

Analisis Performa Algoritma *Hierarchical Clustering* Untuk Identifikasi Pola Kematian Ikan Lele Pada Budidaya Lele Tiga Saudara

¹Agung Amanda Pratama, ²Dewi Suranti, ³Deri Lianda

¹ Mahasiswa, Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Dehasen Bengkulu
Alamat : Jl.Dempo 1 RT 20 RW 05 Sawah Lebar,Kec,Ratu Agung Kota Bengkulu.

email: agungahloy@gmail.com

^{2,3} Dosen Tetap, Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Dehasen Bengkulu
Jl. Meranti Raya No. 32 Kota Bengkulu 38228 Telp. (0736) 22027, 26957 Fax. (0736) 341139;
e-mail: dewisuranti@unived.ac.id derilianda@unived.ac.id

Received: Juni 2026, Revised: Agustus 2026, Accepiend: Oktober 2026)

Abstract- Tiga Saudara Catfish Farm in Bengkulu City has experienced varying mortality rates across different cultivation periods; however, these patterns have not been systematically analyzed using a data-driven approach. This study aims to identify catfish mortality patterns using Agglomerative Hierarchical Clustering (AHC) method. The dataset consisted of 90 days of operational records, including water temperature, pH levels, and feed types. The analytical process followed the Knowledge Discovery in Databases (KDD) framework, encompassing data preprocessing through One-Hot Encoding and Min-Max normalization, Euclidean distance calculation, and the implementation of the AHC algorithm using Single, Average, and Complete Linkage methods. The results indicate that the Average Linkage method achieved the best performance, with a Cophenetic Correlation Coefficient (CCC) value of 0.9789. The optimal number of clusters was determined to be four ($k=4$) based on a Silhouette Score of 0.8175. These four clusters represent distinct operational conditions within the farming system, namely Safe, Alert, Critical, and Recovery conditions. The findings further reveal that the use of alternative feed derived from bread waste contributed to a decline in water pH levels to a range of 6.0–6.8 and a significant increase in fish mortality, resulting in a total loss of 1,465 fish and an estimated financial loss of IDR 1,044,000 per production cycle. Model validity testing through White-Box Testing produced a CCC value of 0.9789 and a Silhouette Score of 0.8175, both exceeding the predetermined acceptance thresholds. Furthermore, system functionality testing through Black-Box Testing on 14 test cases within the AHC pipeline and Power BI dashboard demonstrated a 100% success rate. The clustering results were visualized through an interactive Microsoft Power BI dashboard, providing risk maps, daily mortality trends, and cluster distributions as decision-support tools for catfish farmers. This enables rapid identification of risk conditions and facilitates preventive actions before mass mortality events occur.

Keywords: Data Mining, Agglomerative Hierarchical Clustering, Catfish Mortality, Cophenetic Correlation Coefficient, Silhouette Score, Power BI.

Intisari-Budidaya Lele Tiga Saudara di Kota Bengkulu mengalami variasi tingkat mortalitas yang tidak merata pada setiap periode, namun belum dianalisis secara sistematis berbasis data. Penelitian ini bertujuan untuk mengidentifikasi pola kematian ikan lele menggunakan metode Agglomerative Hierarchical Clustering (AHC). Data yang digunakan merupakan data operasional harian selama 90 hari yang meliputi suhu air, pH, dan jenis pakan. Proses analisis mengikuti tahapan Knowledge Discovery in Databases (KDD), yang mencakup preprocessing data menggunakan One Hot Encoding dan normalisasi Min-Max, perhitungan jarak Euclidean, serta penerapan algoritma AHC

dengan metode linkage Single, Average, dan Complete. Hasil penelitian menunjukkan bahwa metode Average Linkage menghasilkan performa terbaik dengan nilai Cophenetic Correlation Coefficient (CCC) sebesar 0,9789. Jumlah kluster optimal yang diperoleh adalah empat ($k=4$) berdasarkan nilai Silhouette Score sebesar 0,8175. Keempat kluster merepresentasikan kondisi operasional budidaya, yaitu kondisi Aman, Waspada, Bahaya, dan Pulih. Temuan penelitian menunjukkan bahwa penggunaan pakan alternatif berupa limbah roti berkontribusi terhadap penurunan pH air ke rentang 6,0–6,8 dan peningkatan signifikan tingkat kematian ikan hingga total 1.465 ekor dengan estimasi kerugian Rp1.044.000 per siklus. Pengujian validitas model (White Box Testing) menghasilkan nilai CCC 0,9789 dan Silhouette Score 0,8175 yang keduanya melampaui ambang batas yang ditetapkan, sedangkan pengujian fungsionalitas sistem (Black Box Testing) terhadap 14 item uji pada pipeline AHC dan dashboard Power BI seluruhnya menunjukkan hasil berhasil. Hasil klusterisasi divisualisasikan melalui dashboard interaktif Microsoft Power BI yang menampilkan peta risiko, tren mortalitas harian, dan distribusi kluster sebagai alat bantu pemantauan dan pengambilan keputusan bagi pembudidaya ikan lele, sehingga kondisi risiko dapat diidentifikasi secara cepat dan tindakan pencegahan dapat dilakukan sebelum kematian massal terjadi.

Kata Kunci: Data Mining, Agglomerative Hierarchical Clustering, Mortalitas Ikan Lele, Cophenetic Correlation Coefficient, Silhouette Score, Power BI

I. PENDAHULUAN

Perkembangan teknologi informasi (TI) yang pesat telah membawa perubahan signifikan dalam berbagai sektor, termasuk sektor perikanan. Pemanfaatan TI tidak hanya mempercepat proses pengolahan data, tetapi juga meningkatkan ketepatan analisis dan kualitas pengambilan keputusan. Dalam kegiatan budidaya perikanan yang bersifat intensif, data operasional seperti kualitas air, kepadatan ikan, dan pemberian pakan dihasilkan setiap hari. Namun, tanpa pengolahan dan analisis yang tepat, data tersebut belum dimanfaatkan secara optimal untuk mendukung pengelolaan budidaya secara efektif. Kondisi tersebut juga terjadi pada usaha budidaya ikan lele yang umumnya

