JINTECH: Journal of Information Technology Vol. 7, No. 1.
Februari 2026, Halaman : 34–51
E-ISSN : 2746-2331
P-ISSN : 2746-234X

ANALISIS *CLUSTERING* DAERAH PRODUKTIVITAS PADI DI KABUPATEN DELI SERDANG MENGGUNAKAN ALGORITMA *ISOLATION FOREST*

Ardi Wirya Indarto [1)], Asrianda [2)], Sujacka Retno [3)]
[1,2,3)] Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Indonesia
Email korespondensi: ardiwiryaindarto1@gmail.com [1)]

Abstract: Rice is a major food commodity that plays a vital role in national food security. However, differences in rice productivity levels between regions pose a challenge in formulating targeted agricultural policies. This study aims to analyze and cluster rice productivity areas in Deli Serdang Regency using the Isolation Forest algorithm. The data used are rice productivity data from all sub-districts in Deli Serdang Regency for the period 2020–2024, with variables of planted area, harvested area, production volume, and rice productivity. The analysis process is carried out through a web-based system using the Python programming language with the Streamlit framework. The Isolation Forest algorithm is used for clustering and anomaly detection, while cluster quality is evaluated using the Silhouette Score. The results of the 2024 data analysis show that 22 sub-districts in Deli Serdang Regency are divided into four clusters: a high-productivity cluster of 7 sub-districts (31.82%) with an average productivity above 6.2 tons/ha, a medium-productivity cluster of 4 sub-districts (18.18%) with a productivity of 6.0–6.1 tons/ha, a low-productivity cluster of 7 sub-districts (31.82%) with a productivity of around 5.9–6.0 tons/ha, and an anomalous cluster of 4 sub-districts (18.18%). The results of this clustering are expected to assist local governments in determining policies to increase rice productivity more effectively and based on data.
Keywords: Rice, Clustering, Isolation Forest, Anomaly Detection, Deli Serdang.

Abstrak: Padi merupakan komoditas pangan utama yang berperan penting dalam ketahanan pangan nasional. Namun, perbedaan tingkat produktivitas padi antar wilayah menjadi tantangan dalam perumusan kebijakan pertanian yang tepat sasaran. Penelitian ini bertujuan untuk menganalisis dan mengelompokkan daerah produktivitas padi di Kabupaten Deli Serdang menggunakan algoritma *Isolation Forest*. Data yang digunakan merupakan data produktivitas padi dari seluruh kecamatan di Kabupaten Deli Serdang pada periode tahun 2020–2024, dengan variabel luas tanam, luas panen, jumlah produksi, dan produktivitas padi. Proses analisis dilakukan melalui sistem berbasis web menggunakan bahasa pemrograman Python dengan *framework* Streamlit. Algoritma *Isolation Forest* digunakan untuk melakukan klasterisasi sekaligus mendeteksi anomali, sedangkan kualitas klaster dievaluasi menggunakan *Silhouette Score*. Hasil analisis data tahun 2024 menunjukkan bahwa 22 kecamatan di Kabupaten Deli Serdang terbagi ke dalam empat klaster, yaitu klaster produktivitas tinggi sebanyak 7 kecamatan (31,82%) dengan rata-rata produktivitas di atas 6,2 ton/ha, klaster produktivitas sedang sebanyak 4 kecamatan (18,18%) dengan produktivitas 6,0–6,1 ton/ha, klaster produktivitas rendah sebanyak 7 kecamatan (31,82%) dengan produktivitas sekitar 5,9–6,0 ton/ha, serta klaster anomali sebanyak 4 kecamatan (18,18%). Hasil klasterisasi ini diharapkan dapat membantu pemerintah daerah dalam menentukan kebijakan peningkatan produktivitas padi secara lebih efektif dan berbasis data.
Kata kunci: Padi, Klastering, Isolation Forest, Deteksi Anomali, Deli Serdang.

1. Pendahuluan

Padi (*Oryza sativa L.*) merupakan salah satu komoditas pangan utama di dunia, terutama di kawasan Asia, dan menjadi sumber makanan pokok bagi lebih dari setengah populasi global (FAO, 2021). Di Indonesia, padi memiliki peran strategis dalam ketahanan pangan nasional karena beras merupakan konsumsi utama sebagian besar penduduk (Prihatman, 2000). Kabupaten Deli Serdang, sebagai salah satu daerah penghasil padi di Provinsi Sumatera Utara, berkontribusi besar terhadap pasokan beras nasional. Namun, produktivitas padi di wilayah ini menunjukkan variasi yang cukup signifikan, di mana beberapa daerah mampu mencapai hasil panen yang tinggi, sementara daerah lain tertinggal (BPS Deli Serdang, 2022). Ketimpangan ini menunjukkan perlunya strategi pengelolaan pertanian yang lebih tepat guna meningkatkan efisiensi produksi.

Dalam beberapa tahun terakhir, tantangan ketahanan pangan semakin meningkat akibat pertumbuhan populasi, perubahan iklim, dan ketidakpastian kondisi lingkungan yang berdampak langsung pada produksi pertanian (Masdian *et al.*, 2023). Pemerintah Indonesia telah mengimplementasikan berbagai program untuk meningkatkan produktivitas padi, seperti penggunaan varietas unggul dan teknologi pertanian modern (Kementerian Pertanian, 2023). Namun, program tersebut belum merata, sehingga diperlukan pendekatan berbasis data untuk mengidentifikasi daerah yang memerlukan intervensi khusus guna meningkatkan hasil produksi secara lebih optimal.

