala data agar perbedaan rentang nilai yang besar tidak mendistorsi perhitungan jarak pada algoritma *k-means*.

```
scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(X)
```

---

```
X_scaled[:5]
```

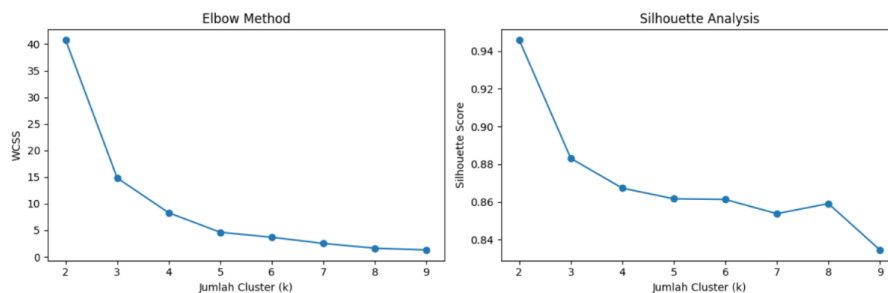
```
array([[1.00000000e+00],
       [1.19106574e-01],
       [1.15393891e-05],
       [3.58757300e-02],
       [3.01131899e-03]])
```

**Gambar 2.** *Source Code* dan Hasil Normalisasi Data Menggunakan *Min-Max Scaling*

Hasil normalisasi menunjukkan bahwa setiap observasi produksi telah berhasil dipetakan ke dalam skala 0 hingga 1. Nilai yang mendekati 1 mencerminkan produksi yang relatif lebih tinggi, sedangkan nilai yang mendekati 0 menunjukkan produksi yang relatif lebih rendah. Dengan skala yang terstandarisasi, perhitungan jarak antar titik data dalam proses *clustering* menjadi lebih proporsional dan akurat.

### 3.3 Penentuan Jumlah Cluster Menggunakan *Elbow Method*

Salah satu tahapan krusial dalam penerapan algoritma *k-means* adalah menetapkan jumlah *cluster* yang paling sesuai. Pemilihan jumlah *cluster* yang keliru berpotensi menghasilkan pengelompokan yang tidak akurat dan sulit diinterpretasikan. Oleh sebab itu, penelitian ini menggunakan *elbow method* sebagai panduan dalam menentukan jumlah *cluster* yang paling optimal.



**Gambar 3.** Grafik Dari *Elbow Method* dan *Silhouette Score*

Grafik pada Gambar 3 menyajikan hasil evaluasi jumlah *cluster* menggunakan *elbow method* dan *silhouette score*. Kurva *elbow method* memperlihatkan bahwa nilai WCSS mengalami penurunan yang sangat tajam dari  $k=2$  menuju  $k=3$ , kemudian penurunannya terus berlanjut namun dengan laju yang semakin melambat hingga  $k=4$ . Di atas  $k=4$ , kurva WCSS cenderung mendatar, mengindikasikan terbentuknya titik siku (*elbow point*) pada  $k=3$ .

Sementara itu, evaluasi *silhouette score* memperlihatkan nilai tertinggi pada  $k=2$  sebesar 0,946, diikuti  $k=3$  sebesar 0,883. Kendati nilai maksimum dicapai pada  $k=2$ , penggunaan dua *cluster* dinilai kurang mampu merepresentasikan keragaman tingkat produksi kopi Arabika di lapangan. Dengan pertimbangan tersebut,  $k=3$  ditetapkan sebagai jumlah *cluster* optimal karena masih menghasilkan nilai *silhouette score* yang tinggi sekaligus memberikan segmentasi yang lebih informatif dan representatif.

**Table 2.** Hasil Evaluasi Penentuan Jumlah *Cluster*

<b>k</b>	<b>WCSS</b>	<b>Silhouette Score</b>
2	40.76	<b>0.946</b>
3	14.83	<b>0.883</b>
4	8.29	<b>0.867</b>
5	4.63	<b>0.862</b>
6	3.69	<b>0.861</b>
7	2.53	<b>0.854</b>
8	1.63	<b>0.859</b>
9	1.31	<b>0.835</b>

Berdasarkan Tabel 2, penurunan WCSS paling signifikan terjadi dari  $k=2$  ke  $k=3$ , yakni dari 40,76 menjadi 14,83. Setelah itu, penurunan masih terjadi namun semakin melambat seiring bertambahnya nilai  $k$ . Pola ini mengonfirmasi terbentuknya *elbow point* pada  $k=3$  sebagai representasi struktur data yang paling optimal. Nilai *silhouette score* sebesar 0,883 pada  $k=3$  mengindikasikan kualitas pengelompokan yang sangat memuaskan, ditandai dengan separasi antar *cluster* yang tegas.