menghadapi tantangan produksi yang kompleks. Salah satu tantangan utama adalah tingginya angka kematian ikan, yang berdampak langsung pada penurunan produktivitas dan keuntungan pembudidaya. Kematian ikan lele tidak terjadi secara acak, melainkan dipengaruhi oleh berbagai faktor yang saling berkaitan, seperti kualitas air, kepadatan tebar, manajemen pakan, dan kondisi lingkungan. Hubungan antar faktor yang kompleks tersebut menyebabkan pembudidaya sering mengalami kesulitan dalam mengidentifikasi penyebab utama kematian ikan secara tepat, sehingga berpotensi menimbulkan kerugian ekonomi. Permasalahan tersebut terlihat pada Budidaya Ikan Lele Tiga Saudara di Kota Bengkulu. Usaha budidaya ini menerapkan sistem intensif dengan jumlah populasi ikan yang besar, sehingga pengelolaan kualitas air dan pakan menjadi aspek yang sangat krusial. Berdasarkan data awal dan hasil pengamatan di lapangan, angka kematian ikan lele menunjukkan variasi tingkat rendah, sedang, dan tinggi yang tidak merata pada setiap periode dan kolam budidaya. Kondisi ini mengindikasikan adanya perbedaan kondisi budidaya yang memengaruhi tingkat mortalitas ikan, namun hingga saat ini belum dianalisis secara sistematis berbasis data. Seiring meningkatnya ketersediaan data operasional budidaya, diperlukan pendekatan analisis yang mampu mengelompokkan data berdasarkan tingkat kemiripan karakteristiknya. Salah satu pendekatan yang banyak digunakan dalam analisis data adalah metode clustering, yang bertujuan untuk menemukan pola dalam sekumpulan data. Penelitian sebelumnya menunjukkan bahwa metode clustering, khususnya *Hierarchical Clustering*, efektif dalam mengelompokkan data dan memberikan pemahaman yang lebih mendalam mengenai pola kejadian serta faktor risiko dalam berbagai bidang, termasuk perikanan. Sebagai contoh, penelitian yang dilakukan oleh Sari et al. (2023) menggunakan metode *Hierarchical Cluster Analysis* untuk mengelompokkan wilayah penangkapan ikan berdasarkan data tangkapan, sehingga menghasilkan beberapa kelompok wilayah dengan karakteristik yang berbeda. Meskipun demikian, penerapan metode *Hierarchical Clustering* dalam analisis pola kematian ikan lele masih relatif terbatas, terutama pada budidaya ikan lele intensif dengan kondisi lokal yang spesifik. Penelitian yang

secara khusus memanfaatkan metode ini untuk menganalisis data mortalitas ikan lele pada tingkat pembudidaya lokal masih jarang dilakukan. Oleh karena itu, diperlukan penelitian yang menerapkan *Hierarchical Clustering* untuk memperoleh pemahaman yang lebih mendalam mengenai pola kematian ikan lele berdasarkan kondisi budidaya yang berbeda.

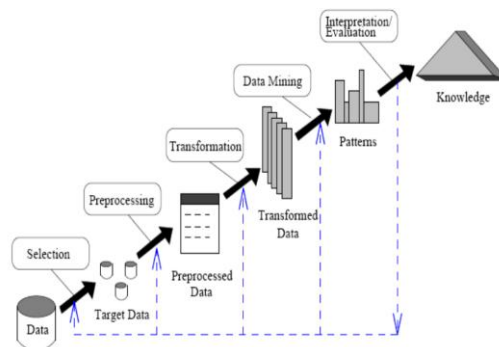
II. TINJAUAN PUSTAKA

Clustering adalah teknik pengelompokan objek data ke dalam beberapa kelompok (klaster) di mana objek dalam satu klaster memiliki kemiripan yang tinggi satu sama lain dibandingkan dengan objek di klaster lain. Dalam konteks *machine learning*, *clustering* dikategorikan sebagai *unsupervised learning* karena bekerja tanpa memerlukan label kelas atau target variabel yang ditentukan sebelumnya (Setiyawan et al., 2025). Tujuan utamanya adalah untuk mengungkap struktur intrinsik yang tersembunyi di dalam data. *Hierarchical Clustering* adalah metode analisis klaster yang mengorganisasikan data dengan cara membangun struktur hierarki berbentuk pohon (dendrogram). Berbeda dengan metode partisional (seperti K-Means), metode ini tidak mewajibkan penentuan jumlah klaster (k) di awal proses. Menurut Wati dan Prasetyo (2024), hasil akhir dari metode ini adalah representasi visual tingkatan kemiripan data, yang memungkinkan peneliti untuk melakukan pemotongan (*cutting*) dendrogram pada level tertentu untuk mendapatkan jumlah klaster yang paling optimal sesuai karakteristik data. Penelitian ini menggunakan pendekatan Agglomerative Hierarchical Clustering (AHC). Algoritma ini bekerja dengan prinsip bottom-up, di mana setiap data pada awalnya dianggap sebagai satu klaster tunggal (singleton). Pada setiap langkah iterasi, dua klaster yang memiliki jarak terdekat akan digabungkan (di-merge) menjadi satu klaster yang lebih besar, dan proses ini berlanjut hingga seluruh data menyatu menjadi satu cluster induk. Fleksibilitas AHC terletak pada variasi metode pengukuran jarak antar-klaster (linkage methods) yang dapat disesuaikan dengan distribusi data (Butar Butar, 2023). Silhouette Score adalah metrik evaluasi lain yang digunakan untuk menilai kualitas clustering. Silhouette Score mengukur seberapa baik suatu titik data berada dalam cluster-nya dibandingkan dengan cluster

lain. Dengan menggunakan metode ini, kita dapat mengevaluasi kualitas clustering untuk berbagai nilai kkk dan memilih nilai yang menghasilkan Silhouette Score tertinggi, yang menunjukkan pemisahan dan kohesi terbaik antar cluster (Putra, A. P., dkk, 2025).

III. METODOLOGI PENELITIAN

Metode penelitian yang digunakan mengacu pada tahapan Knowledge Discovery in Databases (KDD). Metode ini dipilih karena sesuai dengan tujuan penelitian, yaitu menggali pola pengetahuan tersembunyi dari data operasional budidaya ikan lele. Tahapan KDD dilakukan secara sistematis mulai dari data mentah hingga menghasilkan informasi yang dapat diinterpretasikan sebagai dasar pengambilan keputusan.



Gambar 1. Tahapan Knowledge Discovery in Databases (KDD)

(Sumber: Sari R.M et.al 2023)

1. Data Selection

Tahap ini dilakukan untuk memilih atribut yang relevan dari logbook harian budidaya. Atribut yang digunakan meliputi: Tanggal, Suhu Air (°C), pH, Jenis Pakan, dan Jumlah Kematian (ekor). Namun dalam proses clustering, hanya variabel Suhu, pH, dan Jenis Pakan yang digunakan sebagai variabel pembentuk kluster, sedangkan Jumlah Kematian digunakan untuk interpretasi karakteristik kluster.