Salah satu pendekatan yang dapat digunakan adalah *clustering*, yaitu teknik analisis data untuk mengelompokkan wilayah berdasarkan pola produktivitasnya. Dengan mengelompokkan daerah berdasarkan kesamaan karakteristik produktivitas padi, pemerintah dan pemangku kebijakan dapat merancang strategi pertanian yang lebih spesifik dan efektif (Prasetyo *et al.*, 2019). Selain itu, deteksi anomali menggunakan algoritma *Isolation Forest* dapat membantu mengidentifikasi daerah dengan produktivitas yang sangat tinggi maupun sangat rendah. Daerah yang mengalami anomali produksi dapat menjadi indikator adanya faktor eksternal yang perlu diperhatikan, seperti perubahan kondisi tanah, irigasi, atau faktor sosial-ekonomi petani.

Beberapa penelitian sebelumnya telah membahas pengelompokan produktivitas padi menggunakan metode statistik dan *machine learning*. Misalnya, Prasetyo *et al*. (2019) menggunakan metode K-Means *Clustering* untuk mengelompokkan daerah berdasarkan hasil panen di Jawa Timur. Namun, penelitian tersebut belum mengintegrasikan teknik deteksi anomali, yang berperan penting dalam mengidentifikasi wilayah dengan karakteristik produksi ekstrem. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan pendekatan yang lebih komprehensif dengan menggabungkan metode *clustering* dan algoritma *Isolation Forest* untuk menganalisis pola produktivitas padi serta mendeteksi kemungkinan anomali dalam data.

E-ISSN : 2746-2331
P-ISSN : 2746-234X

Berdasarkan urgensi dan kesenjangan penelitian sebelumnya, penelitian ini mengusung judul "**Analisis *Clustering* Daerah Produktivitas Padi di Kabupaten Deli Serdang Menggunakan Algoritma *Isolation Forest***". Hasil penelitian ini diharapkan dapat memberikan wawasan berbasis data kepada pemerintah daerah dan pemangku kebijakan dalam mengidentifikasi wilayah dengan produktivitas tinggi, rendah, maupun anomali. Dengan pendekatan ini, strategi peningkatan produktivitas dapat lebih tepat sasaran, sehingga mendukung ketahanan pangan dan kesejahteraan petani di Kabupaten Deli Serdang. Selain itu, penelitian ini juga dapat menjadi referensi bagi akademisi yang ingin mengembangkan metode serupa dalam konteks lain.

Adapun rumusan masalah dalam penelitian ini adalah:

1. Bagaimana mengklasterisasi hasil produksi padi di Kabupaten Deli Serdang menggunakan algoritma *Isolation Forest*?
2. Bagaimana membangun sebuah sistem klasterisasi hasil produksi padi menggunakan algoritma *Isolation Forest* di Kabupaten Deli Serdang?

Tujuan penelitian ini adalah:

1. Mengetahui bagaimana mengaplikasikan algoritma *Isolation Forest* dalam mengklasterisasi hasil produksi padi di Kabupaten Deli Serdang.
2. Membangun sebuah sistem klasterisasi hasil produksi padi di Kabupaten Deli Serdang menggunakan Python dan Streamlit sebagai antarmuka pengguna.

2. Kajian Kepustakaan

2.1. *Data Mining*

Data mining adalah proses ekstraksi pengetahuan dari kumpulan data besar menggunakan teknik statistik, kecerdasan buatan, dan pembelajaran mesin (Turban *et al*., 2007). Proses ini bertujuan untuk menemukan pola, hubungan, atau informasi tersembunyi dalam data yang dapat digunakan untuk pengambilan keputusan. Tahapan utama dalam *data mining* meliputi:

1. Seleksi Data (*Selection*): Pemilihan data operasional yang relevan sebelum dilakukan analisis lebih lanjut.
2. Pembersihan Data (*Preprocessing/Cleaning*): Penghapusan duplikasi, penanganan data yang hilang, serta perbaikan kesalahan dalam data.
3. Transformasi Data (*Transformation*): Konversi data ke dalam format yang sesuai untuk analisis, termasuk normalisasi.
4. Data Mining: Penerapan algoritma atau metode tertentu untuk menemukan pola atau pengetahuan dari data yang telah diproses.

5. Interpretasi/Evaluasi (*Interpretation/Evaluation*): Interpretasi hasil *data mining* untuk memastikan pola atau informasi yang diperoleh dapat digunakan secara efektif dalam pengambilan keputusan.

Kategori utama dalam *data mining* berdasarkan tugasnya meliputi:

1. Klasifikasi (*Classification*): Mengelompokkan data ke dalam kategori tertentu berdasarkan atribut-atribut yang relevan.
2. Estimasi (*Estimation*): Memprediksi nilai numerik berdasarkan faktor-faktor tertentu.
3. Prediksi (*Prediction*): Memperkirakan nilai atau kejadian di masa mendatang berdasarkan data historis.
4. Asosiasi (*Association*): Menemukan pola hubungan antar variabel dalam *dataset*.
5. Pengelompokan (*Clustering*): Mengelompokkan data berdasarkan kesamaan karakteristiknya, termasuk dalam metode *unsupervised learning*.

2.2. *Clustering*

Clustering adalah teknik dalam *data mining* yang digunakan untuk mengelompokkan objek atau individu berdasarkan kemiripan tertentu. Tujuan utama dari analisis *clustering* adalah membagi data ke dalam kelompok-kelompok yang memiliki perbedaan jelas satu sama lain, sementara setiap anggota dalam satu kelompok memiliki karakteristik yang serupa. Teknik ini diterapkan dalam berbagai bidang seperti pemasaran, segmentasi pasar, prediksi, analisis bisnis, zonasi wilayah, serta pengolahan citra dan pengenalan pola (Harahap *et al*., 2022).

