**JDSP**
**Jurnal Data Science Penusa**

# Data Mining Pengelompokan Bibit Padi Unggul Menggunakan Metode K-Means

**Putri Sintia Wati[1)*], Nuraisana Nuraisana [2)]**
[1,2,)] Program Studi Teknik Informatika, STMIK Pelita Nusantara, Medan
[1,2,)] putrisintiawati42376@gmail.com, nuraisana94@gmail.com

Received: 25 Maret 2025 Accepted: 14 April 2025 Published: 22 Juli 2025
Pg. 707-719

***Putri Sintia Wati**

**Kata Kunci:** Data Mining, Metode *K-Means*, Pengelompokan Bibit Padi Unggul, UD. Maju Jaya Tani.

**JDSP: Jurnal Data Science Penusa** is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

**Abstrak :** UD. Maju Jaya Tani adalah perusahaan dagang di pasar pertanian yang berfokus pada penjualan bibit unggul. UD. Maju Jaya Tani yang belum menerapkan sistem pengelompokan membantu mengelompokan bibit padi atau petani. Sistem pengelompokan memungkinkan usaha dagang dengan mudah memilih bibit padi yang memenuhi kebutuhan para petani. Banyaknya varietas padi dengan masing-masing karakteristiknya menimbulkan permasalahan pada proses penelitian padi, dikarenakan tidak semua varietas dapat sesuai dengan kualitas tanah yang akan ditanam. Salah satu metode pada Data Mining yang digunakan dalam penelitian ini yaitu metode K-Means. Metode K-Means sebagai salah satu teknik clustering dapat digunakan untuk pengelompokan bibit padi unggul. Dengan menggabungkan aspek teknis dan preferensi pada bibit unggul, penelitian ini bertujuan untuk memberikan solusi komprehensif untuk mengembangkan industri beras yang lebih adaptif dan responsif terhadap permintaan pasar. Dalam pembuatan penerapan K-means pada sistem pengelompokan Bibit Padi Unggul yang ada pada UD. Maju Jaya Tani menghasilkan keluaran untuk mengetahui Bibit Padi Unggul dengan melakukan pengelompokan. Dalam menerapkan metode K-Means Clustering pada pengelompokan bibit padi unggul dengan menggunakan 4 kriteria dengan 63 Varietas Padi, dan 2 cluster yaitu bibit padi unggul dan bukan bibit padi unggul. Setelah ditentukan alternatif dan kriteria maka diterapkan perhitungan metode K-Means Clustering dengan hasil akhir bahwa total persentase bibit padi unggul adalah 42,9% dan persentase bukan bibit padi unggul adalah 57,1%, dari pengujian tersebut maka disimpulkan bahwa metode K-Means Clustering dapat digunakan dalam pengelompokan bibit padi unggul.

## PENDAHULUAN

Dengan semakin berkembangnya teknologi yang diciptakan oleh manusia untuk meningkatkan kemampuan dalam menjalankan pekerjaannya, maka seorang manajemen akan banyak dihadapkan pada masalah, baik itu masalah yang sederhana maupun yang kompleks, diperlukan informasi-informasi yang menyeluruh dan akurat, kemampuan menganalisa dan mengolah informasi serta metode penyelesaian yang tepat. Sejalan dengan perkembangan teknologi informasi, semakin bertambah pula kemampuan komputer dalam membantu menyelesaikan permasalahan di berbagai bidang terutama pada bidang pertanian.

Pada ranah pertanian, pemilihan varietas bibit yang unggul sebelum penanaman merupakan suatu langkah tepat untuk optimasi hasil panen maupun pencegahan faktor kerugian. Varietas padi yang khusus berada di daerah tertentu disebut sebagai varietas lokal. Varietas semacam ini hanya cocok ditanam di daerah tertentu, terutama di Indonesia yang memiliki kulatitas tanah yang beriklim tropis sehingga banyak perusahaan dagang yang hanya menjual varietas lokal.

UD. Maju Jaya Tani adalah perusahaan dagang di pasar pertanian yang berfokus pada penjualan bibit unggul. UD. Maju Jaya Tani yang belum menerapkan sistem pengelompokan membantu mengelompokan bibit padi atau petani. Sistem pengelompokan memungkinkan usaha dagang dengan mudah memilih bibit padi yang memenuhi kebutuhan para petani. Banyaknya varietas padi dengan masing-masing karakteristiknya menimbulkan permasalahan pada proses penelitian padi, dikarenakan tidak semua varietas dapat sesuai dengan kualitas tanah yang akan ditanam. Apabila petani hanya asal menebak dan bersikap subjektif, maka selain proses produksi yang kurang efisien, hasil bibit yang dihasilkan juga belum tentu unggul. Oleh karena itu, perlu dibuat sebuah sistem pengelompokan bibit unggul sesuai dengan memperhitungkan segala kriteria yang ada.

Salah satu metode pada Data Mining yang digunakan dalam penelitian ini yaitu metode K-Means. Metode K-Means sebagai salah satu teknik clustering dapat digunakan untuk pengelompokan bibit padi unggul. Dengan menggabungkan aspek teknis dan preferensi pada bibit unggul, penelitian ini bertujuan untuk memberikan solusi komprehensif untuk mengembangkan industri beras yang lebih adaptif dan responsif terhadap permintaan pasar.

**TINJAUAN LITERATUR**

1. Data Mining

Data mining digunakan untuk memastikan keakuratan data yang disimpan dalam database. Data mining adalah proses yang memanfaatkan statistik, matematika, pembelajaran mesin, dan teknik kecerdasan buatan untuk menganalisis, mengidentifikasi, dan mengambil data dari berbagai database besar. (Jelita et al., 2023).