2. Pre-processing (Cleaning)

Tahap ini bertujuan meningkatkan kualitas data sebelum dianalisis. Proses yang dilakukan meliputi pembersihan data dari duplikasi, penanganan missing values menggunakan metode moving average 3 hari sebelumnya untuk mempertahankan karakteristik time-series, serta standarisasi

penulisan kategori pakan agar tidak terjadi inkonsistensi.

3. Transformation

Tahap transformasi dilakukan untuk mengubah data ke dalam format numerik yang dapat diproses oleh algoritma clustering. Variabel kategorikal “Jenis Pakan” ditransformasikan menggunakan metode One Hot Encoding untuk mencegah distorsi jarak dalam perhitungan Euclidean Distance. Selanjutnya dilakukan normalisasi Min-Max pada variabel numerik (Suhu dan pH) untuk menyetarakan skala data ke dalam rentang 0–1, sehingga tidak terjadi dominasi atribut dengan skala lebih besar dalam perhitungan jarak.

4. Data Mining

Pada tahap ini diterapkan teknik clustering menggunakan algoritma Agglomerative Hierarchical Clustering (AHC). Perhitungan jarak antar data menggunakan Euclidean Distance. Untuk menentukan metode terbaik, dilakukan perbandingan tiga metode linkage, yaitu Single Linkage, Average Linkage, dan Complete Linkage.

5. Interpretation atau Evaluation

Tahap akhir meliputi evaluasi dan interpretasi hasil klusterisasi. Jumlah kluster ditentukan berdasarkan analisis dendrogram dan perbandingan nilai Silhouette Score untuk beberapa kandidat jumlah kluster. Validasi struktur dendrogram dilakukan menggunakan Cophenetic Correlation Coefficient (CCC). Setelah kluster optimal diperoleh, karakteristik rata-rata tiap kluster dianalisis untuk diberikan label risiko, seperti “Aman”, “Waspada”, “Kondisi Bahaya”, atau “Pulih”. Hasil akhir divisualisasikan menggunakan dashboard Power BI untuk mempermudah interpretasi oleh pembudidaya.

IV. HASIL DAN PEMBAHASAN

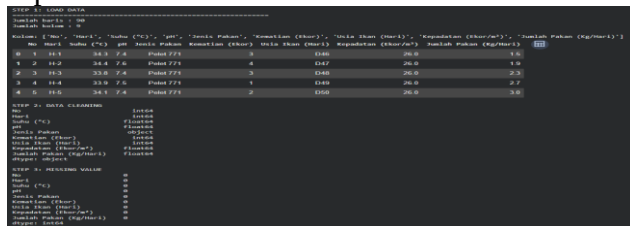
A. Hasil Penelitian

Penelitian ini menghasilkan model pengelompokan pola kematian ikan lele pada Budidaya Lele Tiga Saudara menggunakan algoritma Agglomerative Hierarchical Clustering (AHC). Sistem diimplementasikan menggunakan bahasa pemrograman Python di lingkungan Google Colaboratory dan divisualisasikan menggunakan Microsoft Power

BI Desktop. Hasil setiap tahapan proses dari pipeline yang telah dijalankan dapat dilihat pada penjelasan berikut:

Hasil Load dan Pembersihan Data

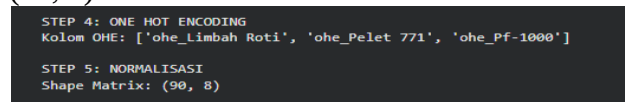
Tahap pertama adalah memuat dataset_mortalitas_lele.csv ke Google Colab. Dataset berhasil dimuat dengan dimensi 90 baris dan 9 kolom. Hasil pemeriksaan missing values menunjukkan nilai nol pada seluruh kolom, sehingga data dinyatakan bersih dan siap diproses.



Gambar 2. Load Data dan Pemeriksaan Missing Value

Hasil One Hot Encoding dan Normalisasi

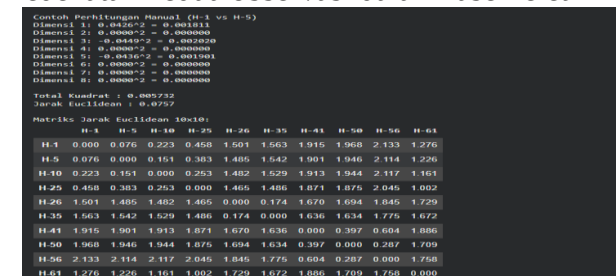
Variabel kategorikal Jenis Pakan diubah ke representasi biner menggunakan One Hot Encoding, menghasilkan tiga kolom baru: ohe_Limbah Roti, ohe_Pelet 771, dan ohe_Pf-1000. Selanjutnya, seluruh variabel numerik dinormalisasi menggunakan metode Min-Max sehingga menghasilkan feature matrix berukuran (90, 8).



Gambar 3. One Hot Encoding dan Matrix Normalisasi

Hasil Perhitungan Matriks Jarak Euclidean

Matriks jarak Euclidean dihitung dari feature matrix 8 dimensi yang telah ternormalisasi. Verifikasi manual untuk pasangan H-1 dan H-5 menghasilkan jarak 0,0757 yang mengonfirmasi kedekatan kedua observasi dalam fase Pelet 771.



Gambar 4. Matriks Jarak Euclidean

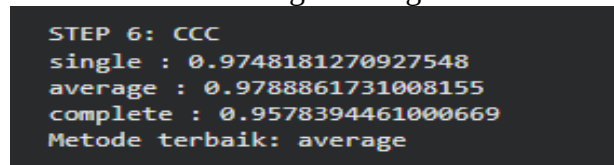
Hasil Perbandingan Metode Linkage dan Nilai CCC

Algoritma AHC dijalankan dengan tiga metode linkage untuk menentukan metode terbaik berdasarkan nilai Cophenetic Correlation Coefficient (CCC). Hasil perbandingan dari Step 6 Google Colab

Tabel 1. Perbandingan Nilai CCC Tiga Metode Linkage

Metode Linkage	Nilai CCC	Kualitas	Keputusan
Single	0,9748	Sangat Baik	Tidak Terpilih
Average	0,9789	Sangat Baik	TERPILIH ✓
Complete	0,9578	Sangat Baik	Tidak Terpilih

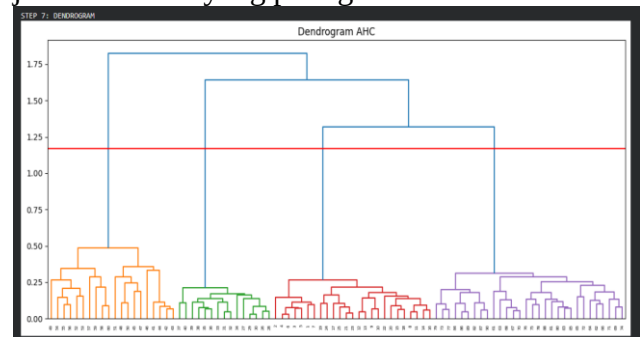
Metode Average Linkage terpilih sebagai metode terbaik dengan nilai CCC tertinggi sebesar 0,9789. Nilai ini jauh melebihi ambang batas minimal 0,70 yang disyaratkan, mengonfirmasi bahwa struktur dendrogram sangat akurat.



