



## Implementasi Clustering Algoritma K-Means Pada Produksi Beras di Provinsi Jawa Timur Tahun 2022

Annisa Lusyani Zahra<sup>1</sup>, Sesilia Tiara Rahayu Ada<sup>2</sup>, Ayu Fakhira Ardini<sup>3</sup>

<sup>1,2,3</sup>Sistem Informasi, UPN Veteran Jawa Timur, Surabaya, Indonesia

Email: <sup>1</sup>20082010153@student.upnjatim.ac.id, <sup>2</sup>20082010167@student.upnjatim.ac.id, <sup>3</sup>20082010169@student.upnjatim.ac.id

### Abstrak

Beras ialah salah satu dari sekian tanaman pangan yang memiliki peranan penting di dalam kehidupan semua masyarakat khususnya bagi masyarakat di Asia. Sebagai salah satu cadangan beras terpenting di Indonesia, Jawa Timur berperan penting dalam produksi beras skala nasional. Oleh karena itu, diperlukan upaya penerapan algoritma K-means untuk meningkatkan efisiensi dan produktivitas sektor pertanian khususnya proses produksi beras. Data yang terkumpul membandingkan nilai centroid data dari dua bilangan cluster yaitu 4 dan 3. Jumlah cluster awal adalah 4. Bila menggunakan metode siku dan siluet, jumlah cluster yang benar adalah 4. Jumlah cluster kedua adalah 3 dan memiliki deskripsi tinggi, sedang, dan rendah. Landasan ilmiah penelitian ini untuk memberikan kontribusi positif terhadap pembangunan sektor pertanian di Jawa Timur, meningkatkan kesejahteraan pedesaan dan mengembangkan kebijakan yang lebih akurat dan efektif untuk menjaga ketahanan pangan nasional.

**Kata Kunci:** Clustering, K-Means, Produksi Beras

### 1. PENDAHULUAN

Beras merupakan hasil dari pengolahan tanaman padi yang telah dipanen, dimana padi yang telah dipanen lalu diproses dengan cara membuang kulit pada bulir-bulir padi yang telah dipisahkan dari tangkainya. Tanaman Padi memiliki nama latin yaitu *Oryza Sativa L* berupa tanaman pangan yang memiliki peranan penting di dalam kehidupan seluruh manusia, dikarenakan padi menjadi salah satu sumber pangan utama khususnya bagi penduduk Asia. Tanaman Padi dapat tumbuh dengan baik di daerah yang memiliki suhu panas dan banyak mengandung uap air dengan curah hujan rata-rata 200 mm/bulan atau lebih [1]. Data Badan Pusat Statistik Produksi Padi dan Beras, mengatakan bahwa Provinsi Jawa Timur merupakan provinsi penghasil padi terbanyak nomor satu di Indonesia



dibandingkan provinsi-provinsi lainnya. Dengan penghasilan padi terbanyak di Indonesia, membuat Provinsi Jawa Timur untuk terus meningkatkan produksi padi demi ketercapaian swasembada pangan khususnya untuk Provinsi Jawa Timur dan umumnya untuk Negara Indonesia. Agar mencegah produksi padi menurun ataupun menjaga agar produksi padi Jawa Timur tetap stabil ataupun meningkat perlu adanya cara untuk mengetahui factor – factor yang dapat mempengaruhi produksi padi di Jawa Timur [2].

Pada zaman yang serba cepat saat ini, perkembangan teknologi dan pertumbuhan penduduk berkembang sangat pesat, yang menimbulkan berbagai macam tantangan, salah satunya yaitu dalam sektor pertanian. Produksi pangan, seperti beras menjadi salah satu aspek penting dalam memenuhi kebutuhan pangan Masyarakat Indonesia. Provinsi Jawa Timur sebagai salah satu lumbung padi utama di Indonesia memiliki peran penting dalam menyumbangkan produksi beras pada sekala nasional. Oleh karena itu, perlu adanya upaya untuk meningkatkan efisiensi dan produktivitas dalam sektor pertanian, khususnya pada proses produksi beras.