Atas dasar kombinasi kedua analisis tersebut, penelitian ini menetapkan  $k=3$  sebagai jumlah *cluster* yang digunakan. Dengan demikian, proses *clustering* selanjutnya dilaksanakan menggunakan algoritma *k-means* dengan tiga *cluster* untuk mengelompokkan wilayah produksi kopi Arabika di Jawa Barat berdasarkan kapasitas produksinya.

### 3.4 Hasil *Clustering* Menggunakan *K-Means*

Setelah jumlah *cluster* optimal ditetapkan pada  $k=3$ , proses pengelompokan dijalankan menggunakan algoritma *k-means* dengan metode inisialisasi *k-means++* guna memperoleh pemilihan titik pusat awal yang lebih berkualitas. Parameter *random\_state* digunakan agar hasilnya tetap konsisten, sedangkan *n\_init* diatur menjadi 10 supaya dapat mendapatkan hasil *clustering* yang terbaik dari beberapa kali percobaan awal.

```

from sklearn.cluster import KMeans

kmeans_final = KMeans(
    n_clusters=k_optimal,
    init='k-means++',
    random_state=42,
    n_init=10
)

df['cluster'] = kmeans_final.fit_predict(X_scaled)

```

**Gambar 4.** Source Code Untuk Penerapan *Algoritma K-Means*

Proses *clustering* diterapkan pada data produksi kopi Arabika yang telah dinormalisasi. Hasilnya berupa label *cluster* yang dilekatkan pada setiap observasi, sehingga masing-masing kabupaten/kota memperoleh kategori *cluster* yang mencerminkan kapasitas produksinya.

**Tabel 3.** Jumlah *Cluster* dan Jumlah Wilayah

<i>Cluster</i>	Jumlah Wilayah
0	<b>294</b>
1	<b>12</b>
2	<b>18</b>

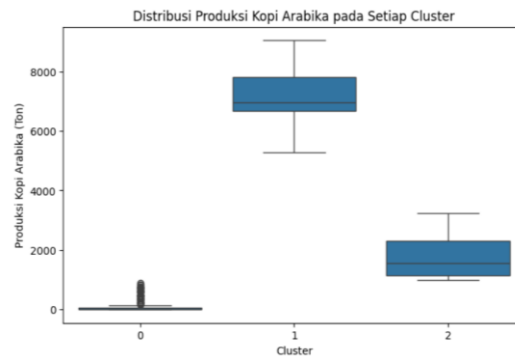
Nilai sentroid yang diperoleh dari proses *clustering* mengungkap perbedaan karakteristik produksi yang signifikan antar *cluster*, pada *cluster* 0 memiliki jumlah anggota paling dominan dengan 294 wilayah, sedangkan *cluster* 1 serta *cluster* 2 masing-masing terdiri dari 12 dan 18 wilayah. Keadaan ini menunjukkan bahwa sebagian besar daerah di Provinsi Jawa Barat memiliki tingkat produksi kopi Arabika yang relatif serupa sehingga terkonsentrasi pada satu kelompok utama, sementara sebagian kecil wilayah memiliki karakteristik produksi yang berbeda secara signifikan. Perbedaan distribusi ini mengindikasikan adanya variasi tingkat produksi kopi Arabika antar wilayah, yang selanjutnya dapat dianalisis lebih lanjut melalui nilai *centroid* sebagai representasi rata-rata produksi pada setiap *cluster*.

**Tabel 4.** Nilai *Centroid* Hasil *Clustering*

<i>Cluster</i>	Rata-rata Produksi (Ton)	Kategori Produksi
<i>Cluster</i> 0	76.20	<b>Rendah</b>
<i>Cluster</i> 2	1835.25	<b>Sedang</b>
<i>Cluster</i> 1	7221.64	<b>Tinggi</b>

