

## PENERAPAN ALGORITMA K-MEANS UNTUK MELAKUKAN KLASTERISASI PADA VARIETAS PADI

Bagas Putra Aryadi <sup>1</sup>, Nirwana Hendrastuty <sup>2</sup>

<sup>1,2</sup> Informatika, Universitas Teknokrat Indonesia,

Jl.ZA. Pagar Alam No.9-11, Bandar Lampung, Indonesia

<sup>1</sup> bagasputraaryadi75@gmail.com, <sup>2</sup> nirwana\_hendrastuty@teknokrat.ac.id

### Abstract

*The aim of this research is to build a model using the K-Means algorithm for clustering rice varieties in Indonesia, with the hope of positively contributing to the increase in rice productivity and quality. The research methodology includes a literature study, data collection of rice varieties, data preprocessing, clustering using the K-Means algorithm, and evaluation of clustering results. The clustering process was carried out using the RapidMiner application, with steps such as importing data, creating a process diagram, executing the process diagram, and analyzing statistical data and K-Means performance vector. The research identified five clusters of rice varieties through the clustering process. The Davies Bouldin index value was -0.48, indicating a good quality of clustering. Clustering can provide valuable information to the agricultural department and farmers about the types of rice available and help in selecting suitable rice varieties based on their growth conditions. The K-Means clustering method was found to be effective in clustering rice varieties based on their characteristics. The five types of rice to be clustered are hybrid rice, superior rice, paddy rice, gogo rice, and swamp rice.*

**Keywords :** Clustering, K-Means algorithm, Rice.

### Abstrak

Penelitian ini bertujuan untuk membangun model menggunakan algoritma K-Means untuk klasterisasi varietas padi di Indonesia, dengan memberikan kontribusi positif bagi peningkatan produktivitas dan kualitas hasil panen padi. Salah satu permasalahan yang menjadi fokus utama penelitian ini adalah penurunan hasil panen padi baik dari segi kuantitas maupun kualitasnya. Metodologi penelitian mencakup studi pustaka, pengumpulan data varietas padi, praproses data, klasterisasi menggunakan algoritma K-Means, dan evaluasi hasil klasterisasi. Hasil klasterisasi dilakukan menggunakan aplikasi RapidMiner, dengan tahap-tahap yang meliputi import data, pembentukan diagram proses, eksekusi diagram proses, dan analisis statistik data serta performance vector K-Means. Hasil penelitian menunjukkan bahwa lima kelompok cluster varietas padi berhasil diidentifikasi melalui proses klasterisasi. Dari hasil pengujian terdapat nilai index davies bouldin sebesar -0.48, yang menandakan kualitas klaster yang baik. Klasterisasi ini dapat memberikan informasi penting kepada pihak dinas pertanian dan petani mengenai jenis-jenis padi yang ada, serta membantu dalam pemilihan varietas padi yang sesuai dengan kondisi lingkungan tumbuhnya. Penerapan metode K-Means clustering memberikan hasil yang baik, menegaskan efektivitasnya dalam mengelompokkan varietas padi berdasarkan karakteristiknya. Jenis padi yang akan di klaster berjumlah 5 yaitu jenis padi hibrida, padi unggul, padi sawah, padi gogo dan padi rawa.

**Kata kunci :** Algoritma K-Means, Klasterisasi, Tanaman Padi

### 1. PENDAHULUAN

Tanaman padi (*Oryza sativa* L.) memiliki peran yang sangat penting dalam memastikan ketahanan pangan di Indonesia, mengingat

sebagian besar penduduk Indonesia mengonsumsi nasi sebagai bahan makanan utama mereka [1]. Ketersediaan dan produktivitas tanaman padi menjadi faktor krusial dalam menjaga kesejahteraan dan

keberlangsungan hidup masyarakat [2]. Namun, tantangan dalam meningkatkan produksi padi telah menjadi fokus utama bagi para pelaku pertanian, mengingat adanya penurunan hasil panen baik dari segi kuantitas maupun kualitasnya [3].

Pada tahun 2017, varietas padi lokal tersebar di lahan seluas 46.281 hektar. Varietas padi lokal yang banyak di budidayakan oleh petani adalah siam pandak, siam mayang dan siam unus. sementara itu, varietas padi unggul tersebar di lahan seluas 140.626 hektar, jauh lebih luas dibandingkan dengan varietas padi lokal [4]. Faktor-faktor yang mempengaruhi petani dalam memilih varietas unggul padi sawah meliputi potensi hasil, umur panen, ketahanan terhadap hama dan penyakit, kerebahan tanaman, harga beli benih, aromatik, tekstur nasi dan kemudahan memperoleh benih.

