Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 82-92
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4523

Penerapan Metode Clustering Dengan K-Means Untuk Memetakan Potensi Tanaman Padi di Sumatera

Irma Sanela[*], Alwis Nazir, Fadhilah Syafria, Elin Haerani, Lola Oktavia

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim, Pekanbaru, Indonesia
Email: [1,*]11950125097@students.uin-suska.ac.id, [2]alwis.nazir@uin-suska.ac.id, [3]fadhilah.syafria@uin-suska.ac.id, [4]elin.haerani@uin-suska.ac.id, [5]lola.oktavia@uin-suska.ac.id
Email Penulis Korespondensi: 11950125097@students.uin-suska.ac.id
Submitted: 02/11/2023; Accepted: 30/11/2023; Published: 30/11/2023

Abstrak−Tanaman padi merupakan sumber utama beras, makanan pokok bagi mayoritas penduduk Indonesia. Meskipun ada alternatif makanan lain, beras tetap tak tergantikan bagi yang terbiasa mengonsumsi nasi. Menurut data FAO (*Food and Agriculture Organization of the United Nations*) tahun 2018, Indonesia merupakan produsen beras terbesar ketiga di dunia dengan total produksi mencapai 59,2 juta ton. Namun, perencanaan tata ruang perkotaan dan pertanian masih belum sepenuhnya terintegrasi, dan ini menghasilkan keputusan yang sering kali bertentangan dalam perencanaan penggunaan lahan pertanian dan perkotaan. Untuk memenuhi kebutuhan beras di Sumatera, diperlukan upaya untuk meningkatkan produksi padi di setiap provinsi. Oleh karena itu, penelitian ini bertujuan untuk memetakan potensi tanaman padi di Sumatera berdasarkan produksi dan hasil panen dari tahun 1993 hingga 2020. Metode yang digunakan dalam penelitian ini adalah K-Means, yang memungkinkan pengelompokkan daerah-daerah potensi padi menjadi tiga kategori, yaitu tinggi, sedang, dan rendah. Hasil penelitian ini menghasilkan tiga kelompok yang dievaluasi menggunakan *Davies Bouldin Index* (DBI) dengan nilai 0.3943. Hasil pengelompokan ini mengindikasikan bahwa Cluster 0 berisi 92 daerah dengan tingkat keberhasilan tinggi, Cluster 2 memiliki 84 daerah dengan tingkat keberhasilan sedang, dan Cluster 1 terdiri dari 48 daerah dengan tingkat keberhasilan rendah. Kategori tingkat keberhasilan rendah ditemukan dalam cluster 1 dengan 84 daerah. Cluster 0 mencakup Aceh, Sumatera Utara, Sumatera Barat, Sumatera Selatan, dan Lampung dengan periode tahun tertentu. Cluster 1 mencakup daerah-daerah lain dengan karakteristik yang berbeda. Cluster 2 mencakup Provinsi Riau, Jambi, dan Bengkulu.

Kata Kunci: Clustering; Data Mining ; K-Means; Python; Padi

Abstract−Rice plants are the primary source of rice, the staple food for the majority of the Indonesian population. Despite the presence of other food alternatives, rice remains irreplaceable for those accustomed to consuming rice. According to data from the Food and Agriculture Organization of the United Nations (FAO) in 2018, Indonesia is the third-largest rice producer in the world, with a total production of 59.2 million tons. However, urban and agricultural spatial planning is not yet fully integrated, resulting in often conflicting decisions in land use planning for agriculture and urban development. To meet the rice demand in Sumatra, efforts are needed to increase rice production in each province. Therefore, this research aims to map the potential for rice cultivation in Sumatra based on production and harvest results from 1993 to 2020. The method used in this study is K-Means, which allows the grouping of rice potential areas into three categories: high, medium, and low. The research results produced three clusters, evaluated using the Davies Bouldin Index (DBI) with a value of 0.3943. The clustering results indicate that Cluster 0 contains 92 areas with a high success rate, Cluster 2 comprises 84 areas with a medium success rate, and Cluster 1 consists of 48 areas with a low success rate. The category of low success rate is found in Cluster 1 with 48 areas. Cluster 0 includes Aceh, North Sumatra, West Sumatra, South Sumatra, and Lampung within certain time periods. Cluster 1 encompasses other areas with different characteristics. Cluster 2 includes the provinces of Riau, Jambi, and Bengkulu.

Keywords: Clustering; Data Mining; K-Means; Python; Rice Crops

1. PENDAHULUAN

Padi adalah sumber bahan pangan yang menghasilkan beras. Bahan makanan ini menjadi makanan utama bagi banyak orang di Indonesia. Meskipun ada alternatif makanan lain yang bisa menggantikan padi, namun nilai padi sangat berharga bagi mereka yang terbiasa mengonsumsi nasi dan sulit digantikan oleh jenis makanan yang berbeda. Padi mengandung nutrisi dan gizi yang esensial bagi kesehatan manusia[1] Beras yang dihasilkan dari pengolahan padi adalah makanan pokok yang menjadi kebutuhan yang sangat penting bagi mayoritas penduduk Indonesia. Berdasarkan data yang dikumpulkan Pada tahun 2018, Menurut (FAO), Indonesia berada di peringkat ketiga dalam produksi beras dunia, jumlah total sekitar 59,2 juta ton. Terdapat peningkatan dalam produksi hasil panen dibandingkan pada tahun 2014, produksi beras yaitu 59 juta ton[2]. Sebagian besar kebijakan terkait sektor pertanian di pulau Sumatera didasarkan pada kebijakan nasional Indonesia. Pada saat ini, terdapat ketidaksepahaman antara kebijakan tata ruang pembangunan dan pembangunan pertanian, sehingga mengakibatkan keputusan yang saling bertentangan baik di tingkat nasional maupun lokal dalam perencanaan tata ruang untuk infrastruktur perkotaan dan lahan pertanian[3].

Hasil produksi padi pada beberapa provinsi di Sumatera pada tahun 2018 yaitu di Sumatera Utara dengan hasil produksi 2.108.284,72 Ton, Aceh dengan hasil produksi 1.861.567,10 Ton, Sumatera Barat dengan hasil produksi 1.483.076,48 Ton, Riau dengan hasil produksi 266.375,53 Ton, Jambi dengan hasil produksi 383.045,74 Ton, Sumatera Selatan dengan hasil produksi 2.994.191,84 Ton, Bengkulu dengan hasil produksi 288.810,52 Ton, dan Lampung dengan hasil produksi 2.488.641,91 Ton [4].