Data mining adalah suatu metode pengolahan data yang bertujuan untuk menemukan pola tersembunyi di dalam data. Data yang diperoleh dari data mining dapat digunakan untuk meningkatkan produktivitas dalam sejumlah besar transaksi. Istilah data mining juga sering disebut sebagai keterbukaan data atau pengenalan pola. Konsep data mining dapat dipandang sebagai node-node yang masih memerlukan pemahaman yang mendalam, mengacu pada informasi mentah yang sedang dihadapi (Riadi & Mesran, 2023).

2. K-means

K-Means membagi data yang ada menjadi satu atau lebih cluster atau kelompok sehingga data yang mempunyai karakteristik yang sama dikelompokkan ke dalam cluster yang sama dan data yang mempunyai karakteristik berbeda dikelompokkan ke dalam kelompok yang lain. Ini adalah metode pengelompokan data non-hierarki (Manihuruk et al., 2020).

3. Penerapan K-Means
   a. Menentukan jumlah cluster (k) pada data set
   b. Menenukan pusat cluster (centroid)
   c. Penentuan nilai centroid pada tahap awal dilakukan secara random, sedangkan pada tahap iterasi digunakan rumus seperti pada persamaan (1) berikut ini :

   $V_{ij} = \frac{1}{Ni}\sum_{k=0}^{Ni} X\ k\ j$ ..... (1)

   Keterangan :
   Vij = centroid rata-rata cluster ke-I untuk variabel ke-j Ni = Jumlah anggota cluster ke-i
   i, k = indeks dari cluster j = indeks dari variabel
   Xkj = nilai data ke-k variabel ke-j untuk cluster tersebut
   d. Pada masing-masing record, hitung jarak terdekat dengan centroid. Jarak centroid yang digunakan adalah Euclidean Distance, dengan rumus seperti dibawah ini:

   $D_e\sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$ ..... (2)

   Keterangan :
   D = Euclidean Distance
   x = Banyaknya objek
   $\sum$p = Jumlah data record

   $d_{Euclidean\,(x,y)} = \sqrt{\sum_i (x_i - y_i)^2}$ ..... (3)

   Keterangan :

De = Euclidean Distance
i = Banyaknya objek
(x,y) = Koordinat objek (
s,t) = Koordinat centroid

e. Kelompokan objek berdasarkan jarak ke centroid terdekat
f. Ulangi langkah ke-2, lakukan iterasi hingga centroid bernilai optimal (Manihuruk et al., 2020).

4. Bibit Padi

Bibit padi merupakan salah satu bahan baku utama yang berperan penting dalam meningkatkan kualitas produk beras. Penggunaan bibit varietas kelas satu yang berkualitas (berlabel) sangat dianjurkan. Hal ini disebabkan oleh karakteristik varietas yang berkualitas, seperti hasil tinggi, tahan terhadap hama dan penyakit, serta nasi yang enak (empuk). Faktor-faktor yang mempengaruhi produksi bibit padi yaitu tersedianya bibit yang bermutu bagi petani pada saat melakukan kegiatan pertanian merupakan syarat penting dalam meningkatkan hasil dan mutu produksi. Sebaiknya gunakan bibit yang berkualitas tinggi agar petani dan produsen bibit tidak dirugikan dan dapat menjamin kualitas bibit yang digunakan.

Padi merupakan komoditas tanaman yang penting di Indonesia. Seiring berjalannya waktu terjadi penyempitan lahan sawah irigasi subur akibat konversi lahan untuk kepentingan non pertanian. Salah satu upaya yang dapat dilakukan adalah penggunaan varietas dan jarak tanam yang tepat, seperti varietas unggul padi gogo (Nanda et al., 2022).

5. Variated Bibit Unggul

Pada ranah pertanian, pemilihan varietas bibit yang unggul sebelum penanaman merupakan suatu langkah tepat untuk optimasi hasil panen maupun pencegahan faktor kerugianVarietas berkualitas tinggi sendiri dikembangkan di Indonesia sebelum tahun 1970. Umumnya varietas yang berkualitas tinggi diberi nama sungai seperti Bengawan, Brantas, dan Citarum. Namun sejak tahun 2008, muncul nama-nama baru untuk memberi nama VUB sesuai dengan karakteristiknya, seperti impari (padi hibrida), impago (padi gogo hibrida), dan impala (padi hibrida lahan basah). (Fitra. R. K. A, 2020).

6. Unified Modeling Language (UML)

UML digunakan untuk menentukan, memvisualisasikan, membuat, dan mendokumentasikan artefak (potongan informasi yang dihasilkan dalam proses pembuatan perangkat lunak, artefak ini dapat berupa model, deskripsi, perangkat lunak, dll) dari sistem perangkat lunak, seperti pemodelan bisnis atau non-perangkat. sistem ini adalah bahasa untuk menjadi Perangkat lunak lainnya. Selain itu, UML adalah bahasa pemodelan yang menggunakan konsep berorientasi objek. UML dikembangkan oleh Grady Booch, James Rumbaugh, dan Ivar Jacobson di bawah bendera Rational Software Corps. UML menyediakan notasi yang membantu memodelkan sistem dari perspektif yang berbeda (Prasetya et al., 2022).

7. PHP (Pre-Hypertext Prepocessor)

PHP adalah bahasa pemrograman yang berjalan di situs web. Biasanya digunakan untuk memproses informasi di Internet. PHP adalah singkatan dari Pre- Hypertext Preprocessor, yang merupakan bahasa pemrograman web sisi server yang terbuka atau gratis. PHP adalah skrip yang tertanam dalam HTML dan ditempatkan di server (Kadarsih & Andrianto, 2022).

8. MySQL

MySQL merupakan sistem manajemen basis data dengan model sistem manajemen basis data relasional "RDBMS". Seperti server SQL. MySQL tersedia secara gratis sebagai sistem open source dan dapat dikembangkan lebih lanjut sesuai kebutuhan. MySQL dapat menangani jutaan query dan ribuan transaksi secara bersamaan, menjadikannya pilihan pertama untuk pengembangan aplikasi berbasis web dan web (Sidharta & Wibowo, 2020).