Clustering termasuk dalam metode *unsupervised learning*, yang berarti proses pengelompokannya tidak memerlukan data dengan label atau target *output*. Dalam *data mining*, terdapat dua jenis metode *clustering* yang umum digunakan, yaitu:

1. Hierarchical Clustering yaitu mengelompokkan data dalam bentuk hirarki bertingkat, di mana setiap objek secara bertahap dikelompokkan berdasarkan kesamaan hingga membentuk hierarki klaster.
2. Non-Hierarchical Clustering yaitu mengelompokkan data tanpa membentuk struktur bertingkat, seperti metode *K-Means* dan DBSCAN yang langsung membagi *dataset* ke dalam beberapa klaster berdasarkan kedekatan karakteristik (Nurdin *et al*., 2024).

2.3. Padi (Oryza sativa L.)

Padi merupakan tanaman pangan utama yang telah dibudidayakan sejak zaman kuno dan banyak ditemukan di wilayah tropis serta subtropis, termasuk Asia, Afrika, Amerika, dan Australia. Secara taksonomi, tanaman padi tergolong dalam

Kingdom: *Plantae*, *Subkingdom*: *Tracheobionta*, Superdivisi: *Spermatophyta*, Divisi: *Magnoliophyta*, Kelas: *Liliopsida*, Subkelas: *Commelinidae*, Ordo: *Cyperales*, Famili: *Poaceae*, Genus: *Oryza L*., dan Spesies: *Oryza sativa L*.

Secara morfologi, padi diklasifikasikan ke dalam tiga varietas utama, yaitu:

1. *Indica*: Umumnya ditemukan di daerah tropis, tanaman tinggi, daun hijau muda, gabah panjang dan ramping.
2. *Japonica*: Tumbuh di daerah beriklim subtropis, tanaman dengan tinggi sedang, daun sempit hijau tua, gabah pendek dan agak bulat.
3. *Javanica*: Banyak ditemukan di Indonesia, tanaman tinggi, daun lebar dan kaku berwarna hijau, gabah panjang, lebar, dan tebal.

Padi dapat tumbuh pada ketinggian 0–1.500 meter di atas permukaan laut (mdpl) dengan kebutuhan sinar matahari langsung dan curah hujan rata-rata sekitar 200 mm per bulan. Jenis tanah ideal adalah tanah sawah dengan pH 5,5–7,5 dan suhu optimal 24–29°C (Prihatman, 2000).

2.4. Algoritma *Isolation Forest*

Isolation Forest adalah salah satu teknik terbaru dalam deteksi anomali yang bekerja dengan mengidentifikasi data yang tidak biasa (anomali) dari suatu kumpulan data. Algoritma ini dibangun berdasarkan konsep *decision trees* dan menggunakan pendekatan unik dengan memilih fitur dan nilai pemisah (*split value*) secara acak antara nilai minimum dan maksimum dari fitur yang dipilih. Metode ini pertama kali diperkenalkan oleh Liu *et al*. (2008) dan didasarkan pada dua asumsi kuantitatif utama:

1. Data anomali berasal dari kategori minoritas, yaitu kategori yang memiliki proporsi jumlah data lebih kecil dibandingkan kategori lainnya.
2. Data anomali memiliki nilai atribut yang sangat berbeda dari data lainnya.

Secara umum, *Isolation Forest* bekerja dengan membangun sebuah *ensemble* dari beberapa *isolation tree* (pohon biner) untuk memisahkan data input. Dalam sebuah *isolation tree*, data normal cenderung terisolasi pada bagian yang lebih dalam dari pohon, sedangkan data anomali cenderung terisolasi lebih dekat ke akar pohon. Hal ini terjadi karena data anomali memiliki karakteristik yang berbeda sehingga lebih mudah dipisahkan dengan sedikit pemisahan (*split*).

Proses pembangunan *isolation tree* dilakukan secara rekursif. Misalkan diberikan dataset dengan *n* elemen data, di mana setiap data memiliki *d* dimensi (atribut). Langkah-langkah pembangunan *isolation tree* adalah sebagai berikut:

1. Memilih atribut (*q*) dan nilai pemisah (*p*) secara acak.
2. Proses pemisahan berlanjut hingga salah satu kriteria berikut terpenuhi:
   a. Tinggi (*depth*) pohon mencapai nilai maksimum.

b. Hanya tersisa satu elemen data dalam node.
c. Semua elemen data dalam node memiliki nilai yang sama.

Setiap *isolation tree* adalah pohon biner (*binary tree*), di mana setiap *node* memiliki maksimal dua anak (*children*). Setelah *isolation tree* dibangun, langkah selanjutnya adalah memberikan peringkat anomali dengan menghitung skor anomali untuk setiap elemen data. Skor anomali dihitung berdasarkan panjang lintasan (*path length*) dari *root node* ke *node* yang merepresentasikan elemen data tersebut. Rumus skor anomali adalah sebagai berikut:

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}}$$

Keterangan:

- x adalah elemen data.
- $h(x)$ adalah panjang lintasan dari *root node* ke node yang merepresentasikan x.
- $E(h(x))$ adalah rata-rata $h(x)$ dari seluruh *isolation tree.*
- $c(n)$ adalah panjang lintasan rata-rata untuk *dataset* dengan *n* elemen.

Interpretasi skor anomali adalah:

1. Jika skor anomali $s(x,n)$mendekati 1, data tersebut dikategorikan sebagai anomali.
2. Jika skor anomali $s(x,n)$kurang dari 0,5, data tersebut dikategorikan sebagai data normal.
3. Jika seluruh *dataset* memiliki skor anomali mendekati 0,5, *dataset* tersebut tidak mengandung anomali.

