

Implementasi K-Means Clustering Kelompok Provinsi Penghasil Bawang Merah di Indonesia Tahun 2023

Anton Maulana Ibrahim^{a*}, Yan Akhmad Fauzi^b

^{a,b} Politeknik Mitra Karya Mandiri

*correspondence : antonmaulanaibrahim2021@gmail.com

Abstract—Shallots are a strategic horticultural commodity that plays a vital role in food security and price stabilization in Indonesia. However, inter-provincial production data is still in the raw form, making it difficult to identify the distribution patterns of high, medium, and low-production areas. This study aims to implement the K-Means Clustering algorithm to group 38 provinces in Indonesia based on 2023 shallot production and productivity data sourced from the Central Statistics Agency (BPS). The method used is a data mining approach through the stages of data selection, preprocessing (Min-Max normalization), data transformation, determining the number of clusters using the Elbow method, and evaluation using the Silhouette Score with the assistance of Orange Data Mining software. The results show the formation of three clusters: a low-medium production cluster dominated by most provinces outside the main centers, a high production cluster consisting of Central Java and East Java as national centers, and a very high production cluster encompassing several provinces such as West Java and West Nusa Tenggara. The results of this grouping provide an overview of the inequality in production distribution between regions and can be the basis for formulating development policies, production equality, and shallot distribution planning in a more targeted and data-based manner.

Index Terms— K-Means Clustering; Data mining; Shallots; Agricultural production; provincial clustering.

Abstrak— Bawang merah adalah komoditas pertanian penting yang berpengaruh besar terhadap keamanan pangan dan harga pasar di Indonesia. Namun, data produksi antarprovinsi masih dalam bentuk mentah, sehingga sulit untuk mengetahui pola penyebaran wilayah yang memiliki produksi tinggi, sedang, atau rendah. Penelitian ini bertujuan untuk menerapkan algoritma K-Means Clustering agar dapat mengelompokkan 38 provinsi di Indonesia berdasarkan data produksi dan produktivitas bawang merah tahun 2023 yang diambil dari Badan Pusat Statistik (BPS). Metode yang dipakai adalah pendekatan data mining dengan langkah-langkah seperti pemilihan data, pra-pemrosesan (normalisasi Min-Max), transformasi data, menentukan jumlah kluster dengan metode Elbow, serta pengevaluasian menggunakan figure score dengan bantuan perangkat lunak Orange Data Mining. Penelitian menunjukkan adanya tiga kelompok, yaitu kelompok produksi rendah hingga menengah yang terdiri dari sebagian besar provinsi di luar pusat utama, kelompok produksi tinggi yang meliputi Jawa Tengah dan Jawa Timur sebagai pusat nasional, serta kelompok produksi sangat tinggi yang mencakup beberapa provinsi seperti Jawa Barat dan Nusa Tenggara Barat. Hasil pengelompokan ini menunjukkan perbedaan besar dalam distribusi produksi antar daerah dan bisa menjadi acuan dalam menyusun kebijakan untuk mengembangkan produksi, meratakan distribusi, serta merencanakan pengembangan bawang merah secara lebih tepat sasaran dan berdasarkan data.

Kata Kunci— K-Means Clustering; Data mining; Bawang merah; Produksi pertanian; Pengelompokan provinsi.

I. PENDAHULUAN

Sektor pertanian menjadi salah satu tulang punggung perekonomian Indonesia, karena memberikan makanan, menyerap tenaga kerja, serta mendorong berbagai aktivitas ekonomi di daerah pedesaan. Beberapa komoditas penting seperti padi, jagung, sayuran, dan tanaman perkebunan berperan besar dalam menjaga ketersediaan makanan nasional dan meningkatkan penghasilan petani [1]. Dari segala jenis sayuran, bawang merah memiliki peran khusus karena digunakan sebagai bumbu utama dalam masakan sehari-hari, bahan dasar industri makanan, serta memiliki harga yang sering berubah-ubah. Kondisi ini membuat informasi tentang penyebaran dan kemampuan produksi bawang merah di antar daerah sangat penting untuk merencanakan produksi, pengiriman, dan menjaga harga tetap stabil.

Data produksi pertanian di Indonesia umumnya terdokumentasi dengan baik di Badan Pusat Statistik (BPS) dan instansi terkait, namun sering kali masih berupa data mentah per provinsi dan per tahun. Data mentah seperti ini sulit langsung dimanfaatkan pembuat kebijakan untuk mengidentifikasi wilayah dengan potensi tinggi, sedang, atau rendah tanpa proses pengolahan yang sistematis [2]. Penelitian pada tanaman

perkebunan menunjukkan bahwa ketika data mentah produksi provinsi diklasterisasi, pemerintah menjadi lebih mudah mengenali provinsi dengan produksi besar maupun kecil sehingga dapat menyusun kebijakan intervensi yang lebih tepat sasaran [1]. Pola yang sama relevan diterapkan pada komoditas bawang merah.