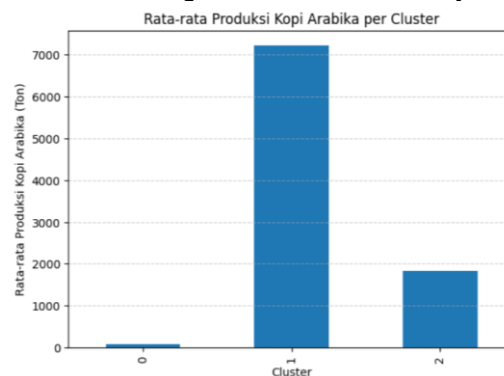
Berdasarkan nilai *centroid* yang diperoleh dari proses *clustering* menggunakan algoritma *k-means*, setiap *cluster* menunjukkan karakteristik tingkat produksi kopi Arabika yang berbeda. *Cluster* pertama merepresentasikan wilayah dengan tingkat produksi rendah dengan rata-rata produksi sekitar 76,20 ton, sedangkan *cluster* kedua menunjukkan wilayah dengan tingkat produksi sedang dengan rata-rata produksi sekitar 1.835,25 ton. Sementara itu, *cluster* ketiga merepresentasikan wilayah dengan tingkat produksi tinggi dengan rata-rata produksi mencapai 7.221,64 ton. Perbedaan nilai *centroid* yang cukup signifikan antar *cluster* menunjukkan adanya variasi tingkat produksi kopi Arabika yang besar antar wilayah di Provinsi Jawa Barat. Hasil pengelompokan ini memberikan gambaran mengenai distribusi produksi kopi Arabika sehingga dapat membantu mengidentifikasi wilayah dengan potensi produksi tinggi maupun wilayah dengan tingkat produksi relatif rendah sebagai dasar dalam perencanaan pengembangan sektor perkebunan kopi Arabika di Provinsi Jawa Barat.

Grafik scatter plot. Visualisasi ini bertujuan untuk memperlihatkan distribusi data produksi kopi Arabika berdasarkan *cluster* yang dihasilkan oleh algoritma *k-means*.



**Gambar 5.** Visualisasi *Boxplot* Setiap *Cluster*

Boxplot distribusi produksi kopi Arabika pada masing-masing *cluster* menunjukkan perbedaan yang sangat jelas antar kelompok. *Cluster 0* memiliki median produksi yang sangat rendah dengan variasi yang relatif kecil, meskipun terdapat beberapa *outlier*. *Cluster 2* menunjukkan tingkat produksi sedang dengan rentang distribusi yang lebih lebar dibanding *cluster 0*. Sementara itu, *Cluster 1* memiliki median produksi tertinggi dengan variasi yang cukup besar, mencerminkan wilayah sentra utama produksi kopi Arabika. Tidak terlihat adanya tumpang tindih distribusi antar *cluster*, yang menunjukkan bahwa hasil *clustering* memiliki pemisahan yang baik dan konsisten dengan evaluasi sebelumnya.



**Gambar 6.** Visualisasi Rata-Rata Produksi per-*Cluster*

Grafik rata-rata produksi per-*cluster* pada Gambar 6 semakin memperjelas perbedaan antar segmen. *Cluster 0* mencatat rata-rata sekitar 76 ton (kategori rendah), *cluster 2* sekitar 1.835 ton (kategori menengah), dan *cluster 1* mencapai sekitar 7.221 ton (kategori tinggi). Perbedaan nilai rata-rata yang cukup jauh antar *cluster* menunjukkan bahwa hasil pengelompokan mampu mengidentifikasi segmentasi produksi secara jelas dan konsisten dengan hasil evaluasi sebelumnya. Karena *centroid* dalam *k-means* merepresentasikan nilai rata-rata dalam ruang fitur yang telah dinormalisasi. Visualisasi mean dalam satuan asli memudahkan interpretasi hasil secara kontekstual dan praktis. Terdapat ketimpangan produksi kopi Arabika antarwilayah, di mana hanya sebagian kecil wilayah yang menjadi sentra produksi utama.

### 3.5 Evaluasi Cluster Menggunakan *Silhouette Score*

Penilaian kualitas hasil pengelompokan dilakukan menggunakan metrik *silhouette score*. Metrik ini mengukur ketepatan posisi suatu titik data dalam *clusternya* dengan cara membandingkan tingkat kedekatannya terhadap sesama anggota *cluster* maupun terhadap anggota *cluster* lain yang berdekatan. Nilai *silhouette score* berkisar antara -1 hingga 1; nilai yang mendekati 1 menandakan kepaduan *cluster* yang tinggi dan separasi yang tegas, nilai mendekati nol mengisyaratkan posisi data yang ambigu di antara dua *cluster*, sedangkan nilai negatif menandakan kemungkinan kesalahan pengelompokan. Berdasarkan hasil evaluasi pada  $k=3$ , diperoleh *silhouette score* global sebesar 0,883, yang mengkonfirmasi bahwa model menghasilkan pengelompokan berkualitas sangat baik dengan separasi antar *cluster* yang jelas dan tingkat keseragaman internal yang kuat.

```

from sklearn.metrics import silhouette_samples
import numpy as np

silhouette_vals = silhouette_samples(X_scaled, df['cluster'])
df['silhouette'] = silhouette_vals

silhouette_per_cluster = df.groupby('cluster')['silhouette'].mean().reset_index()
silhouette_per_cluster
    
```