2.5. Silhouette Score

Silhouette score atau yang biasa disebut *silhouette coefficient* merupakan sebuah metode untuk melakukan pengukuran terhadap kualitas dan kekuatan *cluster*. Metode ini menggabungkan konsep dari *cohesion* yang mengukur sebuah *cluster* yang memiliki hubungan antar objek dan *separation* yang mengukur jarak antara *cluster* yang berbeda (Handoyo, Rumami M, & Nasution, 2014). Berikut merupakan persamaan mengenai perhitungan *silhouette score*:

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}$$

Di mana:

- $a(i)$adalah jarak rata-rata antara titik data $i$dengan semua titik data lain dalam klaster yang sama (kohesi).
- $b(i)$adalah jarak rata-rata antara titik data $i$dengan semua titik data di klaster terdekat berikutnya (separasi).

Nilai *Silhouette Score* berkisar antara -1 hingga 1. Interpretasi nilai tersebut adalah sebagai berikut:

- Nilai mendekati 1: Menunjukkan klaster yang padat (*dense*) dan terpisah dengan baik dari klaster lain (hasil optimal).
- Nilai mendekati 0: Menandakan klaster yang tumpang tindih (*overlapping*), di mana batas antar klaster tidak jelas.
- Nilai negatif (mendekati -1): Menunjukkan penugasan klaster yang salah, di mana sebuah objek lebih mirip dengan klaster lain daripada klaster tempat ia berada.

3. Metode Penelitian

3.1. Tempat dan Waktu Pelaksanaan Penelitian

Penelitian ini dilakukan di kantor Dinas Pertanian, Kabupaten Deli Serdang, yang berlokasi di Tanjung Garbus Satu, Kecamatan Lubuk Pakam, Kabupaten Deli Serdang, Sumatera Utara. Waktu pelaksanaan penelitian ini dimulai dari bulan Agustus 2025 sampai dengan September 2025. Pengambilan data dilakukan secara bertahap untuk memastikan keakuratan informasi terkait produktivitas padi periode 2020–2024.

3.2. Metode Pengumpulan Data

Pendekatan yang digunakan dalam penelitian ini untuk memperoleh data yang diperlukan meliputi:

1. Penelitian Lapangan (*Field Research*): Dilakukan dengan observasi langsung ke Dinas Pertanian Kabupaten Deli Serdang untuk mengumpulkan data terkait produktivitas padi di berbagai wilayah.
2. Penelitian Kepustakaan (*Study Literature*): Melibatkan pengumpulan teori, konsep, dan hasil penelitian terdahulu yang relevan dari jurnal ilmiah, buku, skripsi, serta sumber daring yang kredibel.

Teknik pengumpulan data dilakukan melalui dokumentasi dari Dinas Pertanian Kabupaten Deli Serdang. Data yang dikumpulkan mencakup produktivitas padi dari 22 kecamatan selama periode 2020–2024. Data tersebut disimpan dalam format CSV dengan struktur sebagai berikut:

Tabel 1. Kamus Data pada File CSV Produktivitas Padi

No	Nama Kolom	Tipe Data	Keterangan
1	kecamatan	*String*	Nama kecamatan di Kabupaten Deli Serdang (22 kecamatan).
2	tahun	*Integer*	Tahun periode data (2020-2024).
3	luas_tanam	*Float*	Luas lahan yang ditanami (Hektar).
4	luas_panen	*Float*	Luas lahan yang berhasil dipanen (Hektar).
5	produksi	*Float*	Total hasil panen padi (Ton).

6	produktivitas	*Float*	Hasil kalkulasi (Produksi / Luas Panen) dalam satuan Ton/Ha.

Data kemudian diproses menggunakan pustaka *Pandas* pada Python untuk dibaca ke dalam *DataFrame* dan dipersiapkan untuk analisis lebih lanjut.

3.3. Variabel dan Pra-pemrosesan Data

Variabel yang digunakan dalam penelitian ini meliputi:

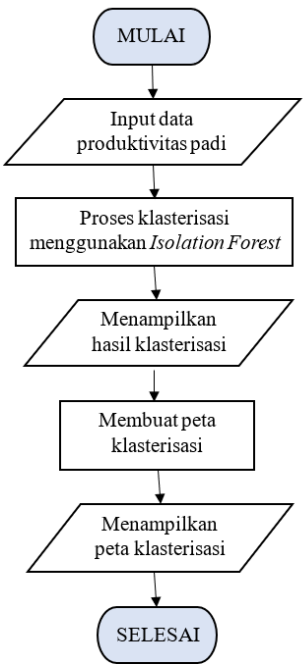
1. Luas tanam (Ha)
2. Luas panen (Ha)
3. Jumlah produksi (Ton)
4. Produktivitas padi (Ton/Ha), dihitung dengan rumus:

$$\text{Produktivitas} = \frac{\text{Produksi (Ton)}}{\text{Luas Panen (Ha)}}$$

Data melalui tahap pra-pemrosesan meliputi pengecekan nilai kosong, duplikasi, dan normalisasi menggunakan teknik Z-Score untuk menyelaraskan skala nilai dalam *dataset*.