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 82-92
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4523

Untuk mengatur kebutuhan beras di Sumatera, perlu dilakukan upaya untuk meningkatkan produksi padi di setiap provinsinya. Data hasil panen atau produksi di setiap provinsi di Sumatera dapat digunakan sebagai acuan untuk memetakan potensi pertanian padi pada daerah tersebut. Penelitian ini bertujuan untuk untuk memetakan daerah potensi padi di Sumatera dari tahun 1993-2020 dengan mengelompokkan potensi tanaman padi dengan menerapkan metode Clustering dengan K-Means.

Hasil dari pengelompokkan tersebut selanjutnya akan diproses dengan metode data mining. Data Mining merupakan salah satu upaya dalam mencari dan menemukan koneksi, pola, dan arah baru yang signifikan dengan menganalisis secara sistematis data dalam jumlah signifikan tersimpan di dalam penyimpanan. Proses ini melibatkan penggunaan teknologi pengenalan pola, dan juga penggunaan pendekatan statistik dan matematika[5]. Metode ini hanya dapat diterapkan pada atribut yang bersifat numerik karena pendekatannya melibatkan pembagian data menjadi beberapa kelompok berdasarkan jarak terdekat[6].

Dalam proses data mining, informasi berharga dapat dihasilkan dari jumlah data yang besar, ini sering dikenal dengan istilah KDD (Penemuan Pengetahuan dalam Basis Data). Penggunaan istilah KDD seringkali terkait dengan konsep menggali informasi tersembunyi dalam basis data yang luas [7]. Salah satu teknik dari data mining yaitu *Clustering*. *Clustering* adalah proses pengelompokan berdasarkan kesamaan kelas yang prinsipnya adalah untuk mengurangi perbedaan antara kelas-kelas yang ada [8]Algoritma dari *clustering* yang digunakan dalam penelitian ini yaitu K-Means. K-Means merupakan metode data clustering non-hirarki yang melakukan pengelompokkan data dalam satu cluster/kelompok atau lebih [9],[10], [11].

Alasan peniliti menggunakan K-Means dalam peniltian ini karena mudah untuk dipahami dan di implementasikan dan sangat umum digunakan oleh peneliti terdahulu sebagai teknik Clustering. Contohnya pada beberapa penelitian yang telah dilakukan sebelumnya. Prasetyaningrum Eka dkk (2023) tentang Perbandingan Algoritma K-Means Dan K-Medoids Untuk Pemetaan Hasil Produksi Buah-Buahan dimana hasil penelitian nya menyatakan bahawa berdasarkan pengujian menggunakan aplikasi Rapidminer versi 9.9 dengan membandingkan Davies Bouldin Index (DBI) dari dua algoritma yang berbeda pada 29 data, kesimpulan yang diambil adalah pada percobaan keempat dengan enam klaster, nilai DBI adalah 0,296. Nilai DBI yang lebih rendah atau mendekati nol menunjukkan hasil klaster yang lebih optimal. Hasil pengujian menunjukkan bahwa algoritma K-Means memiliki nilai DBI yang lebih rendah (0,296) dibandingkan dengan algoritma K-Medoids (0,507). Oleh karena itu, K-Means dianggap sebagai algoritma terbaik untuk mengelompokkan hasil produksi buah-buahan di Kabupaten Kotawaringin Timur [12]. Herna Mahulae tentang Pengelompokan Potensi Produksi Buah-Buahan di Provinsi Sumatera Utara dengan Menerapkan K-*Clustering* dengan hasil dari penelitian tersebut Ini adalah data potensi hasil produksi buah Alpukat, Duku, dan Belimbing yang dikumpulkan dari 30 kabupaten di Sumatera Utara pada tahun 2017[13] Dimas Alif Fajar Fadhillah dkk (2022) menyatakan dalam penelitiannya mengenai Penerapan metode K- Means *Clustering* Pada Pemetaan Lahan Kopi Di Kabupaten Malang bahwa hasil dari penelitian tersebut Dalam hasil penelitian ini, data produksi tanaman kopi di 33 kecamatan setiap tahunnya, dengan total 99 data, menghasilkan pengelompokan dengan jumlah data dalam kategori C1 (rendah) adalah 26 data pada tahun 2018, 24 data pada tahun 2019, dan 24 data pada tahun 2020. Sedangkan pada C2 (kategori sedang) terdapat 3 data pada tahun 2018, 5 data pada tahun 2019, dan 4 data pada tahun 2020. Untuk C3 (kategori tinggi), terdapat 4 data pada tahun 2018, 4 data pada tahun 2019, dan 5 data pada tahun 2020 [14].

Sena Wijayanto dkk (2021) menyatakan dalam penelitiannya mengenai Pengelompokkan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means bahwa Dalam penelitian ini, proses perhitungan menggunakan K-Means berakhir pada iterasi kelima. Hasil dari pengelompokan menunjukkan bahwa C0 mewakili kota/kabupaten dengan produktivitas padi sedang, C1 merupakan representasi kota/kabupaten dengan produktivitas padi yang rendah, sedangkan C2 merupakan representasi kota/kabupaten dengan produktivitas padi yang tinggi. Terdapat 12 daerah yang masuk dalam kelompok C0, 18 daerah dalam C1, dan 5 daerah dalam C2 [15]. Selanjutnya penelitian Lidya dkk (2022) tentang *Clustering* Hasil Panen Berdasarkan Lokasi dan Jenis Bibit. Penelitian ini menggunakan metode clustering k-means dengan Grup 1 dan memiliki centroid sebesar 1.87, 2.12, 2.75, yang terdiri dari 8 data. Grup 1 ini merujuk pada Hasil panen jagung di Binjai Kota mencapai jumlah panen sekitar 501 hingga 1500 ton [16].