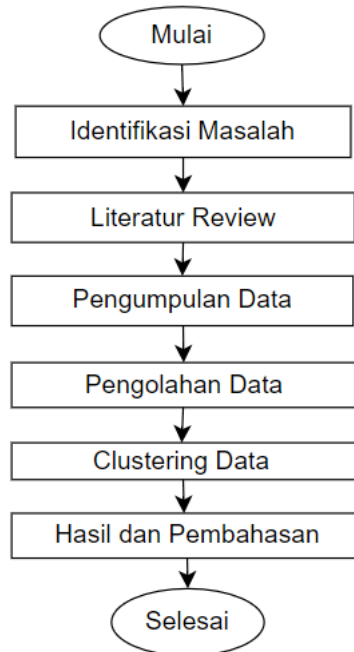
Dalam hal ini, penggunaan teknologi dan pendekatan ilmiah menjadi kunci penting untuk mengoptimalkan produksi beras. Salah satu pendekatan yang dapat diterapkan adalah algoritma clustering, khususnya algoritma K-Means. Clustering dapat membantu mengelompokkan data produksi beras menjadi kelompok-kelompok yang serupa, sehingga memudahkan analisis dan pengambilan keputusan [3]. Algoritma K-Means merupakan salah satu algoritma pembelajaran yang sederhana, dapat menyelesaikan sesuatu masalah dalam mengurangi kesalahan tersebut.

Penelitian ini akan fokus pada implementasi algoritma K-Means pada hasil panen yaitu berupa beras di berbagai kota dan kabupaten Provinsi Jawa Timur pada tahun 2022. Algoritma K-Means dipilih karena keunggulannya dalam mengelompokkan data berdasarkan kesamaan karakteristik, yang dapat memberikan gambaran yang lebih jelas mengenai pola produksi beras [4]. Dengan menerapkan algoritma k-means ini, diharapkan dapat ditemukan pola-pola yang berguna untuk meningkatkan efisiensi produksi, mengidentifikasi potensi perbaikan, dan mendukung pengambilan keputusan strategis dalam pengelolaan sektor pertanian [5]. Demikian dalam penelitian ini, diharapkan dapat memberikan kontribusi positif dalam pengembangan sektor pertanian di Provinsi Jawa Timur,

serta memberikan dasar ilmiah bagi pengambilan kebijakan yang lebih tepat dan efektif dalam rangka meningkatkan kesejahteraan masyarakat petani dan menjaga ketahanan pangan nasional.

## 2. METODE

Pada penelitian ini memiliki tahapan yang dimulai dari Identifikasi masalah, Literatur Review, Pengumpulan Data, Pengolahan Data, Clustering Data, serta hasil dan pembahasan dari pehitungan K-means.



Gambar 1. Tahapan Penelitian

### 2.1. Tahap Identifikasi Masalah

Pada Tahapan ini merupakan tahapan untuk menentukan permasalahan-permasalahan yang akan dijadikan objek untuk penelitian. Dari permasalahan yang telah diidentifikasi, maka selanjutnya akan digunakan untuk mencari alternatif pemecahan permasalahan tersebut. Mengidentifikasi permasalahan dapat membantu untuk memahami ruang

lingkup dan langkah-langkah untuk memecahkan masalah serta penelitian menjadi lebih terarah [6].

## 2.2. Tahap Literatur Review

Pada Tahapan ini melakukan pengumpulan berbagai macam literatur yang berkaitan dengan Algoritma K-Means, Luas lahan dan Produksi Beras di Jawa Timur. Literatur-literatur yang telah diperoleh dapat memperkuat metode pada penelitian ini [7]. Salah satu literatur yang di peroleh untuk dijadikan referensi pada penelitian ini yaitu data dari Badan Pusat Statistik (BPS) Jawa Timur.

## 2.3. Tahap Pengumpulan Data

Data penelitian yang diperoleh serta digunakan merupakan data valid berdasarkan dokumen yang telah dipublikasikan oleh Badan Pusat Statistik Provinsi Jawa Timur melalui situsnya, serta data yang ditampilkan pada situs Basis Data Statistik Pertanian [8]. Data berupa hasil produksi beras setiap kota dan kabupaten dengan satuannya ton pada tahun 2022 serta luas lahan terbarunya di setiap kota dan kabupaten yang berjumlah 9 kota dan 29 kabupaten di Provinsi Jawa Timur [9].

