



Seminar Nasional Ilmu Teknik dan Aplikasi Industri (SINTA)

Homepage: sinta.eng.unila.ac.id



Pemanfaatan Scikit-Fuzzy untuk pengembangan *dataset* sintetis parameter kualitas air budidaya lobster dalam klasifikasi berbasis pembelajaran mesin

Nurrahma Nurrahma ^{a,*}, Herlinawati ^b, Syaiful Alam ^b, Muhammad Fikri ^b

^a Program Studi Teknik Informatika, Universitas Lampung, Jl. Prof. Dr. Ir. Sumantri Brojonegoro No.1, Kota Bandar Lampung, Lampung

^b Program Studi Teknik Elektro, Universitas Lampung, Jl. Prof. Dr. Ir. Sumantri Brojonegoro No.1, Kota Bandar Lampung, Lampung

INFORMASI ARTIKEL

Riwayat artikel:

Diterima 02/11/2025

Direvisi 21/11/2025

Kata kunci:

kualitas air

dataset sintetis

logika fuzzy

Scikit-Fuzzy

pembelajaran mesin

ABSTRAK

Kualitas air merupakan faktor penentu keberhasilan budidaya lobster, sehingga pemantauan dan analisis parameter lingkungan menjadi sangat penting. Namun, keterbatasan data lapangan sering menjadi kendala dalam pengembangan sistem klasifikasi berbasis pembelajaran mesin. Penelitian ini bertujuan mengembangkan *dataset* sintetis parameter kualitas air tambak lobster dengan memanfaatkan logika *fuzzy* menggunakan pustaka Scikit-Fuzzy. Parameter yang digunakan mencakup suhu, pH, kekeruhan, *dissolved oxygen* (DO), dan salinitas. Data diproses menggunakan sistem inferensi *fuzzy* untuk menghasilkan 1.500 data dengan tiga kategori kualitas air, yaitu buruk, sedang, dan baik. *Dataset* sintetis tersebut selanjutnya digunakan dalam pengembangan model pembelajaran mesin melalui tahapan label *encoding*, penyeimbangan data dengan SMOTE, serta pembagian data menjadi 80% untuk pelatihan dan 20% untuk pengujian. Tiga algoritma klasifikasi diterapkan, yaitu *Naïve Bayes*, *Support Vector Machine (SVM)*, dan *Random Forest*. Hasil pengujian menunjukkan bahwa algoritma *Random Forest* memberikan kinerja terbaik dengan akurasi 89,06%, diikuti SVM sebesar 76,60%, sedangkan *Naïve Bayes* menghasilkan akurasi terendah sebesar 61,32%. Temuan ini menunjukkan bahwa *dataset* sintetis berbasis logika *fuzzy* dapat menjadi alternatif sekaligus pelengkap data nyata dalam pengembangan sistem klasifikasi kualitas air pada budidaya lobster. Selain itu, penelitian ini membuka peluang pengembangan lebih lanjut melalui integrasi data lapangan untuk meningkatkan validitas model dan mendukung penerapan praktis di lingkungan budidaya.

1. Pendahuluan

Kualitas air merupakan faktor kunci dalam menentukan keberhasilan budidaya lobster di tambak. Parameter utama seperti suhu, pH, kekeruhan, oksigen terlarut (*dissolved oxygen*), dan salinitas berpengaruh signifikan terhadap pertumbuhan serta tingkat kelangsungan hidup lobster. Perubahan kecil pada parameter-parameter tersebut dapat berdampak langsung pada kesehatan organisme dan produktivitas

tambak. Di Indonesia, pengelolaan budidaya lobster telah diatur melalui Peraturan Menteri Kelautan dan Perikanan Nomor 12/PERMEN-KP/2020, yang menekankan pentingnya pemantauan kondisi lingkungan. Regulasi tersebut merekomendasikan suhu optimal 25—26 °C dan salinitas 30—35 ppt, dengan kualitas perairan yang bebas dari polusi maupun pengaruh air tawar (Prabowo, 2020).

* Penulis korespondensi.

E-mail: nurrahma06@eng.unila.ac.id (N Nurrahma)

Pemanfaatan sistem cerdas dalam bidang akuakultur menawarkan potensi besar untuk meningkatkan efisiensi dan keberlanjutan, tetapi pengembangannya masih terkendala oleh keterbatasan data lapangan (Lu et al., 2025). Pengumpulan data kualitas air memerlukan sensor, peralatan, serta sumber daya yang cukup besar (Jordan and Mitchell, 2015), sehingga *dataset* yang tersedia umumnya terbatas dari segi jumlah maupun variasi. Kondisi ini menjadi hambatan dalam pengembangan model *machine learning*, khususnya untuk tugas klasifikasi yang membutuhkan data beragam dan masif guna mencapai performa optimal. Salah satu pendekatan yang dapat diterapkan untuk mengatasi kendala tersebut adalah pembuatan *dataset* sintesis. Data sintesis, yang dihasilkan secara artifisial melalui algoritma atau simulasi komputer (Bolón-Canedo et al., 2013; Lucini, 2021; Wang et al., 2019), mampu merepresentasikan berbagai kondisi lingkungan berdasarkan pengetahuan domain dan aturan yang telah ditentukan.