Ieannoal Vhallaha dkk (2018) mengenai Pengelompokan Mahasiswa Potensial Drop Out menggunakan Metode *Clustering* K-Means bahwa dalam penelitiannya didapatkam hasil yaitu Dengan melakukan pengelompokan mahasiswa ke dalam beberapa kluster, ditemukan bahwa hasil klastering menunjukkan bahwa angkatan 2014 tergabung dalam kluster 0 dengan jumlah 4 mahasiswa atau 30,77% dari total 13 sampel. Angkatan 2015 tergabung dalam kluster 1 yaitu 4 mahasiswa dan kluster 2 dengan jumlah 2 mahasiswa, yang merupakan 66,7% dari total 9 sampel. Angkatan 2016 tergabung dalam kluster 0 dengan jumlah 2 mahasiswa, dan kluster 1 dengan jumlah 10 mahasiswa atau 50% dari total 24 sampel. Selain itu, angkatan 2017 tergabung pada kluster 2 dengan jumlah 4 mahasiswa atau 22,22% dari total 18 sampel, di mana juga terdapat 4 mahasiswa yang berpotensi dikeluarkan. Dengan metode clustering K-Means dan perangkat lunak dengan atribut SKS Total, IPK, dan Semester, kami dapat mengidentifikasi lebih awal mahasiswa yang berpotensi dikeluarkan [17].

Namun, dalam penelitian ini dilakukan pengembangan sistem clustering menggunakan metode K-Means dengan 224 data produksi tanaman padi dari tahun 1993 hingga 2020. Tujuan utama adalah memanfaatkan metode K-Means untuk mengelompokkan data dengan akurat dan memetakan potensi tanaman padi di Sumatera. Penelitian ini menghasilkan data wilayah dengan produksi dan luas panen yang dikelompokkan menjadi tiga

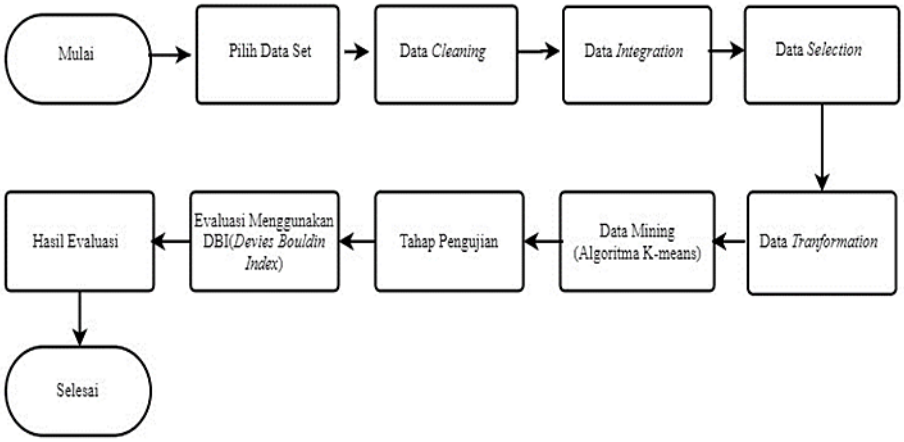
Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 82-92
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4523

cluster yaitu tinggi, sedang, dan rendah. Penelitian bertujuan untuk mengidentifikasi daerah-daerah di Sumatera yang memiliki potensi padi tinggi, sehingga dapat memberikan wawasan kepada daerah dengan potensi sedang dan rendah untuk pengembangan yang lebih intensif.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian yang akan digunakan untuk memetakan potensi padi di Sumatera menggunakan *K-Means* yang dapat dilihat pada gambar 1 sebagai berikut**.**



Gambar 1. Tahapan Penelitian

2.1.1 Pilih Dataset

Penelitian ini dimulai dengan memilih dataset yaitu dataset Tanaman Padi Sumatera yang bersumber dari Kaggle. Data tersebut berisi 8 provinsi tanaman padi tahun 1993-2020 dengan jumlah 224 data dan memiliki 7 atribut yaitu Provinsi, Tahun, Produksi, Luas Panen, Curah Hujan, Kelembapan, dan Suhu rata-rata.

2.1.2 Data Cleaning

Setelah memilih dataset selanjutnya dilakukan proses *cleaning*. Proses *cleaning* ini dilakukan untuk pembersihan data, mengatasi duplikasi, memastikan konsistensi data, dan proses KDD, dilakukan pula perbaikan kesalahan yang mungkin terjadi, termasuk kesalahan cetak atau tipografi. Agar dataset bisa digunakan dalam penelitian. Data yang di cleaning pada penelitian yaitu kesalahan pada penulisan angka di beberapa provinsi pada atribut produksi setelah dilakukan pengecekan pada Badan Pusat Statistik. Berikut pada tabel 1 merupakan data cleaning yang digunakan.

Tabel 1. Data Cleaning

Nama Provinsi	Tahun	Produksi
Aceh	2018	1861567
Aceh	2019	1714438
Aceh	2020	1757313
Sumatera Utara	2020	2040500
Sumatera Barat	2020	1387269
Riau	2020	243685
Jambi	2020	386413
Sumatera Selatan	2020	2743060
Bengkulu	2020	292834
Lampung	2020	265029

Penjelasan mengenai Tabel 1 adalah sebagai berikut.

a. Pada proses cleaning data beberapa provinsi di tahun 2018-2020 terdapat kesalahan dalam penulisan angka dimana pada saat dilakukan pengecekan pada website Badan Pusat Statistik provinsi Aceh tahun 2018 sebelum di cleaning nilai produksi nya yaitu 1751996.94, pada tahun 2019 sebelum di cleaning nilai produksi nya yaitu 1714437.6, dan pada tahun 2020 sebelum di cleaning nilai produksi nya yaitu 317 869,41.
b. Pada provinsi Sumatera Utara 2020 terdapat kesalahan dalam penulisan angka dimana pada saat dilakukan pengecekan pada website Badan Pusat Statistik sebelum di cleaning nilai produksi nya yaitu 2076280.01

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 82-92
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4523

c. Pada provinsi Sumatera Barat tahun 2020 terdapat kesalahan dalam penulisan angka dimana pada saat dilakukan pengecekan pada website Badan Pusat Statistik sebelum di cleaning nilai produksi nya yaitu 1450839.74
d. Pada provinsi Riau tahun 2020 terdapat kesalahan dalam penulisan angka dimana pada saat dilakukan pengecekan pada website Badan Pusat Statistik sebelum di cleaning nilai produksi nya yaitu 269344.05
e. Pada provinsi Jambi tahun 2020 2020 terdapat kesalahan dalam penulisan angka dimana pada saat dilakukan pengecekan pada website Badan Pusat Statistik sebelum di cleaning nilai produksi nya yaitu 374376.27
f. Pada provinsi Sumatera Selatan tahun 2020 terdapat kesalahan dalam penulisan angka dimana pada saat dilakukan pengecekan pada website Badan Pusat Statistik sebelum di cleaning nilai produksi nya yaitu 2696877.46
g. Pada provinsi Bengkulu tahun 2020 terdapat kesalahan dalam penulisan angka dimana pada saat dilakukan pengecekan pada website Badan Pusat Statistik sebelum di cleaning nilai produksi nya yaitu 296925.16
h. Pada provinsi Lampung tahun 2020 terdapat kesalahan dalam penulisan angka dimana pada saat dilakukan pengecekan pada website Badan Pusat Statistik sebelum di cleaning nilai produksi nya yaitu 2604913.29

