

## Pengelompokan Produksi Padi dan Beras Provinsi Jawa Timur dengan Metode Agglomerative Hierarchical Clustering

Hasim Azari<sup>1\*</sup>, Dwi Hartanti<sup>2</sup>, Aprilisa Arum Sari<sup>3</sup>

<sup>1,2,3</sup>Program Studi Teknik Informatika, Universitas Duta Bangsa Surakarta

\*hasimazari24@gmail.com

### Abstrak

Padi merupakan tanaman pangan penghasil beras yang banyak dikonsumsi untuk kebutuhan pokok dan menjadi sumber penghasilan bagi masyarakat Indonesia. Sering kali kebutuhan beras terus meningkat dan tidak bisa hanya mengandalkan produksi dalam negeri sehingga perlu optimalisasi wilayah penghasil padi. Jawa Timur merupakan salah satu provinsi penghasil padi terbesar di Indonesia. Pengelompokan daerah penghasil padi perlu dilakukan guna mengetahui potensi produksi dan daerah yang memerlukan peningkatan. Penelitian ini bertujuan mengelompokkan produksi padi dan beras berdasarkan kabupaten/kota di Provinsi Jawa Timur. Metode yang digunakan adalah Agglomerative Hierarchical Clustering (AHC) dengan algoritma terbaik yaitu average linkage berdasarkan korelasi cophenetic tertinggi sebesar 0.8478. Hasil penelitian menunjukkan 4 cluster : cluster 1 (tinggi) sebanyak 4 kabupaten, cluster 2 (rendah) sebanyak 18 kabupaten, cluster 3 (sedang) sebanyak 7 kabupaten, dan cluster 4 (sangat rendah) sebanyak 9 kota. Uji silhouette coefficient sebesar 0.6347, mendekati 1 sehingga menunjukkan clustering yang baik. Hasil ini dapat digunakan sebagai acuan bagi pemerintah dan masyarakat dalam optimalisasi produksi padi dan beras guna menjaga ketahanan pangan negara.

**Kata kunci:** agglomerative hierarchical clustering, beras, cluster, clustering, data mining, padi

### Abstract

*Paddy is a food crop that produces rice which is widely consumed for basic needs and has become a source of income for the Indonesian people. Often the need for rice continues to increase and cannot only rely on domestic production, so it is necessary to optimize paddy producing areas. Grouping rice-producing areas needs to be done to identify production potentials and areas that require improvement. This research aims to group paddy and rice production based on districts/cities in East Java Province. The method used is Agglomerative Hierarchical Clustering (AHC) with the best algorithm, namely average linkage based on the highest cophenetic correlation of 0.8478. The research results showed 4 clusters: cluster 1 (high) with 4 districts, cluster 2 (low) with 18 districts, cluster 3 (medium) with 7 districts, and cluster 4 (very low) with 9 cities. The silhouette coefficient test is 0.6347, close to 1, thus indicating good clustering. These results can be used as a reference for the government and society in optimizing paddy and rice production to maintain the country's food security.*

**Keywords:** agglomerative hierarchical clustering, rice, cluster, clustering, data mining, paddy

### 1. Pendahuluan

Padi merupakan tanaman pertanian penghasil beras yang kemudian diolah menjadi nasi untuk memenuhi kebutuhan pokok sehari-hari [1]. Selain dapat memenuhi kebutuhan pokok, padi juga menjadi sumber penghasilan utama bagi sebagian masyarakat Indonesia. Hal ini

dikarenakan kondisi geografis dan iklim di Indonesia, dimana sebagian besar wilayahnya memiliki lahan subur dan curah hujan yang cukup untuk mendukung pertumbuhan tanaman padi [2]. Pertanian padi telah memberikan kontribusi besar terhadap perekonomian nasional dan hampir diproduksi di seluruh wilayah Indonesia. Namun,

sering kali konsumsi beras terus meningkat hingga tidak cukup mengandalkan hasil produksi padi dalam negeri, akibatnya harus mengimpor beras agar ketahanan pangan terjaga. Pemerintah perlu mengoptimalkan produksi padi dan beras untuk mencapai kebutuhan masyarakat sehingga mampu mengurangi ketergantungan impor beras [3].