3.4. Skema Sistem

Rancangan sistem menggambarkan tahapan proses pengelompokan wilayah berdasarkan tingkat produktivitas padi dengan memanfaatkan sistem berbasis web. Berikut adalah skema sistem yang dirancang:



Gambar 1. Skema Sistem

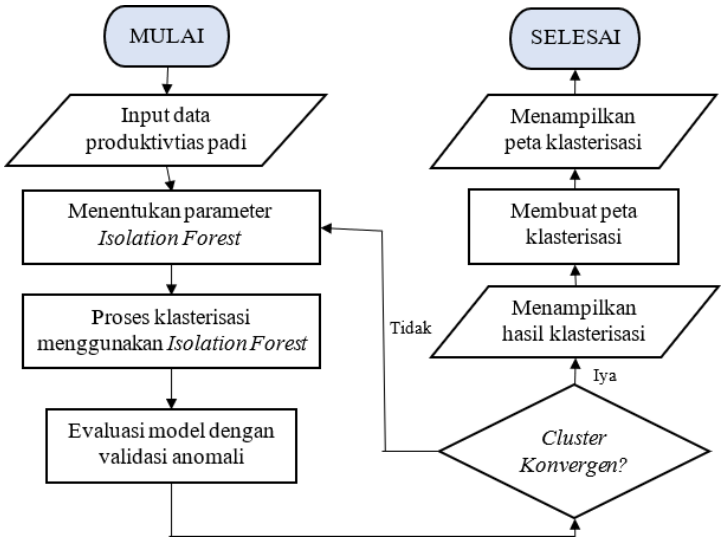
E-ISSN : 2746-2331
P-ISSN : 2746-234X

Keterangan:

1. Input data produktivitas padi dari berbagai wilayah.
2. Proses klasterisasi menggunakan algoritma *Isolation Forest*.
3. Menampilkan hasil klasterisasi dalam bentuk tabel atau grafik.
4. Membuat peta klasterisasi.
5. Menampilkan peta hasil klasterisasi.
6. Proses selesai.

3.5. Skema Isolation Forest

Algoritma *Isolation Forest* diterapkan dengan langkah-langkah sebagai berikut:



Gambar 2. Skema Isolation Forest

Keterangan:

1. Mulai proses klasterisasi.
2. Input data produktivitas padi.
3. Menentukan parameter *Isolation Forest* (jumlah pohon, kontaminasi).
4. Proses klasterisasi menggunakan algoritma *Isolation Forest*.
5. Evaluasi model dengan validasi anomali.
6. Cek konvergensi klaster.
7. Tampilkan hasil klasterisasi.
8. Buat peta klasterisasi.
9. Tampilkan peta klasterisasi.
10. Selesai.

4. Analisa dan Hasil

4.1. Hasil Klasterisasi

Berdasarkan analisis data tahun 2024, algoritma *Isolation Forest* berhasil mengelompokkan 22 kecamatan di Kabupaten Deli Serdang menjadi empat klaster:

Tabel 2. Tabel Hasil Klasterisasi Produktivitas Padi Tahun 2024

No.	Kecamatan	Luas Panen (Ha)	Produksi (Ton)	Produktivitas (Ton/Ha)	Skor Anomali	Klaster
1	Bangun Purba	133	808.64	6.235756	0.0392	Tinggi
2	Batang Kuis	3508	21328.64	6.045322	0.0524	Rendah
3	Beringin	9123	55467.84	6.253846	0.0063	Sedang
4	Biru-biru	1104	6712.32	6.18009	0.1085	Tinggi
5	Deli Tua	35	212.8	6.153529	0.0245	Tinggi
6	Galang	2189	13309.12	5.994026	0.041	Rendah
7	Gunung Meriah	599	3641.92	5.994213	0.1143	Rendah
8	Hamparan Perak	9850	59888	6.056102	-0.005	Anomali
9	Kutalimbaru	616	3745.28	5.964253	0.1029	Rendah
10	Labuhan Deli	9259	56294.72	5.882042	-0.0439	Anomali
11	Lubuk Pakam	3951	24022.08	5.965593	0.0545	Rendah
12	Namo Rambe	1377	8372.16	6.04024	0.0776	Rendah
13	Pagar Merbau	4647	28253.76	6.096085	0.0433	Sedang
14	Pancur Batu	778	4730.24	6.18912	0.1388	Tinggi
15	Pantai Labu	8367	50871.36	6.109921	0.0023	Sedang
16	Patumbak	588	3575.04	6.139226	0.1168	Tinggi
17	Percut S Tuan	10959	66630.72	6.279307	-0.1087	Anomali
18	Sibolangit	786	4778.88	6.187364	0.1392	Tinggi
19	Stm Hilir	986	5994.88	6.251702	0.0914	Tinggi
20	Stm Hulu	712	4328.96	5.842698	0.0207	Rendah
21	Sunggal	4190	25475.2	6.263845	0.0194	Sedang
22	Tanjung Morawa	6182	37586.56	6.030404	-0.0004	Anomali

E-ISSN : 2746-2331
P-ISSN : 2746-234X

Distribusi klaster adalah sebagai berikut:

1. Klaster Tinggi: 7 kecamatan (31,82%) dengan produktivitas >6,2 ton/ha.
2. Klaster Sedang: 4 kecamatan (18,18%) dengan produktivitas 6,0–6,1 ton/ha.
3. Klaster Rendah: 7 kecamatan (31,82%) dengan produktivitas 5,9–6,0 ton/ha.
4. Klaster Anomali: 4 kecamatan (18,18%) dengan pola produktivitas menyimpang.

4.2. Evaluasi Kualitas Klaster

Kualitas klaster yang terbentuk dievaluasi menggunakan *Silhouette Score*. Hasil perhitungan menunjukkan nilai *Silhouette Score* sebesar **0,65**, yang mengindikasikan bahwa klaster yang terbentuk memiliki kohesi yang baik dan pemisahan yang jelas antar klaster. Nilai ini termasuk dalam kategori baik karena mendekati 1.