2.1.3 Data *Integration*

Integrasi data adalah langkah untuk menggabungkan informasi dari berbagai sumber yang berbeda, termasuk basis data, file teks, spreadsheet, atau sumber data lainnya, menjadi satu dataset tunggal yang digunakan dalam proses KDD. Proses ini melibatkan pemilihan data yang relevan dari berbagai sumber, pemetaan atribut, resolusi konflik skema, dan penggabungan data tersebut menjadi satu struktur data yang terpadu. Namun pada penelitian ini tidak memakai proses data *integration* karena hanya mengambil dari 1 sumber data yaitu dataset yang bersumber dari Kaggle tidak mengambil dari sumber data yang lain.

2.1.4 Data *Transformation*

Pada proses ini data *transformation* menggunakan data untuk mengubah atau memodifikasi informasi dalam format yang lebih cocok untuk analisis atau pembuatan model. Data *transformation* berguna untuk meningkatkan kualitas data, membuat data lebih cocok untuk analisis atau pemodelan, mengurangi kompleksitas data, dan meningkatkan interpretabilitas. Tanpa data *transformation*, data yang mentah atau tidak terstruktur bisa sulit untuk diolah dan digunakan dalam keputusan bisnis atau penelitian. Namun pada penelitian ini tidak memakai proses data *transformation*.

2.2 K-Means *Clustering*

Melakukan pendekatan pemodelan dengan menggunakan K-Means. Teknik non-hierarkis yang dikenal sebagai k-means bertujuan untuk membagi data menjadi satu atau lebih kelompok sedemikian rupa sehingga Data dengan ciri-ciri serupa akan digabungkan dalam satu kelompok, sedangkan data yang memiliki karakteristik yang berbeda akan dipisahkan ke dalam kelompok yang berbeda [17]. Agar mencapai hasil yang diinginkan. Data yang tersedia akan menghitung rata-ratanya sebagai langkah awal. Adapun tahapan K-Means sebagai berikut [18]

a. Pilih jumlah cluster yang ingin dibentuk. Jumlah kluster k ditentukan dengan mempertimbangkan faktor-faktor teori dan konsep yang direkomendasikan untuk menentukan berapa banyaknya kelompok yang harus dibentuk.
b. Secara acak, pilih k titik awal centroid pertama. Dalam menentukan centroid awal, objek-objek acak yang tersedia diambil sebanyak k sesuai dengan jumlah kelompok yang telah ditentukan.

$$v = \frac{\sum_{k}^{n}=1\ (Xi)}{N} \text{ , i = 1,2,3, n} \quad (1)$$

Penjelasan rumus diatas adalah sebagai berikut:
v = nilai centroid pada kelompok.
xi = objek ke-I, di mana nilai objek ini berubah sesuai dengan proses iterasi.
n = banyaknya objek/data.

c. Hitunglah jarak antar setiap objek dengan setiap centroid pada kluster. Dengan menggunakan rumus *Euclidean Distance* sebagagai berikut.

$$d\ (x,y) = |x - y| = \sqrt{\sum_{i=1}^{n}(xi - yi)^2} \quad (2)$$

Penjelasan rumus di atas dapat diuraikan sebagai berikut:
d = nilai jarak data ke pusat *cluster*
xi = objek x ke-i
yi = data y ke-i
n = banyaknya objek/data

d. Tempatkan setiap objek dalam centroid yang paling mendekatinya (yang memiliki jarak terkecil). Ini akan membentuk sebuah cluster baru.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 82-92
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4523

e. Mengulangi proses secara berulang dan menentukan lokasi yang baru untuk centroid menggunakan persamaan yang telah ditetapkan.
f. Kembali ke langkah kelima jika posisi centroid yang baru berbeda, sampai menemukan hasil dengan nilai centroid yang sama/data tidak berpindah.

2.3 Tahap Pengujian

Tahap pengujian ini menggunakan Python yang dijalankan pada Google Colab. Hasil dari perhitungan manual akan di uji menggunakan python dengan membandingkan hasil yang didapat.

1. Siapkan lingkungan Python, seperti Google Colab, dan pastikan data yang akan diuji tersedia.
2. Impor data yang akan diuji ke dalam Python, mungkin dari file CSV atau database.
3. Lakukan perhitungan secara manual sesuai rencana pengujian.
4. Lakukan perhitungan yang sama menggunakan Python, dengan memanfaatkan fungsi atau perpustakaan yang sesuai.
5. Bandingkan hasil perhitungan manual dengan hasil perhitungan Python untuk memastikan kesesuaian.
6. Jika ada perbedaan, Analisis hasil untuk memahami sumber masalah yang mungkin terjadi, seperti kesalahan dalam perhitungan.

2.4 Evaluasi

Evaliasi menggunakan model *Davies Bouldin Index* (DBI) selama fase evaluasi cluster ini yang bertujuan untuk membantu dalam pemilihan jumlah cluster yang optimal. Karena DBI memberikan skor yang lebih rendah untuk pengelompokan yang lebih baik, pada tahap ini akan mencoba beberapa nilai yang berbeda dan memilih nilai DBI terendah agar menghasilkan kualitas yang terbaiki. DBI adalah suatu teknik yang diterapkan untuk menilai mutu suatu klaster. Metode ini menilai sejauh mana suatu klaster data bagus dengan menghitung rasio antara jarak antara klaster dengan klaster lainnya [19].

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Proses penelitian ini menggunakan dataset yang bersumber dari Kaggle sebagai informasi dalam penelitian ini. Data tersebut melibatkan beberapa atribut, yaitu: Provinsi, Tahun, Produksi, Luas Panen, Curah Hujan, Kelembapan, dan Suhu Rata-rata. Jumlah data yang diperoleh yaitu 224 data. Pada tabel 2 terdapat data yang digunakan, diantaranya Data Tanaman Padi Sumatera dari tahun 1993-2020 dimana terdapat 8 provinsi yaitu: Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, dan Lampung.

