



Pemetaan Pola Produktivitas Tanaman Pangan Menggunakan Data Warehouse dan K-Means Clustering

Fernando Adi Saputra, Cahyono Budy Santoso*

Fakultas Teknologi dan Desain, Program Studi Sistem Informasi, Universitas Pembangunan Jaya, Tangerang Selatan, Indonesia

Email: ¹fernando.adisaputra@student.upj.ac.id.com, ^{2*}cahyono.budy@upj.ac.id.com

Email Penulis Korespondensi: cahyono.budy@upj.ac.id.com

Abstrak—Produktivitas tanaman pangan antar wilayah menunjukkan perbedaan yang signifikan meskipun memiliki luas areal tanam dan panen yang besar. Ketimpangan ini mencerminkan pola produksi yang belum terstruktur dengan baik, sehingga menyulitkan dalam analisis dan pengambilan keputusan berbasis data. Penelitian ini bertujuan untuk mengevaluasi produktivitas tanaman pangan dengan menggunakan *K-Means*, berdasarkan luas areal tanam dan luas panen dan produktivitas, dengan tujuan mengatasi ketidakaturan pola produksi antar wilayah serta menghasilkan pola pengelompokan yang lebih sistematis dan mudah dianalisis. Metode yang di gunakan yakni KDD (Knowledge Discovery in Database). Data diperoleh dari web resmi <https://jabar.bps.go.id/id> dan <https://opendata.jabarprov.go.id/>, periode 2019-2022. Penerapan data warehouse dengan star schema menghasilkan tabel fakta produktivitas, dan dimensi tahun, tanaman, dan wilayah. Evaluasi metrik menggunakan Davies-Bouldin Index menunjukkan informasi kualitas cluster yang baik. Penelitian ini menghasilkan penerapan data warehouse, clustering menggunakan algoritma K-Means dengan hasil skor Davies-Bouldin Index yaitu 0.419 dengan k optimal = 3, jumlah data pada klaster 0 yaitu 368 data, klaster 1 yaitu 52 data, dan klaster 2 yaitu 12 data. Untuk mempermudah interpretasi dan mendukung pengambilan keputusan berbasis data, business intelligence tools digunakan untuk menampilkan hasil clustering. Dari hasil penelitian ini membantu memberikan rekomendasi strategis dalam meningkatkan produktivitas pertanian tanaman pangan di Jawa Barat melalui intervensi spesifik wilayah.

Kata Kunci: Data Warehouse; Business Intelligence; Knowledge Discovery in Database; K-Means Clustering; Produktivitas Tanaman Pangan

Abstract—The productivity of food crops between regions shows significant differences despite having large areas of planted and harvested areas. These inequalities reflect production patterns that are not yet well structured, which makes it difficult to analyze and make data-driven decisions. This research aims to evaluate the productivity of food crops using *K-Means*, based on planted area and harvest area and productivity, with the aim of overcoming irregular production patterns between regions and producing grouping patterns that are more systematic and easy to analyze. The method used is KDD (Knowledge Discovery in Database). Data obtained from the official <https://jabar.bps.go.id/id> and <https://opendata.jabarprov.go.id/>, website for the 2019-2022 period. The application of data warehouses with star schema yields a Tabel of productivity facts, and dimensions of years, crops and regions. Metric evaluation using the Davies-Bouldin Index showed good cluster quality information. This research resulted in the application of data warehouse, clustering using the *K-Means* algorithm with a Davies-Bouldin Index score of 0.419 with optimal $k = 3$, the amount of data in cluster 0 is 368 data, cluster 1 is 52 data, and cluster 2 is 12 data. To simplify interpretation and support data-driven decision making, business intelligence tools are used to display clustering results. The results of this research help provide strategic recommendations in increasing agricultural productivity of food crops in West Java through region-specific interventions.

Keywords: Data Warehouse; Business Intelligence; Knowledge Discovery in Database; K-Means Clustering; Food Crop Productivity

1. PENDAHULUAN

Laju pertumbuhan penduduk di Provinsi Jawa Barat tidak sebanding dengan peningkatan hasil tanaman pangan, sehingga terjadi ketidakseimbangan yang bisa menyebabkan masalah pangan di beberapa daerah (Purba et al., 2023). Produktivitas tanaman pangan di Jawa Barat sangat penting untuk menjaga ketahanan pangan nasional dan membantu pertumbuhan ekonomi daerah (Prayitno, 2020). Meskipun cakupan areal tanam dan panen cukup besar, namun hal tersebut belum diikuti oleh distribusi produktivitas yang seimbang dengan signifikan antar wilayah dan jenis tanaman seperti padi, jagung, ubi kayu, dan kacang panjang. Menurut data dari BPS Jawa Barat antara 2019 dan 2022, daerah seperti Kabupaten Bandung dan Garut memiliki hasil pertanian yang cukup tinggi, sementara Kota Bekasi dan Kota Banjar memiliki hasil yang lebih rendah (Barat, 2024). Ketidaksihinggaan ini disebabkan oleh kurangnya integrasi data historis, sehingga analisis spasial berbasis data yang akurat, terstruktur, dan visualisasi diperlukan untuk menganalisis pola produktivitas tanaman pangan (Ghita & Trisminingsih, 2021). Ketersediaan data pertanian yang belum terintegrasi dalam satu sistem terstruktur menghalangi analisis pola produktivitas yang bergantung pada variabel produktivitas, luas areal tanam, dan luas panen. Menurut Hidayat & Fitriana (2022) pemetaan area produktif secara menyeluruh menjadi sulit karena tidak ada sistem penyimpanan data yang terintegrasi. Dengan menggunakan data-data yang dikumpulkan melalui <https://jabar.bps.go.id/id> dan <https://opendata.jabarprov.go.id/>, menyediakan data produktivitas untuk setiap jenis tanaman pangan, tetapi data tersebut belum digabungkan menjadi satu kumpulan yang mencakup semua jenis tanaman pangan (Kurniawan Hidayat & Fitriana, 2022).

Data tentang produktivitas belum terintegrasi, tidak ada pengelompokan informasi mengenai luas tanam dan produktivitas berdasarkan daerah, serta belum ada visualisasi yang menunjukkan perbandingan antara luas area tanam dan produktivitas untuk tanaman pangan. Tujuan penelitian ini adalah untuk membuat sistem penyimpanan data yang terintegrasi, mengelompokkan wilayah dan jenis tanaman berdasarkan produktivitas, luas area tanam, dan luas panen dengan data warehouse, *K-Means* serta membuat visualisasi data dengan business intelligence tools. Metode



pemrosesan data mining yang digunakan adalah K-Means clustering karena dapat dengan cepat dan efisien mengelompokkan data dalam jumlah besar (Aldi et al., 2025). Data-data yang telah diolah dan dikelompokkan dengan K-Means, selanjutnya data tersebut di implementasikan menjadi informasi yang mudah dianalisis dengan visualisasi data (Virantika et al., 2022). Dengan business intelligence tools data yang sudah siap dapat divisualisasikan untuk menganalisis pola produktivitas tanaman pangan. Manfaat dari visualisasi ini dapat membantu dan memberikan informasi yang inovatif, menarik, dan informatif sehingga dapat membantu dalam pengambilan sebuah keputusan (Marvaro & Sefina Samosir, 2021).