4.3. Analisis Fenomena Anomali

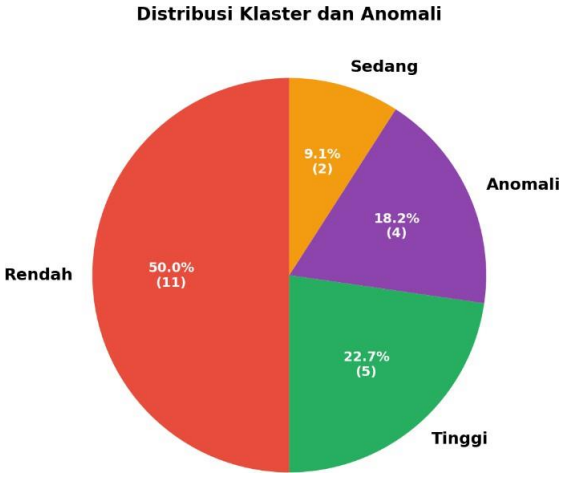
Analisis lebih lanjut terhadap klaster anomali menunjukkan adanya fenomena "Paradoks Anomali", di mana kecamatan dengan luas panen terbesar justru memiliki produktivitas yang relatif rendah. Contohnya adalah Kecamatan Percut Sei Tuan dengan luas panen 10.959 Ha namun produktivitas 6,28 Ton/Ha yang termasuk anomali. Hal ini menunjukkan bahwa luas panen tidak berkorelasi lurus dengan produktivitas, dan kemungkinan terdapat faktor penghambat seperti irigasi yang tidak memadai atau degradasi lahan.

4.4. Visualisasi Hasil

Hasil klasterisasi divisualisasikan dalam berbagai bentuk untuk memudahkan interpretasi:

1. **Diagram Lingkaran (*Pie Chart*)**

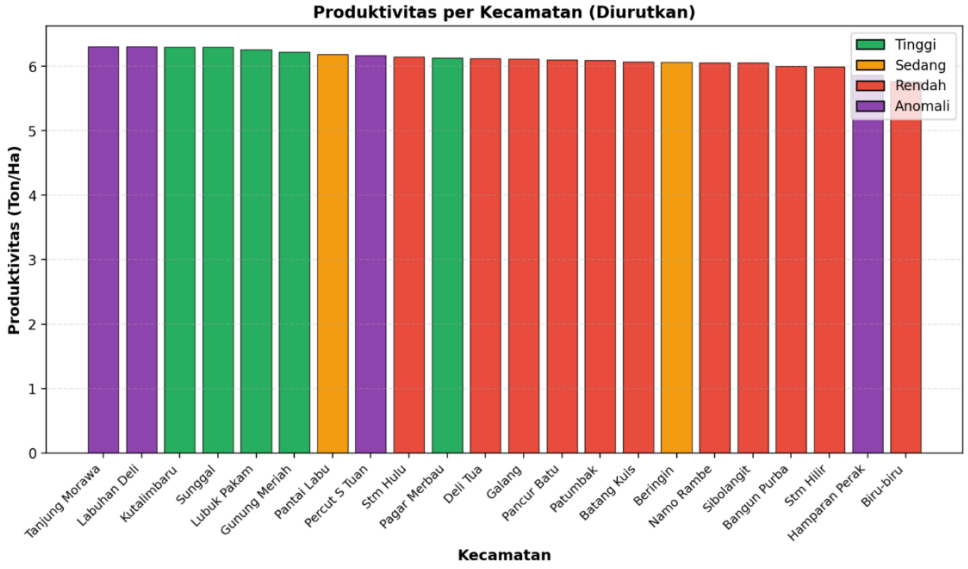
Diagram ini menunjukkan distribusi persentase setiap klaster.



Gambar 3. Grafik Pai

2. **Grafik Batang (*Bar Chart*)**

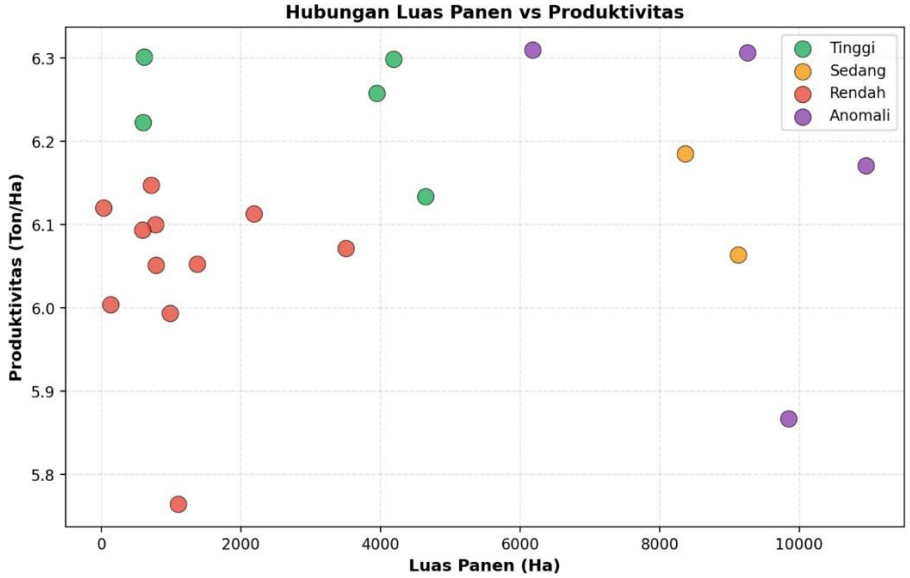
Grafik ini membandingkan produktivitas per kecamatan dengan kode warna sesuai klaster.



Gambar 4. Grafik Batang

3. ***Scatter Plot* (Hubungan Luas Panen vs Produktivitas)**

Scatter plot menunjukkan bahwa tidak ada hubungan linier kuat antara luas panen dan produktivitas.