9. XAMPP

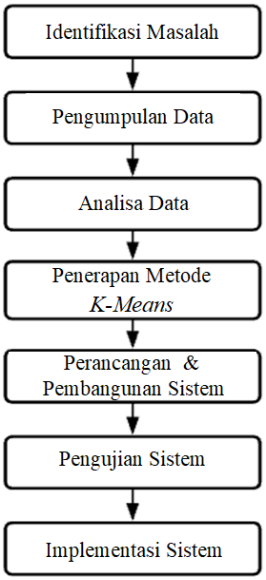
XAMPP merupakan suatu software yang bersifat open source yang merupakan pengembangan dari LAMP (Linux, Apache, MySQL, PHP dan Perl), XAMPP adalah tool pembantu pengembangan paket perangkat lunak berbasis open source yang menggabungkan Apache web server, MySQL, PHP dan beberapa modul lainnya di dalam satu paket aplikasi. (Nopita., Dkk, 2021)

10. WEB

Website adalah keseluruhan halaman web yang terdapat dalam suatu domain yang berisi informasi. Sebuah situs web biasanya terdiri dari banyak halaman web yang saling berhubungan. Website adalah suatu aplikasi berisi dokumen multimedia (teks, gambar, suara, animasi, video) yang menggunakan protokol HTTP (Hyper Text Transfer Protocol) dan diakses melalui perangkat lunak yang disebut browser. (Nopita., Dkk, 2021)

**METODE PENELITIAN**

Metode penelitian diartikan sebagai cara ilmiah untuk mendapatkan data dengan tujuan dan kegunaan tertentu. Kerangka kerja merupakan suatu proses atau langkah-langkah berupa urutan kegiatan yang akan dilakukan dalam suatu penelitian guna menyelesaikan permasalahan yang ada untuk mencapai hasil yang maksimal dan akurat. Tahapan kerangka kerja penelitian dapat dilihat pada gambar 1 berikut :




Gambar 1 Kerangka Kerja Penelitian

**HASIL PENELITIAN**

Penelitian ini mengunakan 63 data varietas padi dalam perhitungan metode K-Means dalam mengelompokan Bibit Padi Unggul yang ada pada UD. Maju Jaya Tani. Kemudian masing-masing data penelitian dimasukkan kedalam tabel 1 dan tabel 2 :

Tabel 1 Data Bibit Padi

| No | Varietas Padi | Warna Gabah | Bentuk Gabah | Kadar Amilosa | Kadar air |
|---|---|---|---|---|---|
| 1 | Inpara 3 | Kuning | Ramping | 22,92 % | 23 % |
| 2 | Inpara 4 | Kuning | Ramping | 3,88 % | 23 % |
| 3 | Inpara 5 | Kuning Coklat | Ramping | 21,65 % | 24 % |
| 4 | Inpari 13 | Kuning | Sedang | 4,29 % | 26 % |
| 5 | Inpari 18 | Coklat | Sedang | 27,6 % | 21 % |
| 6 | Inpari 31 | Coklat | Sedang | 28,48 % | 22 % |
| 7 | Inpari 11 | Coklat | Sedang | 21,42 % | 22 % |
| 8 | Inpari 12 | Kuning Coklat | Sedang | 10,33 % | 25 % |
| 9 | Inpari 17 | Kuning Coklat | Sedang | 5,99 % | 22 % |
| 10 | Inpago 4 | Merah Kuning | Ramping | 26,92 % | 25 % |
| ... | ... | ... | ... | ... | ... |
| 60 | ketan hitam | Kuning | Membulat | 34,38 % | 24 % |
| 61 | pandan wangi | Coklat | Membulat | 11,94 % | 24 % |
| 62 | Hipa 4 | Coklat | Membulat | 11,92 % | 24 % |

| No | Varietas Padi | Warna Gabah | Bentuk Gabah | Kadar Amilosa | Kadar air |
|---|---|---|---|---|---|
| 63 | IR64 | Coklat | Membulat | 31,23 % | 20 % |

Dari data yang telah diperoleh diatas maka akan diambil data yang akan diteliti. Dalam hal ini data yang akan diteliti adalah warna gabah, bentuk gabah, kadar amilosa, dan kadar air. Berikut data yang sudah dikumpulkan dari data diatas tersebut :

Tabel 2 Data Kriteria

| No | Kriteria | Keterangan |
|---|---|---|
| 1 | **Warna Gabah** | - Kuning<br>- Kuning Coklat<br>- Coklat<br>- Merah Kuning<br>- Merah Coklat |
| 2 | **Bentuk Gabah** | - Ramping<br>- Sedang<br>- Membulat |
| 3 | **Kadar Amilosa** | - 1-19%)<br>- 20-25%<br>- >25% |
| 4 | **Kadar air** | - 20-23%<br>- 24-27% |

Kriteria yang digunakan ada 4 dikarenakan data tersebut dapat membantu menilai seberapa unggul bibit padi tiap varietas. Dari data kriteria tersebut maka dijabarkan kembali sehingga mendapatkan nilai dari kriteria yang akan menjadi parameter dan nantinya akan digunakan dalam menghitung pengelompokan Bibit Padi Unggul, berikut adalah table nilai sub kriteria tersebut :

Tabel 3 Data Sub Kriteria Warna Gabah

| No | Sub Kriteria | Nilai | Keterangan |
|---|---|---|---|
| 1 | Kuning | 1 | Berwarna kuning |
| 2 | Kuning Coklat | 2 | Berwarna kuning dan sedikit coklat |
| 3 | Coklat | 3 | Berwarna coklat |
| 4 | Merah Kuning | 4 | Berwarna merah sedikit kuning |
| 5 | Merah Coklat | 5 | Berwarna merah kecoklatan |

Pada sub kriteria ini dibagi menjadi 5 kelompok yaitu seperti yang di jelaskan pada tabel di atas. Jika dilihat dari warna gabah yang berbeda-beda tersebut maka dapat dipastikan nilai atau bobot dari setiap kelompok warna gabah. Semakin gelap warna gabah maka semakin tinggi nilai atau bobot tersebut.