Pada penelitian sebelumnya yang berkaitan dengan K-means clustering produktivitas ubinan tanaman pangan yang dilakukan oleh Cici Armayani, Achmad Fauzi, dan Hermansyah Sembiring. Masalah pada penelitian tersebut adalah melakukan analisis penggunaan data mining untuk mengelompokkan jumlah data produktivitas ubinan tanaman pangan berdasarkan jenis ubinan di kabupaten langkat. Menunjukkan bahwa data dapat dikelompokkan menjadi tiga cluster utama yang menggambarkan pola produktivitas tanaman pangan di berbagai kecamatan sehingga membantu Badan Pusat Statistik membuat keputusan berbasis data merupakan hasil penelitian tersebut (Armayani et al., 2021). Namun, penelitian tersebut belum mengintegrasikan data dalam bentuk data warehouse serta belum memanfaatkan visualisasi berbasis Business Intelligence untuk mendukung analisis yang lebih komprehensif.

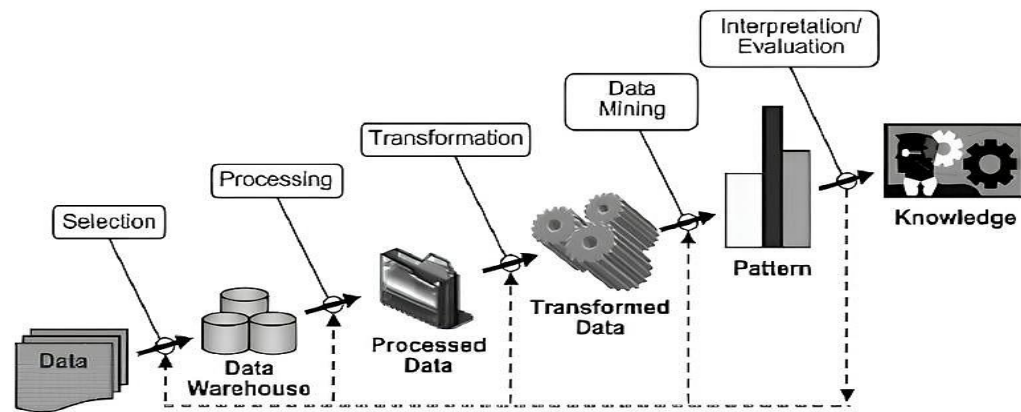
Penelitian sebelumnya yang serupa adalah pengelompokan produktivitas tanaman padi yang dilakukan oleh Wijayanto dan Fathoni (2021). Penelitian ini bertujuan mengelompokkan wilayah di Provinsi Jawa Tengah berdasarkan tingkat produktivitas padi menggunakan algoritma K-Means. Hasil penelitian menunjukkan terbentuk tiga kelompok utama, yaitu tingkat produktivitas tinggi, sedang, dan rendah, yang dapat digunakan sebagai dasar pengambilan keputusan dalam meningkatkan hasil pertanian (Wijayanto & Yoka Fathoni, 2021). Namun, penelitian tersebut hanya menggunakan satu metode clustering tanpa perbandingan dengan metode lain, serta belum memanfaatkan data warehouse sebagai sumber data terintegrasi. Penelitian serupa yakni clustering kelompok sektor perkebunan di Indonesia yang dilakukan oleh Hesti Pratiwi dan Ari Purno Wahyu Wibowo. Masalah pada penelitian ini adalah belum optimalnya pengolahan data perkebunan dalam mengidentifikasi pola produktivitas wilayah secara akurat. Hasil dari penelitian menunjukkan bahwa optimasi K-Means menggunakan PSO mampu meningkatkan kualitas pengelompokan data produksi tanaman sayuran (Pratiwi et al., 2022). Namun, penelitian ini belum menggabungkan pendekatan clustering dengan visualisasi data berbasis Business Intelligence serta belum mengintegrasikan data secara terstruktur menggunakan data warehouse.

Penelitian serupa membahas clustering untuk pemantauan produksi tanaman obat yang dilakukan oleh Miwan Kurniawan Hidayat dan Rina Fitriana. Masalah pada penelitian ini adalah melakukan analisis tentang penggunaan data mining dan sistem intelijen bisnis untuk memetakan hasil produksi tanaman obat di Provinsi Jawa Barat menjadi kelompok dengan metode penyelesaian data warehouse dengan skema bintang dan K-Means Clustering. Bisnis intelijensia berguna untuk menampilkan dashboard interaktif dan membantu dalam pengambilan keputusan strategis untuk meningkatkan produksi dan distribusi tanaman obat (Kurniawan Hidayat & Fitriana, 2022). Namun pada keempat penelitian tersebut belum ada yang merujuk untuk penerapan data warehouse, K-Means, dan visualisasi mengenai luas areal tanam, luas panen, dan produktivitas yang berguna untuk dianalisis sebagai wawasan lebih komprehensif mengenai perbandingan luas areal tanam dan luas panen dengan produktivitas yang dihasilkan.

Sehingga pada penelitian ini bertujuan untuk mengintegrasikan data warehouse dan k-means clustering untuk menganalisis luas areal tanam, luas panen, dan produktivitas berbasis data. Dengan penerapan data warehouse dapat membantu dalam penyimpanan data yang lebih terintegrasi dan terstruktur (Komul, 2025). Data yang telah divisualisasikan dapat membantu para instansi pertanian atau pemerintah daerah dalam untuk dianalisis agar dapat untuk merancang kebijakan intervensi yang lebih terarah dan efisien.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan data dari <https://jabar.bps.go.id/id> dan <https://opendata.jabarprov.go.id/>. Metode yang digunakan yaitu metode KDD (Knowledge Discovery in Database) yang mencakup lima tahapan : selection, preprocessing, transformation, data mining, interpretation/evaluation. Pendekatan Knowledge Discovery in Database (KDD) merupakan suatu proses sistematis yang digunakan untuk mengekstraksi pengetahuan dari data dalam jumlah besar melalui tahapan yang terstruktur. Proses ini memungkinkan identifikasi pola yang valid, baru, dan bermanfaat untuk mendukung pengambilan keputusan berbasis data (Nafi et al., 2024). Tahapan proses Knowledge Discovery in Database (KDD) yang digunakan dalam penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Proses KDD (kajianpustaka.com, 2021)

Pada tahap selection, dilakukan pemilihan data yang relevan, yaitu luas areal tanam, luas panen, dan hasil produksi. Tahap preprocessing dilakukan untuk meningkatkan kualitas data melalui pembersihan data, termasuk penanganan missing value menggunakan metode drop serta penghapusan data duplikat. Selanjutnya, tahap transformation dilakukan dengan mengintegrasikan data ke dalam data warehouse menggunakan skema bintang (star schema) agar data lebih terstruktur. Pada tahap data mining, digunakan algoritma K-Means clustering untuk mengelompokkan data berdasarkan karakteristik produktivitas, sehingga dapat mengatasi ketidakteraturan pola produksi (Mauliadi, 2022). Tahap terakhir adalah evaluation, yang dilakukan menggunakan Davies-Bouldin Index (DBI) untuk menilai kualitas cluster yang dihasilkan. Pendekatan ini memungkinkan data yang semula tidak terstruktur menjadi informasi yang lebih sistematis dan mendukung pemetaan produktivitas secara lebih jelas.