Dalam konteks analisis data modern, data mining dan khususnya teknik clustering telah banyak digunakan untuk mengelompokkan wilayah berdasarkan karakteristik produksi pertanian. Clustering bertujuan membagi sekumpulan objek data ke dalam beberapa kelompok (cluster) sedemikian rupa sehingga objek dalam satu cluster memiliki kemiripan tinggi, sedangkan antarcluster memiliki perbedaan yang jelas [3][4]. Pendekatan ini telah diterapkan untuk pengelompokan produksi padi menurut provinsi, produksi tanaman perkebunan, produksi jagung [5], hingga pemetaan komoditas hortikultura di tingkat kabupaten/kota. Hasil-hasil tersebut menunjukkan bahwa pengelompokan wilayah berbasis data produksi dapat memberikan wawasan spasial yang penting bagi perencanaan sektor pertanian.

Salah satu algoritma clustering non-hirarki yang paling luas digunakan adalah K-Means. Algoritma ini mempartisi data ke dalam K kelompok dengan meminimalkan jarak antara setiap data dan pusat kluster (centroid), sehingga data dalam satu kelompok relatif homogeny [6]. K-Means digemari karena: Struktur algoritme yang sederhana dan mudah dipahami. Waktu komputasi yang relatif cepat untuk data numerik berukuran besar. Hasil cluster yang mudah diinterpretasikan oleh pengambil keputusan [6].

Berbagai penelitian di bidang pertanian telah memanfaatkan K-Means untuk pemetaan produksi dan potensi wilayah. Sudirman dkk. mengelompokkan provinsi di Indonesia berdasarkan produksi padi menggunakan tiga parameter: luas panen, produktivitas, dan produksi, dengan hasil tiga cluster: produksi tinggi, normal, dan rendah. Pendekatan serupa diterapkan pada produksi tanaman perkebunan (kelapa, kopi, kakao) di tingkat provinsi dengan pembentukan dua cluster (produksi banyak dan produksi sedikit), dan hasilnya terbukti membantu pemerintah mengidentifikasi provinsi berpotensi tinggi dan rendah. Di tingkat yang lebih lokal, K-Means juga digunakan untuk memetakan produktivitas jahe per kecamatan, kelapa sawit per blok kebun, serta kelayakan tanaman jagung di suatu kabupaten, semuanya dengan tujuan serupa: mengenali pola wilayah berproduksi tinggi, sedang, dan rendah untuk mendukung perencanaan kebijakan dan alokasi sumber daya [7] [8]

Dalam konteks hortikultura, pemetaan kabupaten/kota di Jawa Barat berdasarkan indikator produksi hortikultura, curah hujan, dan suhu dengan K-Means menunjukkan bahwa pemilihan jumlah kluster yang optimal (misal tiga kluster) dapat mengungkap struktur wilayah dengan produksi tinggi, sedang, dan rendah yang berkaitan dengan kondisi agroklimat. Hasil ini menegaskan bahwa K-Means sangat relevan untuk komoditas hortikultura seperti bawang merah yang juga dipengaruhi oleh kondisi agroekologi.

Sebagai komoditas hortikultura strategis, produksi bawang merah di Indonesia sangat terpusat pada beberapa provinsi utama. Namun, tanpa analisis clustering, sulit untuk melihat secara sistematis bagaimana tingkat produksi antarprovinsi terbagi ke dalam kelompok tinggi, sedang, dan rendah. Pendekatan clustering provinsi telah terbukti efektif dalam menganalisis sektor pertanian secara luas di Indonesia—misalnya untuk memetakan kesenjangan sektor pertanian antarprovinsi tahun 2023, di mana kombinasi kluster hirarki dan non-hirarki (termasuk K-Means) menghasilkan beberapa kluster yang mencerminkan perbedaan produktivitas dan pemanfaatan lahan. Pendekatan serupa pada komoditas khusus seperti padi dan tomat juga memberikan gambaran sebaran wilayah produksi tinggi dan rendah yang dapat dijadikan dasar perencanaan intervensi [9].

Penerapan K-Means pada data produksi bawang merah tahun 2023 di tingkat provinsi, akan dimungkinkan untuk: Mengidentifikasi kelompok provinsi berproduksi tinggi yang menjadi tulang punggung pasokan nasional. Mengenali provinsi dengan produksi sedang dan rendah yang berpotensi ditingkatkan melalui teknologi budidaya, infrastruktur, atau kebijakan harga. Menyusun peta tematik atau sistem pemetaan berbasis web yang menampilkan distribusi cluster, sebagaimana telah dilakukan pada pemetaan produksi padi dan jahe.

Penelitian sebelumnya menunjukkan bahwa hasil clustering semacam ini tidak hanya bersifat deskriptif, tetapi juga dapat dijadikan bahan masukan kebijakan. Misalnya, pada pengelompokan produksi padi per provinsi, cluster dengan produksi rendah direkomendasikan sebagai sasaran sosialisasi dan intervensi untuk meningkatkan produksi. Pada studi lain, peta cluster produksi digunakan sebagai dasar rekomendasi dalam pengalihan fungsi lahan pertanian dan penentuan prioritas pengembangan [6]. Untuk bawang merah, hasil serupa dapat menjadi basis perencanaan stabilisasi pasokan, pengembangan sentra produksi baru, dan penentuan prioritas pembangunan infrastruktur pascapanen [10].