Tabel 4 Data Sub Kriteria Bentuk Gabah

| No | Sub Kriteria | Nilai | Keterangan |
|---|---|---|---|
| 1 | Ramping | 1 | Berbentuk ramping |
| 2 | Sedang | 2 | Berbentuk sedang |
| 3 | Membulat | 3 | Berbentuk membulat |

Pada sub kriteri ini dibagi menjadi 3 bagian yaitu bentuk dari gabah tersebut, jika semakit berisi maka semakin besar nilai atau bobot penilaian tersebut.

Tabel 5 Data Sub Kriteria Kadar Amilosa

| No | Sub Kriteria | Nilai | Keterangan |
|---|---|---|---|
| 1 | 1% S/D 19% | 1 | Beras ketan |
| 2 | 20% S/D 25% | 2 | Beras beramilosa sedang |

| | | | |
|---|---|---|---|
| 3 | >25% | 3 | Beras beramilosa tinggi |

Pada sub kriteria ini memiliki nilai persen yang berbeda-beda yang dimulai dari 1% hingga lebih dari 25% tergantung dari tingkat Kadar Amilosa padi bibit unggul.

Tabel 6 Data Sub Kriteria Kadar air

| No | Sub Kriteria | Nilai | Keterangan |
|---|---|---|---|
| 1 | 20-23% | 1 | Kodisi pada musim kering |
| 2 | 24-27% | 2 | Kodisi pada musim hujan |

Pada sub kriteri ini dibagi menjadi 2 bagian dikarenakan Kadar air optimum dalam penyimpanan bagi sebagian besar benih padi adalah antara 20 % - 27 %.

Dari data tersebut. Yang ingin dicapai adalah bibit unggul dari masing – masing varietas padi. Berikut pengelompokan bibit padi unggul dapat dilihat pada tabel berikut :

Tabel 7 Pengelompokan

| No | Kode | Kelompok |
|---|---|---|
| 1 | C1 | Bibit Padi Unggul |
| 2 | C2 | Bukan Bibit Padi Unggul |

**Metode K-Means**

K-Means termasuk dalam Partitioning Clustering. Semua data harus terdapat pada cluster tertentu, dan semua data yang terdapat pada cluster tertentu pada satu tahapan proses boleh dipindahkan ke cluster lain pada tahapan berikutnya. K-Means membagi data menjadi rentang 'k' yang berbeda. 'k' adalah bilangan bulat positif.

Algoritma K-Means terkenal tidak hanya karena kesederhanaannya tetapi juga kemampuannya dalam mengklasifikasikan data dalam jumlah besar dan outlier dengan sangat cepat. Langkah-langkah algoritma K-Means adalah:

1. Menentukan Pusat Cluster Awal
   Dalam menentukan 'n' pusat cluster awal, dihasilkan angka-angka acak yang mewakili urutan data masukan. Pusat-pusat awal cluster ditentukan dari data itu sendiri dengan mengacak pusat-pusat awal dari data daripada menentukan titik-titik baru.
2. Menghitung Jarak ke Pusat Cluster
   Jarak euclidean digunakan untuk mengukur jarak antara data dengan pusat cluster. Algoritma penghitungan jarak antara data dan pusat cluster
   a. Memperoleh nilai data dan rata-rata cluster
   b. Menghitung jarak Euclidean data ke setiap pusat cluster
3. Mengelompokkan data
   Membandingkan hasil perhitungan jarak dan memilih jarak terpendek antara data dan pusat cluster. Jarak ini menunjukkan bahwa data berada dalam kelompok yang sama dengan pusat cluster terdekat. Algoritma pengelompokan data :
   a. Mendapatkan nilai jarak setiap pusat cluster dengan data
   b. Carilah nilai jarak minimum
   c. Kelompokkan data berdasarkan pusat cluster dengan jarak terkecil.
4. Tentukan pusat cluster baru.
   Untuk memperoleh pusat cluster baru dapat dihitung dari nilai mean anggota cluster dan pusat cluster. Jika hasil yang diperoleh tidak konvergen, maka digunakan pusat cluster baru untuk melakukan iterasi berikutnya. Proses iterasi berhenti ketika jumlah maksimum iterasi yang dimasukkan oleh pengguna tercapai atau ketika hasil yang diperoleh menyatu (pusat cluster baru sama dengan pusat cluster lama).
   Algoritma K-Means sangat terkenal karena kemudahan dan juga serta kemampuannya untuk mengklasifikasi data besar dan outlier dengan sangat cepat. Dan berikut ini adalah langkah-langkah algoritma K-Means :

a. Cari jumlah anggota tiap cluster
b. Hitung pusat baru dengan rumus :

Pusat cluster baru :

ini adalah rumus untuk menentukan 2 pusat awal cluster

$$\frac{x1+x2+x3.....+xn+xp}{Jumlah+1} \quad \text{.....(1)}$$

Dimana :

x1, x2, x3, ...xn = anggota cluster

xp = pusat lama

**Perhitungan K-Means**

Contoh perhitungan manual menggunakan metode K-Means dengan 63 data untuk mengetahui proses pengelompokan data. Berikut adalah tahapan-tahapan dalam pengelompokan bibit padi unggul. Dalam tahap ini akan dijelaskan langkah-langkah pengoperasian algoritma K-Means secara manual:

Diketahui :

Jumlah Cluster = 2,
Jumlah data = 63,
Jumlah Kriteria = 4

Menentukkan kriteria yang nantinya digunakan sebagai parameter nilai buat masing-masing varietas, dengan kriteria nya sebagai berikut :