Untuk mengetahui pengaruh dari faktor-faktor yang menentukan keputusan petani memilih varietas unggul padi sawah digunakan metode data mining, yang memungkinkan penemuan pola atau aturan tertentu dari sejumlah besar data [5]. Salah satu teknik yang umum digunakan dalam data mining adalah *algoritma k-means*, yang dapat digunakan untuk melakukan klusterisasi atau pengelompokan data [6].

Pada penelitian terdahulu yang telah dilakukan oleh Yani dengan judul "Implementasi Metode *Fuzzy C-Means* Pada Sistem Clustering Data Varietas Padi" menggunakan uji coba perhitungan secara manual menggunakan *MS. Excel* dan *Matlab* [7]. Penelitian yang dilakukan oleh Rani dan Afiani dengan judul "Penerapan *K-Means Clustering* Untuk Mengetahui Varietas Padi Unggul Produksi Balai Pengkajian Teknologi Pertanian Jawa Timur" ini dibangun menggunakan *software Microsoft Visual Studio 2010* dan *Microsoft SQL Server 2008* [8].

Jenis padi yang akan di klaster yaitu jenis padi hibrida, padi unggul, padi sawah, padi gogo dan padi rawa dan jumlah data nya sebanyak 303 data varietas padi. Dengan demikian, sistem yang memanfaatkan metode ini dapat membantu petani dalam memilih varietas padi yang sesuai dengan kondisi lingkungan tumbuhnya. Oleh karena itu, penelitian ini bertujuan untuk membangun model menggunakan metode *k-means* untuk melakukan klusterisasi pada varietas padi yang dapat memberikan kontribusi positif bagi peningkatan produktivitas dan kualitas hasil panen padi di Indonesia.

## 2. METODOLOGI PENELITIAN

Tahapan penelitian ini mencakup studi pustaka untuk mengumpulkan informasi terkini, pengumpulan data varietas padi, dan pengolahan data untuk persiapan analisis. Selanjutnya, pengolahan data set terdiri dari data selection, data cleaning dan pre praprocessing. Setelah itu ada TF-IDF. Selanjutnya klusterisasi varietas padi dilakukan menggunakan metode *K-Means Clustering* untuk mengelompokkan varietas berdasarkan karakteristiknya. Penelitian ini merupakan penelitian terapan yang diarahkan untuk mendapat informasi guna memecahkan permasalahan yang ada.

### 2.1. Studi Pustaka

Tahap awal dalam penelitian ini adalah melakukan studi pustaka yang berfokus pada penelitian-penelitian sebelumnya [9] terkait penggunaan algoritma *K-Means*, serta untuk mengonfirmasi landasan teoritis yang digunakan. Pengumpulan data dan informasi terkait dilakukan melalui berbagai sumber seperti jurnal, media online, buku, dan penelitian terdahulu yang relevan dengan topik penelitian ini. Studi pustaka ini menjadi landasan untuk memperkuat penelitian dan memperoleh pemahaman yang lebih mendalam mengenai masalah yang diteliti.

### 2.2. Pengumpulan Data

Setelah melalui studi pustaka, tahap berikutnya adalah analisis data dan atribut untuk mendapatkan pemahaman yang lebih mendalam tentang penelitian ini. Data yang digunakan diambil dari Dinas Pertanian Kabupaten Tanggamus dan buku varietas padi, tersaji dalam format file *.xls* dengan total 303 data dan 5 atribut. Atribut yang digunakan dalam analisis ini terdiri dari berbagai informasi mengenai varietas padi, dengan klusterisasi kelas yang didefinisikan sebagai padi hibrida, padi unggul, padi sawah, padi gogo dan padi rawa. Tahap ini penting sebagai persiapan untuk proses klusterisasi menggunakan metode *K-Means*.

### 2.3. Pengolahan Data Set

Pengolahan data set dilakukan untuk memastikan kebersihan dan keterandalan data yang akan digunakan dalam analisis lebih lanjut. Proses ini meliputi pembersihan data untuk menghilangkan nilai yang hilang, menghapus atribut yang tidak relevan atau tidak digunakan dalam analisis, serta mengatasi adanya duplikat data yang mungkin ada. Langkah-langkah ini

bertujuan untuk memastikan kualitas data yang optimal sebelum dilakukan proses selanjutnya, sehingga hasil analisis yang dihasilkan dapat lebih akurat dan dapat dipercaya [10]. Berikut penjelasan dari tahapan yang dilalui :

a. *Data selection*

*Data selection* adalah proses menganalisis data-data yang relevan dari *database* karena sering ditemukan bahwa tidak semua data dibutuhkan dalam proses data mining [11].

b. *Data Cleaning*

*Data cleaning* yaitu melakukan pembersihan data terhadap *noise* yang ditemukan berupa *missing value*, inkonsisten data dan *redundant data*. Pada penelitian ini *data cleaning* dilakukan secara manual dengan menggunakan excel 2019

c. *Data Transformation*

*Data transformation* yaitu pada tahap ini data yang sudah bersih tersebut perlu ditransformasi terlebih dahulu. Data yang sudah ditransformasi dapat diolah menggunakan data mining.