Implementasi K-Means menuntut penentuan jumlah kluster K yang tepat. Berbagai studi di bidang pertanian menggunakan metode seperti Elbow, Davies-Bouldin Index (DBI), dan Silhouette Index untuk

mengevaluasi kualitas pengelompokan dan memilih K yang optimal . Pada clustering padi per provinsi, DBI digunakan untuk menilai kualitas cluster dan dikaitkan dengan pemilihan jumlah cluster yang paling baik. Penelitian lain pada data pertanian menggunakan kombinasi K-Means dan indeks mutu cluster untuk menunjukkan bahwa K-Means dapat menghasilkan struktur cluster yang jelas antara wilayah berproduksi tinggi dan rendah. Pendekatan evaluasi ini dapat diadaptasi pada data produksi bawang merah guna memastikan bahwa jumlah cluster yang digunakan benar-benar merepresentasikan pola nyata di lapangan [11].

Penelitian ini menggunakan data BPS Tahun 2023 dan agribisnis tahun 2023 menggambarkan kondisi terkini sektor pertanian dan pola produksi setelah berbagai dinamika ekonomi dan pasca-pandemi. Selaras dengan penelitian lain yang juga memetakan sektor pertanian Indonesia tahun 2023 menggunakan pendekatan clustering, sehingga hasilnya dapat dibandingkan atau dikontekstualisasikan dengan pola sektor pertanian yang lebih umum.

Memberikan dasar untuk memantau perubahan struktur produksi bawang merah dari waktu ke waktu, jika di kemudian hari dilakukan studi lanjutan dengan rentang tahun yang lebih panjang seperti pada penelitian padi dan tanaman lainnya [12]. Dengan demikian, implementasi K-Means clustering untuk menentukan kelompok provinsi penghasil bawang merah di Indonesia tahun 2023 memiliki urgensi akademik dan praktis. Dari sisi akademik, penelitian ini memperkaya literatur penerapan data mining dan clustering pada komoditas hortikultura spesifik, melengkapi studi serupa pada padi, jagung, tomat, jahe, dan tanaman perkebunan [13]. Dari sisi praktis, hasil pengelompokan diharapkan dapat menjadi dasar bagi pemerintah pusat dan daerah, maupun pelaku rantai pasok, untuk merumuskan strategi pengembangan bawang merah yang lebih terarah, efisien, dan berbasis bukti.

II. TINJAUAN PUSTAKA

K-Means adalah metode clustering non-hierarkis. Metode ini mengelompokkan data ke dalam beberapa cluster. Tujuannya adalah agar data dengan karakteristik serupa berada di cluster yang sama. Ini juga meminimalkan variasi di dalam cluster [6]. Dalam studi kelayakan penanaman jagung di Lampung Selatan, K-Means digunakan untuk mengelompokkan kecamatan berdasarkan potensi produksi jagung. Dengan cara ini, diperoleh kelompok wilayah dengan potensi tinggi dan rendah sebagai dasar prioritas pengembangan.

Data mining adalah proses mengolah data dalam jumlah besar untuk menemukan pola, hubungan, serta informasi yang bermanfaat. Dalam proses ini, terdapat berbagai teknik yang dapat digunakan, salah satunya adalah clustering, yaitu metode untuk mengelompokkan data berdasarkan kemiripan karakteristik yang dimilikinya. [14]. Orange Data Mining adalah perangkat lunak open-source yang digunakan untuk analisis data, visualisasi, dan machine learning. Orange dikembangkan oleh University of Ljubljana dan dirancang untuk memudahkan pengguna dalam melakukan eksplorasi data tanpa harus memiliki kemampuan pemrograman yang tinggi.

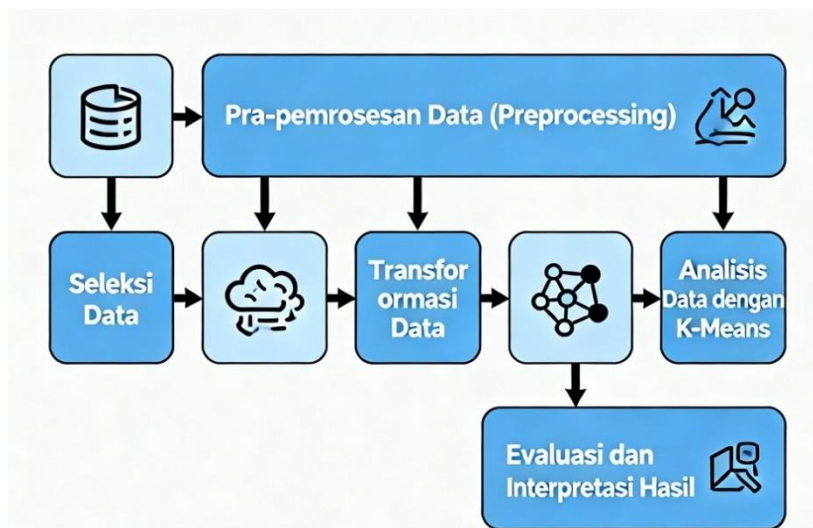
Clustering adalah teknik analisis data yang bertujuan untuk mengelompokkan objek ke dalam beberapa kelompok (cluster) sehingga objek dalam satu kelompok memiliki tingkat kemiripan yang tinggi, sedangkan objek antar kelompok memiliki perbedaan yang signifikan [15].