cluster	silhouette
0	0.914021
1	0.778912
2	0.447952

Gambar 7. Hasil Silhouette Score per Cluster

Secara per-cluster, cluster 0 memperoleh nilai 0,914 dan cluster 1 sebesar 0,779, keduanya masuk kategori sangat baik. Sementara cluster 2 memperoleh nilai 0,448 yang masuk kategori cukup baik.

### 3.6 Penyajian Hasil

Pada tahap ini bertujuan untuk menyajikan hasil analisis kepada pengguna atau pemangku kepentingan agar dapat dimanfaatkan sebagai dasar pengambilan keputusan.

Tabel 4. Hasil Clustering Wilayah Produksi Kopi Arabika

Kategori Cluster	Wilayah
Produksi Rendah	Kabupaten Cirebon, Kota Bandung, Kabupaten Karawang, Kabupaten Sukabumi, Kota Bekasi, Kabupaten Indramayu, Kota Cimahi, Kabupaten Cianjur, Kabupaten Bandung Barat, Kota Depok, Kabupaten Subang, Kabupaten Bogor, Kota Tasikmalaya, Kabupaten Kuningan, Kabupaten Purwakarta, Kota Bogor, Kabupaten Pangandaran, Kabupaten Majalengka, Kota Banjar, Kabupaten Tasikmalaya, Kabupaten Bekasi, Kota Sukabumi, Kabupaten Ciamis, Kota Cirebon
Produksi Sedang	Kabupaten Bandung
Produksi Tinggi	Kabupaten Garut, Kabupaten Bandung Barat, Kabupaten Sumedang

Penerapan *k-means clustering* pada data produksi kopi Arabika di Jawa Barat berhasil membagi daerah ke dalam tiga segmen berdasarkan kapasitas produksinya: rendah, menengah, dan tinggi. Sebagian besar wilayah masuk dalam kategori produksi rendah, satu wilayah dalam kategori menengah, dan beberapa wilayah lainnya dalam kategori tinggi.

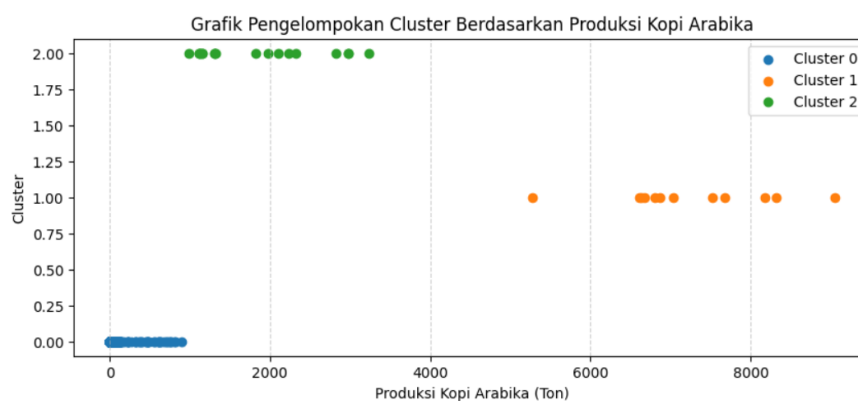
Hasil *clustering* menunjukkan sebagian besar pada wilayah berada pada kategori Produksi Rendah, sedangkan hanya satu wilayah termasuk dalam kategori Produksi Sedang dan beberapa wilayah lainnya berada pada kategori Produksi Tinggi. Kondisi ini menunjukkan bahwa distribusi produksi kopi Arabika di Jawa Barat masih belum merata dan cenderung terkonsentrasi pada wilayah tertentu.

Dominasi wilayah pada kategori Produksi Rendah mengindikasikan bahwa pengembangan produksi kopi Arabika di Jawa Barat masih memiliki potensi untuk ditingkatkan, khususnya pada wilayah yang belum menjadi sentra produksi utama.