Gambar 5. Perbandingan Nilai CCC Tiga Metode Linkage

Hasil Dendrogram AHC

Dendrogram dihasilkan menggunakan metode Average Linkage pada Step 7 Google Colab. Garis potong (cut-off) ditentukan pada jarak 1,17 (garis merah) yang secara visual menunjukkan adanya 4 kelompok data (k=4). Longest gap yang paling jelas terlihat antara jarak penggabungan iterasi ke-6 dan ke-7 mengonfirmasi k=4 sebagai jumlah kluster yang paling natural.



Gambar 6. Dendrogram AHC

Nilai Silhouette Score dihitung untuk kandidat jumlah kluster k=2 hingga k=6. Hasil dari Step 8 Google Colab menunjukkan bahwa k=4 menghasilkan nilai tertinggi sebesar 0,8175.

Tabel 2. Perbandingan Silhouette Score untuk Berbagai Nilai k

Jumlah Klaster (k)	Silhouette Score	Keterangan
k = 2	0,4994	Kurang baik – struktur klaster tidak optimal
k = 3	0,6355	Cukup baik – fase Waspada & Pulih tidak terpisah
k = 4	0,8175	Optimal – pemisahan 4 klaster sangat baik
k = 5	0,7334	Menurun – sub-klaster mulai tumpang tindih
k = 6	0,7254	Menurun – terlalu banyak fragmentasi

```

STEP 8: SILHOUETTE
k= 2 = 0.499448279787574
k= 3 = 0.6354899637835157
k= 4 = 0.8174680424045341
k= 5 = 0.7333700043621532
k= 6 = 0.7254467906442924
k optimal: 4

```

Gambar 7. Perbandingan Silhouette Score

Hasil Klasterisasi Final dan Pelabelan Risiko
Model AHC final dijalankan dengan Average Linkage dan k=4 pada Step 9–11 Google Colab. Distribusi data ke setiap klaster dan label risiko yang diberikan berdasarkan karakteristik rata-rata mortalitas

Tabel 3. Hasil Klasterisasi Final — Distribusi dan Karakteristik Klaster

Klaster	Label Risiko	n Hari	Suhu (°C)	pH	Mortalitas Ekor/Hari	Pakan Dominan
K-1	AMAN	25	33,76	7,38	2,68	Pelet 771 (H-1 s.d. H-25)
K-2	WASPADA	15	34,58	7,17	7,80	PF-1000 (H-26 s.d. H-40)
K-3	BAHAYA	20	36,86	6,38	52,20	Limbah Roti (H-41 s.d. H-60)
K-4	PULIH	30	34,03	7,24	7,90	Pelet 771 (H-61 s.d. H-90)

```

STEP 9: CLUSTER FINAL
cluster_id_num
0 30
1 25
2 20
3 15
Name: count, dtype: int64

STEP 10: KARAKTERISTIK
No Hari Suhu (°C) pH Kematian (Ekor) \
cluster_id_num
0 50.5 50.5 36.850000 6.380000 52.20
1 75.5 75.5 34.026667 7.240000 7.90
2 33.0 33.0 34.580000 7.172223 7.80
3 13.0 13.0 33.760000 7.380000 2.68

cluster_id_num Usia Ikan (Hari) Kepadatan (Ekor/m³) Jumlah Pakan (Kg/Hari) \
0 95.5 24.340000 32.130000
1 120.5 22.540000 26.276667
2 70.0 25.233333 18.360000
3 58.0 25.950000 6.128000

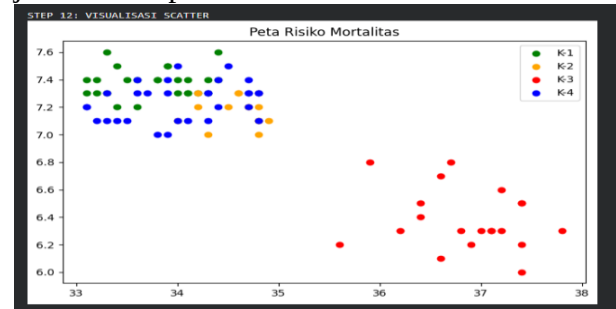
cluster_id_num ohe_Limbah_Roti ohe_Pelet_771 ohe_PF-1000
0 1.0 0.0 0.0
1 0.0 1.0 0.0
2 0.0 0.0 1.0
3 0.0 1.0 0.0

STEP 11: LABEL
cluster_id status_risiko
0 Aman
1 Pulih
2 Bahaya
3 Waspada
Name: count, dtype: int64

```

Gambar 8. Distribusi Klaster dan Label Risiko**Hasil Visualisasi Scatter Plot (Peta Risiko)**

Scatter plot pada Gambar 4.8 dihasilkan pada Step 12 Google Colab untuk menampilkan posisi setiap hari observasi dalam ruang dua dimensi Suhu (sumbu X, °C) versus pH (sumbu Y), di mana setiap titik diwarnai berdasarkan klaster risiko yang telah terbentuk. Dari total 90 titik data, terlihat pola pengelompokan yang sangat jelas dan terpisah antar klaster.