III. METODE

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan data mining. Teknik data mining digunakan untuk menemukan pola atau struktur tersembunyi dalam data produksi bawang merah antar provinsi di Indonesia. Metode clustering dipilih karena tujuan penelitian adalah mengelompokkan provinsi berdasarkan kemiripan karakteristik produksinya tanpa label kelas sebelumnya. Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Indonesia tahun 2023. Dataset mencakup seluruh provinsi di Indonesia dengan variabel yang relevan terhadap produksi bawang merah, antara lain: - Produksi bawang merah (ton) dan - Produktivitas (kuintal/hektar)

Data diambil dari publikasi resmi BPS sehingga memiliki validitas dan reliabilitas yang dapat dipertanggungjawabkan. Teknik pengumpulan data dilakukan dengan dokumentasi, yaitu mengunduh dan mengolah data statistik produksi bawang merah tahun 2023 dari situs resmi BPS (<https://www.bps.go.id>). Data kemudian disusun dalam format spreadsheet (.xlsx atau .csv) agar dapat diproses menggunakan perangkat lunak data mining.

Tahapan-tahapan penelitian dengan judul Implementasi K-Means Clustering untuk Menentukan Kelompok Provinsi Penghasil Bawang Merah di Indonesia Tahun 2023 adalah sebagai berikut:



Gambar 1. Alur Penelitian

Pada gambar 1. Menunjukkan tahapan-tahapan penelitian diawali dengan seleksi data yang berkaitan langsung dengan produksi bawang merah dipilih dari dataset BPS. Provinsi yang tidak memiliki data lengkap akan dievaluasi untuk ditangani pada tahap pra-pemrosesan. Pra-pemrosesan Data (Preprocessing) Tahap ini bertujuan untuk meningkatkan kualitas data, meliputi: Pemeriksaan data hilang (missing values), Penghapusan duplikasi data, Penyeragaman format data numeric, Normalisasi data menggunakan metode Min-Max Normalization, agar setiap variabel memiliki skala yang sebanding dan tidak mendominasi proses clustering. Tahap selanjutnya Transformasi Data, Data ditransformasikan ke dalam format yang sesuai untuk algoritma K-Means, yaitu seluruh atribut bersifat numerik dan berada pada skala yang seragam. Analisis data menggunakan algoritma K Means. Algoritma K-Means digunakan untuk mengelompokkan provinsi ke dalam beberapa cluster berdasarkan kemiripan karakteristik produksi bawang merah. Langkah-langkah algoritma K-Means adalah sebagai berikut:

1. Menentukan jumlah cluster (k)
2. Menentukan pusat cluster (centroid) secara acak
3. Menghitung jarak setiap data ke centroid menggunakan Euclidean Distance
4. Mengelompokkan data ke centroid terdekat
5. Memperbarui centroid berdasarkan rata-rata anggota cluster
6. Mengulangi proses hingga centroid stabil atau tidak mengalami perubahan signifikan

Penentuan jumlah cluster (k) dilakukan dengan mempertimbangkan metode Elbow Method untuk memperoleh hasil pengelompokan yang optimal.

Proses pengolahan dan analisis data pada penelitian ini dilakukan menggunakan perangkat lunak Orange Data Mining. Pemilihan Orange sebagai alat analisis didasarkan pada beberapa pertimbangan metodologis dan teknis sebagai berikut:

1. Mendukung Algoritma K-Means Clustering

Orange menyediakan implementasi algoritma K-Means Clustering yang stabil dan terintegrasi dengan fitur evaluasi seperti Silhouette Score. Hal ini memudahkan peneliti dalam melakukan proses pengelompokan data provinsi berdasarkan kemiripan produksi bawang merah secara efisien dan sistematis.

2. Antarmuka Visual Berbasis Workflow

Orange menggunakan sistem visual programming berbasis drag-and-drop widget yang memungkinkan peneliti membangun alur analisis (*workflow*) secara terstruktur. Dengan pendekatan ini, tahapan seperti: Input data, Seleksi atribut, Normalisasi data, Proses clustering, Visualisasi hasil, dapat dilakukan secara transparan dan terdokumentasi dengan baik.

3. Kemudahan Preprocessing Data

Orange menyediakan fitur preprocessing seperti: Penanganan missing value, Normalisasi (*Min-Max Scaling*, *Z-Score*), Transformasi atribut. Hal ini penting karena algoritma K-Means sensitif terhadap skala data, sehingga normalisasi menjadi tahapan yang wajib dilakukan sebelum proses

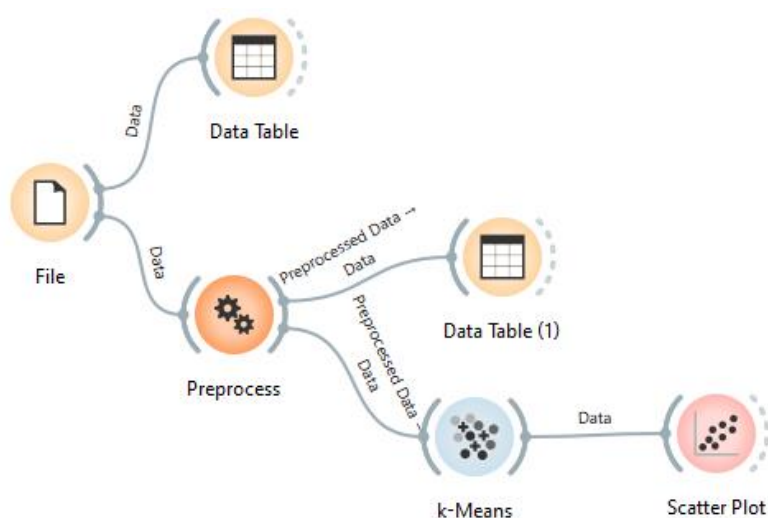
clustering.