**Tabel 5.** Evaluasi Model *Silhouette Score*

<i>Cluster</i>	<i>Silhouette Score</i>	Kategori Kualitas
0	0.914	Sangat Baik
1	0.779	Sangat Baik
2	0.448	Cukup Baik
Global	0.883	Sangat Baik

Merujuk pada hasil evaluasi menggunakan metrik *silhouette score*, model secara keseluruhan memperoleh skor sebesar 0,883 yang tergolong dalam kategori sangat baik ( $\geq 0,50$ ). Capaian ini menunjukkan bahwa algoritma *k-means* berhasil membentuk struktur *cluster* yang solid, di mana setiap kelompok memiliki kohesi internal yang kuat sekaligus separasi yang tegas terhadap kelompok lainnya. Pada level per-*cluster*, *cluster 0* mencatatkan nilai 0,914 dan *cluster 1* sebesar 0,779 keduanya masuk dalam kategori sangat baik, sementara *cluster 2* memperoleh nilai 0,448 yang terklasifikasi dalam kategori cukup baik. Meskipun angkanya lebih kecil dibandingkan kelompok lain, nilai tersebut menunjukkan bahwa cara pengelompokan datanya masih tepat dan tidak ada tumpang tindih antar kelompok yang berarti, sehingga jumlah kelompok yang digunakan dalam penelitian ini bisa dianggap sudah tepat untuk mewakili pola produksi kopi Arabika. Selanjutnya kebagian grafiki pengelompokan *cluster*.

**Gambar 8.** Grafik Pengelompokan *Cluster* Berdasarkan Produksi

Visualisasi hasil *clustering* ditampilkan dalam bentuk *scatter plot* yang menggambarkan distribusi nilai produksi kopi Arabika berdasarkan metode *k-means*. Sumbu horizontal (X) menunjukkan nilai produksi kopi Arabika (ton), sedangkan sumbu vertikal (Y) menunjukkan label *cluster*. Setiap titik merepresentasikan wilayah kabupaten/kota dengan warna yang berbeda sesuai dengan kelompok *cluster* yang terbentuk. Berdasarkan grafik tersebut, terlihat bahwa pengelompokan data terbentuk dengan pemisahan yang cukup jelas antar *cluster*, di mana *cluster 0* berada pada rentang produksi terendah, *cluster 2* pada rentang produksi menengah, dan *cluster 1* pada tingkat produksi tertinggi. Tidak terdapat tumpang tindih yang signifikan antar *cluster*, sehingga menunjukkan bahwa proses *clustering* mampu membentuk kelompok data yang relatif homogen dengan perbedaan yang jelas antar *cluster*. Visualisasi ini juga memperkuat hasil evaluasi *silhouette score* yang menunjukkan kualitas *clustering* dalam kategori sangat baik serta mengindikasikan adanya perbedaan tingkat produksi kopi Arabika yang cukup signifikan antar wilayah.

Tabel 6. Karakteristik dan Interpretasi per-*cluster*

<i>Cluster</i>	Kategori Produksi	Karakteristik Utama	Implikasi Pengembangan
<i>Cluster</i> 0	Produksi Rendah	Sebagian besar wilayah dengan tingkat produksi kopi Arabika relatif rendah dan belum menjadi sentra utama produksi.	Perlu peningkatan produktivitas melalui pengembangan budidaya, perluasan lahan potensial, dan dukungan kebijakan.
<i>Cluster</i> 2	Produksi Sedang	Wilayah dengan tingkat produksi yang relatif stabil dan berada pada kategori menengah.	Memiliki potensi untuk ditingkatkan menjadi wilayah produksi tinggi melalui peningkatan teknologi dan pengelolaan produksi.
<i>Cluster</i> 1	Produksi Tinggi	Wilayah dengan tingkat produksi kopi Arabika tertinggi dan berperan sebagai sentra utama produksi.	Dapat difokuskan pada penguatan rantai nilai, peningkatan kualitas, dan pengembangan industri hilir kopi.