2.1 Selection

Tahapan yang pertama adalah selection digunakan sebagai proses pemilihan data. Dataset yang digunakan berasal dari data sekunder melalui situs web resmi dari <https://jabar.bps.go.id/id> dan <https://opendata.jabarprov.go.id/>. Pemilihan data ini bertujuan untuk memastikan bahwa hanya atribut yang memiliki keterkaitan dengan permasalahan yang akan dianalisis yang digunakan, sehingga dapat meningkatkan efektivitas proses data mining dan menghasilkan informasi yang lebih akurat (Fauzi et al., 2025). Data yang digunakan adalah data tanaman pangan mengenai luas areal tanam, luas panen dan produktivitas di daerah kabupaten kota Jawa Barat dari tahun 2019-2022. Isi dataset terdiri dari 432 dengan 11 atribut. Atribut tersebut terdiri dari : id, kode_provinsi, kode_kabupaten_kota, jenis_tanaman, luas_areal_tanam, luas_panen, satuan, jumlah_produkivitas, satuan2 dan tahun. Lalu penghapusan atribut dataset yang tidak diperlukan adalah kode_provinsi, kode_kabupaten_kota, nama_provinsi, dapat dilihat pada Tabel 1. Lalu atribut-atribut tersebut di seleksi untuk membuang atribut yang tidak perlu sehingga menjadi 9 atribut yang terdiri dari : id, nama_kabupaten_kota, jenis_tanaman, luas_areal_tanam, luas_panen, satuan, jumlah_produkivitas, satuan, dan tahun. Hasil dari seleksi ditampilkan pada Tabel 2.

Tabel 1. Dataset Awal

id	kode_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	Jenis_Tanaman	luas_areal_tanam	luas_panen	Satuan	Jumlah_produkivitas	satuan2	tahun
1	32	3201	KABUPATEN BOGOR	Ubi Kayu	3584	4011	He kta r	297.09	KUINTAL PER HEKTAR	2019
2	32	3202	KABUPATEN SUKABUMI	Ubi Kayu	3727	4554	He kta r	294.79	KUINTAL PER HEKTAR	2019
3	32	3203	KABUPATEN CIANJUR	Ubi Kayu	3594	3513	He kta r	254.35	KUINTAL PER HEKTAR	2019
4	32	3204	KABUPATEN BANDUNG	Ubi Kayu	2995	2300	He kta r	246.38	KUINTAL PER HEKTAR	2019

Tabel 2. Hasil Akhir Seleksi Data

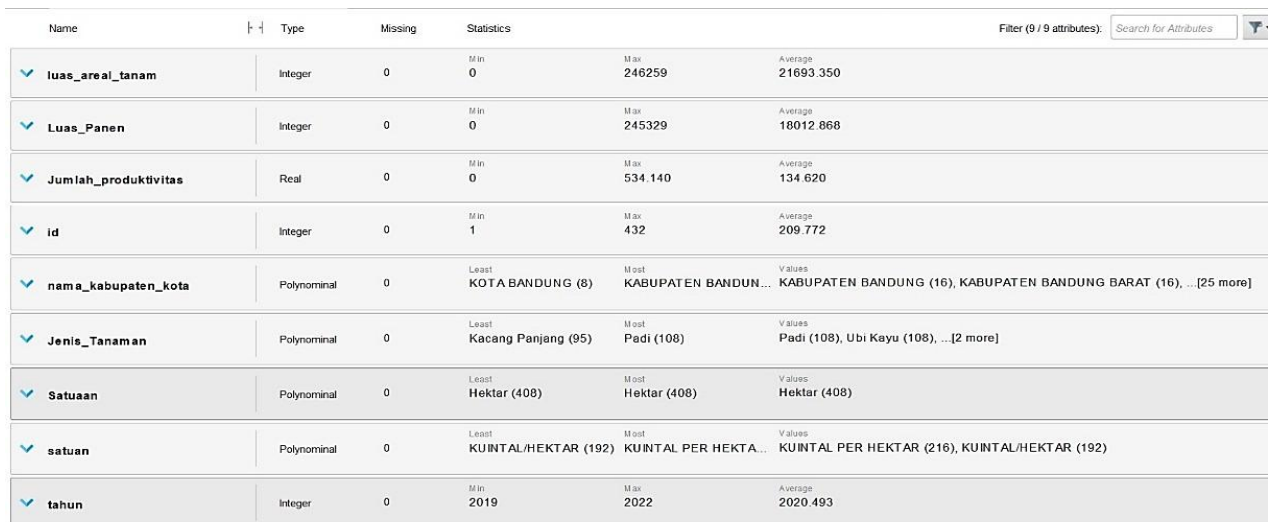
id	nama_kabupaten_kota	Jenis_Tanaman	luas_areal_tanam	Luas_Panen	Satuan	Jumlah_produkivitas	satuan	tahun
1	KABUPATEN	Ubi Kayu	3584	4011	Hekt	297.09	KUINTAL PER	201

id	nama_kabupaten_kota	Jenis_Tanaman	luas_areal_tanam	Luas_Panen	Satuan	Jumlah_produk_tivitas	satuan	tahun
2	BOGOR KABUPATEN SUKABUMI	Ubi Kayu	3727	4554	ar Hektar	294.79	HEKTAR KUINTAL PER HEKTAR	9 2019
3	KABUPATEN CIANJUR	Ubi Kayu	3594	3513	Hektar	254.35	KUINTAL PER HEKTAR	2019
4	KABUPATEN BANDUNG	Ubi Kayu	2995	2300	Hektar	246.38	KUINTAL PER HEKTAR	2019

Berdasarkan Tabel 1 memiliki banyak atribut yang tidak penting untuk melakukan proses data mining maka dari data tersebut atribut-atribut yang tidak penting di seleksi dan dihapus untuk menemukan atribut yang sesuai dengan kebutuhan dalam data mining. Pada Tabel 2 merupakan hasil dari seleksi atribut yang tidak dibutuhkan dalam proses data mining dalam pemrosesan data mining terdapat tiga variabel utama untuk di clustering, yaitu luas areal tanam, luas panen, dan produktivitas. Pemilihan atribut yang relevan pada tahap selection ini bertujuan untuk memastikan bahwa proses analisis dapat berjalan secara optimal serta menghasilkan pengelompokan data yang lebih akurat.

2.2 Preprocessing Data

Tahap kedua yaitu preprocessing data, pada tahap ini data yang sudah dilakukan seleksi selanjutnya akan di proses untuk menemukan dan menghilangkan data missing value. Pada tahap preprocessing, dilakukan proses pembersihan data untuk memastikan kualitas dataset yang digunakan dalam analisis. Penanganan missing value dilakukan dengan teknik drop, yaitu menghapus baris data yang memiliki nilai kosong pada atribut yang digunakan. Pemilihan metode ini bertujuan untuk menghindari distorsi hasil analisis yang dapat disebabkan oleh data yang tidak lengkap. Preprocessing Data diperlukan untuk pembersihan data sehingga data dapat diolah menjadi dataset yang dapat diproses (Agung et al., 2023). Hasil proses dapat dilihat pada Gambar 2.