4. Visualisasi dan Evaluasi Model

Orange memiliki berbagai widget visualisasi seperti: Scatter Plot, Silhouette Plot, Data Table, Distance Map; Fitur ini membantu peneliti dalam: Mengevaluasi kualitas cluster, Menginterpretasikan hasil pengelompokan, Menyajikan hasil analisis secara informative.

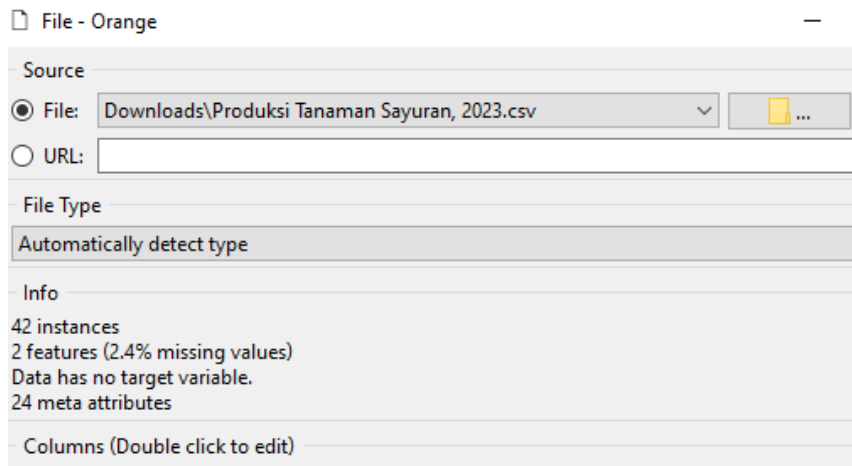
IV. HASIL DAN PEMBAHASAN

Hasil penerapan algoritma K-Means Clustering pada data produksi bawang merah tahun 2023 berasal dari Badan Pusat Statistik (BPS). Pengolahan data dilakukan dengan menggunakan Orange Data Mining melalui beberapa tahap, yaitu seleksi data, normalisasi, proses clustering, dan evaluasi kualitas cluster menggunakan metode silhouette. Hasil pengelompokan kemudian diinterpretasikan untuk mengidentifikasi pola dan karakteristik provinsi berdasarkan tingkat produksinya. Untuk mengimplementasikan algoritma K-Means Clustering terhadap data produksi bawang merah tahun 2023 dibuat workflow yang terdiri dari beberapa tahapan utama sebagai berikut:.



Gambar 2. Workflow

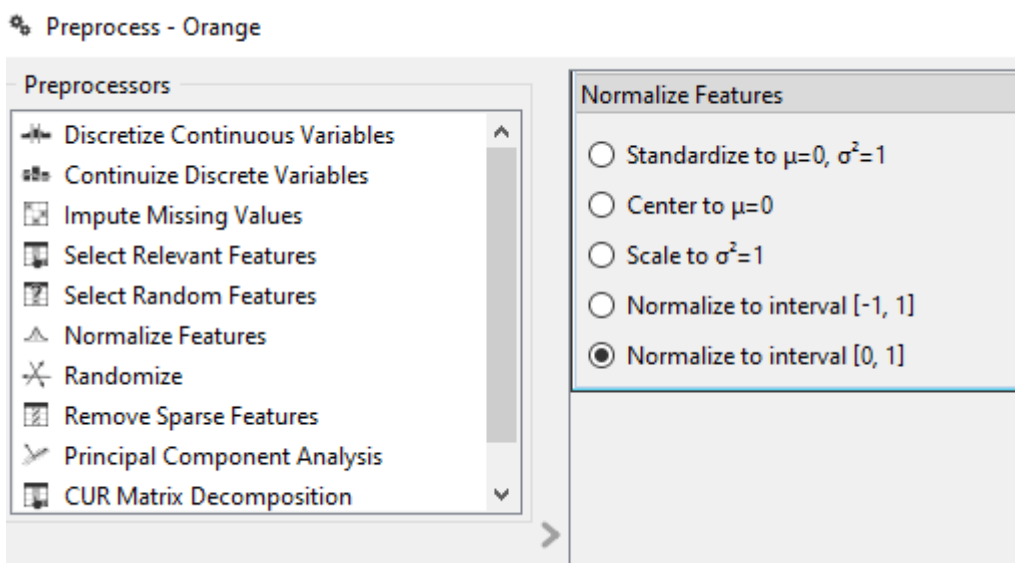
Widget File digunakan untuk mengimpor dataset produksi bawang merah yang telah diperoleh dari Badan Pusat Statistik (BPS). Dataset umumnya berformat .csv dan berisi data 38 provinsi beserta variabel penelitian seperti produksi, luas panen, dan produktivitas. Dataset diambil dari Tahap ini merupakan proses awal untuk memasukkan data ke dalam sistem Orange.