Tabel 2. Dataset Penelitian

Provinsi	Produksi	Luas Panen	Cutrah Hujan	Kelembapan	Suhu Rata-rata
Aceh	1329536	323589	1627	82	26
Aceh	1299699	329041	1521	82	27
Aceh	1382905	339253	1476	83	26
Aceh	1419128	348223	1557	83	26
Aceh	1368074	337561	1339	82	26
Aceh	1404580	365892	1465	83	27
Aceh	1478712	359817	1778	83	26
Aceh	1486909	336765	1975	91	27
Aceh	1547499	295212	1689	69	29
Aceh	1314165	315131	1297	69	29
Aceh	1246614	367636	1507	71	29
---	---	---	---	---	---
---	---	---	---	---	---
Lampung	2164089	464103	1706	78	27
Lampung	2650290	545149	2211	76	25

3.2 Penerapan K-Means

Berikut adalah langkah-langkah dari K-Means.

a. Tentukan berapa banyak kluster yang ingin dibuat. Jumlah kelompok k ditentukan dengan mempertimbangkan pertimbangan teori dan konsep yang direkomendasikan. Untuk menentukan berapa banyak kelompok yang harus dibentuk.
b. Secara acak, pilih k titik awal centroid pertama. Dalam menentukan centroid awal, objek-objek acak yang tersedia diambil sebanyak k sesuai dengan jumlah kelompok yang telah ditentukan, seperti terlihat pada tabel 3 berikut.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 82-92
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4523

Tabel 3. Centroid awal (Iterasi-1)

Centroid	Produksi	Hasil Panen	Curah Hujan	Kelembapan	Suhu Rata-rata
C0	1329536	323589	1627	82	26
C1	2918152	754569	2549	84	26
C2	436297	146133	2738	86	26

c. Melakukan perhitungan jarak terpendek atau perbedaan antara setiap kelompok dengan setiap datauntuk mengukur jarak terdekat antara data dan centroid awal masing-masing data. Berikut adalah langkah-langkah perhitungan untuk menentukan jarak antara setiap data dan kelompok pertama (C0).

$$d(1,0) = \sqrt{\begin{matrix}(1329536-1329536)^2+(323589-323589)^2+\\(1627-1627)^2+(82-82)^2\,(26-26)^2\end{matrix}} = 0$$

1. Perhitungan pada data ke 1 cluster kedua (C1)

$$d(1,1) = \sqrt{\begin{matrix}(1329536-2918152)^2+(323589-754569)^2+\\(1627-2549)^2+(82-84)^2(26-26)^2\end{matrix}} = 1646039$$

2. Perhitungan pada data ke 1 cluster ketiga (C2)

$$d(1,2) = \sqrt{\begin{matrix}(1329536-436297)^2+(323589-146133)^2+\\(1627-2738)^2+(82-86)^2(26-26)^2\end{matrix}} = 91069$$

d. Setelah menghitung jarak antara data dan centroid, langkah selanjutnya adalah mengelompokkan data berdasarkan nilai terkecil (jarak terpendek). Berikut tabel 4 hasil dari pengelompokkannya.

Tabel 4. Hasil Pengelompokkan

Provinsi	C0	C1	C2	Jarak Terdekat	Cluster
Aceh	0	1646039	910696	0	0
Aceh	30331	1673459	882564	30331	0
Aceh	55620	1590431	966107	55620	0
Aceh	92917	1553123	1003394	92917	0
Aceh	40994	1605191	951239	40994	0
Aceh	86146	1562681	992909	86146	0
Aceh	153512	1492587	1064092	153512	0
Aceh	157924	1490979	1067767	157924	0
Aceh	219802	1445579	1121158	219802	0
Aceh	17547	1663094	893988	17547	0
Aceh	93895	1715738	840047	93895	0
---	---	---	---	---	---
---	---	---	---	---	---
Lampung	846300	808073	1756807	808073	1
Lampung	1339208	340010	2249662	340010	1

Tabel 5. Output Cluster (Iterasi Kedua)

Centroid	Hasil
C0	74
C1	66
C2	84

Berdasarkan perhitungan menggunakan rumus Euclidean Distance pada data pertama, dapat disimpulkan bahwa data tersebut memiliki jarak terdekat dengan cluster 0. Dengan demikian, data awal termasuk dalam kelompok cluster 0. Proses perhitungan ini kemudian berlanjut hingga data ke-224, di mana setiap data akan menemukan cluster yang memiliki jarak terdekat.

e. Untuk iterasi selanjutnya, nilai centroid diperbarui dengan menghitung rata-rata dari semua atribut data yang termasuk dalam kelompok tersebut. Berikut perhitungan nilai centroid yang diperbarui.

1. Rata-rata pada cluster Pertama (C0)

$C0$ (Produksi) $= \frac{128413867}{74} = 1735323$ $C0$ (Luas Panen) $= \frac{130591906}{74} = 413404$

$C0$ (Curah Hujan) $= \frac{172263}{74} = 2328$ $C0$ (Kelembapan) $= \frac{5963}{74} = 81$

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 82-92
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4523

C0 (Suhu rata-rata) = $\frac{1977}{74}$ = 27

2. Rata-rata pada cluster kedua (C1)

C1 (Produksi) = $\frac{208488447}{66}$ = 1735323 *C1* (Luas Panen) = $\frac{42382496}{66}$ =413404

C1(Curah Hujan) = $\frac{161737}{66}$ = 2328 *C1* (Kelembapan) = $\frac{5320}{66}$ = 81

C1 (Suhu rata-rata) = $\frac{2256}{84}$ = 27

3. Rata-rata pada cluster ketiga (C2)

C1 (Produksi) = $\frac{39330496}{84}$ = 468220 *C1* (Luas Panen) = $\frac{10879990}{84}$ = 129524

C1(Curah Hujan) = $\frac{215357}{84}$ = 2564 *C1* (Kelembapan) = $\frac{6849}{84}$ = 82

C1 (Suhu rata-rata) = $\frac{2256}{84}$ = 27

Tabel 6. Centroid baru (Itierasi-2)

Centroid	Produksi	Hasil Panen	Curah Hujan	Kelembapan	Suhu Rata-rata
C0	1735323	413404	2328	81	27
C1	3158916	642159	2451	81	27
C2	39330496	129524	2564	82	27