Hasil segmentasi menggunakan *k-means* membagi daerah penghasil kopi Arabika di Jawa Barat ke dalam tiga kelompok yang mencerminkan perbedaan kapasitas produksi antar wilayah. *cluster* 0 mengidentifikasi kawasan dengan volume produksi kopi Arabika paling rendah; mayoritas kabupaten dan kota masuk dalam kelompok ini, mengindikasikan bahwa sebagian besar wilayah belum berkembang menjadi sentra produksi yang signifikan. *Cluster* 2 menempati posisi menengah dengan tingkat produksi yang relatif stabil, sekaligus menyimpan potensi yang dapat dikembangkan lebih jauh melalui intervensi yang tepat. Adapun *cluster* 1 merepresentasikan wilayah dengan kapasitas produksi tertinggi yang berperan sebagai pusat utama produksi kopi Arabika di Jawa Barat. Perbedaan karakteristik yang nyata antar ketiga kelompok ini menegaskan adanya ketimpangan produktivitas yang perlu disikapi secara strategis, sekaligus menjadikan hasil pengelompokan ini sebagai fondasi empiris yang kuat dalam merancang kebijakan pengembangan komoditas kopi Arabika yang lebih presisi dan berbasis data.

### 3.7 Pembahasan

Hasil analisis menunjukkan bahwa algoritma *k-means clustering* mampu menjawab permasalahan yang diidentifikasi di bagian pendahuluan, yakni belum optimalnya pemanfaatan data produksi kopi Arabika sebagai landasan pengambilan kebijakan berbasis bukti. Sebelumnya, data produksi cenderung hanya disajikan dalam format statistik deskriptif tanpa penggalan pola yang memadai. Melalui proses *clustering*, data berhasil dikategorikan ke dalam tiga segmen produksi rendah, menengah, dan tinggi yang membuktikan bahwa data produksi tidak hanya berfungsi sebagai dokumen statistik, melainkan dapat diolah menjadi informasi strategis untuk memetakan kondisi asli pada tiap wilayah. Adanya segmentasi ini memudahkan pemerintah daerah dan pemangku kepentingan dalam menentukan prioritas kebijakan, seperti intervensi di wilayah berproduksi rendah, penguatan di wilayah menengah, serta optimalisasi sentra produksi tinggi.

Perbedaan nilai centroid antar *cluster* menunjukkan adanya kesenjangan produksi yang cukup besar antarwilayah di Provinsi Jawa Barat. *Cluster* produksi rendah memiliki rata-rata produksi sebesar 76,20 ton, *cluster* produksi sedang sebesar 1.835,25 ton, sedangkan *cluster* produksi tinggi mencapai 7.221,64 ton. Selisih nilai tersebut mengindikasikan bahwa produksi kopi Arabika di Jawa Barat masih terkonsentrasi pada wilayah tertentu dan belum tersebar secara merata. Temuan ini penting karena memberikan dasar empiris bahwa strategi pengembangan tidak dapat dilakukan dengan pendekatan yang sama untuk seluruh daerah. Wilayah dengan produksi rendah membutuhkan intervensi pada aspek budidaya, akses sarana produksi, dan peningkatan kapasitas petani, sedangkan wilayah dengan produksi tinggi lebih

membutuhkan penguatan kualitas produk, efisiensi rantai pasok, dan pengembangan industri hilir.

Dari perspektif kualitas model, *silhouette score* global sebesar 0,883 membuktikan bahwa hasil pengelompokan sangat memuaskan, dengan kohesi internal *cluster* yang tinggi dan separasi antar kelompok yang tegas. Secara rinci, *cluster* 0 memperoleh nilai 0,914 dan *cluster* 1 sebesar 0,779 keduanya masuk kategori sangat baik, sementara *cluster* 2 memperoleh nilai 0,448 yang masuk kategori cukup baik. Meskipun salah satu *cluster* memiliki nilai lebih rendah, secara keseluruhan model tetap representatif dalam menggambarkan pola distribusi produksi kopi Arabika di Jawa Barat, dan penggunaan tiga *cluster* dinilai sudah tepat untuk merepresentasikan struktur data yang ada.

Temuan penelitian ini sejalan dengan berbagai studi terdahulu yang membuktikan efektivitas algoritma *k-means* di berbagai sektor, seperti pada sektor Sosial [18], sektor ekonomi dan usaha tani [9]. Putra mengonfirmasi relevansi metode yang sama untuk klasifikasi pasar ekspor kopi Indonesia [10]. Studi di Jawa Tengah menunjukkan bahwa *k-means* mampu mengidentifikasi segmen wilayah secara objektif berdasarkan data produktivitas padi [11]. Temuan serupa diperoleh pada analisis produksi beras di Sumatera Utara [12] dan pemetaan potensi padi di beberapa provinsi [13]. Membangun di atas temuan-temuan tersebut, penelitian ini memperluas aplikasi *k-means* ke konteks produksi kopi Arabika tingkat kabupaten/kota di Jawa Barat, sehingga memperkaya bukti empiris bahwa metode *clustering* relevan dan andal untuk analisis spasial di sektor pertanian.