## 2.4. Klasterisasi

Klasterisasi adalah proses pengelompokan data ke dalam kelompok-kelompok yang memiliki karakteristik atau atribut yang serupa di dalamnya [12]. Dalam konteks penelitian ini, klasterisasi dilakukan menggunakan metode *K-Means Clustering*. Metode ini membagi data ke dalam beberapa kelompok berdasarkan kesamaan antara objek-objek data, di mana setiap objek data akan diatributkan ke kelompok tertentu yang memiliki pusat klaster atau *centroid* yang terdekat dengannya. Tujuan utama dari klasterisasi adalah untuk mengidentifikasi pola atau struktur yang tersembunyi dalam data, sehingga memudahkan pemahaman dan pengambilan keputusan berdasarkan karakteristik dari masing-masing kelompok data yang terbentuk [13].

## 2.5. K-Means Clustering

Metode *K-Means Clustering* adalah teknik pengelompokan data ke dalam *cluster-cluster* yang memiliki titik pusat yang berbeda-beda untuk setiap *cluster* [14]. Tujuan dari *K-Means*, yaitu untuk meminimalkan dari fungsi objektif yang diatur dalam proses pengelompokan, pada umumnya akan berusaha meminimalkan variasi didalam suatu kelompok dan memaksimalkan variasi antar kelompok [15]. Dalam penelitian ini, peneliti telah menetapkan penggunaan 5 *cluster*. Proses *K-Means Clustering* terdiri dari tiga tahapan, pertama adalah mengumpulkan data varietas padi, kedua adalah pra proses data yang terdiri dari *data selection*, *data cleaning*, dan *data*

*transformation*, ketiga adalah permodelan menggunakan *algoritma k-means* yang terdiri dari permodelan *algoritma k-means* di aplikasi *rapidminer*, membuat diagram proses dan eksekusi diagram proses [16].

## 2.6. TF-IDF

Metode TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah sebuah metode yang digunakan dalam pengolahan teks atau dokumen untuk mengevaluasi seberapa penting sebuah kata dalam sebuah dokumen terhadap kumpulan dokumen yang lain [17]. *Term Frequency* (TF) mengukur seberapa sering sebuah kata muncul dalam sebuah dokumen. Ini dihitung dengan cara menghitung jumlah kemunculan kata tersebut dalam dokumen tersebut, kemudian dibagi dengan total jumlah kata dalam dokumen tersebut. *Inverse Document Frequency* (IDF) mengukur seberapa unik sebuah kata tersebut di dalam kumpulan dokumen. Kata-kata yang muncul di banyak dokumen akan memiliki IDF yang rendah, sedangkan kata-kata yang jarang muncul akan memiliki IDF yang tinggi. IDF dihitung dengan cara membagi jumlah total dokumen dengan jumlah dokumen yang mengandung kata tersebut, kemudian mengambil logaritma dari hasil tersebut. Dengan menggabungkan TF dan IDF, kita mendapatkan nilai TF-IDF yang menggambarkan seberapa penting suatu kata dalam sebuah dokumen terhadap kumpulan dokumen secara keseluruhan. Semakin tinggi nilai TF-IDF, semakin penting kata tersebut dalam dokumen tersebut [18].

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan Data

Dataset yang diperoleh dari Dinas Pertanian Tanggamus, Kota Agung terdiri dari 303 baris data yang mencakup lima varietas padi, yaitu padi hibrida, padi unggul, padi sawah, padi gogo, dan padi rawa. Data ini merupakan hasil yang dapat digunakan untuk analisis dan penelitian lebih lanjut terkait karakteristik dan pemilihan varietas padi. Pembagian data untuk data training sebesar 70% dan untuk data testing sebesar 30%.