## 2.4. Tahap Pengolahan Data

Data yang diolah akan menghasilkan daerah mana saja yang berpotensi menghasilkan produksi beras lebih banyak dari daerah-daerah lainnya, dan dilanjutkan dengan menganalisis hasil dari olah data tersebut seperti dapat mengetahui alasan mengapa produksi beras sedikit serta penanggulangan atau solusi yang dapat diterapkan untuk mengatasi hal tersebut dan hal semacam lainnya [10].

## 2.5. Tahap Clustering

Clustering adalah proses pengelompokan beberapa data menjadi suatu kelompok-kelompok data yang mempunyai karakteristik sama [11]. Penelitian ini menggunakan algoritma K-Means suatu algoritma iteratif yang berupaya membagi suatu kumpulan data menjadi kumpulan data yang telah ditentukan sedemikian rupa sehingga kumpulan data tersebut tidak tumpang tindih [12]. Dalam metode pengimplementasian K-Means, data dibagi menjadi kelompok atau klaster, apabila karakter yang dimiliki

sama oleh setiap masing-masing data akan dikelompokkan di dalam kluster yang sama, begitupun sebaliknya apabila masing-masing data memiliki karakter yang berbeda akan berbeda pula kelompoknya atau klasternya.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Pengumpulan Data

Berdasarkan tahapan-tahapan pada metode penelitian, dalam melakukan clustering diawali dengan tahapan dikumpulkannya data. Data ini didapatkan dari Badan Pusat Statistik Provinsi Jawa Timur dan Basis Data Statistik Pertanian, datanya berupa jumlah produksi beras per tahun dan luas lahan berdasarkan kota atau kabupaten pada tahun 2022, dapat dilihat dibawah ini pada tabel 1.

**Tabel 1.** Data Produksi Beras dan Luas Lahan Provinsi Jawa Timur 2022

Kota/Kabupaten	Produksi per Tahun (Ton)	Luas Lahan (Ha)
Kabupaten Pacitan	52215.06	17831.47
Kabupaten Ponorogo	207532.72	62964.37
Kabupaten Trenggalek	66841	22026.97
Kabupaten Tulungagung	119651.42	39531.25
Kabupaten Blitar	124424.09	34137.74
Kabupaten Kediri	97499.45	29379.86
Kabupaten Malang	156831.2	45901.91
Kabupaten Lumajang	173704.72	55042.5
Kabupaten Jember	350708.35	118489.25
Kabupaten Banyuwangi	266887.02	76748.03
Kabupaten Bondowoso	137817.3	48240.7
Kabupaten Situbondo	81778.89	26784.22
Kabupaten Probolinggo	107060.91	33503.12
Kabupaten Pasuruan	142770.43	47384.82
Kabupaten Sidoarjo	112331.58	31368.74
Kabupaten Mojokerto	162733.96	49395.45
Kabupaten Jombang	194689.98	54896.68
Kabupaten Nganjuk	217385.07	70331.77
Kabupaten Madiun	231876.72	73061.53
Kabupaten Magetan	153997.9	43965.49
Kabupaten Ngawi	436494.35	128585.79
Kabupaten Bojonegoro	406670.62	133665.5
Kabupaten Tuban	288097.19	85288.02
Kabupaten Lamongan	521919.84	150903.4
Kabupaten Gresik	237459.53	63287.2

Kota/Kabupaten	Produksi per Tahun (Ton)	Luas Lahan (Ha)
Kabupaten Bangkalan	114400.05	40717.81
Kabupaten Sampang	99601.72	34880.84
Kabupaten Pamekasan	62158.54	20497.42
Kabupaten Sumenep	133142.3	41835.41
Kota Kediri	5888.24	1864.13
Kota Blitar	3009.59	728.19
Kota Malang	6219.47	1718.11
Kota Probolinggo	4679.36	1323.99
Kota Pasuruan	4211.51	1393.98
Kota Mojokerto	2906.13	833.03
Kota Madiun	6613.73	2194.58
Kota Surabaya	4687.1	1447.98
Kota Batu	3904.84	1059.45

Pada tabel 1 diatas terdiri dari 38 kota atau kabupaten yang terdapat di Provinsi Jawa Timur disertai dengan jumlah produksi per tahun dalam satuan ton dan luas lahan dalam satuan hektare.