Kontribusi ilmiah penelitian ini terletak pada integrasi terpadu antara *k-means*, *elbow method*, dan *silhouette score* dalam menganalisis distribusi produksi kopi Arabika secara regional. Pendekatan ini menunjukkan bahwa analisis *data mining* dapat dimanfaatkan tidak hanya untuk klasifikasi bisnis dan pemasaran, tetapi juga untuk mendukung perencanaan sektor pertanian berbasis data. Selain itu, model yang dihasilkan bersifat replikatif dan dapat diterapkan pada komoditas lain, seperti teh, kakao, padi, maupun hortikultura di wilayah berbeda. Dari sisi praktis, hasil penelitian dapat digunakan sebagai dasar penentuan prioritas pembinaan, alokasi bantuan, pengembangan sentra produksi, serta penyusunan kebijakan pemerataan produksi kopi Arabika di Jawa Barat secara lebih tepat sasaran, berkelanjutan, dan adaptif terhadap karakteristik masing-masing wilayah.

#### 4. Simpulan

Penelitian ini mengkaji penerapan metode *k-means clustering* untuk mengelompokkan wilayah penghasil kopi Arabika di Provinsi Jawa Barat berdasarkan pola sebaran kapasitas produksinya. Analisis yang dilakukan berhasil membagi seluruh kabupaten dan kota ke dalam tiga segmen yang berbeda, yakni segmen berproduksi rendah, menengah, dan tinggi, sesuai dengan volume kopi Arabika yang dihasilkan masing-masing wilayah. Sebagian besar daerah terkonsentrasi pada segmen produksi rendah, sementara hanya segelintir wilayah yang masuk ke dalam kategori produksi tinggi dan berperan sebagai sentra utama penghasil kopi Arabika. Penilaian kualitas model menggunakan *silhouette score* menghasilkan skor global sebesar 0,883, yang mengonfirmasi bahwa struktur pengelompokan yang terbentuk memiliki kualitas yang sangat memuaskan, ditandai dengan kohesi internal yang kuat dan separasi antar kelompok yang tegas. Temuan ini secara empiris membuktikan adanya ketimpangan kapasitas produksi kopi Arabika yang nyata antar daerah di Jawa Barat. Secara keseluruhan, penelitian ini menghadirkan landasan analitis yang kokoh untuk mendukung perumusan strategi pengembangan komoditas kopi Arabika yang lebih presisi, berbasis bukti, dan berorientasi keberlanjutan, guna mendorong pemerataan produksi sekaligus memperkuat daya saing kopi Arabika Jawa Barat di tingkat regional maupun nasional.

#### Daftar Referensi

- [1] C. N. Maghfiroh, "Proyeksi Produktivitas Kopi Robusta dan Arabika 2024-2033 Berdasarkan Status Pengusahaan," *Agrosaintifika: Jurnal Ilmu-Ilmu Pertanian*, vol. 7, no. 1, pp. 7-16, 2024.
- [2] I. Andriyani and M. M. Ubaidillah, "Penilaian Indikasi Geografis Pegunungan Hyang Argopuro dan Kesesuaian Lahannya untuk Budidaya Kopi," *agriTECH*, vol. 42, no. 2, p. 131, Jul. 2022, doi: 10.22146/agritech.60195.
- [3] N. Oktaviani *et al.*, "Perubahan Iklim Mikro dan Produksi Kopi Arabika (*Coffea arabica* L.) pada Daerah Aktivitas Geothermal PLTP Kamojang di Kabupaten Bandung," *Jurnal Agrikultura*, vol. 35, no. 3, pp. 400-412, 2024.