### 3.2 Pra Proses Data

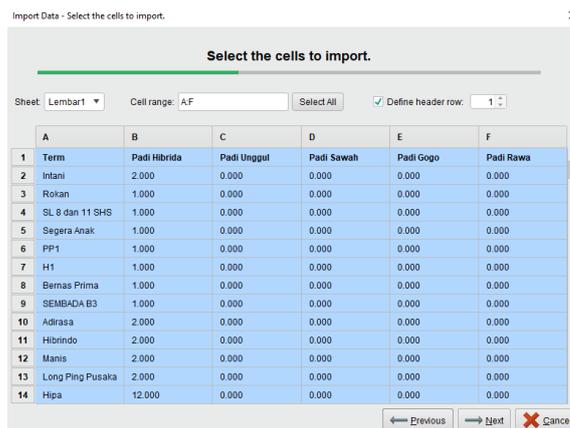
Pra proses data merupakan tahap penting dalam penelitian ini, yang melibatkan persiapan data sebelum dilakukan proses klasterisasi. Tahapan pra proses data terdiri dari *data selection*, *data cleaning*, dan *data transformation*. Pertama,

dalam *data selection*, dilakukan pemilihan data yang relevan dan diperlukan untuk proses klusterisasi. Kemudian, pada tahap *data cleaning*, dilakukan pembersihan data dengan menghapus baris data yang mengandung *missing value*. Diketahui bahwa dari total 303 baris data yang ada, tidak terdapat *missing value* dalam dataset ini. Selanjutnya, pada tahap *data transformation*, dilakukan transformasi data untuk mempersiapkannya agar siap untuk proses klusterisasi. Data dalam penelitian ini terdiri dari tipe *categorical* dan *numerical*, sehingga transformasi dapat dilakukan baik menggunakan *spreadsheet* Ms. Excel maupun menggunakan komponen yang tersedia dalam platform *Rapidminer*. Tahap ini penting untuk memastikan konsistensi dan kualitas data sebelum dilakukan analisis lebih lanjut menggunakan algoritma klusterisasi.

### 3.3 Pemodelan Menggunakan Algoritma K-Means

#### 1. Pemodelan Algoritma K-Means di Aplikasi RapidMiner:

Pemodelan metode *K-Means* di aplikasi *RapidMiner* dilakukan dengan langkah-langkah berikut, yang mencakup pembentukan desain proses dan hasil atau output eksekusi. Pembentukan Diagram Proses, langkah pertama adalah membentuk diagram proses, yang dimulai dengan *import* data dan dilanjutkan dengan tahap-tahap analisis menggunakan *algoritma K-Means*. Kemudian dilakukan import data dengan mengklik tombol "*Import Data*" di *Repository View*, kemudian memilih lokasi file pada *My Computer*, dan menentukan kolom yang akan diinput dengan menentukan rentang sel (misalnya, kolom A sampai F ditulis sebagai A:F). Setelah itu, format kolom ditentukan dengan menetapkan tipe data atau aturan atribut. Akhirnya, data disimpan di *local repository* data. Tampilan *import* data ke aplikasi *rapidminer* ditunjukkan pada gambar 1.



Gambar 1. Import Data Padi Di Aplikasi Rapidminer

## 2. Membuat Diagram Proses

Membuat diagram proses dilakukan dengan menambahkan operator dan menghubungkan operator di area proses *view RapidMiner*. Langkah-langkahnya adalah sebagai berikut [19]:

#### 1. Menambahkan Operator

Operator-operator yang diperlukan untuk analisis menggunakan *algoritma K-Means* ditambahkan ke dalam proses *view RapidMiner*. Ini melibatkan pemilihan operator yang sesuai dari panel operator dan menariknya ke dalam area proses *view*.

#### 2. Menghubungkan Operator:

Setelah operator-operator telah ditambahkan ke dalam proses *view*, langkah selanjutnya adalah menghubungkan operator tersebut. Ini dilakukan dengan menarik koneksi dari operator satu ke operator lainnya, sesuai dengan alur logika dan proses analisis yang diinginkan.

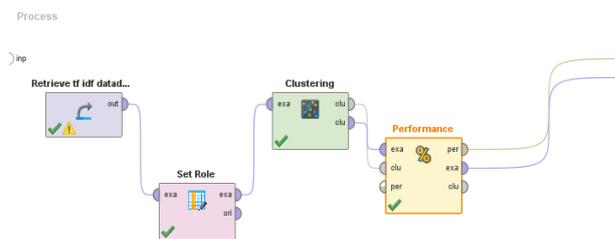
Dengan membuat diagram proses ini, kita dapat mengatur urutan operasi dan menggambarkan alur kerja dari awal hingga akhir analisis menggunakan *algoritma K-Means* pada data varietas padi. Operator yang digunakan untuk membentuk *design* proses metode *K-Means* ditunjukkan pada tabel 1.