Name	Type	Missing	Statistics	Filter (9 / 9 attributes):
luas_areal_tanam	Integer	0	M/in 0, Max 246259, Average 21693.350	Search for Attributes
Luas_Panen	Integer	0	M/in 0, Max 245329, Average 18012.868	
Jumlah_produk_tivitas	Real	0	M/in 0, Max 534.140, Average 134.620	
id	Integer	0	M/in 1, Max 432, Average 209.772	
nama_kabupaten_kota	Polynomial	0	Least KOTA BANDUNG (8), Most KABUPATEN BANDUN...	Values KABUPATEN BANDUNG (16), KABUPATEN BANDUNG BARAT (16), ... [25 more]
Jenis_Tanaman	Polynomial	0	Least Kacang Panjang (95), Most Padi (108)	Values Padi (108), Ubi Kayu (108), ... [2 more]
Satuan	Polynomial	0	Least Hektar (408), Most Hektar (408)	Values Hektar (408)
satuan	Polynomial	0	Least KUINTAL/HEKTAR (192), Most KUINTAL PER HEKTA...	Values KUINTAL PER HEKTAR (216), KUINTAL/HEKTAR (192)
tahun	Integer	0	M/in 2019, Max 2022, Average 2020.493	

Gambar 2. Hasil *Preprocessing Missing Value*

Berdasarkan Gambar 2, dataset yang digunakan memuat beberapa atribut penting, antara lain luas areal tanam, luas panen, jumlah produktivitas, identitas wilayah, jenis tanaman, satuan, serta tahun. Atribut-atribut tersebut terdiri dari tipe data numerik dan kategorikal, yang menunjukkan bahwa data memiliki karakteristik yang beragam dan dapat dianalisis dari berbagai aspek. Seluruh atribut pada dataset tidak menunjukkan adanya nilai kosong, sehingga dapat disimpulkan bahwa data telah memenuhi kelengkapan yang diperlukan untuk proses analisis lebih lanjut. Selain itu, informasi statistik seperti nilai minimum, maksimum, dan rata-rata pada atribut numerik memperlihatkan adanya perbedaan nilai antar data, yang mengindikasikan variasi produktivitas pada setiap wilayah dan jenis tanaman. Variasi tersebut menjadi dasar penting dalam proses pengelompokan data, karena memungkinkan identifikasi pola produktivitas yang berbeda. Dengan kondisi data yang lengkap dan bervariasi, dataset ini dinilai telah siap untuk digunakan pada tahap data mining menggunakan metode K-Means.

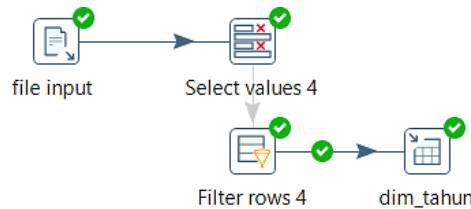
2.3 Transformation Data

Pada tahap ketiga ini yaitu transformation data berfokus untuk pembentukan data warehouse dengan star schema yang terdiri dari 1 tabel fakta dan 3 tabel dimensi dan transformasi ETL dengan tools pentahoo. Penggunaan tools pentahoo dalam transformation data memungkinkan integrasi data dari berbagai sumber sekaligus melakukan pembersihan dan standarisasi sehingga kualitas data menjadi lebih baik (Saputra et al., 2025). Pembangunan data warehouse ditampilkan pada Gambar 3.



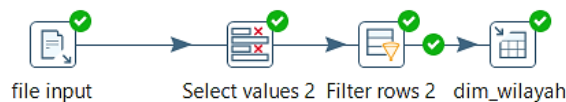
Gambar 3. Data Warehouse Star Schema

Setelah pembangunan data warehouse selanjutnya melakukan proses ETL pada tabel fakta dan dimensi. Proses ETL pada Tabel dim_tahun diawali dengan membaca file data mentah yang berisikan data tanaman pangan, lalu menambahkan select values, penambahan filter rows dan menambahkan penulisan pada tabel dimensi_tahun. Alur Proses ETL dimensi_tahun ditampilkan dengan Gambar 4.



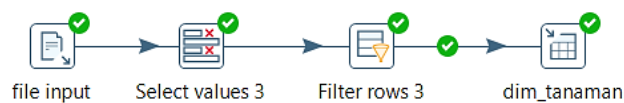
Gambar 4. Proses ETL dim_tahun

Proses ETL pada Tabel dim_wilayah diawali dengan membaca file data mentah yang berisikan data tanaman pangan, lalu menambahkan select values, penambahan filter rows dan menambahkan penulisan pada tabel dim_wilayah. Alur proses ETL dim_wilayah ditampilkan dengan Gambar 5.



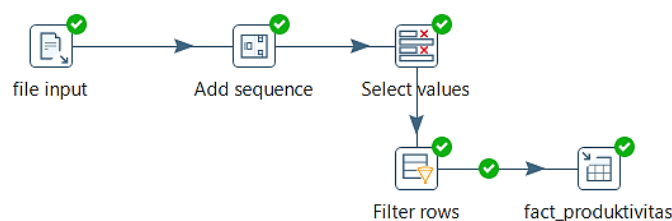
Gambar 5. Proses ETL dim_wilayah

Proses ETL dim_tanaman diawali dengan membaca file data mentah yang berisikan data tanaman pangan, lalu menambahkan select values, penambahan filter rows dan menambahkan penulisan pada tabel dim_tanaman. Alur Proses ETL dim_tanaman ditampilkan pada Gambar 6.



Gambar 6. Proses ETL dim_tanaman

Proses ETL fakta_produkktivitas diawali dengan membaca file data mentah yang berisikan data tanaman pangan, lalu selanjutnya menambahkan add sequence, select values, penambahan filter rows dan menambahkan penulisan pada tabel fakta_produkktivitas. Alur Proses ETL fakta_produkktivitas ditampilkan pada Gambar 7.



Gambar 7. Proses ETL fakta_produkktivitas

Proses ETL bermanfaat sebagai kebersihan data dan intergritas data sehingga dapat menjamin kualitas data (Ghita & Trisminingsih, 2021).

2.4 Data Mining

Data Mining yang digunakan dalam penelitian ini yaitu K-Means Clustering dengan menggunakan software Rapid Miner. Hasil jumlah cluster yang terbaik berdasarkan nilai DBI. Nilai skor yang dihasilkan pada DBI semakin kecil maka cluster yang dihasilkan juga baik (Kurniawan Hidayat & Fitriana, 2022). Pencarian nilai cluster terbaik dilakukan dengan cara bertahap untuk menentukan hasil nilai DBI yang maksimal. Hasil skor DBI ditampilkan pada Tabel 3.

Tabel 3. *Davies Bouldin Index*

Jumlah Cluster	Davies Bouldin Index
2	0.447
3	0.419
4	0.557

Cluster yang paling optimal dengan nilai skor paling kecil yaitu ditunjukkan pada jumlah cluster 3 dengan nilai 0.419. Dalam penelitian ini software yang digunakan dalam melakukan data mining menggunakan Rapid Miner. Semakin kecil hasil skor yang DBI maka hasil clustering semakin valid dan optimal (Yufita et al., 2025). Keunggulan dari Rapid Miner yaitu efektif untuk data mining data yang banyak dalam waktu cepat lalu tools yang disediakan juga begitu banyak (Hafizah et al., 2024).