Provinsi Jawa Timur merupakan salah satu wilayah penghasil padi terbesar di Indonesia. Menurut Badan Pusat Statistik Jawa Timur pada tahun 2023, produksi padi kualitas GKG (Gabah Kering Giling) di Jawa Timur mencapai 9,71 ton. Produksi ini meningkat dari tahun 2022 sebanyak 9,52 ton tetapi menurun dari tahun sebelumnya sebesar 9,78 ton. Sedangkan produksi beras di Jawa Timur pada tahun 2023 mencapai 5,60 ton sama seperti produksi padi mengalami peningkatan dari tahun 2022 namun menurun dari tahun sebelumnya. Sebagai upaya dalam mengoptimalkan agar stabil produksi padi dan beras, pemerintah daerah perlu melakukan kajian dalam pengelompokan daerah-daerah penghasil padi untuk mengetahui potensi produksinya serta daerah yang masih perlu peningkatan [4].

Metode dalam *data mining* dapat digunakan untuk mengelompokkan data yang memiliki kesamaan ke dalam sejumlah kelompok atau *cluster* [5]. Salah satu metode *clustering* yang dapat digunakan adalah *Agglomerative Hierarchical Clustering* (AHC). Metode ini mengelompokkan

data menjadi *cluster* sendiri lalu digabung dengan *cluster* yang lebih besar hingga mencapai jumlah *cluster* yang ingin dibentuk [6].

Dari uraian diatas, maka peneliti tertarik untuk melakukan pengelompokan produksi padi dan beras berdasarkan kabupaten/kota di Provinsi Jawa Timur menggunakan metode *Agglomerative Hierarchical Clustering* (AHC). Adapun manfaat dari penelitian ini dapat dijadikan acuan baik bagi pemerintah maupun masyarakat umum dalam menyukseskan optimalisasi wilayah penghasil padi khususnya di Jawa Timur yang dipetakan dalam sejumlah *cluster*, sehingga dapat turut berperan menjaga ketahanan pangan negara.

## 2. Tinjauan Pustaka

### 2.1. Penelitian Terkait

Beberapa penelitian terkait yang mendukung penelitian ini adalah :

- Penelitian pertama oleh NP Dharshinni dan Ciok Fandi (2022) dengan judul “Penerapan Metode *K-Medoids Clustering* Untuk Mengelompokkan Ketahanan Pangan” telah disampaikan bahwa Sumatera Utara mengalami penurunan dan kenaikan tidak stabil dalam ketersediaan hasil pangan. Maka, dilakukan pengelompokan hasil tanaman pangan di Sumatera Utara dengan *K-Medoids*. Dari penelitian tersebut menghasilkan sebanyak 3 *cluster* [7].