Metode logika *fuzzy* dinilai sesuai dengan tujuan tersebut karena kemampuannya dalam merepresentasikan ketidakpastian dan variasi alami pada parameter lingkungan (Arigo et al., 2024). Dengan mendefinisikan fungsi keanggotaan dan aturan berbasis pengetahuan pakar, sistem *fuzzy* dapat menghasilkan nilai parameter kualitas air yang realistis untuk digunakan dalam pelatihan model klasifikasi.

Penelitian ini mengusulkan pemanfaatan pustaka Scikit-Fuzzy berbasis Python untuk membangun sistem inferensi *fuzzy* yang menghasilkan *dataset* sintesis parameter kualitas air tambak lobster. Fokus penelitian mencakup perancangan variabel *fuzzy*, penyusunan aturan inferensi, serta konstruksi *dataset* sintesis yang dapat digunakan untuk klasifikasi kualitas air. Kontribusi utama penelitian ini adalah penyediaan pendekatan sistematis dalam pembuatan *dataset* sintesis berbasis logika *fuzzy*, yang diharapkan dapat menjadi alternatif maupun pelengkap data empiris guna mendukung penelitian dan penerapan sistem cerdas pada akuakultur, khususnya budidaya lobster.

2. Studi pustaka

2.1. Kualitas Air Budidaya Lobster

Kualitas lingkungan perairan merupakan faktor kunci yang menentukan keberlanjutan dan keberhasilan budidaya benih lobster (Erlania et al., 2014). Perairan yang kaya akan plankton menyediakan sumber nutrisi penting bagi larva *phyllosoma* dan *puerulus*, sehingga mendukung pertumbuhan serta perkembangan pada fase awal kehidupan lobster (Junaidi et al., 2011). Kesuburan perairan menjadi salah satu parameter utama yang harus diperhatikan, karena ekosistem tambak lobster terdiri atas komponen abiotik dan biotik yang saling

berinteraksi melalui aliran energi dan siklus hara. Ketidakseimbangan interaksi tersebut dapat menyebabkan terganggunya stabilitas ekosistem.

Penilaian kualitas air umumnya mencakup parameter fisik, kimia, dan biologis. Parameter fisik meliputi suhu dan kekeruhan, sedangkan parameter kimia meliputi pH, oksigen terlarut, dan salinitas. Sementara itu, parameter biologis mencakup keanekaragaman mikroorganisme yang berperan dalam menjaga keseimbangan ekosistem perairan. Status kualitas air menggambarkan tingkat pencemaran atau kondisi lingkungan perairan pada periode tertentu. Perubahan signifikan pada parameter kualitas air dapat berdampak negatif terhadap kesehatan organisme akuatik dan menurunkan produktivitas sistem budidaya.

2.2. Dataset Sintesis pada Penelitian

Dalam pengembangan sistem berbasis pembelajaran mesin, ketersediaan *dataset* yang besar, representatif, dan berkualitas merupakan prasyarat penting untuk memperoleh model yang akurat. Namun, dalam bidang akuakultur, proses pengumpulan data sering kali terkendala oleh keterbatasan biaya, waktu, serta ketersediaan perangkat sensor (Jordan and Mitchell, 2015). Untuk mengatasi kendala tersebut, *dataset* sintesis dapat dimanfaatkan sebagai alternatif atau pelengkap terhadap data nyata. *Dataset* sintesis dihasilkan secara artifisial melalui simulasi, aturan berbasis pengetahuan, atau model statistik tertentu, sehingga mampu merepresentasikan variasi kondisi lingkungan secara lebih komprehensif (Lucini, 2021).

2.3. Logika Fuzzy untuk Representasi Ketidakpastian

Logika *fuzzy* merupakan salah satu pendekatan komputasi cerdas yang dikembangkan untuk mengatasi ketidakpastian dan ambiguitas dalam proses pengambilan keputusan. Konsep utamanya terletak pada fungsi keanggotaan (*membership functions*) yang mengubah data numerik menjadi variabel linguistik, seperti "rendah", "sedang", atau "tinggi" (Ross, 2016). Melalui basis aturan (*rule base*) yang dinyatakan dalam bentuk pernyataan "JIKA-MAKA" (*IF-THEN*), sistem *fuzzy* dapat melakukan proses inferensi yang menyerupai pola penalaran manusia (Pedrycz and Gomide, 1998).

Dalam konteks evaluasi kualitas air, logika *fuzzy* dinilai efektif karena mampu merepresentasikan variasi alami yang sulit dijelaskan melalui batasan numerik konvensional. Oleh karena itu, pendekatan ini dianggap ideal untuk menghasilkan *dataset* sintesis yang realistis dan mencerminkan kondisi lingkungan yang sebenarnya.