2.5 Interpretation/Evaluation

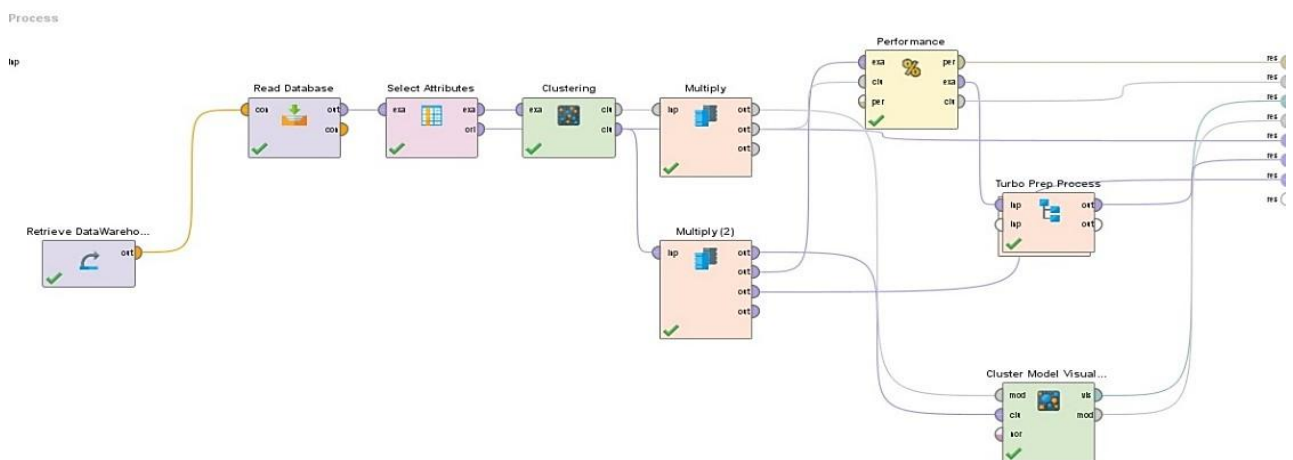
Tahap selanjutnya yaitu Interpretation/Evaluation pada tahap ini membuat dashboard visualisasi dari hasil data-data yang telah di proses data warehouse dan K-Means clustering. Pembuatan dashboard visualisasi menggunakan tools Business Intelligence (BI) yaitu Power BI. Manfaat dashboard visualisasi data dapat membantu dalam pengambilan keputusan yang lebih tepat, mampu memberikan gambaran informasi yang spasial dan temporal (Heru Stiawan et al., 2022).

3. HASIL DAN PEMBAHASAN

Analisis produktivitas tanaman menggunakan algoritma K-Means dan penerapan data warehouse akan menghasilkan data yang baik dan akurat. Dengan penerapan K-Means dan Data warehouse dalam analisis produktivitas tanaman pangan, dapat memberikan sebuah informasi mengenai produktivitas tanaman pangan dalam pendukung pengambilan sebuah keputusan. Dengan memberikan sebuah informasi yang lengkap mengenai informasi wilayah produktivitas, perbandingan antara luas areal tanam, luas panen dan produktivitas, dapat membantu dalam pengambilan keputusan yang tepat dan cepat. Data set yang udah diolah melalui data warehouse dan proses ETL data disimpan kedalam database PostgreSQL agar data data dapat terintergrasi.

3.1 Data Mining

Dalam pemrosesan clustering software yang digunakan yaitu Rapid Miner. Proses menggunakan beberapa operators, yaitu Read Database, Select Attributes, Clustering (K-Means), Multiply, Performance, Cluster Model Visualizer. Tampilan Rangkain Operator ditampilkan pada Gambar 8.



Gambar 8. Skema Data Mining

Jumlah cluster yang telah ditentukan dari hasil skor DBI terkecil yaitu 3 cluster, dengan masing masing cluster 0 yang berisikan 368 data, cluster 1 sebanyak 52 data dan cluster 2 sebanyak 12 data. Visualisasi hasil dari jumlah cluster pada RapidMiner Studio ditampilkan pada Gambar 9.

Number of Clusters: 3

Cluster 0

368

luas_areal_tanam is on average 83.33% smaller, luas_panen is on average 83.00% smaller, jumlah_produkktivitas is on average 9.13% larger

Cluster 1

52

luas_areal_tanam is on average 383.99% larger, luas_panen is on average 344.57% larger, jumlah_produkktivitas is on average 52.97% smaller

Cluster 2

12

luas_panen is on average 1,052.30% larger, luas_areal_tanam is on average 891.66% larger, jumlah_produkktivitas is on average 50.59% smaller

Gambar 9. Visualisasi Jumlah Cluster

Hasil clustering dihitung dengan average pada setiap variable nya. Hasil clustering akhir ditampilkan pada Tabel 4 berikut.

Tabel 4. Hasil Clustering

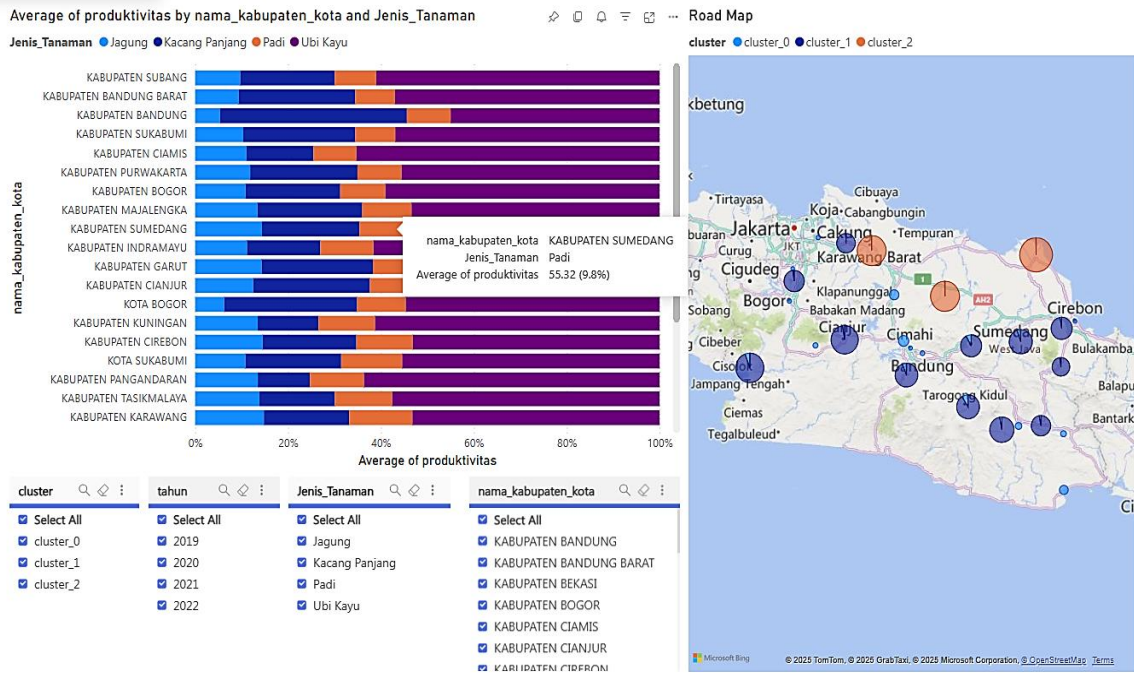
cluster	nama_kabupaten_kota	Jenis_Tanaman	luas_areal tanam	luas_panen	produktivitas
cluster_0	KABUPATEN BOGOR	Ubi Kayu	3584	4011	297.09
cluster_0	KABUPATEN SUKABUMI	Ubi Kayu	3727	4554	294.79
cluster_0	KOTA DEPOK	Padi	102	84	63.3
cluster_0	KOTA CIMAHI	Padi	268	70	71.13
cluster_1	KABUPATEN GARUT	Jagung	78667	77543	73.39
cluster_1	KABUPATEN GARUT	Jagung	63048	60444	84.54
cluster_1	KABUPATEN CIREBON	Padi	84108	87700	70.21
cluster_1	KABUPATEN BOGOR	Padi	82470	54987	68.48
cluster_2	KABUPATEN SUBANG	Padi	183030	169416	56.39
cluster_2	KABUPATEN KARAWANG	Padi	182349	181915	58.37
cluster_2	KABUPATEN INDRAMAYU	Padi	246259	227051	58.12

Berdasarkan pengolahan data dengan metode K-Means, diperoleh hasil pengelompokan yang menggambarkan karakteristik masing-masing cluster yaitu;

1. Cluster 0 berisi luas areal tanam dan luas panen tanaman padi dan ubi kayu di wilayah kabupaten Bogor, kabupaten Sukabumi, kota Depok, kota Cimahi kecil namun produktivitas nya tinggi.
2. Cluster 1 berisi luas areal tanam dan luas panen tanaman padi dan jagung di wilayah kabupaten Garut, kabupaten Cirebon, kabupaten Bogor sedang namun produktivitasnya sedang.
3. Cluster 2 berisi luas areal tanam dan luas panen tanaman padi di wilayah kabupaten Karawang dan kabupaten Indramayu besar namun produktivitasnya rendah.