- Penelitian kedua oleh Irma Sanela dan kawan-kawan (2023) dengan judul “Penerapan Metode *Clustering* dengan *K-Means* untuk Memetakan Potensi Tanaman Padi di Sumatera” telah disampaikan bahwa perlu adanya pemetaan potensi tanaman padi di Sumatera untuk membantu upaya mengoptimalkan produksi padi di setiap provinsi. Dalam penelitian tersebut menggunakan metode *K-Means* telah dihasilkan sebanyak 3 *cluster* untuk kategori produksi padi dari tinggi, sedang, hingga rendah [8].
- Penelitian ketiga oleh Nur Mutmainnah Djafar, dkk (2021) dengan judul “Pengelompokan Produksi Perkebunan Menurut Kabupaten/Kota Jawa Tengah Tahun 2020 Menggunakan *Hierarchical Clustering*” disampaikan bahwa perkebunan di Indonesia memiliki peranan penting dalam peningkatan ekonomi nasional dan produktivitas perkebunan perlu dioptimalkan terutama di wilayah Jawa Tengah. Oleh sebab itu, dilakukan penerapan *data mining* berupa pengelompokan produksi perkebunan Provinsi Jawa Tengah. Hasil dari penelitian memperoleh 3 *cluster* untuk memetakan hasil produksi perkebunan dari tinggi, sedang, hingga rendah [9].
- Penelitian keempat oleh Ardiana Fatma Dewi dan Kurnia Ahadiyah (2022) dengan judul “*Agglomerative Hierarchy Clustering* Pada Penentuan Kelompok Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Pendidikan” bertujuan menemukan algoritma analisis hierarki *cluster* terbaik dan didapatkan algoritma *average linkage* dipilih sebagai algoritma terbaik karena menghasilkan nilai korelasi *cophenetic* terbesar dibandingkan algoritma lainnya. Hasil penelitian tersebut menetapkan jumlah *cluster* optimal sebanyak 2 *cluster* [10].
- Penelitian yang terakhir oleh Della Tri Cahaya, Desi Puspita, dan Riduan Syahri (2024) berjudul “Penerapan Metode *K-Means Clustering* untuk Pengelompokan Potensi Padi di Kota Pagar Alam” bertujuan untuk mengklasifikasikan potensi padi menggunakan Algoritma *K-Means* di Kota Pagar Alam. Hasil penelitian tersebut diperoleh sebanyak 3 *cluster* potensi padi dari tinggi, sedang, hingga rendah [11].

## 2.2. Landasan Teori

### 1. Data Mining

*Data mining* adalah serangkaian teknik ekstraksi pengetahuan yang berguna untuk menemukan pola khusus dari kumpulan data sehingga menghasilkan informasi nilai tambah yang sebelumnya tidak diketahui [12][13].

## 2. Pengelompokan Data

Pengelompokan data atau *clustering* merupakan salah satu metode dalam *data mining* yang digunakan untuk mengidentifikasi kelompok (*cluster*) dengan cara mengelompokkan sejumlah data berdasarkan kemiripan serupa. *Clustering* bermanfaat untuk menemukan kelompok data yang sebelumnya tidak diketahui dan dapat diidentifikasi karakteristik dari masing-masing kelompok data. [14].

### 3. Agglomerative Hierarchical Clustering (AHC)

*Agglomerative Hierarchical Clustering* (AHC) merupakan metode pengelompokan data hierarki dengan menggabungkan dua kelompok atau lebih berdasarkan kemiripan melalui perhitungan jarak antar data [15]. Berikut ini algoritma yang banyak digunakan dalam *agglomerative* [10] :

#### a. Single Linkage

Pada *single linkage*, pengelompokan diperoleh berdasarkan jarak antar data terkecil atau terdekat. Berikut ini persamaan *single linkage* :

$$d_{(UV)W} = \min \{d_{UW}, d_{VW}\} \quad (1)$$

Dengan  $d_{UW}$  adalah jarak terkecil dari U dan W,  $d_{VW}$  adalah jarak terkecil dari V dan W, sedangkan  $d_{(UV)W}$  adalah jarak antara *cluster* UV dan W.

#### b. Complete Linkage

Pada *complete linkage* adalah kebalikan dari *single linkage*, dimana pengelompokan diperoleh berdasarkan jarak maksimum. Berikut ini persamaan dari *complete linkage* :

$$d_{(UV)W} = \max \{d_{UW}, d_{VW}\} \quad (2)$$

#### c. Average Linkage

Pada *average linkage*, hasil pengelompokan diperoleh berdasarkan jarak rata-rata dari semua data. Berikut ini persamaan dari *average linkage* :

$$d_{(UV)W} = \frac{d_{UW} + d_{VW}}{N_{(UV)}N_W} \quad (3)$$

Dengan  $N_{(UV)}$  dan  $N_W$  adalah banyaknya data pada *cluster* (UV) dan W.

## 4. Jarak Euclidean

Jarak *Euclidean* merupakan matriks perhitungan jarak antar data dimana semakin kecil nilai jaraknya maka semakin dekat kemiripan data tersebut. Berikut ini persamaannya [16] :

$$d_{ij} = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \quad (4)$$

Keterangan :

$d_{ij}$  : Jarak antar data ke-i dan ke-j.

p : Jumlah variabel yang digunakan.