TABEL 1. OPERATOR PROSES KLASTERISASI K-MEANS

No	Operator Dan Fungsi	Parameter
1	Import Data atau Read XLS Befungsi untuk memasukkan atau menggunakan dataset yang berada di file Ms Excel agar dapat digunakan sebagai data proses	Lokasi file
2	Set Role	Set role

No	Operator Dan Fungsi	Parameter
	Berfungsi untuk mendefinisikan atribut target atau label	atribut status sebagai label
3	K-Means Berfungsi menghasilkan model klasifikasi	k = 5 atau mengikuti parameter di <i>optimize parameter grid</i>
4	Apply Model Berfungsi melakukan operasi/hitungan pada data testing berdasarkan suatu model	Tidak ada
5	Performance Berfungsi penilaian performance suatu model klasifikasi	<i>Main criterion = accuracy</i>

Hasil dari pembentukan diagram proses ditunjukkan pada gambar 2.



Gambar 2. Hasil Pembentukan Diagram Proses

### 3. Eksekusi diagram proses

Setelah pembuatan diagram proses selesai, langkah selanjutnya adalah eksekusi diagram tersebut. Proses eksekusi ini melibatkan menjalankan diagram proses yang telah dirancang untuk melakukan analisis menggunakan *algoritma K-Means*. Hasil atau output dari proses eksekusi diagram proses model *K-Means* mencakup berbagai informasi yang penting untuk evaluasi dan interpretasi hasil klusterisasi.

Output yang diperoleh dari proses eksekusi meliputi data hasil klusterisasi, yang memberikan informasi tentang pengelompokan varietas padi ke dalam cluster-cluster yang dihasilkan oleh *algoritma K-Means*. Selain itu, hasil statistik data juga disajikan, termasuk nilai rata-rata, standar deviasi, dan statistik lainnya yang relevan terkait dengan hasil klusterisasi [20]. Selain itu, hasil optimasi parameter seperti jumlah *cluster* yang

optimal dan titik-titik pusat kluster juga dapat diperoleh, bersamaan dengan nilai performa dari model klusterisasi yang dihasilkan. Yang diperlihatkan pada Gambar 3.

Row No.	id	Term	cluster	Padi Hibrida	Padi Unggul	Padi Sawah	Padi Gogo	Padi Rawe
1	1	Intani	cluster_0	2	0	0	0	0
2	2	Rokan	cluster_0	1	0	0	0	0
3	3	SL 8 dan 11...	cluster_0	1	0	0	0	0
4	4	Segera Anak	cluster_0	1	0	0	0	0
5	5	PP1	cluster_0	1	0	0	0	0
6	6	H1	cluster_0	1	0	0	0	0
7	7	Bemas Prima	cluster_0	1	0	0	0	0
8	8	SEMBADA B3	cluster_0	1	0	0	0	0
9	9	Adrasa	cluster_0	2	0	0	0	0
10	10	Hibrido	cluster_0	2	0	0	0	0
11	11	Manis	cluster_0	2	0	0	0	0
12	12	Long Ping Pu...	cluster_0	2	0	0	0	0
13	13	Hipa	cluster_2	12	0	0	0	0
14	14	Hipa Jatim	cluster_0	3	0	0	0	0

Gambar 3. Tampilan Data Untuk Algoritma K-Means di Rapidminer

Setelah proses klusterisasi menggunakan *algoritma K-Means*, langkah berikutnya adalah menganalisis statistik data yang dihasilkan [21]. Tampilan statistik data dari *algoritma K-Means* memberikan informasi yang penting untuk pemahaman lebih lanjut tentang karakteristik dari masing-masing kluster yang terbentuk pada Gambar 4.

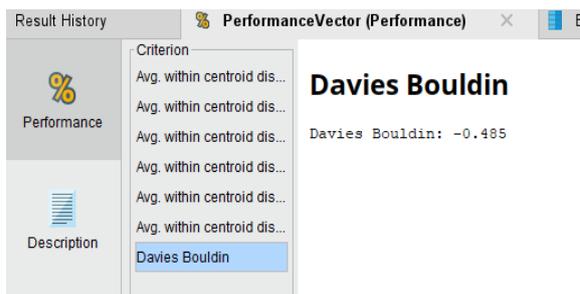
Name	Type	Missing	Statistics	Filter (8/8 attributes)	Search for Attributes
id	Integer	0	Min 1, Max 169, Average 85		
Term	Polynomial	0	Label: Woyia (1), Adi (1), Adrasa		
cluster	Nominal	0	Label: cluster_2 (1), cluster_0 (163)		
Padi Hibrida	Integer	0	Min 0, Max 12, Average 0.225		
Padi Unggul	Integer	0	Min 0, Max 25, Average 0.225		
Padi Sawah	Integer	0	Min 0, Max 25, Average 0.876		
Padi Gogo	Integer	0	Min 0, Max 5, Average 0.207		

Gambar 4. Tampilan Statistik Data Algoritma K-Means

Setelah melihat tampilan statistik data dari *algoritma K-Means*, langkah selanjutnya adalah melihat *performance vector K-Means* [22]. Tampilan ini memberikan informasi tentang evaluasi performa dari model klusterisasi yang dihasilkan. Dapat dilihat pada gambar 5.