### 3.2. Olah Data

Tahapan pengolahan data diawali dengan melakukan eksplorasi data untuk dilihat apakah terdapat missing value di dalam data tersebut, dihasilkan bahwa tidak terdapatnya missing value dalam data tersebut. Lalu dilanjutkan dengan menentukan berapa cluster yang akan digunakan, penelitian ini akan menganalisis dan membandingkan 2 jumlah cluster yang akan digunakan. Jumlah cluster pertama yaitu 4 cluster, dimana ini didapatkan dari hasil perhitungan dengan menggunakan metode Elbow dan Silhouette yang didapatkannya jumlah cluster yang tepat adalah 4. Jumlah cluster kedua yaitu 3 yang nanti memiliki keterangan tinggi, sedang, dan rendah.

#### 3.2.1. Cluster 4

Tabel 2. Nilai Centroid Cluster 4

Cluster	Produksi per Tahun (Ha)	Luas Lahan (Ha)
0	127849.7	40138.0
1	428948.3	132911.0
2	234846.9	69511.1
3	18611.2	6076.6

Dapat dilihat pada Table 2, rentang dari cluster 0 sampai dengan cluster 3 memiliki angka yang berbeda pada produksi per tahun dan luas lahan di masing-masing clusternya. Cluster 0 memiliki arti hanya dengan memiliki luas lahan yang sedikit dapat menghasilkan produksi per tahun yang cukup banyak, cluster 1 memiliki arti bahwa memiliki luas lahan yang besar menghasilkan produksi per tahun yang banyak juga, cluster 2 memiliki arti luas lahan yang cukup dan menghasilkan produksi per tahun yang banyak, cluster 3 memiliki arti luas lahan yang sedikit juga menghasilkan produksi per tahun yang sedikit juga.

### 3.2.2. Cluster 3

**Tabel 3.** Nilai Centroid Cluster 3

Cluster	Produksi per Tahun (Ha)	Luas Lahan (Ha)
0	400778.1	123386.4
1	159589.9	48828.8
2	23470.3	7669.5

Dapat dilihat pada table 3, rentang dari cluster 0 sampai dengan cluster 2 memiliki angka yang berbeda pada produksi per tahun dan luas lahan di masing-masing clusternya. Cluster 0 memiliki arti luas lahan yang banyak juga menghasilkan produksi per tahun yang banyak juga, cluster 1 memiliki arti luas lahan yang cukup banyak juga memiliki produksi per tahun yang cukup banyak juga, cluster 2 memiliki arti luas lahan yang sedikit juga menghasilkan produksi per tahun yang sedikit.

### 3.3. Clustering

Tahapan clustering atau pengelompokan kota dan kabupaten di beberapa cluster dilihat berdasarkan dari jumlah produksi per tahun dan luas lahannya, berikut dapat dilihat hasil akhir dari tahap clustering ini yaitu pada table 4 dan 5.

**Tabel 4.** Clustering pada cluster 4

Kota/Kabupaten	Cluster
Pacitan	3
Ponorogo	2
Trenggalek	3
Tulungagung	0
Blitar	0
Kediri	0

Kota/Kabupaten	Cluster
Malang	0
Lumajang	0
Jember	1
Banyuwangi	2
Bondowoso	0
Situbondo	0
Probolinggo	0
Pasuruan	0
Sidoarjo	0
Mojokerto	0
Jombang	2
Nganjuk	2
Madiun	2
Magetan	0
Ngawi	1
Bojonegoro	1
Tuban	2
Lamongan	1
Gresik	2
Bangkalan	0
Sampang	0
Pamekasan	3
Sumenep	0
<b>Kota</b>	
Kediri	3
Blitar	3
Malang	3
Probolinggo	3
Pasuruan	3
Mojokerto	3
Madiun	3
Surabaya	3
Batu	3