$X_{ik}$  : Nilai data ke-i pada variabel ke-k.

$X_{jk}$  : Nilai data k-j pada variabel ke-k.

## 5. Koefisien Korelasi Cophenetic

Koefisien Korelasi *Cophenetic* adalah pengukuran seberapa baik hierarki *cluster* memiliki hubungan dengan data asli pada matriks jarak *euclidean*. Korelasi *Cophenetic* dapat digunakan untuk uji validasi tingkat keoptimalan algoritma *clustering* dalam menghasilkan *cluster*. Berikut ini persamaannya [17] :

$$r_{coph} = \frac{\sum_{i < k} (d_{ik} - \bar{d})(d_{c_{ik}} - \bar{d}_c)}{\sqrt{[\sum_{i < k} (d_{ik} - \bar{d})^2][\sum_{i < k} (d_{c_{ik}} - \bar{d}_c)^2]}} \quad (5)$$

Keterangan :

$r_{coph}$  : Korelasi *Cophenetic*.

$d_{ik}$  : Jarak *Euclidean* antar objek  $i$  dan  $k$

$\bar{d}$  : Rata-rata  $d_{ik}$

$d_{c_{ik}}$  : Jarak *cophenetic* antara objek  $i$  dan  $k$

$\bar{d}_c$  : Rata-rata  $d_{c_{ik}}$

### 6. Koefisien Silhouette

Koefisien *Silhouette* adalah metode yang digunakan untuk menguji seberapa baik hasil pengelompokan data. Berikut ini persamaan dari *Silhouette Coefficient* [18] :

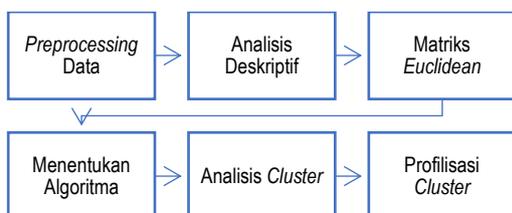
$$Si = \frac{(bi-ai)}{\text{Max}(ai,bi)} \quad (6)$$

Dengan  $ai$  adalah rata-rata jarak  $i$  terhadap objek dalam klaster sama,  $bi$  adalah rata-rata jarak  $i$  terhadap objek pada klaster lain. Nilai *silhouette* berada antara -1 hingga 1, pengelompokan data semakin baik jika nilainya mendekati 1..

## 3. Metode Penelitian

### 3.1. Tahapan Penelitian

Berikut ini adalah tahapan penelitian yang dilakukan dengan metode *Agglomerative Hierarchical Clustering* (AHC) :



Gambar 1. Tahapan Penelitian

1. *Preprocessing* data merupakan tahap persiapan data agar layak diproses seperti

pemilihan data dan variabel serta perubahan format data agar dapat dilakukan proses perhitungan.

2. Analisis deskriptif sebagai gambaran umum data berupa statistik dasar mulai dari nilai terendah, nilai tertinggi hingga rata-rata.
3. Menghitung matriks *euclidean* yang akan digunakan dalam algoritma *clustering* untuk mencari kemiripan antar data.
4. Menentukan algoritma *Agglomerative* terbaik berdasarkan korelasi *cophenetic* tertinggi.
5. Setelah menemukan algoritma terbaik, berikutnya melakukan analisis *cluster* untuk menentukan jumlah *cluster*. Kemudian, dilakukan uji *silhouette coefficient* untuk menentukan *cluster* yang optimal. Proses pembentukan *cluster* dilakukan dengan memotong *dendogram* sehingga jumlah anggota pada setiap *cluster* dapat ditentukan.
6. Profilisasi *cluster* sebagai langkah terakhir untuk mengetahui karakteristik *cluster*.