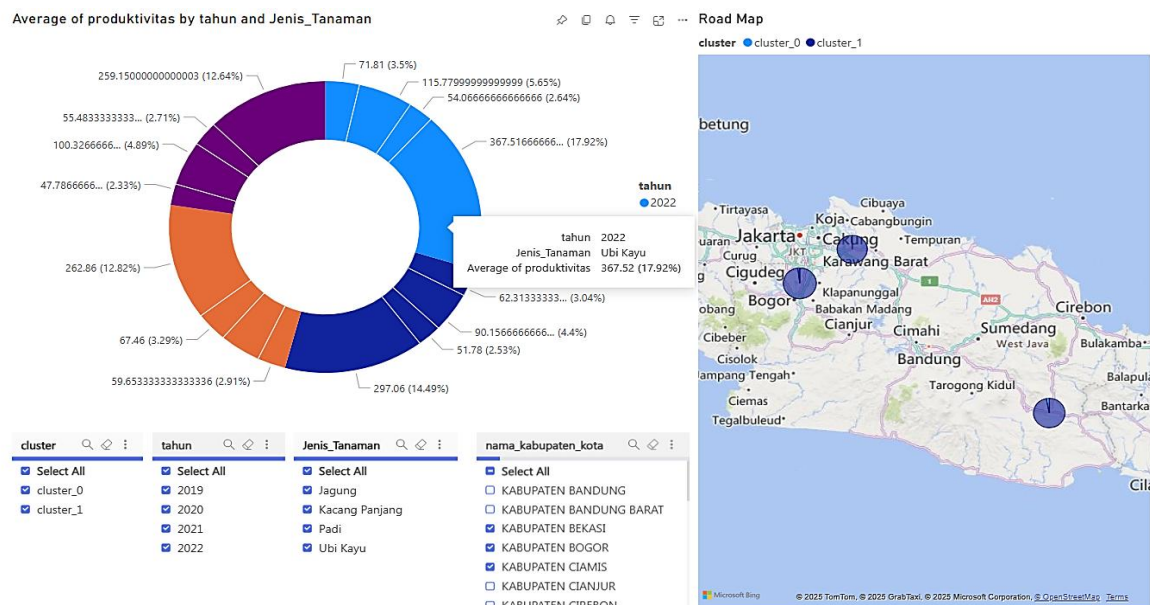
Dashboard visual dirancang untuk membantu kebutuhan, data hasil akhir di implementasikan pada Power BI. Visual yang disajikan yaitu dashboard dengan informasi-informasi mengenai average produktivitas berdasarkan nama kabupaten dan jenis tanaman dalam bentuk stacked bar chart, informasi average produktivitas berdasarkan tahun dan jenis tanaman dalam bentuk donut bar chart, informasi average luas areal tanam dan produktivitas berdasarkan jenis tanaman, cluster dan tahun dalam bentuk scatter chart, informasi data Tabel mengenai cluster, nama kabupaten dan kota, jenis tanaman, luas areal tanam, luas panen, produktivitas dan tahun, informasi average luas area tanam dan produktivitas berdasarkan nama kabupaten kota dalam bentuk bullet chart. Pada halaman dashboard dilengkapi dengan informasi seperti peta wilayah dan fitur filter.

Visual stacked bar chart ditampilkan pada Gambar 10. Grafik ini menampilkan perbandingan antar kategori dengan membagi dan menampilkan informasi data dalam bentuk batang untuk memudahkan pemahaman dan analisis data. Grafik batang berisikan mengenai nama kabupaten kota dan jenis tanaman. Visual donut bar chart ditampilkan dengan Gambar 11 menunjukkan grafik dengan tampilan donut mengenai perbandingan data dengan ukuran relatif berbagai komponen mencerminkan kepentingan relatifnya. Donut bar chart digunakan untuk menampilkan presentase average produktivitas berdasarkan jenis tanaman dan tahun. Visual Scatter chart ditampilkan dengan Gambar 12 menunjukkan grafik dengan tampilan titik-titik pada bidang koordinat mengenai persebaran data dengan titik-titik koordinat. Scatter chart digunakan untuk menampilkan average luas areal tanam dan produktivitas berdasarkan jenis tanaman dan tahun.



Gambar 10. Dashboard Stacked Bar Chart

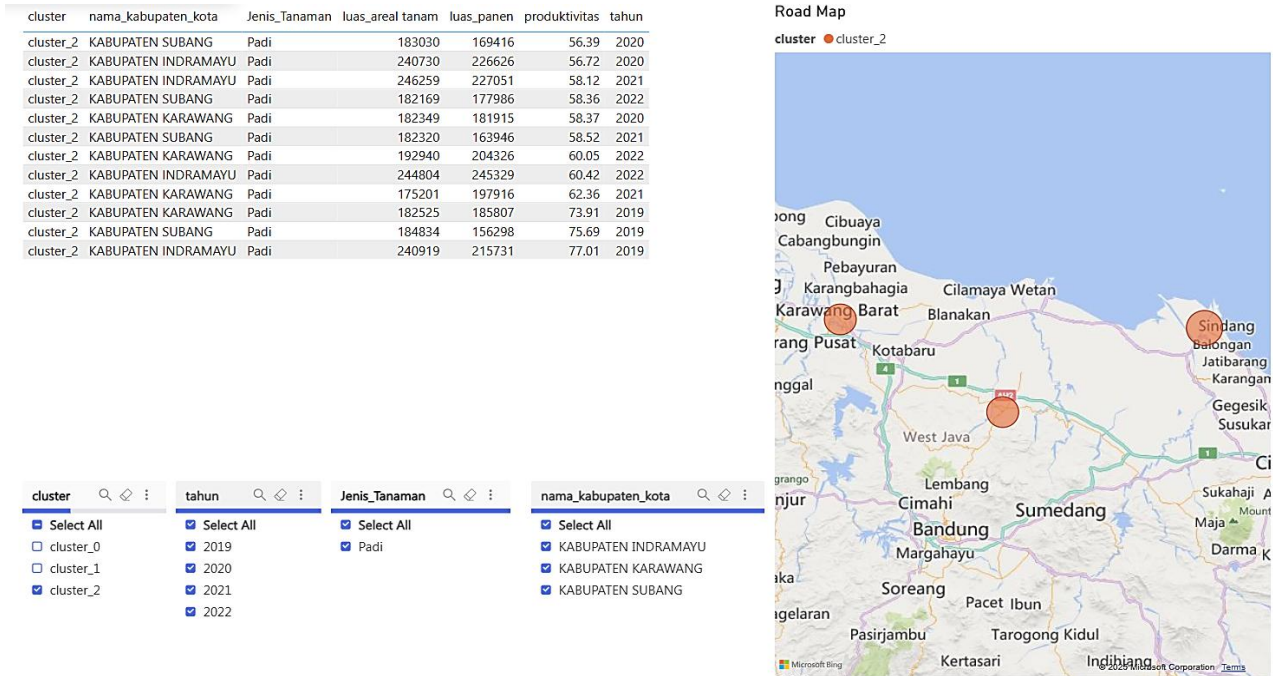
Berdasarkan Gambar 10, terlihat bahwa rata-rata produktivitas tanaman pangan berbeda pada setiap wilayah dan jenis tanaman. Komoditas seperti padi dan ubi kayu cenderung mendominasi kontribusi produktivitas di sebagian besar wilayah, sedangkan komoditas lain seperti jagung dan kacang panjang memiliki proporsi yang lebih kecil. Hal ini menunjukkan adanya variasi pola produksi antar daerah. Selain itu, hasil clustering yang ditampilkan pada peta menunjukkan bahwa wilayah terbagi ke dalam beberapa kelompok dengan karakteristik produktivitas yang berbeda. Beberapa wilayah berada pada cluster dengan tingkat produktivitas lebih tinggi, sementara wilayah lainnya berada pada kelompok dengan produktivitas yang lebih rendah. Persebaran ini mengindikasikan bahwa tidak semua wilayah memiliki tingkat produksi yang merata.