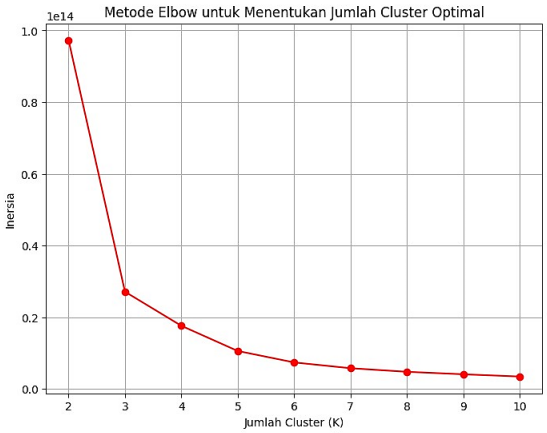
f. Lakukan perhitungan ulang jarak antara data dan centroid yang diperbaharui. Jika masih ada data yang berpindah, maka proses dari langkah c hingga e diulang sampai iterasi terakhir memiliki data yang serupa dengan iterasi sebelumnya. Hasil dari perhitungan dapat terlihat pada tabel 7.

Tabel 7. Output Cluster Iterasi-5 (iterasi berakhir)

Centroid	Hasil
C0	92
C1	48
C2	84

3.3 Pengujian Menggunakan *Python* Pada K-Means

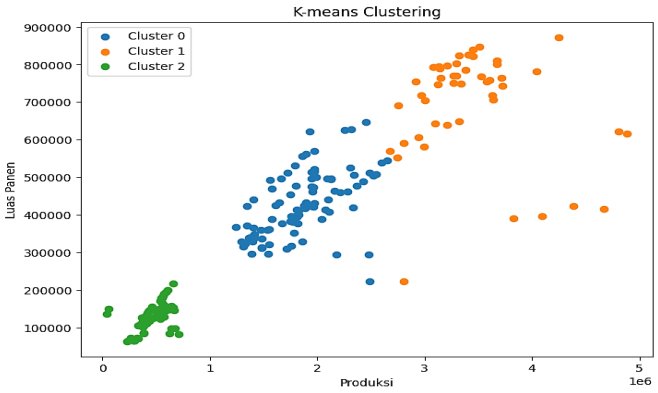
Python, sebuah bahasa pemrograman open-source, didesain dengan fokus pada beberapa aspek kunci, termasuk efisiensi perangkat lunak, kemudahan dalam pengembangan, integrasi yang mulus antara komponen-komponen, dan kemampuan untuk mengeksekusi program di berbagai platform. Python digunakan dalam berbagai konteks, seperti pengembangan sistem, pembuatan antarmuka pengguna, scripting web, perhitungan numerik, dan banyak aplikasi lainnya. Popularitasnya yang tidak dapat disangkal menjadi bukti keberhasilan bahasa ini [20]. *Python* digunakan dalam berbagai konteks, seperti pengembangan sistem, pembuatan antarmuka pengguna, scripting web, perhitungan numerik, dan banyak lagi. Sebagai bukti popularitasnya yang tak terbantahkan [21]. Pada tahapan ini penerapan Python menggunakan Google Colabs. Peneliti menggunakan metode elbow untuk membantu peneliti dalam menentukan nilai k yang optimum, Didapatkan nilai k yang optimal adalah 3.



Gambar 2. Hasil Elbow Method

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 82-92
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4523

Kemudian peneliti melakukan inisialisasi k pusat cluster (centroid) secara random. Selanjutnya Menghitung jarak data pada tiap-tiap centroid, kemudian mengklaster objek berdasarkan jarak ke centroid minimum (terdekat). Berikut merupakan hasil clustering menggunakan *Python* dengan nilai k = 3 yang dapat dilihat pada gambar 3



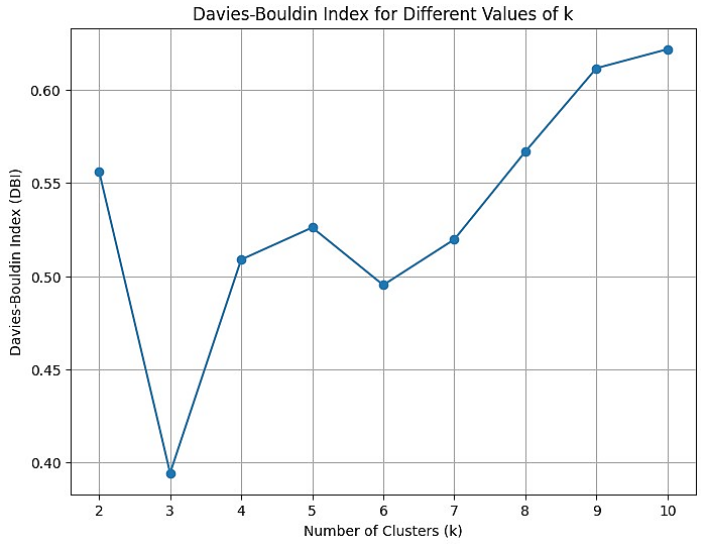

```
Cluster 0: 92 data
Cluster 1: 48 data
Cluster 2: 84 data
```

Gambar 3. Grafik Hasil Clustering k= 3

Gambar 3 menjelaskan bahwa potensi padi di sumatera pada C0 yaitu 92 daerah penghasil padi sedang dengan jumlah keseluruhan produksi 1865913.783, Luas Panen 428435.26, pada C1 48 daerah pada penghasil padi tertinggi dengan jumlah keseluruhan Produksi 3442463.479, Luas Panen 699132.438. pada C2 yaitu 84 daerah penghasil padi terendah dengan jumlah keseluruhan Produksi 468220.202, dan Luas Panen 129523.690. C0 dengan jumlah 92 yaitu pada provinsi Aceh dari tahun 1993-2020, Sumatera Utara dar tahun 2018-2020, Sumatera Barat dari tahun 1993-2016, dan juga tahun 2018-2020, Sumatera Selatan 1993-2006 dan 2019, Lampung tahun 1993-2008 dan 2018-2019, untuk C1 dengan jumlah 48 yaitu pada provinsi Sumatera Utara 1993-2017, Sumatera Barat tahun 2017, Sumatera Selatan tahun 2007-2018 dan 2020, Lampung tahun 2009-2017 dan 2020, untuk C2 dengan jumlah 84 yaitu pada provinsi Riau tahun 1993-2020, Jambi 1993-2020 dan Bengkulu dari tahun 1993-2020. Dari gambar 3 terlihat bahwa tren pada data cenderung naik, maka semakin besar luas panen maka hasil produksi pada tiap daerah akan semakin meningkat atau semakin banyak.