Tabel 8 Parameter Bibit Padi Unggul

| No | Varietas Padi | K1 | K2 | K3 | K4 |
|---|---|---|---|---|---|
| 1 | Inpara 3 | 1 | 1 | 2 | 1 |
| 2 | Inpara 4 | 1 | 1 | 1 | 1 |
| 3 | Inpara 5 | 2 | 1 | 2 | 2 |
| 4 | Inpari 13 | 1 | 2 | 1 | 2 |
| 5 | Inpari 18 | 3 | 2 | 3 | 1 |
| 6 | Inpari 31 | 3 | 2 | 3 | 1 |
| 7 | Inpari 11 | 3 | 2 | 2 | 1 |
| 8 | Inpari 12 | 2 | 2 | 1 | 2 |
| 9 | Inpari 17 | 2 | 2 | 1 | 1 |
| 10 | Inpago 4 | 4 | 1 | 3 | 2 |
| ... | ... | ... | ... | ... | ... |
| 60 | ketan hitam | 1 | 3 | 3 | 2 |
| 61 | pandan wangi | 3 | 3 | 1 | 2 |
| 62 | Hipa 4 | 2 | 2 | 1 | 2 |
| 63 | IR64 | 3 | 3 | 3 | 1 |

**Iterasi ke-1**

1. Penentuan pusat awal cluster

   Untuk penentuan awal pusat cluster di asumsikan sebagai berikut :

   Diambil data ke - 1 sebagai pusat Cluster Ke - 1 :

   C1 = 1; 1; 2; 1

   Diambil data ke - 2 sebagai pusat Cluster Ke - 2 :

   C2 = 1; 1; 1; 1

2. Perhitungan matrix jarak pusat cluster

   Untuk mengukur jarak antara data dengan pusat cluster digunakan Euclidian distance, kemudian akan didapatkan matrik jarak sebagai berikut:

   Rumus Euclidian distance:

   $$d_{Euclidean\ (x,y)} = \sqrt{\sum_i (x_i - y_i)^2} \quad \text{.....(2)}$$

   Dimana :

   x = pusat cluster

y = data

Sebagai contoh, perhitungan jarak ke A-1 terhadap pusat cluster adalah :

$C1 = \sqrt{(1-1)^2 + (1-1)^2 + (2-2)^2 + (1-1)^2} = 0$

$C2 = \sqrt{(1-1)^2 + (1-1)^2 + (1-2)^2 + (1-1)^2} = 1$

Dan seterusnya dilanjutkan untuk data ke 2, 3, ... n

Setiap kolom pada matrik menunjukkan nilai jarak data terhadap pusat cluster. Baris pertama pada matrik menunjukkan nilai jarak data terhadap titik pusat cluster pertama, baris kedua pada matrik menunjukkan nilai jarak data terhadap titik pusat cluster kedua dan seterusnya.

3. Matrix Pengelompokkan data

   Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat cluster, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat cluster terdekat.

   Berikut hasil matriks pengelompokkan group, nilai 1 berarti data tersebut berada dalam group. Dan dari hasil matrix pengelompokkan maka didapat pembagain kelompok varietas sebagai berikut :

Tabel 9 Parameter Matriks Hasil Ke-1

| No | Kode | K1 | K2 | K3 | K4 | C1 | C2 | Hasil |
|---|---|---|---|---|---|---|---|---|
| 1 | A001 | 1 | 1 | 2 | 1 | 0 | 1 | C1 |
| 2 | A002 | 1 | 1 | 1 | 1 | 1 | 0 | C2 |
| 3 | A003 | 2 | 1 | 2 | 2 | 1,4142136 | 1,7320508 | C1 |
| 4 | A004 | 1 | 2 | 1 | 2 | 1,7320508 | 1,4142136 | C2 |
| 5 | A005 | 3 | 2 | 3 | 1 | 2,4494897 | 3 | C1 |
| 6 | A006 | 3 | 2 | 3 | 1 | 2,4494897 | 3 | C1 |
| 7 | A007 | 3 | 2 | 2 | 1 | 2,236068 | 2,4494897 | C1 |
| 8 | A008 | 2 | 2 | 1 | 2 | 2 | 1,7320508 | C2 |
| 9 | A009 | 2 | 2 | 1 | 1 | 1,7320508 | 1,4142136 | C2 |
| 10 | A010 | 4 | 1 | 3 | 2 | 3,3166248 | 3,7416574 | C1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 60 | A060 | 1 | 3 | 3 | 2 | 2,4494897 | 3 | C1 |
| 61 | A061 | 3 | 3 | 1 | 2 | 3,1622777 | 3 | C2 |
| 62 | A062 | 2 | 2 | 1 | 2 | 2 | 1,7320508 | C2 |
| 63 | A063 | 3 | 3 | 3 | 1 | 3 | 3,4641016 | C1 |

Total C1 = 31

C2 = 32

4. Penentuan pusat cluster baru

   Setelah diketahui anggota tiap-tiap cluster, kemudian pusat cluster baru dihitung berdasarkan data anggota tiap-tiap cluster sesuai dengan rumus pusat anggota cluster.