Criterion	Avg. within centroid distance
Avg. within centroid distance	Avg. within centroid distance: -1.166

Gambar 5. Tampilan Performance Vector K-Means



Gambar 6. Hasil dari Davies Bouldin

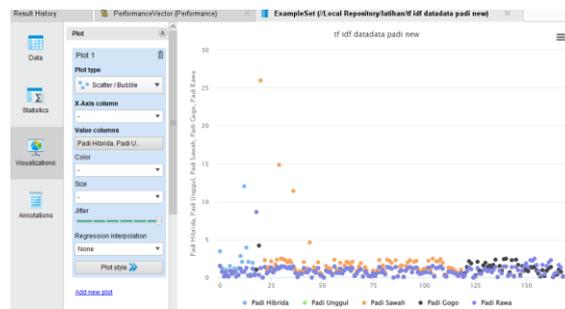
Gambar 6 menunjukkan hasil evaluasi *Davies Bouldin* dari data varietas padi, yang dinyatakan sebagai -0.485. Indeks *Davies Bouldin* merupakan algoritma evaluasi yang mengukur kualitas kluster dengan memperhatikan jarak intra-cluster yang rendah (tingkat kesamaan dalam kluster) dan jarak antara-cluster yang tinggi (tingkat perbedaan antara kluster) [23]. Semakin rendah nilai indeks *Davies Bouldin*, semakin baik kualitas kluster yang dihasilkan, karena menandakan adanya kesamaan yang tinggi dalam kluster dan perbedaan yang signifikan antara kluster [24]. Dengan nilai -0.485, dapat disimpulkan bahwa kluster yang dihasilkan memiliki tingkat kesamaan yang tinggi dalam kluster (jarak intra-cluster rendah) dan perbedaan yang signifikan antara kluster (jarak antara-cluster tinggi), sehingga menunjukkan kualitas kluster yang baik dalam analisis varietas padi.

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
Padi Hibrida	0.160	0	12	0	0
Padi Unggul	0	25	0	0	6.500
Padi Sawah	0.607	25	0	12	0
Padi Gogo	0.194	0	0	0	2.500
Padi Rawa	0.123	0	0	0	4

Gambar 7. Centroid Table

Pada Gambar 7, terdapat tabel *centroid* yang menampilkan nilai *centroid* dari masing-masing *cluster* dalam bentuk tabel. Tabel *centroid* ini memberikan gambaran tentang lokasi titik pusat dari setiap kluster yang terbentuk selama proses klusterisasi menggunakan *algoritma K-Means* [25]. Informasi ini sangat berguna untuk memahami karakteristik dari masing-masing kluster dan membantu interpretasi hasil klusterisasi secara lebih mendalam. Dengan

mengetahui nilai *centroid* dari setiap kluster, pengguna dapat melihat bagaimana data dalam kluster tersebut dipusatkan di sekitar titik-titik ini, serta memahami perbedaan karakteristik antara kluster satu dengan yang lainnya. Hal ini memungkinkan pengguna untuk mengambil keputusan yang lebih tepat dalam analisis data dan mengidentifikasi pola-pola yang muncul di dalam setiap kluster.



Gambar 8 Visualization

Pada Gambar 8, terdapat data penyebaran dari kluster yang dihasilkan selama proses klusterisasi. Data ini menggambarkan jumlah observasi atau data yang terdistribusi ke dalam setiap kluster. Dari data tersebut, terlihat bahwa kluster dengan penyebaran tertinggi adalah kluster padi sawah, yang memiliki jumlah observasi sebanyak 25. Sementara itu, kluster dengan penyebaran terendah adalah kluster padi rawa, yang tidak memiliki observasi sama sekali (angka 0). Selain itu, terdapat total lima kluster yang diidentifikasi dalam proses klusterisasi ini, yaitu padi hibrida, padi unggul, padi sawah, padi gogo, dan padi rawa. Data penyebaran ini memberikan gambaran yang jelas tentang distribusi data ke dalam masing-masing kluster, sehingga membantu untuk memahami proporsi dan distribusi data dalam analisis kluster.