### 3.2. Jenis dan Sumber Data

Jenis data yang digunakan adalah data sekunder yang bersumber dari Badan Pusat Statistik (BPS) Jawa Timur berupa produksi padi (GKG) dan beras menurut kabupaten/kota pada tahun 2023

### 3.3. Lokasi Penelitian

Lokasi penelitian ini berfokus pada wilayah kabupaten/kota di Jawa Timur. Pemilihan lokasi

tersebut dikarenakan Jawa Timur merupakan salah satu provinsi dengan hasil produksi padi dan beras terbesar di Indonesia

#### 4. Hasil dan Pembahasan

##### 4.1. Hasil Penelitian

###### 1. *Preprocessing* Data

Pada tahap *preprocessing* data, diperoleh hasil berupa data yang siap untuk diproses. Berikut ini hasil dari *preprocessing* data :

Tabel 1. Hasil *Preprocessing* Data

No.	Kabupaten / Kota	Produksi Padi	Produksi Beras
1	Pacitan	92993	53695.84
2	Ponorogo	392994	226922.69
3	Trenggalek	114875	66330.87
--	--	--	--
38	Kota Batu	4863	2807.87

###### 2. Analisis Deskriptif

Dengan melakukan analisis deskriptif dapat diperoleh gambaran umum produksi padi dan beras di Provinsi Jawa Timur pada tahun 2023 yang ditampilkan pada tabel statistik berikut ini :

Tabel 2. Statistik Deskriptif

Produksi	Min	Max	Rata-rata
Padi	3293	798705	255543,74
Beras	1901,31	461188,28	147556,09

Berdasarkan statistik deskriptif, dapat diketahui bahwa produksi padi terendah sebesar 3293 ton yaitu Kota Mojokerto. Sedangkan produksi padi tertinggi sebanyak 798705 ton yaitu Kabupaten Lamongan. Selanjutnya hasil produksi beras

terendah juga terdapat di Kota Mojokerto sebanyak 1901,31 ton sedangkan produksi paling tinggi masih di Kabupaten Lamongan sebesar 461188,28 ton.

###### 3. Matriks *Euclidean*

Perhitungan jarak antar data menggunakan persamaan (4) untuk menentukan bagaimana data dikelompokkan dengan metode *agglomerative*. Berikut ini hasil perhitungan matriks *Euclidean* :

Tabel 3. Matriks *Euclidean*

Data	1	2	3
2	346421.91		
3	25267.88	321154.03	
4	164560.33	181861.58	139292.45
--	--	--	--

Berdasarkan matriks *Euclidean* diatas, dapat diketahui bahwa jarak antara data ke-1 dan data ke-2 adalah 346421.91, antara data ke-1 dan data ke-3 adalah 25267.88, dan seterusnya untuk seluruh pasangan jarak antar data.

###### 4. Menentukan Algoritma Terbaik

Setelah perhitungan jarak antar data selesai, selanjutnya mempertimbangkan algoritma *agglomerative* untuk pembentukan *cluster* melalui koefisien korelasi *cophenetic*. Algoritma yang memberikan nilai koefisien korelasi *cophenetic* tertinggi dianggap sebagai algoritma terbaik. Berikut ini hasil koefisien korelasi *cophenetic* :

Tabel 4. Hasil Koefisien Korelasi *Cophenetic*



Adapun daftar anggota kabupaten/kota pada setiap *cluster* dapat dilihat pada tabel berikut ini :

Tabel 7. Daftar Anggota *Cluster*

<b>Cluster</b>	<b>Anggota Cluster</b>
1	Lamongan, Ngawi, Bojonegoro, dan Jember.
2	Pacitan, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Bondowoso, Situbondo, Probolinggo, Pasuruan, Sidoarjo, Mojokerto, Magetan, Bangkalan, Sampang, Pamekasan, dan Sumenep.
3	Ponorogo, Banyuwangi, Jombang, Nganjuk, Madiun, Tuban, dan Gresik.
4	Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, dan Kota Batu

#### 6. Profilisasi *Cluster*

Setelah analisis pembentukan *cluster* selesai dilakukan, maka tahap terakhir adalah profilisasi *cluster* untuk memahami bagaimana karakteristik dari masing-masing *cluster* yang telah terbentuk.