   Karena C1 memiliki 31 anggota maka cluster baru untuk C1 menjadi:

   P1 = (1 + 2 + 3 + ... + 3)/31 = 2,193548387

   P2 = (1 + 1 + 2 +... + 3)/31 = 1,774193548

   P3 = (2 + 2 + 3 + ... + 3)/31 = 2,709677419

   P4 = (2 + 2 + 1 + ... + 1)/31 = 1,548387097

   C1 = (2,193548387; 1,774193548; 2,709677419; 1,548387097)

   Dan begitu juga untuk C2. Karena C2 hanya mempunyai 33 anggota, maka cluster baru untuk C2 menjadi :

   P1 = (1 + 1 + 3 + ... + 2)/33 = 2,78125

   P2 = (1 + 2 + 2 + ... + 2)/33 = 1,90625

   P3 = 1 + 1 + 2 + ... + 1)/33 = 1

   P4 = (1 + 2 + 1 + ... + 2)/33 = 1,4375

   C2 = (2,78125; 1,90625; 1; 1,4375)

**Iterasi ke-2**

Untuk proses Iterasi ke-2, ulangi lagi langkah ke 2 (kedua) seperti Iterasi Ke-1 yaitu untuk mencari perhitungan matrix jarak pusat cluster hingga posisi data tidak mengalami perubahan, dan untuk Iterasi ke-2 matrix jarak pusat cluster nya menjadi seperti ini :

$C1 = \sqrt{(2{,}1935-1)^2 + (1{,}7742-1)^2 + (2{,}7097-2)^2 + (1{,}5484-1)^2}$

$C1 = 1{,}68175618$

$C2 = \sqrt{(2{,}7812-1)^2 + (1{,}9062-1)^2 + (1-2)^2 + (1{,}4375-1)^2}$

$C2 = 2{,}277179588$

Dan seterusnya dilanjutkan untuk data ke 2, 3, ... n

Langkah selanjutnya sama dengan langkah pada nomer 3 iterasi ke-1 diatas, jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat cluster, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat cluster terdekat

Berikut ini akan ditampilkan data matriks pengelompokkan group, nilai 1 berarti data tersebut berada dalam group. Dan dari hasil matrix pengelompokkan itu kita mendapatkan pembagain kelompok alternatif sebagai berikut :

Tabel 10 Parameter Matriks Hasil Ke-2

| No | Kode | P1 | P2 | P3 | P4 | C1 | C2 | Hasil |
|---|---|---|---|---|---|---|---|---|
| 1 | A001 | 1 | 1 | 2 | 1 | 1,68175618 | 2,277179588 | C1 |
| 2 | A002 | 1 | 1 | 1 | 1 | 2,290776875 | 2,045860913 | C2 |
| 3 | A003 | 2 | 1 | 2 | 2 | 1,159496823 | 1,657723401 | C1 |
| 4 | A004 | 1 | 2 | 1 | 2 | 2,145343189 | 1,870306626 | C2 |
| 5 | A005 | 3 | 2 | 3 | 1 | 1,042289963 | 2,061079056 | C1 |
| 6 | A006 | 3 | 2 | 3 | 1 | 1,042289963 | 2,061079056 | C1 |
| 7 | A007 | 3 | 2 | 2 | 1 | 1,227079136 | 1,117160183 | C2 |
| 8 | A008 | 2 | 2 | 1 | 2 | 1,793153821 | 0,967236721 | C2 |
| 9 | A009 | 2 | 2 | 1 | 1 | 1,819938136 | 0,900303768 | C2 |
| 10 | A010 | 4 | 1 | 3 | 2 | 2,037371958 | 2,573528099 | C1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 60 | A060 | 1 | 3 | 3 | 2 | 1,793153821 | 2,947125188 | C1 |
| 61 | A061 | 3 | 3 | 1 | 2 | 2,297806944 | 1,249218506 | C2 |
| 62 | A062 | 2 | 2 | 1 | 2 | 1,793153821 | 0,967236721 | C2 |
| 63 | A063 | 3 | 3 | 3 | 1 | 1,593104287 | 2,331425932 | C1 |

Karena C1 memiliki 29 anggota maka cluster baru untuk C1 menjadi: hitung hingga didapat persamaan**,** pada iterasi ke-3 = Iterasi ke-4 memiliki anggota yang sama maka tidak perlu dilakukan iterasi/perulangan lagi. Hasil clustering telah mencapai stabil dan konvergen. Sehingga didapat hasil akhir dan dapat dilihat pada tabel 4.15 dibawah ini.

Tabel 11 Hasil Perhitungan K-Means

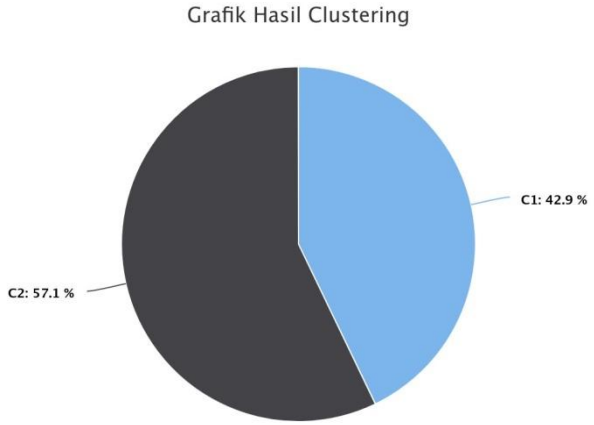
| No | Kode | Varietas Padi | Hasil | Keterangan |
|---|---|---|---|---|
| 1 | A001 | Inpara 3 | C1 | Bibit Padi Unggul |
| 2 | A002 | Inpara 4 | C2 | Bukan Bibit Padi Unggul |
| 3 | A003 | Inpara 5 | C1 | Bibit Padi Unggul |
| 4 | A004 | Inpari 13 | C2 | Bukan Bibit Padi Unggul |
| 5 | A005 | Inpari 18 | C1 | Bibit Padi Unggul |
| 6 | A006 | Inpari 31 | C1 | Bibit Padi Unggul |
| 7 | A007 | Inpari 11 | C2 | Bukan Bibit Padi Unggul |
| 8 | A008 | Inpari 12 | C2 | Bukan Bibit Padi Unggul |
| 9 | A009 | Inpari 17 | C2 | Bukan Bibit Padi Unggul |
| 10 | A010 | Inpago 4 | C1 | Bibit Padi Unggul |
| ... | ... | ... | ... | ... |
| 60 | A060 | ketan hitam | C1 | Bibit Padi Unggul |

| No | Kode | Varietas Padi | Hasil | Keterangan |
|---|---|---|---|---|
| 61 | A061 | pandan wangi | C2 | Bukan Bibit Padi Unggul |
| 62 | A062 | Hipa 4 | C2 | Bukan Bibit Padi Unggul |
| 63 | A063 | IR64 | C1 | Bibit Padi Unggul |