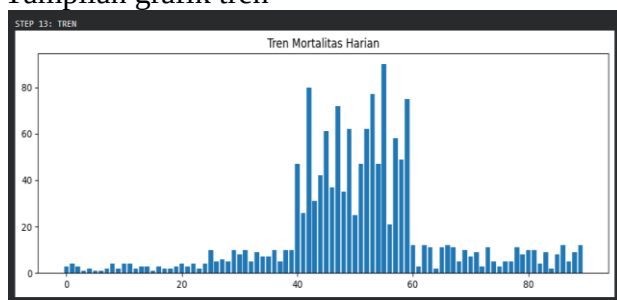
**Gambar 9. Peta Risiko Mortalitas (Scatter Plot Suhu vs pH)**

Klaster K-1 (Aman) ditampilkan dengan warna hijau dan terkonsentrasi di area kiri-atas grafik, yaitu pada rentang suhu 33,1–34,8°C dan pH 7,2–7,6. Pengelompokan yang rapat di area ini mencerminkan kondisi parameter air yang stabil dan konsisten selama 25 hari fase awal budidaya dengan pakan Pelet 771. Klaster K-2 (Waspada) ditampilkan dengan warna kuning dan berada di area tengah-kiri grafik, pada rentang suhu 34,0–34,9°C dan pH 7,0–7,3. Posisi K-2 yang berada tepat di bawah K-1 secara visual mengonfirmasi bahwa fase PF-1000 (H-26 hingga H-40) merupakan peralihan — parameter air mulai menyimpang ke arah yang lebih berisiko namun belum mencapai kondisi kritis. Klaster K-3 (Bahaya) ditampilkan dengan warna merah dan terpisah jauh di area pojok kanan-bawah grafik, pada rentang suhu 35,6–37,8°C dan pH 6,0–6,8. Pemisahan spasial yang sangat ekstrem ini secara visual langsung menunjukkan anomali yang terjadi selama fase Limbah Roti (H-41 hingga H-60). Klaster K-4 (Pulih) ditampilkan dengan warna biru dan menempati area yang sebagian besar beririsan dengan K-1, yaitu pada rentang suhu 33,1–34,8°C dan pH 7,0–7,5. Meskipun parameter suhu dan pH K-4 sudah kembali mendekati kondisi K-1, algoritma AHC tetap berhasil memisahkan keduanya berdasarkan kombinasi variabel multivariat lainnya. Secara keseluruhan, scatter plot ini mengonfirmasi secara visual bahwa hasil klasterisasi AHC sangat bermakna: batas alami antar klaster

terlihat jelas terutama antara K-3 (Bahaya) dengan ketiga klaster lainnya, yang konsisten dengan nilai Silhouette Score tinggi sebesar 0,8175 dan CCC sebesar 0,9789 yang dihasilkan model.

Hasil Visualisasi Tren Mortalitas Harian

Grafik tren mortalitas harian dari Step 13 Google Colab menampilkan fluktuasi kematian ikan sepanjang 90 hari siklus budidaya. Lonjakan tajam kematian terlihat jelas pada fase Limbah Roti (H-41 hingga H-60), dengan puncak tertinggi pada H-56 sebesar 90 ekor/hari. Tampilan grafik tren



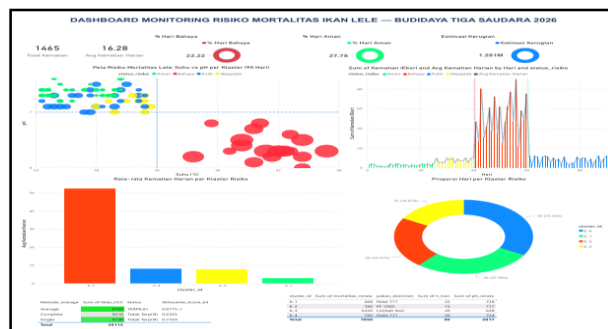
Gambar 10. Tren Kematian Harian Selama 90 Hari Siklus Budidaya

Berdasarkan grafik tersebut, terdapat empat pola yang sangat kontras sesuai fase pergantian pakan. Pada fase pertama (H-1 s.d. H-25) dengan pakan Pelet 771, batang grafik terlihat sangat rendah dan stabil di bawah 5 ekor/hari, mencerminkan kondisi budidaya yang terkendali. Memasuki fase kedua (H-26 s.d. H-40) dengan pakan PF-1000, batang mulai meningkat secara bertahap ke kisaran 5–10 ekor/hari sebagai indikator awal penurunan kondisi kualitas air. Lonjakan paling dramatis terjadi pada fase ketiga (H-41 s.d. H-60) saat pakan diganti ke Limbah Roti, di mana tinggi batang melonjak secara ekstrem dengan puncak tertinggi pada H-56 sebesar 90 ekor/hari. Volatilitas tinggi juga tampak jelas pada fase ini dengan batang yang naik-turun tajam antara 21–90 ekor/hari. Pada fase keempat (H-61 s.d. H-90) setelah pakan dikembalikan ke Pelet 771, batang grafik menurun namun belum kembali ke level awal, berkisar antara 2–12 ekor/hari, yang mengindikasikan adanya efek residual pasca fase kritis.

Hasil Dashboard Power BI

File hasil_clustering_lele.csv yang diekspor dari Step 14 Google Colab diimpor ke Microsoft Power BI Desktop. Dashboard interaktif yang dihasilkan menampilkan empat jenis visualisasi utama: (1) Scatter plot peta risiko Suhu vs pH,

Card KPI yang memuat total kematian 1.465 ekor, rata-rata 16,28 ekor/hari, persentase hari bahaya 22,2%, dan estimasi kerugian; (3) grafik tren mortalitas 90 hari; serta (4) donut chart proporsi klaster. Tampilan dashboard Power BI

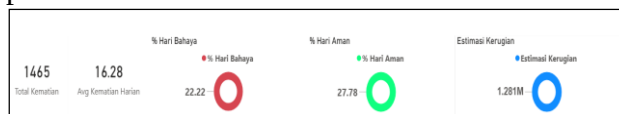


Gambar 11. Tampilan Dashboard Power BI

Dashboard Power BI yang dihasilkan terdiri dari lima fitur visualisasi utama yang saling terintegrasi dan dapat difilter secara interaktif. Berikut adalah penjelasan lengkap hasil masing-masing fitur tersebut:

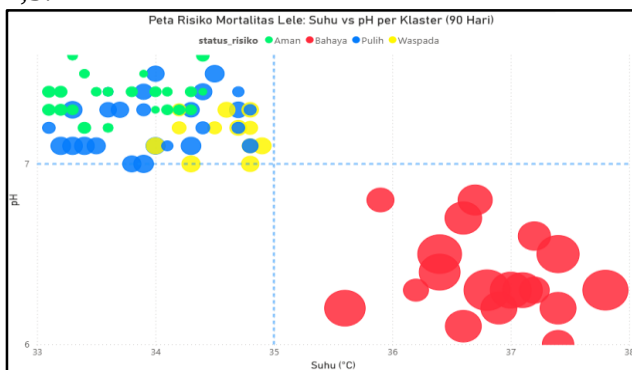
Card KPI (Key Performance Indicator)

Dashboard menampilkan lima Card KPI di bagian atas yang merangkum metrik utama siklus budidaya secara sekilas. Card pertama menampilkan Total Kematian sebesar 1.465 ekor sebagai akumulasi kematian seluruh siklus 90 hari. Card kedua menampilkan Rata-rata Kematian Harian sebesar 16,28 ekor/hari. Card ketiga menampilkan persentase % Hari Bahaya sebesar 22,22%, yang berarti 20 dari 90 hari berada di klaster K-3 (Bahaya) — meskipun hanya 22% waktu, fase ini menyumbang 1.044 ekor atau 71,3% dari total kematian. Card keempat menampilkan % Hari Aman sebesar 27,78%, yang berarti 25 hari budidaya berlangsung dalam kondisi ideal klaster K-1. Card kelima menampilkan Estimasi Kerugian sebesar Rp 1,281 juta yang dihitung berdasarkan total 1.465 ekor dikalikan estimasi harga per ekor. Kelima Card ini memungkinkan pembudidaya menangkap kondisi kritis siklus budidaya hanya dalam satu pandangan tanpa perlu menelusuri data harian.