#### 4. Kesimpulan dan Saran

Berdasarkan hasil implementasi dan pengujian algoritma K-Means untuk melakukan klusterisasi pada varietas padi, beberapa kesimpulan dapat diambil. Pertama, dari proses klusterisasi diperoleh lima kelompok cluster dengan karakteristik masing-masing, termasuk padi hibrida, padi unggul, padi sawah, padi gogo, dan padi rawa. Kedua, hasil klusterisasi ini dapat digunakan untuk memberikan informasi kepada pihak dinas pertanian dan petani mengenai berbagai jenis padi yang ada, sehingga membantu dalam pengambilan keputusan terkait pemilihan varietas padi yang sesuai dengan kebutuhan dan

kondisi lingkungan. Ketiga, penerapan metode K-Means clustering dalam penelitian ini menunjukkan hasil yang baik, menegaskan bahwa metode ini efektif dalam mengelompokkan varietas padi berdasarkan karakteristiknya. Kesimpulan ini memberikan kontribusi positif dalam upaya meningkatkan produktivitas dan kualitas hasil panen padi di Indonesia.

#### Daftar Pustaka:

- [1] R. Tsaniyah, A. Sholihah, A. B. Putra, S. Fitri, and A. Wati, "Analisis Produksi Padi Jawa Timur menggunakan Sistem Dinamik Mendukung Upaya Ketahanan Pangan," *Progresif*, vol. 20, no. 1, pp. 304–313, 2024, doi: <http://dx.doi.org/10.35889/progresif.v20i1.1563>.
- [2] W. A. Saputro, I. Firdauzi, and F. A. Harahap, "Potensi dan Ketersediaan Bahan Pangan Alternatif dalam Mendukung Ketahanan Pangan di Kabupaten Banyumas," *Jurnal Pertanian Agros*, vol. 25, no. 2, pp. 1208–1218, 2023.
- [3] I. M. Sakir and D. Desinta, "Pemanfaatan Refugia Dalam Meningkatkan Produksi Tanaman Padi Berbasis Kearifan Lokal," *Jurnal Lahan Suboptimal*, vol. 7, no. 1, pp. 97–105, 2019, doi: [10.33230/jlso.7.1.2018.367](https://doi.org/10.33230/jlso.7.1.2018.367).
- [4] Syahrullah, L. Aphrodyanti, and Mariana, "Kerusakan Beras oleh Sitophilus Oryzae L. dari Beberapa Varietas Padi," *Proteksi Tanaman Tropika*, vol. 2, no. 3, pp. 136–142, 2019.
- [5] I. Cinar and M. Koklu, "Identification of Rice Varieties Using Machine Learning Algorithms," *Tarim Bilimleri Dergisi*, vol. 28, no. 2, pp. 307–325, 2022, doi: [10.15832/ankutbd.862482](https://doi.org/10.15832/ankutbd.862482).
- [6] M. R. Larijani, E. A. Asli-Ardeh, E. Kozegar, and R. Loni, "Evaluation of image processing technique in identifying rice blast disease in field conditions based on KNN algorithm improvement by K-means," *Food Science and Nutrition*, vol. 7, no. 12, pp. 3922–3930, 2019, doi: [10.1002/fsn3.1251](https://doi.org/10.1002/fsn3.1251).
- [7] F. I. Nurjanah, Andi Farmadi, "Implementasi Metode Fuzzy C-Means Pada Sistem Clustering Data Varietas Padi," vol. 01, no. 01, pp. 23–32, 2018.
- [8] F. Rani and K. Afiani, "Penerapan K-Means Clustering Untuk Mengetahui Varietas Padi Unggul Produksi Balai Pengkajian Teknologi Pertanian Jawa Timur," *Jurnal Mahasiswa Teknik Informatika*, vol. 2, no. 1, pp. 336–343, 2018.
- [9] H. Snyder, "Literature review as a research methodology: An overview and guidelines," *Journal of Business Research*, vol. 104, no. March, pp. 333–339, 2019, doi: [10.1016/j.jbusres.2019.07.039](https://doi.org/10.1016/j.jbusres.2019.07.039).
- [10] M. R. Ondang and O. Kembuan, "Analisis Pola Lapangan Usaha Per Daerah Menggunakan Algoritma K-Means Clustering Pada Kabupaten Minahasa Selatan," *INNOVATIVE*, vol. 3, no. 6, pp. 7672–7686, 2023.
- [11] I. Rahmianti, "Analisis Kelayakan Pemberian Kredit Koperasi Dengan Metode Data Mining Decision Tree," *Jurnal Informatika dan Rekayasa Elektronik*, vol. 5, no. 2, pp. 153–161, 2022, doi: [10.36595/jire.v5i2.663](https://doi.org/10.36595/jire.v5i2.663).
- [12] V. Y. Kiselev, T. S. Andrews, and M. Hemberg, "Challenges in unsupervised clustering of single-cell RNA-seq data," *Nature Reviews Genetics*, vol. 20, no. 5, pp. 273–282, 2019, doi: [10.1038/s41576-018-0088-9](https://doi.org/10.1038/s41576-018-0088-9).
- [13] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *Multidisciplinary scientific journal*, vol. 2, no. 2, pp. 226–235, 2019, doi: [10.3390/j2020016](https://doi.org/10.3390/j2020016).
- [14] A. Amrullah, I. Purnamasari, B. N. Sari, and A. Voutama, "ANALISIS CLUSTER FAKTOR PENUNJANG PENDIDIKAN MENGGUNAKAN ALGORITMA K-MEANS (STUDI KASUS : KABUPATEN KARAWANG)," *Jurnal Informatika dan Rekayasa Elektronik*, vol. 5, no. 2, pp. 244–252, 2022.
- [15] K. Ademariana, F. R. Lumbanraja, and R. Andrian, "Jurnal Pepadun Clustering K-Means Jenis Kata Pada Laporan Kegiatan Kuliah Kerja Nyata ( KKN ) Universitas Lampung Menggunakan Word2vec © 2021 Ilmu Komputer Unila Publishing Network all rights reserved Jurnal Pepadun," vol. 3, no. 2, pp. 221–228, 2021.
- [16] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics (Switzerland)*, vol. 9, no. 8, pp. 1–12, 2020, doi: [10.3390/electronics9081295](https://doi.org/10.3390/electronics9081295).
- [17] S. Qaiser and R. Ali, "Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents," *International Journal of Computer Applications*, vol. 181, no. 1, pp. 25–29, 2018, doi: [10.3390/electronics9081295](https://doi.org/10.3390/electronics9081295).