Setelah di ketahui hasil pengelompokan bibit padi unggul menggunakan metode K-Means di dapatlah sebanyak 27 varietas padi dengan kualitas Bibit Padi Unggul dan sebanyak 36 varietas padi dengan kualitas Bukan Bibit Padi Unggul. Sehingga dapat diketahui persentase tiap varietas adalah.

$$Persentase\ Bibit\ unggul = \frac{Total\ Varietas}{Bibit\ Unggul} x100\%$$
$$Persentase\ Bibit\ unggul = \frac{27}{63} x100\%$$
$$Persentase\ Bibit\ unggul = 42{,}9\%$$

$$Persentase\ Bukan\ Bibit\ unggul = \frac{Total\ Varietas}{Bukan\ Bibit\ Unggul} x100\%$$
$$Persentase\ Bukan\ Bibit\ unggul = \frac{36}{63} x100\%$$
$$Persentase\ Bukan\ Bibit\ unggul = 57{,}1\%$$




Gambar 2 Grafik Hasil Clustering

Hasil yang telah diperolah akan disortir akan mempermudah dalam pengelompokan. Hasilnya dapat dilihat pada tabel 12 dan tabel 13 dibawah ini.
Tabel hasil bibit padi unggul sebanyak 27 data.

Tabel 12 Hasil Bibit Padi Unggul

| No | Kode | Varietas Padi | Hasil | Keterangan |
|---|---|---|---|---|
| 1 | A001 | Inpara 3 | C1 | Bibit Padi Unggul |
| 2 | A003 | Inpara 5 | C1 | Bibit Padi Unggul |
| 3 | A005 | Inpari 18 | C1 | Bibit Padi Unggul |
| 4 | A006 | Inpari 31 | C1 | Bibit Padi Unggul |
| 5 | A010 | Inpago 4 | C1 | Bibit Padi Unggul |
| 6 | A015 | Inpago 10 | C1 | Bibit Padi Unggul |
| 7 | A016 | Conde | C1 | Bibit Padi Unggul |
| 8 | A020 | Aen Muti | C1 | Bibit Padi Unggul |
| 9 | A021 | Intani 1 | C1 | Bibit Padi Unggul |
| 10 | A022 | Intani 2 | C1 | Bibit Padi Unggul |
| ... | ... | ... | ... | ... |
| 25 | A059 | ketan putih | C1 | Bibit Padi Unggul |

| No | Kode | Varietas Padi | Hasil | Keterangan |
|---|---|---|---|---|
| 26 | A060 | ketan hitam | C1 | Bibit Padi Unggul |
| 27 | A063 | IR64 | C1 | Bibit Padi Unggul |

Tabel hasil bibit padi tidak unggul sebanyak 36 data.

Tabel 13 Hasil Bibit Padi Tidak Unggul

| No | Kode | Varietas Padi | Hasil | Keterangan |
|---|---|---|---|---|
| 1 | A002 | Inpara 4 | C2 | Bukan Bibit Padi Unggul |
| 2 | A004 | Inpari 13 | C2 | Bukan Bibit Padi Unggul |
| 3 | A007 | Inpari 11 | C2 | Bukan Bibit Padi Unggul |
| 4 | A008 | Inpari 12 | C2 | Bukan Bibit Padi Unggul |
| 5 | A009 | Inpari 17 | C2 | Bukan Bibit Padi Unggul |
| 6 | A011 | Inpago 5 | C2 | Bukan Bibit Padi Unggul |
| 7 | A012 | Inpago 7 (Beras Merah) | C2 | Bukan Bibit Padi Unggul |
| 8 | A013 | Inpago 8 | C2 | Bukan Bibit Padi Unggul |
| 9 | A014 | Inpago 9 | C2 | Bukan Bibit Padi Unggul |
| 10 | A017 | Situ Bagendit | C2 | Bukan Bibit Padi Unggul |
| ... | ... | ... | ... | ... |
| 34 | A058 | ketan merah | C2 | Bukan Bibit Padi Unggul |
| 35 | A061 | pandan wangi | C2 | Bukan Bibit Padi Unggul |
| 36 | A062 | Hipa 4 | C2 | Bukan Bibit Padi Unggul |

**PEMBAHASAN**

Dalam pembuatan penerapan K-means pada sistem pengelompokan Bibit Padi Unggul yang ada pada UD. Maju Jaya Tani menghasilkan keluaran untuk mengetahui Bibit Padi Unggul dengan melakukan pengelompokan. Hasil program ini menunjukkan bahwa sistem yang dibangun dengan berbasis web dapat mempermudah UD. Maju Jaya Tani dalam melakukan pengelompokan Bibit Padi Unggul, dengan sistem ini UD. Maju Jaya Tani dapat mengambil tindakan menggunakan metode K-means.

Pada data mining pengelompokan bibit padi unggul menggunakan metode k-means *clustering* memiliki kelemahan dan kelebihan dalam proses kerja sistem tersebut. Adapun kelemahan dan kelebihan sistem ini sebagai berikut:

1. Kelebihan Sistem
   Kelebihan dari sistem yang dibangun adalah :
   a. Sistem ini dapat mengelola banyak data varietas padi yang akan digunakan untuk pengelompokan bibit padi unggul.
   b. Sistem yang dibangun ini dapat membantu UD. Maju Jaya Tani dalam pengelompokan bibit padi unggul secara cepat.
   c. Sistem dibuat dengan tampilan sederhana dan mudah digunakan dimana hanya menginput data – data yang dibutuhkan kemudian sistem akan mengolahnya sehingga menghasilkan hasil bibit padi unggul.
2. Kelemahan Sistem
   Kelemahan dari sistem yang dibangun adalah :
   a. Aplikasi ini belum dilengkapi keamanan data yang baik karena belum terdapat algoritma pengaman data
   b. Sistem ini hanya bisa untuk memprediksi bibit padi unggul saja. Belum terdapat fitur-fitur yang dapat membantu UD. Maju Jaya Tani dalam menentukan hal lainnya.