Gambar 12. Card KPI Peta Risiko Mortalitas (Bubble Chart Suhu vs pH)

Fitur ini menampilkan 90 titik data harian dalam bidang dua dimensi dengan sumbu X adalah Suhu Air (°C) dan sumbu Y adalah pH Air. Setiap titik berbentuk gelembung (bubble) di mana ukuran gelembung mencerminkan jumlah kematian harian — semakin besar gelembung, semakin tinggi jumlah kematian pada hari tersebut. Warna gelembung disesuaikan dengan klaster risiko: hijau untuk K-1 (Aman), kuning untuk K-2 (Waspada), merah untuk K-3 (Bahaya), dan biru untuk K-4 (Pulih). Hasil visualisasi menunjukkan pemisahan klaster yang sangat jelas. Klaster K-3 (Bahaya) terkonsentrasi di pojok kanan-bawah pada suhu 35,6–37,8°C dan pH 6,0–6,8, dengan ukuran gelembung paling besar mencerminkan mortalitas massal hingga 90 ekor/hari. Klaster K-1 (Aman) mengelompok di area kiri-atas pada suhu 33,1–34,4°C dan pH 7,2–7,6 dengan gelembung kecil. Klaster K-2 (Waspada) berada di zona tengah suhu 34,0–34,9°C dan pH 7,0–7,3, sedangkan K-4 (Pulih) tersebar di suhu 33,1–34,8°C dengan pH yang sudah kembali ke kisaran normal 7,0–7,5.

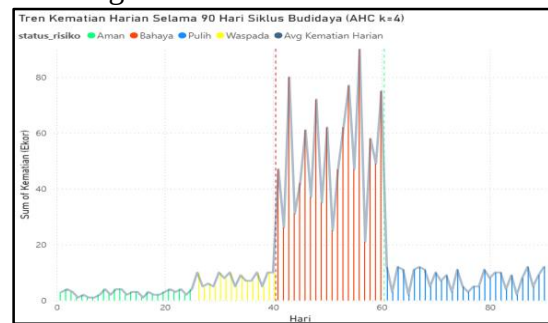


Gambar 13. Peta Risiko Mortalitas (Suhu vs pH)

a) Grafik Tren Mortalitas Harian (Line Chart 90 Hari)

Fitur ini menampilkan grafik garis yang memvisualisasikan fluktuasi jumlah kematian ikan setiap hari selama siklus 90 hari secara kronologis, dilengkapi dengan garis rata-rata mortalitas sebagai referensi. Sumbu X merepresentasikan hari observasi (H-1 s.d. H-90) dan sumbu Y merepresentasikan jumlah kematian harian (ekor). Garis tren menunjukkan tiga pola yang sangat kontras: (1) garis landai stabil di bawah 5 ekor/hari selama 25 hari fase K-1 (Pelet 771); (2) lonjakan mendadak dari 10 ekor menjadi 47 ekor hanya dalam satu malam pada peralihan H-40 ke H-41 saat pergantian ke Limbah Roti;

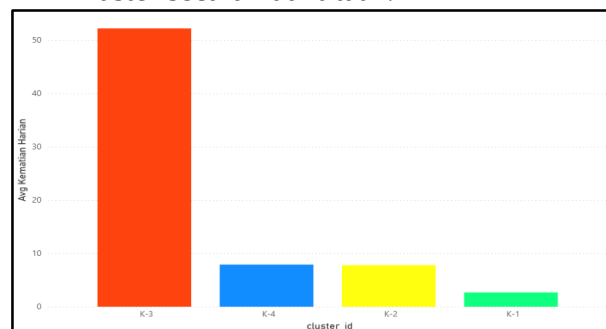
(3) volatilitas ekstrem antara 21–90 ekor/hari selama 20 hari fase K-3 dengan puncak tertinggi 90 ekor/hari di H-56; dan (4) penurunan bertahap selama 30 hari fase K-4 dengan rata-rata 7,90 ekor/hari yang belum kembali ke level K-1. Grafik tren ini merupakan fitur paling informatif untuk memahami dampak langsung pergantian jenis pakan terhadap mortalitas ikan secara kronologis.



Gambar 14. Grafik Tren Mortalitas Harian

b) Bar Chart Rata-rata Kematian Harian per Klaster Risiko

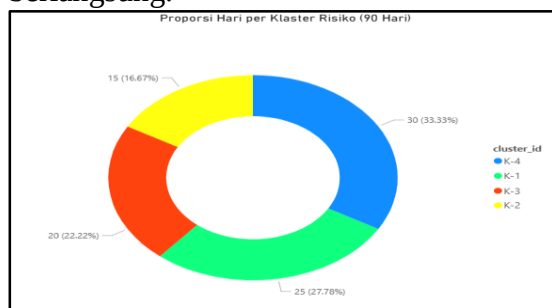
Fitur ini menampilkan diagram batang yang membandingkan rata-rata kematian harian antar keempat klaster risiko secara berdampingan. Hasil bar chart menunjukkan kesenjangan yang sangat mencolok: K-3 (Bahaya) memiliki batang paling tinggi dengan rata-rata 52,20 ekor/hari, disusul K-4 (Pulih) sebesar 7,90 ekor/hari, K-2 (Waspada) sebesar 7,80 ekor/hari, dan K-1 (Aman) sebagai yang terendah hanya 2,68 ekor/hari. Perbandingan ini secara visual menunjukkan bahwa rata-rata kematian harian di klaster K-3 hampir 20 kali lipat dibandingkan K-1, dan hampir 7 kali lipat dibandingkan K-2 dan K-4. Bar chart ini memperkuat validitas pelabelan risiko tiap klaster secara kuantitatif.