Gambar 3. Mengunggah Dataset

Widget Data Table pertama digunakan untuk menampilkan dan memverifikasi data mentah sebelum dilakukan pengolahan lebih lanjut. Pada tahap ini dilakukan pengecekan: Kelengkapan data, Kesesuaian tipe data (numerik/kategorik) Konsistensi penulisan nama provinsi Tahap ini penting untuk memastikan tidak terdapat kesalahan input sebelum proses preprocessing.

Widget Preprocess digunakan untuk melakukan tahapan pra-pemrosesan data. Dalam penelitian ini, preprocessing meliputi: Normalisasi data (misalnya Min-Max Scaling) Penyesuaian atribut numeric Pembersihan data jika diperlukan, Penulis melakukan pengerjaan dengan 1 Dataset yang kemudian diolah dengan preprocessed data yang kemudian menghasilkan dengan normalisasi yang kemudian diproses menggunakan algoritma k-means dan menghasilkan hasil clustering yang terdiri dari 3 cluster



Gambar 4. Preprocessing data

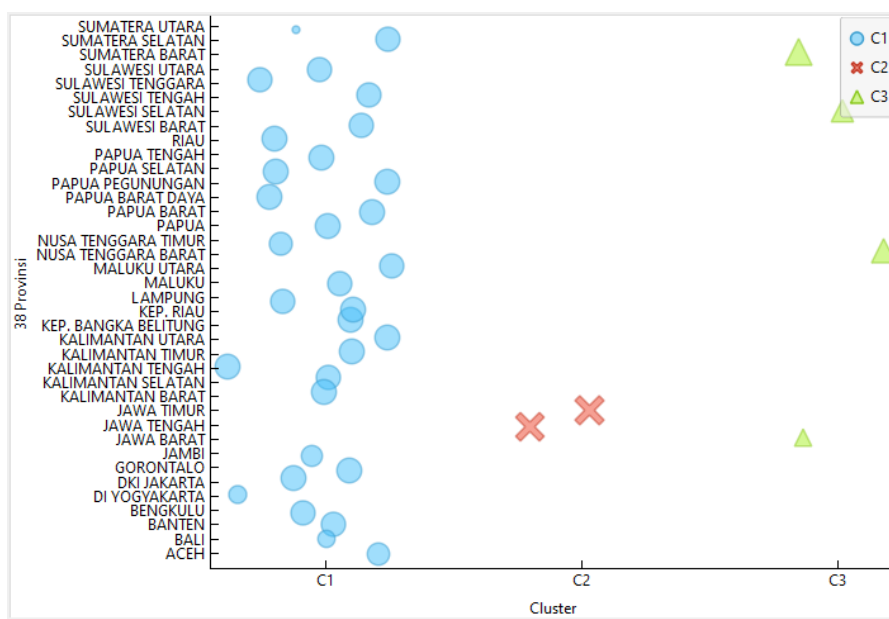
Tahapan ini sangat penting karena algoritma K-Means sensitif terhadap skala data. Jika tidak dilakukan normalisasi, variabel dengan nilai besar (misalnya produksi dalam ton) dapat mendominasi proses pengelompokan. Gambar4 tersebut menunjukkan tampilan widget Preprocess pada Orange Data Mining, khususnya bagian Normalize Features, yang digunakan dalam penelitian ini sebelum proses K-Means Clustering dilakukan. preprocessing dilakukan karena algoritma K-Means menggunakan perhitungan jarak (Euclidean Distance) yang sangat sensitif terhadap perbedaan skala data

Data hasil dari preprocessing yang ditandai dengan berubahnya nilai angka pada data

	Produksi Tanaman Sayuran	38 Provinsi
1	0.00688735	ACEH
2	0.0330364	SUMATERA UTARA
3	0.117828	SUMATERA BARAT
4	0.000162701	RIAU
5	0.00926894	JAMBI
6	0.000602952	SUMATERA SELATAN
7	0.000337996	BENGKULU
8	0.00110466	LAMPUNG
9	8.86546e-05	KEP. BANGKA BELITUNG
10	8.56323e-06	KEP. RIAU
11	0	DKI JAKARTA
12	0.0903446	JAWA BARAT
13	0.241327	JAWA TENGAH
14	0.0165991	DI YOGYAKARTA
15	0.244137	JAWA TIMUR
16	0.000433199	BANTEN
17	0.0178186	BALI
18	0.1071	NUSA TENGGARA BARAT
19	0.00574693	NUSA TENGGARA TIMUR
20	5.49054e-05	KALIMANTAN BARAT
21	3.62678e-05	KALIMANTAN TENGAH
22	0.000243296	KALIMANTAN SELATAN
23	0.000126937	KALIMANTAN TIMUR
24	0.000176805	KALIMANTAN UTARA
25	0.00158823	SULAWESI UTARA

Gambar 5. Data Transformation

Pada data hasil cluster terbagi menjadi 3 yaitu terbesar menengah hingga terendah