**KESIMPULAN**

Berdasarkan penelitian dan pembahasan pada bab-bab sebelumnya maka dapat di ambil Kesimpulan sebagai berikut:

1. Dalam menerapkan metode K-Means *Clustering* pada pengelompokan bibit padi unggul dengan menggunakan 4 kriteria dengan 63 Varietas Padi, dan 2 cluster yaitu bibit padi unggul dan bukan bibit padi unggul. Setelah ditentukan alternatif dan kriteria maka diterapkan perhitungan metode K-Means *Clustering* dengan hasil akhir bahwa total persentase bibit padi unggul adalah 42,9% dan persentase bukan bibit padi unggul adalah 57,1%, dari pengujian tersebut maka disimpulkan bahwa metode K-Means *Clustering* dapat digunakan dalam pengelompokan bibit padi unggul.

2. Dalam merancang dan membangun sistem pengelompokan bibit padi unggul menggunakan metode K-Means Clustering adalah dengan merancang diagram menggunakan UML diagram yaitu untuk membangun *usecase diagram, activity diagram,* dan *class diagram*. Dan selanjutnya dibangun dengan Bahasa pemrograman menggunakan aplikasi *sublime text* untuk membuat *PHP* dan kemudian membangun database menggunakan aplikasi *MYSQL* sehingga aplikasi pengelompokan bibit padi unggul dapat selesai dan bisa diimplementasikan.

**REFERENCES**

Fadillah, N., Nofriansyah, D., & Taufik, F. (2022). Implementasi Metode K-Means Untuk Mengelompokkan Sarung Tangan. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, *1*(6), 845. https://doi.org/10.53513/jursi.v1i6.6444

Fitra, R. K. A. (2020). PENERAPAN K-MEANS CLUSTERING UNTUK MENGETAHUI VARIETAS PADI UNGGUL PRODUKSI BALAI PENGKAJIAN TEKNOLOGI PERTANIAN JAWA TIMUR : *Jurnal Mahasiswa Teknik Informatika*, *1*(1), 508–515. BPTP Jawa Timur.

Jelita, T., Buaton, R., Simajuntak, M., & Kaputama, S. (2023). Pengelompokan Bidang Usaha Terhadap Bantuan Produktif Usaha Mikro (BPUM) Berdasarkan Wilayah Deli Serdang Menggunakan Metode Clustering K-Means (Studi Kasus: Dinas Koperasi Dan UMKM Kabupaten Deli Serdang). *Journal of Computer Science and Information Technology E-ISSN*, *3*(2), 50.

Kadarsih, K., & Andrianto, S. (2022). Membangun Website SMA PGRI Gunung Raya Ranau Menggunakan PHP dan MYSQL. *JTIM: Jurnal Teknik Informatika Mahakarya*, *03*(2), 37–44.

Manihuruk, N. A., Zarlis, M., Irawan, E., & Tambunan, H. S. (2020). Penerapan Data Mining Dalam Mengelompokkan Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, *4*(1), 29–34. https://doi.org/10.30865/komik.v4i1.2575

Nanda, A. P., Sucipto, & Anggraeni, E. Y. (2022). Sistem Pendukung Keputusanuntuk Menentukan Bibit Padi Terbaikmenggunakan MetodeSimple Additive Weighting (Saw). *Jurnal Cendikia*, *22*(1), 1–6.

Nopita. L, Zaid Romegar. M & Ali Subhan. A (2021). Company Profile Pada Kantor Desa Lumpatan 1 Berbasis Web. *Jurnal Nasional Ilmu Komputer*; e-ISSN: 2746-1343.

Pitrawati, & Sanjaya, A. (2021). Rekayasa Perangkat Lunak Perhitungan Harga Pokok Produksi Metode Full Costing Pada Umkm Mitra Cake Di Bandar Lampung. *Jurnal Informasi dan Komputer*, *9*(2), 154–162. https://doi.org/10.35959/jik.v9i2.243

Prasetya, A. F., Sintia, S., & Putri, U. L. D. (2022). Perancangan Aplikasi Rental Mobil Menggunakan Diagram UML (Unified Modelling Language). *Jurnal Ilmiah Komputer ….*

Riadi, R., & Mesran. (2023). Penerapan Data Mining Menggunakan Algoritma K- Means Untuk Analisa Penjualan Parfume. *Journal of Informatics, Electrical and Electronics Engineering*, *2*(4), 138–145. https://doi.org/10.47065/jieee.v2i4.1181

Sidharta, K., & Wibowo, T. (2020). Studi Efisiensi Sumber Daya Terhadap Efektivitas Penggunaan Database : Studi Kasus SQL Server Dan MySQL. *Conference on Business, Social Sciences and Innovation Technology*, *1*(1), 508–515. http://journal.uib.ac.id/index.php/cbssit

Triyandana, G., Putri, L. A., & Umaidah, Y. (2022). Penerapan Data Mining Pengelompokan Menu Makanan

dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode K-Means. *Journal of Applied Informatics and Computing*, *6*(1), 40–46. https://doi.org/10.30871/jaic.v6i1.3824

Yendrianof, D., Romindo, Sari, A. N., Tantriawan, H., Putri, E. E., Manuhutu, M. A., Turaina, R., Defiariany, Putri, N. E., Priyantono, T., Jamaludin, Simarmata, J., Rismayani, & Aisa, S. (2022). *Analisis dan Perancangan Sistem Informasi* (A. Karim (ed.); Cetakan 1). Yayasan Kita Menulis.