- 10.5120/ijca2018917395.
- [18] T. Dogan and A. K. Uysal, "On Term Frequency Factor in Supervised Term Weighting Schemes for Text Classification," *Arabian Journal for Science and Engineering*, vol. 44, no. 11, pp. 9545–9560, 2019, doi: 10.1007/s13369-019-03920-9.
- [19] Sumarjono and M. A. Saputra, "Penerapan Data Mining Untuk Prediksi Ujuk Kerja Operasional Penambangan Batubara," *Tpt Perhapi*, pp. 245–258, 2022, [Online]. Available: <https://www.prosiding.perhapi.or.id/index.php/prosiding/article/view/293%0Ahttps://www.prosiding.perhapi.or.id/index.php/prosiding/article/view/293/369>
- [20] V. Virtusena, A. Johar, and A. Wijanarko, "Pengelompokan Potensi Kelulusan Mahasiswa Fakultas Teknik Unib Menggunakan Algoritme K-Means (Studi Kasus: Fakultas Teknik Universitas Bengkulu)," *Rekursif: Jurnal Informatika*, vol. 9, no. 2, pp. 206–225, 2021, doi: 10.33369/rekursif.v9i2.17073.
- [21] R. A. Indraputra and R. Fitriana, "K-Means Clustering Data COVID-19," *Jurnal Teknik Industri*, vol. 10, no. 3, pp. 275–282, 2020, doi: 10.25105/jti.v10i3.8428.
- [22] W. E. Susanto and D. Riana, "Komparasi Algoritma," *Jurnal Speed*, vol. 8, no. 3, pp. 18–27, 2016.
- [23] S. Ramadhani, D. Azzahra, and T. Z, "Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student's Thesis," *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, vol. 13, no. 1, pp. 24–33, 2022, doi: 10.31849/digitalzone.v13i1.9292.
- [24] A. A. Vergani and E. Binaghi, "A soft davis-bouldin separation measure," *IEEE International Conference on Fuzzy Systems*, vol. 2018-July, pp. 1–8, 2018, doi: 10.1109/FUZZ-IEEE.2018.8491581.
- [25] A. Primandana, S. Adinugroho, and C. Dewi, "Optimasi Penentuan Centroid pada Algoritme K-Means Menggunakan Algoritme Pillar (Studi Kasus: Penyandang Masalah Kesejahteraan Sosial di Provinsi ...," *... Teknologi Informasi dan Ilmu ...*, vol. 3, no. 11, pp. 10678–10683, 2020, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/6748/3264>