Berikut ini adalah hasil dari profilisasi *cluster* :

Tabel 8. Profilisasi *Cluster*

<b>Cluster</b>	<b>Rata-rata produksi padi</b>	<b>Rata-rata produksi beras</b>
1	723161.25	417567.82
2	210021.00	121005.81
3	424213.43	244949.45
4	7571.67	4371.91

<b>Keterangan</b>	
	Tinggi
	Rendah
	Sedang
	Sangat Rendah

Berdasarkan profilisasi *cluster* di atas, dapat diketahui bahwa semakin tinggi hasil produksi padi maka produksi beras juga semakin tinggi karena pada dasarnya kedua variabel ini berasal dari hasil tanaman pangan yang sama.

Pada tabel 8 dapat diketahui karakteristik masing-masing *cluster* berdasarkan rata-rata produksi padi dan beras. Pada *cluster* 1 memiliki rata-rata hasil produksi padi dan beras paling tinggi dibandingkan *cluster* lain sebanyak 723161.25 ton untuk produksi padi dan 417567.82 ton untuk produksi beras. Adapun pada *cluster* 2 merupakan kelompok dengan hasil produksi yang rendah dengan rata-rata produksi padi sebesar 210021 ton dan rata-rata produksi beras sebesar 121005.81 ton. *Cluster* 3 merupakan *cluster* sedang dengan rata-rata produksi padi sebanyak 424213.43 ton dan rata-rata produksi beras sebanyak 244949.45 ton. Sedangkan *cluster* terakhir yaitu *cluster* 4 memiliki rata-rata hasil produksi padi sebanyak 7571.67 ton dan beras sebanyak 4371.91 ton. *Cluster* 4 merupakan *cluster* dengan produksi padi dan beras paling rendah dibandingkan *cluster* lainnya.

Profilisasi *cluster* juga ditampilkan dalam bentuk visualisasi pemetaan *cluster*. Berikut ini merupakan pemetaan dari hasil pengelompokan produksi padi dan beras berdasarkan kecamatan/kota di Provinsi Jawa Timur pada tahun 2023 :



Gambar 3. Visualisasi pemetaan *cluster*

#### 4.2. Pembahasan

Berdasarkan hasil *clustering* yang dilakukan, diperoleh sejumlah *cluster* produksi padi dan beras di Provinsi Jawa Timur berdasarkan kabupaten/kota pada tahun 2023. Berikut ini interpretasi hasil *clustering* :

- Pada *cluster* 1 sebanyak 4 kabupaten merupakan *cluster* tinggi dalam memproduksi padi dan beras dimana yang tertinggi adalah Kabupaten Lamongan kemudian Kabupaten Ngawi, Kabupaten Bojonegoro dan Kabupaten Jember.
- Pada *cluster* 2 merupakan *cluster* dengan produksi padi dan beras rendah sebanyak 18 kabupaten diantaranya : Pacitan, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Bondowoso, Situbondo, Probolinggo, Pasuruan, Sidoarjo, Mojokerto, Magetan, Bangkalan, Sampang, Pamekasan, dan Sumenep. Pada *cluster* ini dapat menjadi perhatian yang lebih terkait optimalisasi wilayah karena memiliki jumlah anggota

paling banyak dibandingkan *cluster* lainnya dan hasil produksinya juga rendah

- Adapun *cluster* 3 dengan produksi padi dan beras sedang sebanyak 7 kabupaten meliputi Ponorogo, Banyuwangi, Jombang, Nganjuk, Madiun, Tuban, dan Gresik.
- Terakhir, *cluster* 4 merupakan *cluster* sangat rendah karena seluruh anggotanya yang berjumlah 9 didominasi oleh wilayah perkotaan diantaranya : Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, dan Kota Batu. Kota Mojokerto merupakan wilayah paling rendah dalam memproduksi padi dan beras dibandingkan wilayah lainnya.