Gambar 5. Grafik Titik

E-ISSN : 2746-2331
P-ISSN : 2746-234X

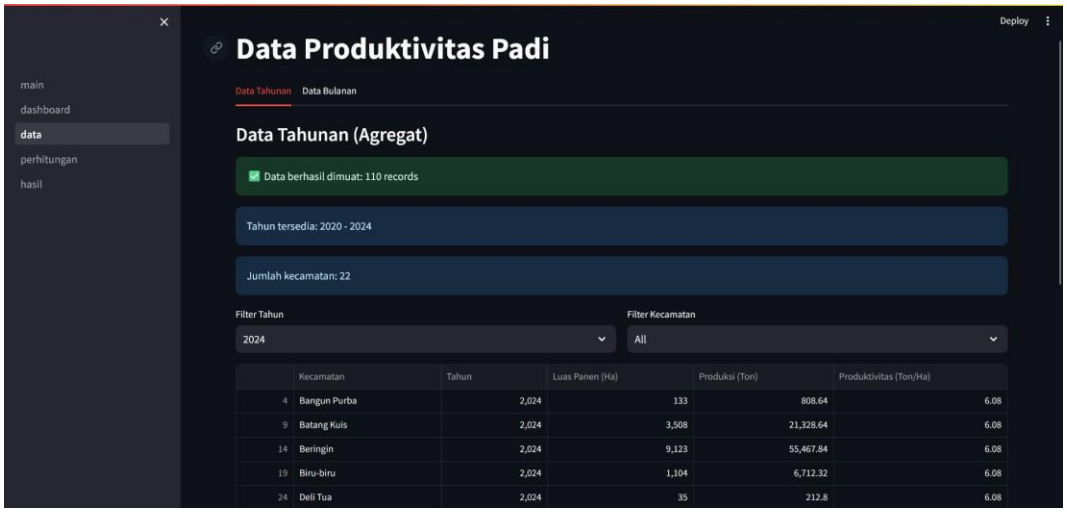
4.5. Implementasi Sistem

Sistem berbasis web berhasil dibangun dengan antarmuka yang terdiri dari beberapa halaman:

1. Halaman Dashboard

Gambar 6. Tampilan Halaman Dashboard

2. Halaman Tampilan Data



Gambar 7. Tampilan Halaman Data Tahunan

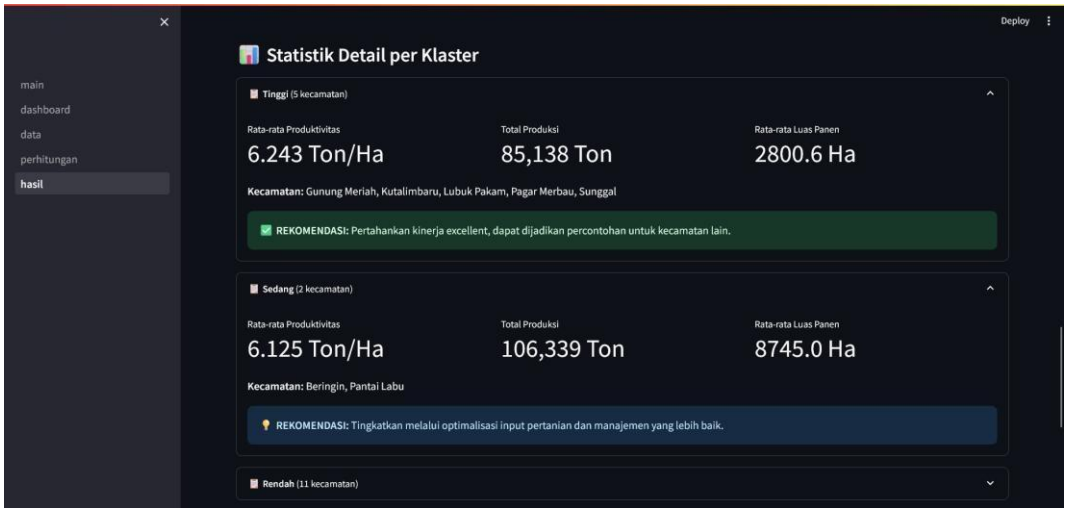
3. Halaman Perhitungan Algoritma *Isolation Forest*

Gambar 8. Tampilan Perhitungan Klasterisasi dan Anomali

4. Halaman Hasil Klasterisasi

Gambar 9. Tampilan Hasil Analisis dalam Bentuk Tabel

E-ISSN : 2746-2331
P-ISSN : 2746-234X



Gambar 10. Tampilan Statistik Detail per Klaster

4.6. Pengujian Sistem

Pengujian sistem dilakukan untuk memverifikasi fungsionalitas. Hasil pengujian menunjukkan bahwa semua fitur berjalan sesuai harapan:

Tabel 3. Hasil Pengujian Sistem

No.	Skenario Pengujian	Deskripsi Pengujian	Hasil yang Diharapkan	Hasil Aktual	Status
1	Login Admin Valid	Memasukkan *username* dan *password* untuk akses sistem.	Admin berhasil masuk ke halaman *dashboard*.	Sesuai Harapan	Valid
2	*Import* Data CSV	Memilih dan memuat file CSV produktivitas padi ke dalam sistem.	Sistem berhasil membaca file dan menampilkan data dalam format tabel (*DataFrame*).	Sesuai Harapan	Valid
3	Filter Data	Melakukan filter berdasarkan tahun atau kecamatan.	Tabel data menyaring informasi secara dinamis sesuai pilihan pengguna.	Sesuai Harapan	Valid
4	Eksekusi Analisis	Menekan tombol "Jalankan Analisis" dengan parameter tertentu.	Algoritma *Isolation Forest* menghitung skor anomali dan menentukan klaster secara otomatis.	Sesuai Harapan	Valid