Gambar 6. Visualisasi Data Clustering

	Cluster	Silhouette	Produksi Tanaman Sayuran	38 Provinsi
1	C1	0.604346	0.028211	ACEH
2	C1	0.558227	0.135319	SUMATERA UTARA
3	C3	0.618342	0.482632	SUMATERA BARAT
4	C1	0.611549	0.000666434	RIAU
5	C1	0.601389	0.0379661	JAMBI
6	C1	0.611127	0.00246973	SUMATERA SELATAN
7	C1	0.611382	0.00138445	BENGKULU
8	C1	0.610638	0.00452474	LAMPUNG
9	C1	0.611162	0.000363134	KEP. BANGKA BELITUNG
10	C1	0.611695	3.50755e-05	KEP. RIAU
11	C1	0.611704	0	DKI JAKARTA
12	C3	0.588312	0.370057	JAWA BARAT
13	C2	0.64071	0.988491	JAWA TENGAH
14	C1	0.590828	0.0679907	DI YOGYAKARTA
15	C2	0.642704	1	JAWA TIMUR
16	C1	0.61129	0.00177441	BANTEN
17	C1	0.588846	0.0729859	BALI
18	C3	0.610453	0.438687	NUSA TENGGARA BARAT
19	C1	0.605685	0.0235398	NUSA TENGGARA TIMUR
20	C1	0.611652	0.000224896	KALIMANTAN BARAT
21	C1	0.611669	0.000148555	KALIMANTAN TENGAH
22	C1	0.611472	0.000996556	KALIMANTAN SELATAN
23	C1	0.611583	0.000519942	KALIMANTAN TIMUR
24	C1	0.611536	0.000724206	KALIMANTAN UTARA
25	C1	0.610158	0.00650547	SULAWESI UTARA
26	C1	0.61012	0.00666022	SULAWESI TENGAH
27	C3	0.604478	0.415585	SULAWESI SELATAN

Gambar 7. Hasil Klastering

Berdasarkan Gambar 6 dan Gambar 7 visualisasi klusterisasi produksi bawang merah dari 38 provinsi di Indonesia, data terbagi menjadi 3 kluster (C1, C2, dan C3) yang menggambarkan tingkat produksi berbeda. Kluster C1 (Produksi Rendah – Menengah) Ditandai dengan titik biru. Merupakan kluster dengan jumlah provinsi terbanyak. Berisi mayoritas provinsi seperti: Aceh, Bali, Banten, Bengkulu, DI Yogyakarta, DKI Jakarta, Gorontalo, Sebagian besar provinsi di Kalimantan, Sulawesi, Maluku, Nusa Tenggara, Papua, dan Sumatera. Karakteristik: Produksi bawang merah relatif rendah hingga menengah dibanding provinsi lain. Umumnya merupakan daerah non-sentra utama bawang merah atau memiliki keterbatasan lahan serta produksi skala kecil–menengah. Kluster C2 (Produksi Tinggi) Ditandai dengan tanda silang merah Berisi sedikit provinsi, yaitu: Jawa Tengah, Jawa Timur Karakteristik: Merupakan sentra utama produksi bawang merah nasional. Kedua provinsi ini memiliki produksi jauh lebih tinggi dibanding provinsi lain, sehingga terpisah jelas dari kluster lainnya. Kluster C3 (Produksi Sangat Tinggi / Outlier). Ditandai dengan segitiga hijau Berisi beberapa provinsi dengan produksi sangat tinggi, antara lain: Jawa Barat ,Nusa Tenggara Barat, Sumatera Barat (beberapa provinsi lain dengan posisi ekstrem pada grafik) Karakteristik: Produksi lebih tinggi dibanding mayoritas provinsi (C1), tetapi berbeda pola dari C2. Bisa dikategorikan sebagai daerah penghasil besar namun bukan yang paling dominan seperti Jawa Tengah dan Jawa Timur.

V. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma K-Means Clustering untuk mengelompokkan 38 provinsi di Indonesia berdasarkan data mengenai produksi dan produktivitas bawang merah tahun 2023 yang diperoleh dari Badan Pusat Statistik (BPS). Proses analisis dilakukan melalui langkah-langkah penambangan data yang teratur, termasuk pemilihan data, pra-pengolahan dengan normalisasi Min-Max, transformasi data, penetapan jumlah kelompok menggunakan metode Elbow, dan penilaian kualitas kelompok menggunakan Skor Silhouette dengan software Orange Data Mining. Hasil pengelompokan menunjukkan bahwa data produksi bawang merah terbagi menjadi tiga kelompok utama: kelompok produksi rendah hingga menengah (C1), kelompok produksi tinggi (C2), dan kelompok produksi sangat tinggi (C3).

Kelompok C1 memiliki jumlah provinsi terbanyak dan didominasi oleh provinsi yang tidak termasuk pusat utama bawang merah nasional. Tingkat produksinya relatif lebih rendah atau menengah. Kelompok C2 terdiri dari provinsi dengan produksi yang sangat dominan, yaitu Jawa Tengah dan Jawa Timur, yang menjadi tulang punggung pasokan bawang merah nasional. Di sisi lain, kelompok C3 mencakup provinsi

dengan produksi yang tinggi tetapi tidak sebesar kelompok C2, seperti Jawa Barat, Nusa Tenggara Barat, dan Sumatera Barat, yang berfungsi sebagai pusat pendukung produksi nasional.