Hasil *clustering* tersebut dapat mendukung optimalisasi ketahanan pangan khususnya hasil produksi padi dan beras di Provinsi Jawa Timur. Pemerintah dapat menggunakan hasil ini untuk menerapkan kebijakan yang lebih tepat sasaran. Misalnya, menerapkan program yang dapat menjaga kestabilan produksi di wilayah dengan produksi tinggi dan meningkatkan produksi di wilayah dengan produksi rendah

#### 5. Kesimpulan

Dari penelitian yang dilakukan telah berhasil mengelompokkan produksi padi dan beras di Provinsi Jawa Timur Tahun 2023. Algoritma *average linkage* dengan korelasi *cophenetic*

tertinggi dipilih sebagai algoritma terbaik. Diperoleh 4 *cluster* optimal dengan uji *silhouette coefficient* sebesar 0.6347 dimana *cluster* 1 (tinggi) sebanyak 4 kabupaten, *cluster* 2 (rendah) sebanyak 18 kabupaten, *cluster* 3 (sedang) sebanyak 7 kabupaten dan *cluster* 4 (sangat rendah) sebanyak 9 kota.

Hasil dari penelitian ini dapat berkontribusi dalam memberikan wawasan terkait pola produksi padi dan beras di Jawa Timur yang dipetakan dalam sejumlah *cluster*. Selain itu, hasil penelitian dapat digunakan oleh pemerintah daerah Jawa Timur untuk mengidentifikasi wilayah yang memerlukan tindakan khusus, seperti daerah dengan produksi padi dan beras yang rendah.

Adapun keterbatasan dalam penelitian ini terdapat pada data yang digunakan hanya mencakup variabel produksi padi dan beras di Jawa Timur sehingga diperlukan pengembangan berikutnya. Misalnya, menambah variabel penting yang mempengaruhi produksi padi dan beras seperti luas panen dan faktor lainnya. Selain itu dapat menggunakan metode *clustering* lain untuk membandingkan hasil yang lebih akurat.

## 6. Daftar Pustaka

- [1] P. Trisnawati and A. I. Purnamasari, "Penerapan Pengelompokan Produktivitas Hasil Pertanian Menggunakan Algoritma K-Means," *Infotek : Jurnal Informatika dan Teknologi*, vol. 6, no. 2, pp. 249–257, Jul. 2023, doi: 10.29408/jit.v6i2.10198.
- [2] S. Wijayanto and M. Yoka 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, Oct. 2021, doi: 10.5281/3918.jupiter.2021.10.
- [3] Luth Fimawahib, Imam Rangga Bakti, and Asep Supriyanto, "Algoritma K-Medoids untuk Pengelompokan Produksi Padi dan Beras sebagai Upaya Optimalisasi Ketahanan Pangan di Provinsi Riau," *SATIN - Sains dan Teknologi Informasi*, vol. 8, no. 2, pp. 13–24, Dec. 2022, doi: 10.33372/stn.v8i2.877.
- [4] Syahla Anisah and Lisnur Wachidah, "Penerapan Metode Non-Hierarchical Clustering K-Means untuk Pengelompokan Kabupaten/Kota di Provinsi Jawa Barat Berdasarkan Faktor Produksi Padi Tahun 2021," *Bandung Conference Series: Statistics*, vol. 3, no. 2, pp. 323–332, Jul. 2023, doi: 10.29313/bcss.v3i2.8148.
- [5] P. A. Wisnuadi, B. I. A. Nugraha, H. K. Setiawan, and D. Hartanti, "Penerapan K-means Clustering pada Penjualan dan Pajak BBM di DKI Jakarta," in *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB) 2023*, 2023.
- [6] A. Rozaq, "Implementation of K-Means and Agglomerative Hierarchical Methods to House Clusterization," *Jurnal Media Informatika Budidarma*, vol. 6, no. 2, p. 933, Apr. 2022, doi: 10.30865/mib.v6i2.3573.
- [7] N. P. Dharshinni and C. Fandi, "Penerapan Metode K-Medoids Clustering Untuk Mengelompokkan Ketahanan Pangan," *Jurnal Media Informatika Budidarma*, vol. 6, no. 4, p. 2301, Oct. 2022, doi: 10.30865/mib.v6i4.4939.
- [8] I. Sanela, A. Nazir, F. Syafria, E. Haerani, and L. Oktavia, "Penerapan Metode Clustering Dengan K-Means Untuk Memetakan Potensi Tanaman Padi di Sumatera," *Journal of Computer System*