Dapat dilihat pada table 4 didapatkan rata-rata di wilayah kabupaten Provinsi Jawa Timur masuk ke dalam cluster 0, sedangkan di wilayah kotanya semuanya masuk ke dalam cluster 3. Ada beberapa wilayah kabupaten yang memasuki cluster 1 yaitu kabupaten Jember, Ngawi, Bojonegoro, dan Lamongan, wilayah kabupaten yang memasuki cluster 2 yaitu Kabupaten Ponorogo, Banyuwangi, Jombang, Nganjuk, Madiun, Tuban, dan Gresik, lalu wilayah kabupaten yang memasuki cluster 3 yaitu Kabupaten Trenggalek dan Pamekasan.



**Tabel 5.** Clustering pada cluster 3

<b>Kota/Kabupaten</b>	<b>Cluster</b>
Pacitan	2
Ponorogo	1
Trenggalek	2
Tulungagung	1
Blitar	1
Kediri	1
Malang	1
Lumajang	1
Jember	0
Banyuwangi	1
Bondowoso	1
Situbondo	2
Probolinggo	1
Pasuruan	1
Sidoarjo	1
Mojokerto	1
Jombang	1
Nganjuk	1
Madiun	1
Magetan	1
Ngawi	0
Bojonegoro	0
Tuban	0
Lamongan	0
Gresik	1
Bangkalan	1
Sampang	1
Pamekasan	2
Sumenep	1
<b>Kota</b>	
Kediri	2
Blitar	2
Malang	2
Probolinggo	2
Pasuruan	2
Mojokerto	2
Madiun	2
Surabaya	2
Batu	2

Dapat dilihat pada table 5 didapatkan rata-rata di wilayah kabupaten provinsi jawa timur masuk ke dalam cluster 1, sedangkan di wilayah kotanya masuk ke dalam cluster 2. Ada beberapa wilayah kabupaten yang masuk ke dalam cluster 0 yaitu Kabupaten Jember, Ngawi, Bojonegoro, Tuban, dan Lamongan, wilayah kabupaten yang masuk ke dalam cluster 2 yaitu Kabupaten Pacitan, Trenggalek, Situbondo, dan Pamekasan.

### 3.4. Analisis Data

Pada hasil nilai centroid data pada table 2 yaitu dengan jumlah clusternya 4 dan table 3 dengan jumlah clusternya 3, didapatkan hasil kesimpulan berupa keterangan dari setiap clusternya. Berikut keterangan pada masing-masing cluster, dengan jumlah cluster 4 dan 3.

Jumlah cluster 4 memiliki keterangan pada setiap clusternya yaitu:

- Cluster 0: luas lahan sedang, produksi per tahun tinggi.
- Cluster 1: luas lahan tinggi, produksi per tahun tinggi.
- Cluster 2: luas lahan sedang, produksi per tahun sedang.
- Cluster 3: luas lahan rendah, produksi per tahun rendah.

Jumlah cluster 3 memiliki keterangan pada setiap clusternya yaitu:

- Cluster 0: luas lahan tinggi, produksi per tahun tinggi.
- Cluster 1: luas lahan sedang, produksi per tahun sedang.
- Cluster 2: luas lahan rendah, produksi per tahun rendah.

Dapat disimpulkan dari keterangan di atas bahwa dengan jumlah cluster 4 memiliki spesifikasi yang detail atau menyeluruh dibandingkan dengan jumlah cluster 3, dimana suatu wilayah di daerah kabupaten dan kota Provinsi Jawa Timur tidak hanya dilihat dari luas lahan tinggi dan produksi per tahun tinggi tetapi dilihat juga perbandingan antara luas lahan sedang dan produksi per tahun tinggi. Tentunya akan berbeda apabila suatu wilayah memiliki luas lahan yang tinggi tetapi produksi per tahunnya kalah dengan wilayah yang hanya memiliki luas lahan yang sedang.