5	Visualisasi Klaster	Menampilkan peta dan grafik hasil klasterisasi.	Peta interaktif menampilkan wilayah berdasarkan kategori warna klaster yang berbeda.	Sesuai Harapan	Valid
6	*Export* Hasil	Mengunduh hasil analisis dalam format file CSV.	Sistem menghasilkan file unduhan yang berisi data produktivitas beserta label klaster terpilih.	Sesuai Harapan	Valid

5. Kesimpulan dan Saran

5.1. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa:

1. Algoritma *Isolation Forest* berhasil mengelompokkan 22 kecamatan di Kabupaten Deli Serdang ke dalam empat klaster: tinggi, sedang, rendah, dan anomali dengan kualitas klaster yang baik (Silhouette Score = 0,65).
2. Sistem berbasis web yang dibangun menggunakan Python dan Streamlit berhasil memfasilitasi proses analisis data produktivitas padi dengan antarmuka yang intuitif dan fitur visualisasi yang lengkap.
3. Hasil klasterisasi dapat menjadi acuan bagi pemerintah daerah dalam merumuskan kebijakan pertanian yang lebih tepat sasaran, terutama dalam mengalokasikan sumber daya untuk daerah dengan produktivitas rendah atau anomali.

5.2. Saran

Untuk pengembangan penelitian selanjutnya, disarankan:

1. Mengintegrasikan variabel tambahan seperti faktor iklim, jenis tanah, penggunaan pupuk, dan teknologi pertanian untuk analisis yang lebih komprehensif.
2. Membandingkan performa algoritma *Isolation Forest* dengan metode klasterisasi lain seperti DBSCAN atau *Gaussian Mixture Models*.
3. Mengembangkan fitur prediksi produktivitas berdasarkan data historis dan faktor lingkungan.
4. Melakukan validasi hasil dengan melibatkan ahli pertanian atau pihak Dinas Pertanian untuk memastikan relevansi temuan dengan kondisi lapangan.

Daftar Pustaka

Badan Pusat Statistik Deli Serdang. (2022). *Statistik daerah Kabupaten Deli Serdang 2022*. BPS.

Barbariol, T., & Susto, G. A. (2021). TiWS-iForest: Isolation Forest in weakly supervised and Tiny ML scenarios. *Information Sciences, 610*. https://doi.org/10.1016/j.ins.2022.07.129

Chang, T.-T. (1985). The ethnobotany of rice in island Southeast Asia. *Asian Perspectives, 26*(1), 69–76.

Dewa, R., & Windarto, W. (2024). Deteksi anomali jaringan menggunakan Isolation Forest pada log Wazuh dengan pemberitahuan WhatsApp di PT XYZ. *KRESNA: Jurnal Riset dan Pengabdian Masyarakat, 4*(2), 208–216. https://doi.org/10.36080/kresna.v4i2.170

Food and Agriculture Organization. (2021). *The state of food security and nutrition in the world 2021*. Food and Agriculture Organization of the United Nations.

Harahap, L. M., Fuadi, W., Rosnita, L., Darnila, E., & Meiyanti, R. (2022). Klastering sayuran unggulan menggunakan algoritma K-Means. *Jurnal Teknik Informatika dan Sistem Informasi, 8*(3). https://doi.org/10.28932/jutisi.v8i3.5277

Kementerian Pertanian Republik Indonesia. (2023). *Program intensifikasi pertanian 2023*. Kementerian Pertanian RI.

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, 413–422.

Masdian, A. R., Bashit, N., & Hadi, F. (2023). Analisis Produktivitas Padi Menggunakan Algoritma *Machine Learning Random Forest* Di Kabupaten Batang Tahun 2018-2022. *Elipsoida: Jurnal Geodesi dan Geomatika, 6(1),* 43-51.

Nurdin, Bustami, Meiyanti, R., & Fahada, A. (2024). Clustering types of capture fisheries products using the K-Means clustering algorithm. *Journal of Theoretical and Applied Information Technology, 15*(17). www.jatit.org

Prasetyo, A., Wibowo, R., & Lestari, D. (2019). Clustering daerah produktivitas padi di Jawa Timur menggunakan algoritma K-Means. *Jurnal Ilmu Komputer, 10*(1), 23–34.

Prihatman, K. (2000). Budidaya Padi Pendayagunaan dan Pemasyarakatan Ilmu Pengetahuan dan Teknologi. *Penebar Swadaya*. Jakarta.

Sunarto, M., Kurniawan, D., Siswanto, E., & Huda, H. (2023). Deteksi anomali menggunakan Extended Isolation Forest (EIF). *Teknik: Jurnal Ilmu Teknik dan Informatika, 1*(2), 96–111. https://doi.org/10.51903/teknik.v1i2.324

Triana, O. (2024). Deteksi anomali jaringan menggunakan algoritma *Isolation*

Forest. *Jurnal Dunia Data, 1*(5).

Wibawa, I., & Karyawati, A. (2023). *Isolation Forest* dengan exploratory data analysis pada anomaly detection untuk data transaksi. *Jurnal Nasional Teknologi Informasi dan Aplikasinya, 1*(3), 803–810. https://doi.org/10.24843/JNATIA.2023.v01.i03.p04

Wijayanto, A., Sugiharto, A., & Santoso, R. (2024). Identifikasi dini curah hujan berpotensi banjir menggunakan algoritma Long Short-Term Memory (LSTM) dan *Isolation Forest*. *Jurnal Teknologi Informasi dan Ilmu Komputer, 11*, 637–646. https://doi.org/10.25126/jtiik.938718

Zulfikar, A., Rahmani, F., & Azizah, N. (2023). Deteksi anomali menggunakan *Isolation Forest* belanja barang persediaan konsumsi pada Satuan Kerja Kepolisian Republik Indonesia. *Jurnal Manajemen Perbendaharaan, 4*(1), 1–15. https://doi.org/10.33105/jmp.v4i1.435