- and Informatics (JoSYC), vol. 5, no. 1, pp. 82–92, Nov. 2023, doi: 10.47065/josyc.v5i1.4523.
- [9] N. M. Djafar, L. N. Wijayanti, A. R. Elprilita, and E. Widodo, “Pengelompokan Produksi Perkebunan Menurut Kabupaten/Kota Jawa Tengah Tahun 2020 Menggunakan Hierarchical Clustering,” *Journal of Mathematics Education and Science*, vol. 4, no. 2, pp. 59–66, Oct. 2021, doi: 10.32665/james.v4i2.230.
- [10] A. F. Dewi and K. Ahadiyah, “Agglomerative Hierarchy Clustering Pada Penentuan Kelompok Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Pendidikan,” *Zeta - Math Journal*, vol. 7, no. 2, pp. 57–63, Nov. 2022, doi: 10.31102/zeta.2022.7.2.57-63.
- [11] D. Tri Cahaya, D. Puspita, and R. Syahri, “Penerapan Metode K-Means Clustering untuk Pengelompokan Potensi Padi di Kota Pagar Alam,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 2187–2193, Apr. 2024, doi: 10.36040/jati.v8i2.9432.
- [12] R. W. Abdullah, D. Hartanti, H. Permatasari, A. W. Septyanto, and Y. A. Bagaskara, “Penerapan Data Mining untuk Memprediksi Jumlah Produk Terlaris Menggunakan Algoritma Naive Bayes Studi Kasus (Toko Prapti),” *Jurnal Ilmiah Informatika Global*, vol. 13, no. 1, Mar. 2022, doi: 10.36982/jiig.v13i1.2060.
- [13] A. Arum and P. Pramono, “Penerapan Algoritma Deep Belief Networks (DBNs) Untuk Prediksi Kanker Serviks,” *DutaCom*, vol. 17, no. 1, pp. 50–57, Feb. 2023, doi: 10.47701/dutacom.v17i1.3790.
- [14] T. P. Sari, A. L. Hananto, E. Novalia, T. Tukino, and S. S. Hilabi, ‘Implementasi Algoritma K-Means dalam Analisis Klasterisasi Penyebaran Penyakit Hiv/Aids’, *Infotek : Jurnal Informatika dan Teknologi*, vol. 6, no. 1, pp. 104–114, Jan. 2023, doi: 10.29408/jit.v6i1.7423.
- [15] Z. Muttaqin, “Implementasi Unsupervised Learning pada Nilai Jasmani Kesamaptaaan Sekolah Polisi Negara dengan Metode Clustering Analysis,” *PROSISKO: Jurnal Pengembangan Riset dan Observasi Sistem Komputer*, vol. 10, no. 1, pp. 18–23, Apr. 2023, doi: 10.30656/prosisko.v10i1.6269.
- [16] L. P. W. Adnyani and P. R. Sihombing, “Analisis Cluster Time Series dalam Pengelompokan Provinsi di Indonesia Berdasarkan Nilai PDRB,” *Jurnal Bayesian : Jurnal Ilmiah Statistika dan Ekonometrika*, vol. 1, no. 1, pp. 47–54, Mar. 2021, doi: 10.46306/bay.v1i1.5.
- [17] A. Septianingsih, “Pemetaan Kabupaten Kota di Provinsi Jawa Timur Berdasarkan Tingkat Kasus Penyakit Menggunakan Pendekatan Agglomeratif Hierarchical Clustering,” *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 3, no. 2, pp. 367–386, Aug. 2022, doi: 10.46306/lb.v3i2.139.
- [18] C. Dewi, E. P. . Siam, G. A. . Wijayanti, M. . Putri, N. . Aulia, and R. . Nooraeni, “Comparison of DBSCAN and K-Means Clustering for Grouping the Village Status in Central Java 2020”, *Jurnal Matematika, Statistika dan Komputasi (JMSK)*, vol. 17, no. 3, pp. 394-404, May 2021