3.4 Evaluasi

Tahapan evaluasi digunakan untuk menentukan jumlah kluster yang optimal. Evaluasi menerapkan metode DBI digunakan untuk menilai data. Evaluasi dilakukan dalam rentang kluster dari 2 hingga 10 untuk mengamati nilai jarak antara kluster. Dan untuk mengetahui apakah menggunakan 3 cluster merupakan pilihan yang terbaik. Hasil evaluasi ini dapat dilihat pada gambar 4.



4

Gambar 4. Hasil Evaluasi DBI

Jumlah Kluster (k) = 2: Davies-Bouldin Index (DBI) = 0.5564
Jumlah Kluster (k) = 3: Davies-Bouldin Index (DBI) = 0.3943

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 82-92
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4523

Jumlah Kluster (k) = 4: Davies-Bouldin Index (DBI) = 0.5089
Jumlah Kluster (k) = 5: Davies-Bouldin Index (DBI) = 0.5261
Jumlah Kluster (k) = 6: Davies-Bouldin Index (DBI) = 0.4953
Jumlah Kluster (k) = 7: Davies-Bouldin Index (DBI) = 0.5198
Jumlah Kluster (k) = 8: Davies-Bouldin Index (DBI) = 0.5669
Jumlah Kluster (k) = 9: Davies-Bouldin Index (DBI) = 0.6117
Jumlah Kluster (k) = 10: Davies-Bouldin Index (DBI) = 0.6219
Kluster Terbaik (k) adalah: 3

Gambar 4 menjelaskan bahwa cluster terbaik yaitu berada pada cluster ke 3 dengan nilai DBI 0.3943 dimana dinyatakan cluster terbaik karena nilai DBI nya lebih kecil dan medekati 0

3.5 Pembahasan

Dalam pemetaan produksi tanaman padi di Sumatera, hasilnya dapat dianalisis menggunakan model K-Means yang telah dibangun. Kualitas kluster dibentuk melalui tahapan clustering K-Means dapat dinilai dengan mengevaluasi hasil kluster tersebut. Dalam penelitian ini, Davies Bouldin Index (DBI) digunakan sebagai alat evaluasi data. Evaluasi hasil pengelompokan dalam penelitian ini menghasilkan tiga kelompok dengan skor 0.3943. Untuk melihat perbandingan nilai DBI antar cluster pada saat tahap evaluasi dengan menggunakan nilai DBI dapat dilihat pada tabel 8 berikut.

Tabel 8. Perbandingan Nilai DBI

Jumlah Cluster	Nilai DBI
2	0.5564
3	0.3943
4	0.5089
5	0.5261
6	0.4953
7	0.5198
8	0.5669
9	0.6117
10	0.6219

Dalam Tabel 8, cluster terbaik dengan nilai DBI sekitar 0.3943 adalah cluster ketiga, yang mendekati nilai DBI terbaik. Cluster ini mencakup Aceh, Sumatera Utara, Sumatera Barat, Sumatera Selatan, dan Lampung untuk C0 dengan total produksi padi sekitar 1.865.913,783dan luas panen sekitar 428.435,26. Cluster 1 (C1) mencakup daerah-daerah lain dengan total produksi padi tertinggi sekitar 3.442.463,479 dan luas panen sekitar 699.132.438,0. Sementara Cluster 2 (C2) mencakup Riau, Jambi, dan Bengkulu dengan total produksi padi terendah sekitar 468.220,202 dan luas panen sekitar 129.523,690. Cluster ketiga dianggap yang terbaik karena nilai DBI mendekati 0.

Tabel 9. Hasil Keseluruhan Tiap Cluster

Nama	C0	C1	C2
Produksi	1.865.913.783	3.442.463.479	468.220.202
Luas Panen	428.435.261	699.132.438	129.523.690
Curah Hujan	2.406.141	2.346.708	2.563.762
Kelembapan	80.522	80.875	81.548
Suhu Rata-rata	26.641	27	26.940

Pada tabel 9 dapat dijelaskan bahwa semakin tinggi luas panen maka semakin besar hasil produsi, kemudian Curah Hujan, Kelembapan, dan Suhu Rata-rata sangat berpengaruh pada pertumbuhan padi yang baik, Namun jika dilihat dari faktor untuk pertumbuhan padi, curah hujan yang baik dalam pertumbuhan padi yaitu sekitar 50-200 mm, Kelembapan yang baik sekitar 50-75% dan untuk suhu rata-rata yaitu sekitar 25-30 ºC.

4. KESIMPULAN

Dalam penelitian ini telah berhasil mengelompokkan potensi tanaman padi di Sumatera dengan menerapkan algoritma k-means. Penelitian ini menggunakan 224 data dengan 6 parameter yaitu Provinsi, Produksi, Luas Panen, Curah Hujan, Kelembapan, dan Suhu rata-rata. Dimana hasil yang didapat dengan menggunakan perhitungan manual dan pengujian menggunakan python adalah sama yaitu terbentuk 3 cluster dalam kategori tingkat keberhasilan sedang terdapat dalam cluster 0 dengan 92 daerah, sementara kategori tingkat keberhasilan tinggi ada di cluster 2 dengan 48 daerah. Di sisi lain, kategori tingkat keberhasilan rendah ditemukan dalam cluster 1 dengan 84 daerah. Cluster 0 mencakup Aceh, Sumatera Utara, Sumatera Barat, Sumatera Selatan, dan

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 82-92
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4523

Lampung dengan periode tahun tertentu. Cluster 1 mencakup daerah-daerah lain dengan karakteristik yang berbeda. Cluster 2 mencakup Provinsi Riau, Jambi, dan Bengkulu. Hasil evaluasi yang dilakukan dengan menerapkan metode DBI (Davies Bouldin Index) dinilai baik dengan hasil yang terbentuk yaitu 3 kluster. K=3 merupakan kluster yang terbaik setelah eksperimen dilaksanakan evaluasi menggunakan k=2, sampai dengan k=10 dengan nilai 0.3943. Nilai DBI pada K-Means lebih kecil dibandngkan algoritma lainnya sehingga K-Means dapat dintayakan lebih baik. Pada grafik tren pada data cenderung naik, maka semakin besar luas panen maka hasil produksi pada tiap daerah akan semakin meningkat atau semakin banyak. Penelitian ini memiliki kelebihan, termasuk kemudahan implementasi, dukungan library, skalabilitas, preprocessing data, sumber daya online, dan integrasi dengan machine learning yang memudahkan peneliti dalam memetakan potensi tanaman padi dengan efisien. Disarankan untuk penelitian selanjutnya menambahkan jumlah data pada atribut provinsi agar hasil pemetaan tanaman padi di sumatera akan semakin akurat.