Gambar 15. Bar Chart Rata-rata Kematian Harian per Klaster Risiko

c) Donut Chart Proporsi Klaster

Fitur ini menampilkan diagram donat yang menggambarkan distribusi proporsi hari budidaya pada masing-masing klaster risiko dari total 90 hari. Hasil donut chart menunjukkan: K-1 (Aman) sebesar 27,8% atau 25 hari, K-2 (Waspada) sebesar 16,7% atau 15 hari, K-3 (Bahaya) sebesar 22,2% atau 20 hari, dan K-4 (Pulih) sebesar 33,3% atau 30 hari. Dari visualisasi ini terlihat bahwa lebih dari separuh siklus (55,5% atau 50 hari) dihabiskan dalam kondisi tidak ideal (K-2, K-3, dan K-4 gabungan), dengan K-3 (Bahaya) menyumbang lebih dari seperlima total waktu budidaya. Donut chart ini memberikan gambaran makro yang cepat bagi pembudidaya untuk menilai kualitas keseluruhan siklus budidaya yang telah berlangsung.



Gambar 16. Donut Chart Proporsi Hari per Klaster Risiko

d) Slicer Interaktif (Filter Dinamis)

Dashboard dilengkapi dengan slicer yang berfungsi sebagai panel filter interaktif dengan tiga opsi penyaringan: filter berdasarkan Status Risiko (Aman, Waspada, Bahaya, Pulih), filter berdasarkan Jenis Pakan (Pelet 771, PF-1000, Limbah Roti), dan filter berdasarkan rentang hari. Ketika salah satu filter dipilih, seluruh visualisasi pada dashboard — scatter plot, card KPI, grafik tren, dan donut chart — secara otomatis memperbarui tampilannya hanya untuk data yang sesuai filter tersebut. Sebagai contoh, saat filter ‘Bahaya’ dipilih, scatter plot hanya menampilkan 20 titik data K-3, card KPI berubah menampilkan total kematian 1.044 ekor dengan rata-rata 52,2 ekor/hari, dan grafik tren hanya menampilkan fluktuasi H-41 s.d. H-60. Fitur slicer ini menjadikan dashboard bukan sekadar laporan statis, melainkan alat eksplorasi data mandiri yang dapat digunakan pembudidaya tanpa keahlian pemrograman untuk menganalisis kondisi spesifik yang ingin ditinjau.

status_risiko	Jenis Pakan	Hari
☐ Aman	☐ Limbah Roti	☐ 1
☐ Bahaya	☐ Pelet 771	☐ 2
☐ Pulih	☐ Pf-1000	☐ 3
☐ Waspada		☐ 4
		☐ 5
		☐ 6
		☐ 7
		☐ 8

Gambar 17. Donut Chart Proporsi Hari per Klaster Risiko

Berdasarkan pengujian black box yang telah dilakukan terhadap seluruh tahapan pipeline AHC dan dashboard Power BI, dapat disimpulkan bahwa tidak terdapat kesalahan (error) pada setiap fungsi yang diuji. Seluruh 14 item pengujian menunjukkan hasil berhasil, sehingga sistem klasterisasi mortalitas ikan lele menggunakan algoritma Agglomerative Hierarchical Clustering ini dapat berfungsi sebagaimana mestinya.

V. PENUTUP

A. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai analisis pola kematian ikan lele pada Budidaya Lele Tiga Saudara menggunakan algoritma Agglomerative Hierarchical Clustering (AHC), maka dapat diambil kesimpulan sebagai berikut:

1. Dataset yang digunakan berjumlah 90 observasi harian dengan tujuh variabel utama: suhu air, pH, jenis pakan, usia ikan, kepadatan tebar, jumlah pakan, dan jumlah kematian harian. Tidak terdapat missing value pada seluruh atribut. Variabel kategorikal jenis pakan diubah menggunakan One Hot Encoding menjadi tiga kolom biner, dan variabel numerik dinormalisasi dengan Min-Max Normalization ke rentang [0, 1] menghasilkan feature matrix (90, 8) sebagai input AHC.
2. Perbandingan tiga metode linkage menghasilkan nilai CCC: Single = 0,9748, Average = 0,9789 (tertinggi), dan Complete = 0,9578. Metode Average Linkage terpilih karena menghasilkan CCC tertinggi yang jauh melampaui ambang batas 0,70, mengonfirmasi bahwa struktur dendrogram sangat akurat merepresentasikan kemiripan data asli.
3. Jumlah klaster optimal ditetapkan k=4 berdasarkan nilai Silhouette Score tertinggi 0,8175 (melampaui target $\geq 0,5$), yang masuk kategori ‘Sangat Baik’ ($\geq 0,70$). Hal ini berarti

pemisahan keempat klaster memiliki batasan yang jelas, akurat, dan bermakna secara matematis.

4. Algoritma AHC berhasil membagi 90 hari observasi ke dalam empat klaster risiko: K-1 (Aman, 25 hari, suhu 33,76°C, pH 7,38, mortalitas 2,68 ekor/hari, pakan Pelet 771); K-2 (Waspada, 15 hari, suhu 34,58°C, pH 7,17, mortalitas 7,80 ekor/hari, pakan PF-1000); K-3 (Bahaya, 20 hari, suhu 36,86°C, pH 6,38, mortalitas 52,20 ekor/hari, pakan Limbah Roti); dan K-4 (Pulih, 30 hari, suhu 34,03°C, pH 7,24, mortalitas 7,90 ekor/hari, kembali ke Pelet 771). Total kematian seluruh siklus mencapai 1.465 ekor, dengan K-3 menyumbang 1.044 ekor (71,3%) hanya dalam 20 hari.
5. Tiga faktor dominan penyebab kematian massal yang berhasil diidentifikasi adalah: (a) pergantian pakan secara mendadak ke Limbah Roti pada H-41 yang memicu lonjakan mortalitas dari 10 menjadi 47 ekor dalam satu malam; (b) penurunan pH secara drastis ke rentang 6,0–6,8 akibat akumulasi bahan organik tidak terstandar; dan (c) kenaikan suhu ekstrem hingga 36,86°C rata-rata yang bersinergi dengan kondisi pH rendah memperburuk kondisi fisiologis ikan hingga puncak 90 ekor/hari di H-56.
6. Klaster K-4 (Pulih) menangkap fenomena efek residual pasca-krisis: meskipun pakan telah dikembalikan ke Pelet 771 dan parameter kualitas air sudah normal sejak H-61, mortalitas harian masih rata-rata 7,90 ekor/hari — hampir tiga kali lipat kondisi K-1. Hal ini menunjukkan adanya kerusakan organ dan kontaminasi substrat kolam yang bersifat residual dengan total 237 ekor kematian tambahan selama 30 hari pemulihan.
7. Uji Validitas Model (White Box Testing) menghasilkan CCC 0,9789 dan Silhouette Score 0,8175, keduanya melampaui target yang ditetapkan. Uji Fungsionalitas (Black Box Testing) terhadap 14 item pengujian pipeline AHC dan dashboard Power BI seluruhnya menunjukkan hasil berhasil, sehingga sistem dinyatakan berfungsi dengan benar sesuai tujuan penelitian.
8. Hasil klasterisasi divisualisasikan melalui dashboard interaktif Microsoft Power BI yang menampilkan scatter plot peta risiko Suhu vs pH, grafik tren mortalitas 90 hari, Card KPI

(total kematian 1.465 ekor, rata-rata 16,28 ekor/hari, 22,2% hari bahaya), dan donut chart proporsi klaster. Dashboard dilengkapi slicer interaktif sehingga dapat digunakan pembudidaya tanpa keahlian pemrograman.