Gambar 11. Dashboard Donut Bar Chart

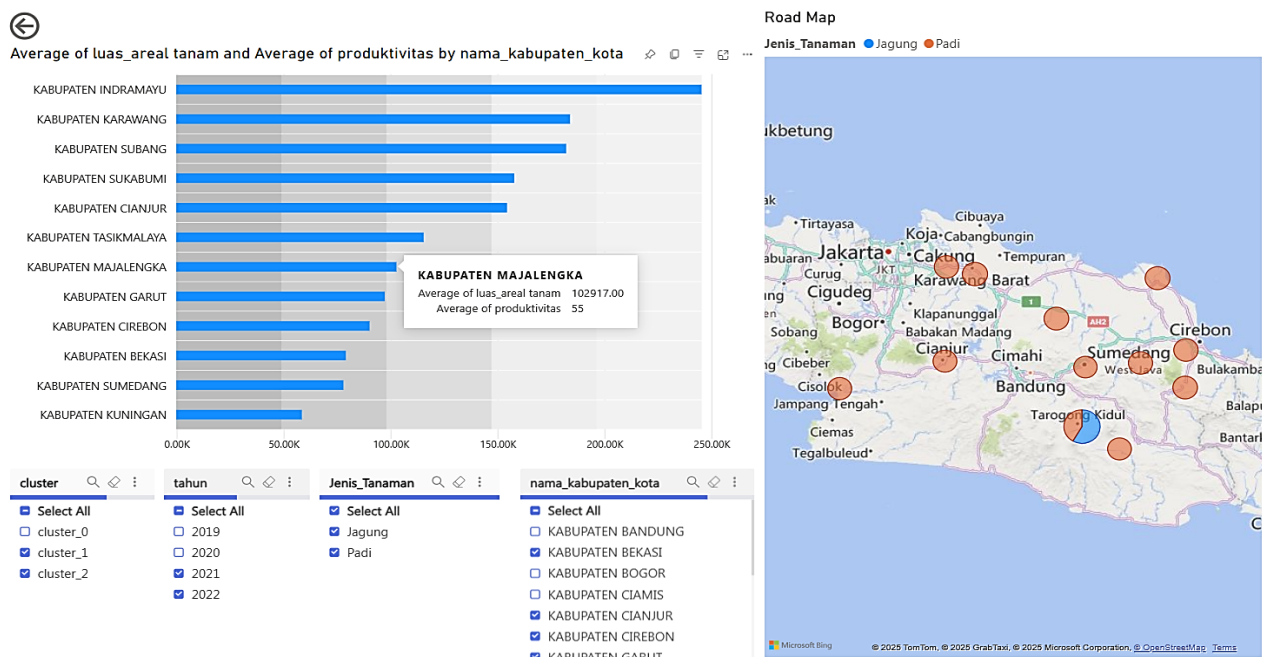
Berdasarkan Gambar 11, terlihat bahwa rata-rata produktivitas tanaman pangan mengalami variasi pada setiap jenis tanaman dan tahun pengamatan. Komoditas seperti ubi kayu dan padi menunjukkan kontribusi produktivitas yang relatif lebih besar dibandingkan komoditas lainnya, sedangkan jagung dan kacang panjang memiliki proporsi yang lebih kecil. Hal ini mengindikasikan adanya perbedaan tingkat produksi antar jenis tanaman dari waktu ke waktu. Selain itu, distribusi data pada diagram menunjukkan bahwa tidak semua jenis tanaman memiliki tingkat produktivitas yang stabil setiap tahunnya. Fluktuasi nilai rata-rata produktivitas mencerminkan adanya dinamika produksi yang dipengaruhi oleh berbagai faktor, seperti kondisi wilayah dan karakteristik komoditas.

Visual Tabel dashboard pada Gambar 12 menunjukkan data dengan bentuk Tabel yang lengkap dipadu dengan road map dan filter untuk mempermudah pembacaan. Tabel dashboard digunakan untuk memberikan tampilan data mengenai cluster, nama kabupaten kota, jenis tanaman, luas areal tanam, luas panen, produktivitas dan tah Visual bullet chart pada Gambar 13 menunjukkan data dengan tampilan batang horizontal (bar) dengan elemen visual tambahan mengenai perbandingan dengan optimal. Bullet chart digunakan untuk memberikan tampilan average luas areal tanam dan produktivitas berdasarkan nama kabupaten kota.



Gambar 12. Dashboard Tabel

Berdasarkan Gambar 12, terlihat bahwa data yang termasuk dalam cluster_2 didominasi oleh komoditas padi dengan nilai luas areal tanam, luas panen, dan produktivitas yang relatif tinggi. Wilayah seperti Kabupaten Indramayu, Karawang, dan Subang muncul secara konsisten dalam cluster ini, yang menunjukkan bahwa daerah tersebut memiliki kontribusi produksi yang cukup besar dibandingkan wilayah lainnya. Nilai produktivitas yang ditampilkan juga menunjukkan variasi antar tahun, namun secara umum berada pada rentang yang lebih tinggi dibandingkan cluster lainnya. Hal ini mengindikasikan bahwa cluster_2 dapat dikategorikan sebagai kelompok wilayah dengan tingkat produktivitas tinggi.



Gambar 13. Dashboard Bullet Chart



Berdasarkan Gambar 13, terdapat halaman dashboard yang terdiri dari bullet chart, filter dan map. Pada bagian bullet chart terdapat daftar wilayah di daerah Jawa Barat salah satunya kabupaten Majalengka pada gambar tersebut terdapat average luas areal tanam dan average produktivitas. Filter terdapat cluster, tahun, jenis tanaman, dan nama kabupaten/kota. Map menunjukkan hasil data yang sesuai dengan filter yang diaktifkan.