REFERENCES

[1] Edy, *Pengantar Teknologi Budidaya Tanaman Serelia : Jagung dan Padi*. PT Nas Media Pustaka, 2022.

[2] F. Marisa *et al.*, "Digitasi Produktivitas Panen Padi Berbasis K-Means Clustering," *SMARTICS Journal*, vol. 7, no. 1, pp. 21–26, 2021, doi: 10.21067/smartics.v7i1.5270.

[3] B. R. Aprildahani, C. T. H. Permana, and S. T. E. W. Hutama, "Kebutuhan Lahan Pertanian Minimum untuk Kesejahteraan Petani di Pulau Sumatera," *Journal of Science and Applicative Technology*, vol. 5, no. 1, pp. 116–125, Mar. 2021, doi: 10.35472/jsat.v5i1.409.

[4] Badan Pusat Statistik, "Luas Panen, Produksi, dan Produktivitas Padi Menurut Provinsi 2020," 12 Juli 2021.

[5] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Denga Algoritma K-Means," *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 2, pp. 100–108, Jun. 2021, [Online]. Available: http://jim.teknokrat.ac.id/index.php/JTSI

[6] T. Hartati and Y. Arie Wijaya, "Analisis Data Lalu Lintas Jaringan Di Kantor Cangehgar Cyber Operation Center Menggunakan Algritma K-Means Network Traffic Data Analysis At Cangehgar Cyber Operation Center Office Using K-Means Algorithm," *Jurnal Ilmiah NERO*, vol. 7, no. 1, pp. 75–84, 2022.

[7] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means Pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Teknika*, vol. 21, no. 1, pp. 60–64, 2018, doi: 10.18196/st.211211.

[8] F. Indriyani and Infriani E, "Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means (Clustering Sales Data At Outdoor Equipment Stores Using K-Means Method)," *Jurnal Informatika (JUITA)*, vol. 7, no. 2, pp. 109–113, Nov. 2019.

[9] R. Pradena Harjono, A. Magdalena, and I. Pakereng, "Penerapan Metode K-Means Clustering Untuk Analisis Potensi Lahan Pangan Pada Provinsi Kalimantan Selatan," *Jurnal Sains Komputer & Informatika (J-SAKTI*, vol. 7, no. 1, pp. 332–338, Mar. 2023.

[10] M. A. Sembiring *et al.*, "Penerapan Metode Algoritma K-Means Clustering Untuk Pemetaan Penyebaran Penyakit Demam Berdarah Dengue (DBD)," *Journal of Science and Social Research*, no. 3, pp. 336–341, 2021, [Online]. Available: http://jurnal.goretanpena.com/index.php/JSSR

[11] A. Setiadi and E. Delima Sikumbang, "K-Means Clustering Dalam Penerimaan Karyawan Baru," *Informatics For Educators And Professionals*, vol. 4, no. 2, pp. 103–112, Jun. 2020.

[12] E. Prasetyaningrum and P. Susanti, "Perbandingan Algoritma K-Means Dan K-Medoids Untuk Pemetaan Hasil Produksi Buah-Buahan," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, pp. 1775–1783, 2023, doi: 10.30865/mib.v7i4.6477.

[13] H. Mahulae, "Pengelompokan Potensi Produksi Buah-Buahan Di Provinsi Sumatera Utara dengan Menerapkan K-Clustering (Studi Kasus : Dinas Tanaman Pangan dan Holtikultura)," *JURIKOM (Jurnal Riset Komputer)*, vol. 7, no. 2, pp. 312–325, Apr. 2020, doi: 10.30865/jurikom.v7i2.2122.

[14] D. Fadhillah, A. Faisol, and N. Vendyansyah, "Penerapan Metode K-Means Clustering Pada Pemetaan Lahan Kopi Di Kabupaten Malang," *Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 1, pp. 162–170, Feb. 2022.

[15] S. Wijayanto, M. Yoka Fathoni, J. DI Panjaitan No, K. Purwokerto Selatan, K. Banyumas, and J. Tengah, "Pengelompokkan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means," *Jurnal JUPITER*, vol. 13, no. 2, pp. 212–219, Oct. 2021.

[16] Lidya, R. Buaton, and Nurhayati, "Clustering Hasil Panen Berdasarkan Lokasi dan Jenis Bibit (Studi Kasus: Dinas Pangan Dan Pertanian Kota Binjai)," *Jurnal Informatika Kaputama (JIK)*, vol. 6, no. 3, pp. 336–344, Aug. 2022.

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 1, November 2023, Page 82-92
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i1.4523

[17] I. Vhallah and J. Santony, "Pengelompokan Mahasiswa Potensial Drop Out menggunakan Metode Clustering K-Means," *SMARTICS Journal*, vol. 2, no. 2, pp. 572–577, 2018, [Online]. Available: http://jurnal.iaii.or.id

[18] Teguh Pribadi, Rahmad Irsyada, Hastie Audytra, and Doni Abdul Fatah, "Implementasi Algoritma K-Means Untuk Klasterisasi Potensi Desa Pada Sektor Produksi Pertanian Di Kabupaten Bojonegoro," *Jurnal SimanteC*, vol. 9, pp. 20–28, 2020.

[19] S. Nanda Saputra, E. Haerani, L. Oktavia, and F. Syafria, "Penerapan Algoritma K-Means Pada Clustering Penerima Bantuan Pangan Non Tunai (BPNT) Application of K-Means Algorithm on Clustering Recipients of Non-Cash Food Assistance (NCFA)," *Journal of Computing Engineering, System and Science)*, vol. 8, no. 2, pp. 438–449, 2023, [Online]. Available: www.jurnal.unimed.ac.id

[20] E. Irfiani, S. Sulistia Rani, S. Nusa Mandiri Jl Kramat Raya No, and J. Pusat, "Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita," *Jurnal Sistem Dan Teknologi Informasi*, vol. 6, no. 4, pp. 17–27, Oct. 2018.

[21] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *Jurnal Sains dan Manajemen*, vol. 9, no. 1, pp. 95–100, Mar. 2021, [Online]. Available: www.bps.go.id