B. Saran

Adapun saran yang dapat diberikan berdasarkan hasil penelitian ini untuk pengembangan ke depan adalah sebagai berikut:

1. Diperlukan pengembangan lebih lanjut dengan menambahkan variabel kualitas air yang lebih lengkap seperti kadar oksigen terlarut (DO), kadar amonia (NH₃), dan turbiditas air agar model klasterisasi dapat memberikan gambaran kondisi budidaya yang lebih komprehensif.
2. Disarankan untuk mengimplementasikan sistem monitoring berbasis IoT yang mengirimkan data suhu dan pH secara real-time ke dashboard Power BI, sehingga pembudidaya dapat menerima peringatan dini ketika kondisi mulai memasuki zona K-2 (Waspada) sebelum berkembang menjadi K-3 (Bahaya).
3. Penelitian selanjutnya disarankan melakukan perbandingan performa AHC dengan metode clustering lain seperti K-Means, DBSCAN, atau Fuzzy C-Means pada dataset yang sama untuk mengetahui metode paling optimal dalam mendeteksi pola kematian ikan lele.
4. Dataset penelitian sebaiknya diperluas mencakup minimal 3 siklus budidaya (270 hari) agar pola klaster yang terbentuk lebih representatif dan dapat digunakan sebagai model prediksi risiko untuk siklus berikutnya.
5. Diperlukan pemeliharaan model dan dashboard secara berkala, terutama setelah penambahan data siklus baru, agar sistem dapat terus digunakan secara berkelanjutan dan disesuaikan dengan perkembangan kondisi budidaya.

DAFTAR PUSTAKA

- [1] Afdan, M. (2023). Pengaruh Kualitas Perairan Terhadap Produksi Ikan Lele Dumbo (*Clarias gariepinus*). *Jurnal Biologi Pubmedia*, 1(1), 1–10.
- [2] Butar, R. P. B. (2023). Analisis hierarchical dan non-hierarchical clustering untuk pengelompokan potensi ekonomi kelautan

- Indonesia 2021. *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, 11(3), 543-553.
- [3] Enterprise, J. (2023). *Python untuk Programmer Pemula*. Jakarta: Elex Media Komputindo.
- [4] Florentia, et al. (2025). Teknik Manajemen Kualitas Air dan Tingkat Kelangsungan Hidup Ikan Lele Dumbo (*Clarias gariepinus*) yang Dibudidayakan dengan Sistem Bioflok. *Jurnal Maiyah*, 4(1), 9–21.
- [5] Head, T., et al. (2023). Interactive Computing with Jupyter and Google Colab for Reproducible Research. *Journal of Data Science and Analytics*, 5(2), 112-125.
- [6] Herlina, A., et al. (2024). Sistem Monitoring Kualitas Air Kolam Lele Berbasis IoT dengan Sistem Tenaga Hibrida. *Jurnal Rekayasa Elektro Sriwijaya*, 6(1), 45-52.
- [7] Kusuma, N. P. D., Tangguda, S., & Lau, J. R. (2024). Analisis Kualitas Air dan Hubungannya dengan Pertumbuhan Ikan Lele Sangkuriang pada Padat Tebar Berbeda. *Journal Galung Tropika*, 13(2), 256–267.
- [8] Permana, I., & Salisah, F. N. S. (2022). Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation: the effect of data normalization on the performance of the classification results of the backpropagation algorithm. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, 2(1), 67-72.
- [9] Pratama, A. Y., Arifin, A. Z., & Wibowo, A. (2024). Implementasi Algoritma Machine Learning Menggunakan Python untuk Prediksi Kualitas Udara. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 8(2), 145-154.
- [10] Primawestri, M., Sumardianto, S., & Kurniasih, R. A. (2023). Karakteristikstik Ikan Lele (*Clarias gariepinus*) dengan Perbedaan Rasio Daging dan Tulang. *Jurnal Ilmu dan Teknologi Perikanan*, 5(1), 44-51.
- [11] Putra, A. P., Tshivana, J., & Rilvani, E. (2025). Perbandingan Teoritis dan Eksperimen Algoritma K-Means dan K-Medoids Dalam Klasterisasi Data. *Kohesi: Jurnal Sains dan Teknologi*, 10(2), 61-70.
- [12] Raya, A. M., et al. (2025). Modeling Mortality Rates Through Hierarchical Clustering for Three-Way Data. *ResearchGate Preprint*.
- [13] Sari, R. M., Wahyuni, S., & Rizka, A. (2023). Implementasi Algoritma K-Means Untuk Guru Berprestasi. *Jurnal Minfo Polgan*, 12(1), 531-538.
- [14] Setiyawan, D. T., Berlilana, & Barkah, A. S. (2025). Analisis Pola Penyebaran Penyakit Menggunakan Pendekatan Clustering Hierarkis dan K-Means. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 10(4), 4251-4260.
- [15] Suhirman, S., & Wintolo, H. (2019). System for Determining Public Health Level Using the Agglomerative Hierarchical Clustering Method. *Compiler*, 8(1), 95-104.
- [16] Susilo, M., & Kurniati, R. (2021). Rancang Bangun Website Toko Online Menggunakan Metode Agile Development Berbasis Python Framework Flask. *Jurnal JTik (Jurnal Teknik Informatika Kaputama)*, 5(1), 123-130.
- [17] The Pandas Development Team. (2024). *Pandas documentation*. Zenodo.
- [18] Virtanen, P., et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17, 261–272.
- [19] Waskom, M. L. (2021). Seaborn: Statistical Data Visualization. *Journal of Open Source Software*, 6(60), 3021.
- [20] Wati, F., & Prasetyo, I. D. (2024). *Penerapan Algoritma Hierarchical Clustering Untuk Visualisasi Angka Kelahiran di DKI Jakarta* (Skripsi). Universitas Mercu Buana.