Temuan ini menunjukkan adanya konsentrasi produksi bawang merah yang cukup besar di area tertentu, terutama di Pulau Jawa. Hal ini menggambarkan ketidakmerataan distribusi produksi antarprovinsi. Informasi ini penting sebagai dasar untuk merumuskan kebijakan pertanian. Kebijakan ini dapat membantu pemerataan pengembangan pusat produksi, meningkatkan produktivitas di daerah yang berklaster rendah, serta merencanakan distribusi dan stabilisasi harga. Dari sisi akademik, penelitian ini menunjukkan bahwa metode K-Means efektif untuk menemukan pola sebaran produksi hortikultura berdasarkan data numerik. Dari sudut pandang praktis, hasil pengelompokan dapat menjadi acuan bagi pemerintah dan pemangku kepentingan dalam merencanakan strategi pengembangan bawang merah yang lebih terarah, efisien, dan berbasis pada bukti untuk tahun-tahun mendatang.

VI. DAFTAR PUSTAKA

- [1] M. Isaura, F. D. Aryasatyawan, and S. P. Wulandari, "Analisis Kesenjangan Regional di Sektor Pertanian Indonesia Tahun 2023 Melalui Pendekatan Cluster," vol. 2, 2024.
- [2] N. D. Salsabila, K. Aulisari, and H. Z. Zahro, "Penerapan Algoritma K-Means Untuk Klasterisasi Produktivitas Tanaman Jahe," *Infotek J. Inform. dan Teknol.*, vol. 8, no. 1, pp. 228–238, 2025.
- [3] M. M. Cortez, "Data mining tools | rapidminer : K-means method on clustering of rice crops by province as efforts to stabilize food crops in Indonesia," 2018, doi: 10.1088/1757-899X/420/1/012089.
- [4] E. Sugiatna, A. M. Ibrahim, and I. A. Hadi, "Implementasi Algoritma Klasifikasi C4 . 5 Untuk Memprediksi Kelayakan Pembelian Kendaraan To Predict the Eligibility of Vehicle Purchases)," vol. 1, no. 2, pp. 124–132, 2019.
- [5] D. F. Pasaribu, I. S. Damanik, E. Irawan, and H. S. Tambunan, "Memanfaatkan Algoritma K-Means Dalam Memetakan Potensi Hasil Produksi Kelapa Sawit PTPN IV Marihat," vol. 2, no. 1, pp. 11–20, 2021.
- [6] W. Wahyudi and E. Sabara, "Desain Dan Implementasi Media Pembelajaran Mikrokontroler Berbasis Hybrid Learning Menggunakan Wokwi Simulation," *J. Media Elektr.*, vol. 19, no. 3, pp. 186–193, 2022, doi: 10.59562/metrik.v19i3.5473.
- [7] R. Lakui, G. Sastro, and T. H. Setiawan, "Clustering of Regencies and Cities in West Java Province Based on Horticultural Indicators Using the K-Means Method," vol. 3, no. 4, pp. 159–164, 2025.
- [8] P. Yang, "基于 K-Means 聚类算法的地区农业差异性 分类研究 Research on the Classification of Regional Agricultural Differences Based on K-Means Clustering Algorithm," vol. 15, no. April, pp. 498–504, 2025, doi: 10.12677/hjas.2025.154062.
- [9] D. Abdullah, S. S. Ansari, S. Ahmar, and R. R. Rahmat, "The application of K - means clustering for province clustering in Indonesia of the risk of the COVID - 19 pandemic based," *Qual. Quant.*, vol. 56, no. 3, pp. 1283–1291, 2022, doi: 10.1007/s11135-021-01176-w.
- [10] N. Hidayah and R. S. Adiandri, "Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency," 2021, doi: 10.1088/1742-6596/1751/1/012038.
- [11] A. N. Fuad and M. S. Zuhrie, "Rancang Bangun Sistem Monitoring Dan Pengontrolan Ph Nutrisi Pada Hidroponik Sitem Nutrient Film Technique (NFT) Menggunakan Pengendali Pid Berbasis Arduino UNO." [Online]. Available: www.dfrobot.com
- [12] I. Mujahidin, S. H. Hasanah, U. Terbuka, F. C-means, and D. Index, "Comparative analysis of k-means, k-medoids, and fuzzy c-means for clustering provinces in indonesia based on rice production in 2024 1,2," vol. 14, pp. 356–365, 2025, doi: 10.14710/j.gauss.14.2.356-365.
- [13] Pattharaporn Thongnim; Ekkapot Charoenwanit; Thanaphon Phukseng, "Cluster Quality in Agriculture: Assessing GDP and Harvest Patterns in Asia and Europe with K-Means and Silhouette Scores," 2023, [Online]. Available: <https://ieeexplore.ieee.org/document/10423469>
- [14] S. Setyaningtyas, B. I. Nugroho, and Z. Arif, "Tinjauan Pustaka Sistematis Pada Data Mining :

- Studi Kasus Algoritma K-Means Clustering,” vol. 10, no. 2, pp. 52–61, 2022.
- [15] N. Wakhidah, “Clustering Menggunakan K-Means Algorithm (K-Means Algorithm Clustering)”.