4. KESIMPULAN

Penelitian ini berhasil mengintegrasikan konsep data warehouse dengan metode clustering menggunakan algoritma K-Means untuk menganalisis produktivitas tanaman pangan di Jawa Barat secara lebih terstruktur dan komprehensif. Melalui pendekatan Knowledge Discovery in Database (KDD), data yang sebelumnya tersebar dan tidak terintegrasi berhasil diolah melalui tahapan seleksi, preprocessing, hingga transformasi menggunakan Pentaho Data Integration sehingga menghasilkan data yang bersih, konsisten, dan siap dianalisis. Hasil pengujian menggunakan Davies-Bouldin Index menunjukkan bahwa jumlah cluster optimal adalah tiga dengan nilai sebesar 0,419 yang menandakan kualitas pengelompokan yang baik. Setiap cluster mampu merepresentasikan karakteristik produktivitas yang berbeda berdasarkan luas areal tanam, luas panen, dan hasil produksi, sehingga memberikan gambaran yang jelas mengenai kondisi produktivitas di setiap wilayah. Selain itu, pemanfaatan visualisasi melalui tools business intelligence seperti Microsoft Power BI membantu menyajikan informasi secara lebih interaktif dan mudah dipahami, sehingga dapat mendukung pengambilan keputusan berbasis data. Meskipun demikian, penelitian ini masih memiliki keterbatasan pada penggunaan variabel yang terbatas dan belum mempertimbangkan faktor eksternal seperti kondisi iklim atau kualitas tanah. Oleh karena itu, penelitian selanjutnya disarankan untuk menambahkan variabel yang lebih beragam serta mengombinasikan metode clustering lain agar menghasilkan analisis yang lebih akurat dan mendalam dalam mendukung kebijakan peningkatan produktivitas pertanian.

REFERENCES

- Agung, A., Daniswara, A., & Nuryana, I. K. D. (2023). *Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru*. 05, 97–100. <https://doi.org/10.26740/jinacs.v5n01.p97-100>
- Aldi, M. A., Fatah, Z., Informasi, T., Ibrahimy, U., Timur, S. J., Informasi, S., Ibrahimy, U., & Timur, S. J. (2025). *Implementasi K-Means Cluster Ing Dalam Pengelompokan Data Kunjungan*. 2(1), 13–19. <https://doi.org/10.69714/3hhfj353>
- Armayani, C., Fauzi, A., & Sembiring, H. (2021). Implementasi Data Mining Pengelompokan Jumlah Data Produktivitas Ubinan Tanaman Pangan Berdasarkan Jenis Ubinan Dengan Metode Clustering Dikab Langkat (Studi Kasus: Badan Pusat Statistik Langkat). *Jurnal Informatika Kaputama (JIK)*, 5(1), 185–196. <https://doi.org/10.59697/jik.v5i1.318>
- Barat, B. P. S. P. J. (2024). *Potensi Pertanian Provinsi Jawa Barat: Menilik Kekuatan Komoditas Padi, Kopi, dan Domba*. <https://share.google/mjR9ZM9dosbFuT772>
- Fauzi, R. A., Rohana, T., Nurlaelasari, E., & Wahiddin, D. (2025). *Journal of Computer Networks , Architecture and High Performance Computing Bitcoin Price Prediction Using Long Short Term Memory Journal of Computer Networks , Architecture and High Performance Computing*. 7(3), 659–671. <https://doi.org/10.47709/cnahpc.v7i3.5945>
- Ghita, F., & Trisminingsih, R. (2021). Pengujian Data Warehouse SOLAP untuk Komoditas Pertanian Indonesia. *Jurnal Ilmu Komputer Dan Agri-Informatika*, 8(1), 42–56. <https://doi.org/10.29244/jika.8.1.42-56>
- Hafizah, A. S., Zulham, & Eka, M. (2024). *Analisis Menggunakan Algoritma Fp-Growth Untuk Mencari*. 18, 1386–1400. <https://doi.org/10.46576/wdw.v18i4.5204>
- Heru Stiawan, M Najibuloh Muzzaki, Sucipto, Anita Sari Wardani, Rina Firliana, M Iqbal Khalid, Adhi Wicak Milbar Gamas, & Shandy Arshad Busro. (2022). Model Visualisasi Informasi Dashboard Pada Pemetaan Tanaman Obat Dan Langka Kabupaten Kediri Menggunakan Microsoft Power Bi. *Jurnal Informatika Teknologi Dan Sains*, 4(4), 366–371. <https://doi.org/10.51401/jinteks.v4i4.2056>
- Komul, M. P. (2025). *Pemanfaatan Data Warehouse dalam Mengatasi Tantangan Pengelolaan Informasi di Manajemen Klinik*. 26(1), 78–86. <https://doi.org/10.37817/tekinfo.v26i1.4684>
- Kurniawan Hidayat, M., & Fitriana, R. (2022). Penerapan Sistem Intelijensia Bisnis Dan K-Means Clustering Untuk Memantau Produksi Tanaman Obat. *Jurnal Teknologi Industri Pertanian*, 32(2), 204–219. <https://doi.org/10.24961/j.tek.ind.pert.2022.32.2.204>
- Marvaro, E., & Sefina Samosir, R. (2021). Penerapan Business Intelligence dan Visualisasi Informasi di CV. Mitra Makmur Dengan Menggunakan Dashboard Tabelau. *KALBISCIENTIA Jurnal Sains Dan Teknologi*, 8(2), 37–46. <https://doi.org/10.53008/kalbiscientia.v8i2.197>
- Mauliadi, R. (2022). *Jurnal Informatika Ekonomi Bisnis Data Mining Menggunakan Algoritma K-Means Clustering dalam Analisis Tingkat Potongan Harga terhadap Harga Jual Sepeda Motor Honda*. 4, 7–9. <https://doi.org/10.37034/infeb.v4i4.156>
- Nafi, K., Sholehah, R., & Rahim, N. (2024). *Unveiling Insights : A Knowledge Discovery Approach to Comparing Topic Modeling Techniques in Digital Health Research*. 8(1), 111–125. <https://doi.org/10.29407/intensif.v8i1.22058>



- Pratiwi, H., Purno, A., Teknik, F., Widyatama, U., Kidul, C., & Bandung, K. (2022). *Implementasi Algoritma K-Means Untuk Mengklaster*. 6(1), 39–48. <https://doi.org/10.35145/joisie.v6i1.2252>
- Prayitno, G. P. (2020). Ketahanan Pangan Kabupaten/Kota Provinsi Jawa Barat. *Agribusiness Journal*, 14(1). <https://doi.org/10.15408/aj.v14i1.16320>
- Purba, S. F., Yulianti, A., Raphael, Y., & Khotimah, H. (2023). Determinan Kesejahteraan Petani Tanaman Pangan di Provinsi Jawa Barat. *Jurnal Ilmu Pertanian Indonesia*, 29(1), 59–67. <https://doi.org/10.18343/jipi.29.1.59>
- Saputra, A. K., Laksitowening, K. A., Herdiani, A., & Telkom, U. (2025). *Performance Analysis Of Extract , Transform , And Load Methods For Business Intelligence In E-Learning Systems Using Pentaho*. 6(1). <https://doi.org/10.52436/1.jutif.2025.6.1.4173>
- Virantika, E., Kusnawi, K., & Ipmawati, J. (2022). Evaluasi Hasil Pengujian Tingkat Clusterisasi Penerapan Metode K-Means Dalam Menentukan Tingkat Penyebaran Covid-19 di Indonesia. *Jurnal Media Informatika Budidarma*, 6(3), 1657. <https://doi.org/10.30865/mib.v6i3.4325>
- Wijayanto, S., & Yoka Fathoni, M. (2021). Pengelompokan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means. *Jupiter*, 13(2), 212–219. <https://doi.org/10.5281/3918.jupiter.2021.10>
- Yufita, A., Kurniawan, R., Wijaya, A. Y., & Suprpti, T. (2025). *Optimalisasi Klasterisasi Tenaga Kesehatan Menggunakan K-Means dan Davies Bouldin Indexs*. <https://doi.org/10.20961/ijai.v9i2.96645>