## 4. KESIMPULAN

Kesimpulan yang didapatkan dari penelitian ini bahwa implementasi clustering menggunakan algoritma k-means dapat dilakukan dengan berdasarkan produksi per tahun dan luas lahan pada masing-masing kota

serta kabupaten di Provinsi Jawa Timur, serta membandingkan nilai centroid data dari kedua jumlah cluster yaitu 4 dan 3. Tentunya didapatkan hasil yang berbeda setelah clustering atau pengelompokan berdasarkan kedua jumlah cluster tersebut, dimana rata-rata kota memiliki nilai produksi per tahun dan luas lahan yang sedikit dibandingkan dengan kabupaten, daerah kota selalu terletak di cluster akhir atau cluster 2 dan 3, hal ini memiliki arti bahwa wilayah kota lebih memfokuskan untuk pembangunan infrastruktur dibandingkan pertaniannya.

## REFERENSI

- [1] Junaidi, "Usaha Peningkatan Produksi Padi (*Oryza Sativa* L) dengan Penambahan N pada Perlakuan Dosis Pupuk Kandang," *J. Agroteknologi dan Agribisnis (AGRINIKA)*, vol. 2, no. 1, 2018.
- [2] M. Ishaq et al., "Analisis Faktor – Faktor yang Mempengaruhi Produksi Padi di Provinsi Jawa Timur menggunakan Regresi Semiparametrik Spline," *J. Sains dan Seni ITS*, vol. 6, no. 1, 2017.
- [3] D. N. Batubara et al., "Analisis Metode K-MEANS Pada Pengelompokan Keberadaan Area Resapan Air Menurut Provinsi," *Seminar Nasional Sains dan Teknologi Informasi (SENSASI)*, vol. 2, no. 1, 2019.
- [4] P. Alkhairi and A. P. Windarto, "Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara," *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, vol. 1, no. 1, 2019.
- [5] D. F. Pasaribu et al., "Memanfaatkan Algoritma K-Means Dalam Memetakan Potensi Hasil Produksi Kelapa Sawit PTPN IV Marihat," *BIOS: J. Teknologi Informasi dan Rekayasa Komputer*, vol. 2, no. 1, 2021.
- [6] H. Kurniawan and S. Defit, "Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal," *J. Appl. Computer Sci. and Tech.*, vol. 1, no. 2, pp. 80-89, 2020.
- [7] M. G. Sadewo et al., "Algoritma K-Means Dalam Mengelompokkan Desa/Kelurahan Menurut Keberadaan Keluarga Pengguna Listrik dan Sumber Penerangan Jalan Utama Berdasarkan Provinsi," *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, vol. 1, no. 1, 2019.
- [8] W. S. Utami, N. Pratiwi, and F. Muhammad, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Clustering Perokok Usia

- Lebih dari 15 Tahun," *Bull. Inf. Tech. (BIT)*, vol. 4, no. 4, pp. 501-507, 2023.
- [9] Ramadhani and M. Megawati, "Implementasi Algoritma K-Means Untuk Klustering Data Produktivitas Kelapa Sawit," *Indonesian J. Inf. Research and Software Eng. (IJIRSE)*, vol. 3, no. 1, 2023, pp. 56-64.
- [10] K. Rahmat, M. S. Hasibuan, and R. Hasibuan, "Klasterisasi Wilayah Prioritas Vaksin Menggunakan Algoritma K-Means Clustering," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 3, pp. 1585-1592, 2023.
- [11] T. Tendean and W. Purba, "Analisis Cluster Provinsi Indonesia Berdasarkan Produksi Bahan Pangan Menggunakan Algoritma K-Means," *J. Sains dan Teknologi*, vol. 1, no. 2, pp. 5-11, 2020.
- [12] Y. S. Sy, "Klasterisasi Pasien Rawat Inap Peserta BPJS Berdasarkan Jenis Penyakit Menggunakan Algoritma K-Means," *J. Sistim Informasi dan Teknologi*, pp. 33-37, 2023.