- [4] T. W. Rahmawati, S. I. Santoso, and S. Nurfadillah, "Analisis Trend Luas Lahan dan Produksi Kopi di Indonesia," *Agromedia: Berkala Ilmiah Ilmu-Ilmu Pertanian*, vol. 42, pp. 145–153, 2024.
- [5] A. Savira, N. N. Azizah, N. A. A. Anhar, R. S. Habsah, and V. A. Mumtazah, "Analisis Faktor-Faktor yang Mempengaruhi Kuantitas Kopi Indonesia Terhadap Ekspor ke Amerika Serikat Periode 2015-2021," *Jurnal Economina*, vol. 2, no. 1, pp. 1242–1253, 2023, doi: 10.55681/economina.v2i1.278.
- [6] F. A. Saputra and A. Iskandar, "Data Mining Penerapan Asosiasi Apriori Dalam Penentuan Pola Penjualan," *Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 4, pp. 778–788, Aug. 2023, doi: 10.47065/josyc.v4i4.4043.
- [7] D. Selvia and A. Ramadhanu, "Implementasi Metode K-Means Clustering dengan Teknik Pengolahan Citra untuk Mengidentifikasi Jenis Sepatu," *Jurnal Informatika Teknologi dan Sains*, vol. 7, no. 1, pp. 361–366, 2025.
- [8] L. M. Harahap, W. Fuadi, L. Rosnita, E. Darnila, and R. Meiyanti, "Klastering Sayuran Unggulan Menggunakan Algoritma K-Means," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 3, pp. 567-579, Dec. 2022, doi: 10.28932/jutisi.v8i3.5277.
- [9] T. Santoso, A. Darmawan, N. Sari, M. A. F. Syadza, E. C. B. Himawan, and W. A. Rahman, "Clusterization of Agroforestry Farmers using K-Means Cluster Algorithm and Elbow Method," *Jurnal Sylva Lestari*, vol. 11, no. 1, pp. 107–122, Jan. 2023, doi: 10.23960/jsi.v11i1.646.
- [10] I. K. N. Putra, "Implementasi Metode K-Means Clustering Pada Ekspor Kopi Berdasarkan Negara Tujuan," *Jurnal Ilmu Komputer dan Bisnis*, vol. 14, no. 1, pp. 45–55, May 2023, doi: 10.47927/jikb.v14i1.332.
- [11] S. Wijayanto and M. Y. Fathoni, "Pengelompokan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means," *JUPITER: Jurnal Penelitian Ilmu dan Teknologi Komputer*, vol. 13, no. 2, pp. 212–219, 2021, doi: 10.5281/3918.jupiter.2021.10.
- [12] I. Febriani, M. Safii, and O. Alfina, "Implementasi Data Mining Peningkatan Produksi Beras Menggunakan Metode K-Means Clustering," *Majalah Ilmiah METHODODA*, vol. 12, no. 3, pp. 258–268, Dec. 2022, doi: 10.46880/methododa.Vol12No3.pp258-268.
- [13] S. H. Riono, Rizdania, and H. M. Indahsari, "Analisis Clustering Produktivitas Padi Daerah- Daerah di Tiga Provinsi Besar Pulau Jawa Menggunakan Algoritma K-Means Berbasis Python," *SMARTICS Journal*, vol. 10, no. 2, pp. 39–48, Nov. 2024, doi: 10.21067/smartics.v10i2.11826.
- [14] S. S. Hilabi *et al.*, "Pemanfaatan Data Analitik dalam Big Data: Studi Kasus Implementasi di Pemerintahan," *Jurnal Teknik Informatika dan Sistem Informasi*, Vol. 12, No. 1, pp. 378-390, 2025.
- [15] B. Huda *et al.*, "Analisis Sentimen E-Learning X Terhadap Antarmuka Pengguna Menggunakan Kombinasi Multinomial Naive Bayes dan Pendekatan Design Thinking," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, Vol. 11, No. 4, Agustus 2024, pp. 895-902, doi: 10.25126/jtiik.1147678.
- [16] B. O. Ndasak, P. Batarius, Y. C. H. Siki, "Klasterisasi Data Hasil Produksi Pertanian dan Peternakan Provinsi Nusa Tenggara Timur Menggunakan Metode K-Means," *JIP: Jurnal Informatika Polinema.*, Volume 9, Edisi 4, pp. 415-426, Agustus 2023.
- [17] M. R. Syahkur, D. Hartama, and S. Solikhun, "Evaluasi Jumlah Cluster pada Algoritma K-Means++ Menggunakan Silhouette dan Elbow dengan Validasi Nilai DBI dalam Mengelompokkan Gizi Balita," *JST (Jurnal Sains dan Teknologi)*, vol. 13, no. 3, pp. 487–496, Oct. 2024, doi: 10.23887/jstundiksha.v13i3.86419.
- [18] B. Susarianto, & T. Nizami, "Penentuan Penerima Bantuan Sosial Beras untuk Masyarakat Miskin Menggunakan Metode K-Means." *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 12, no. 2, pp. 772-783, 2023.