2.4. Pustaka Scikit-Fuzzy

Scikit-Fuzzy merupakan pustaka *open-source* berbasis Python yang menyediakan berbagai alat untuk membangun sistem inferensi *fuzzy*. Pustaka ini mendukung proses fuzzifikasi, perancangan fungsi keanggotaan, penyusunan basis aturan, hingga tahap defuzzifikasi (Scikit-fuzzy, n.d.).

Keunggulan utama Scikit-Fuzzy terletak pada fleksibilitasnya serta integrasi yang baik dengan ekosistem komputasi ilmiah Python, seperti NumPy dan SciPy. Dalam penelitian ini, Scikit-Fuzzy digunakan sebagai perangkat utama untuk menghasilkan *dataset* sintesis parameter kualitas air yang relevan bagi kebutuhan budidaya lobster.

3. Metodologi

Penelitian ini berfokus pada pengembangan *dataset* sintesis parameter kualitas air budidaya lobster dengan menggunakan pendekatan logika *fuzzy*. Metode ini bertujuan untuk mengatasi keterbatasan ketersediaan data aktual melalui pembuatan *dataset* yang bersifat realistis dan memiliki variasi representatif terhadap kondisi lingkungan perairan. Implementasi sistem dilakukan dengan memanfaatkan pustaka Scikit-Fuzzy dalam lingkungan pemrograman Python.

3.1. Pendefinisian Variabel Fuzzy

Tahap awal penelitian ini adalah pendefinisian variabel masukan (*antecedent*) dan variabel keluaran (*consequent*) yang digunakan dalam sistem inferensi *fuzzy*. Pada penelitian ini, variabel masukan terdiri atas lima parameter utama kualitas air, yaitu:

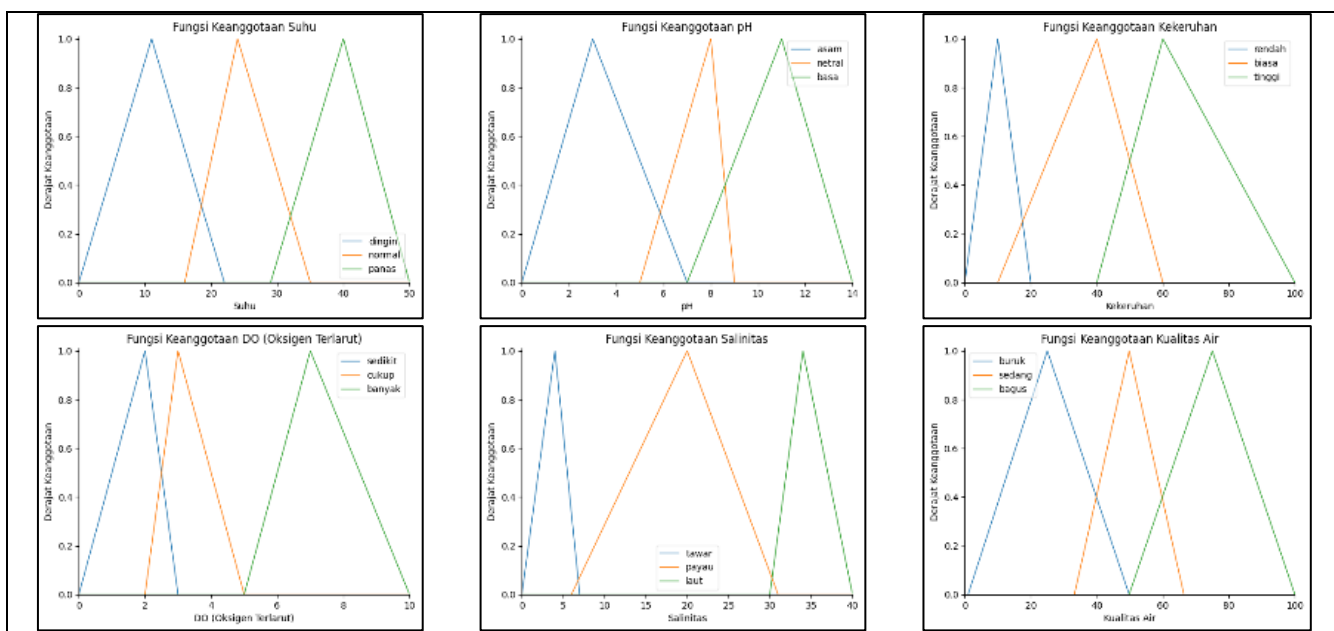
- Suhu (°C): direpresentasikan dalam rentang [0, 50].
 - pH: direpresentasikan dalam rentang [0, 14].
 - Kekeruhan (NTU): direpresentasikan dalam rentang [0, 100].
 - *Dissolved Oxygen* (DO) (mg/L): direpresentasikan dalam rentang [0, 10].
 - Salinitas: direpresentasikan dalam rentang [0, 40].
- Sementara itu, variabel keluaran adalah Kualitas Air, yang dalam skala [0, 100]. Implementasi pendefinisian variabel-variabel tersebut dilakukan menggunakan pustaka Scikit-Fuzzy.

3.2. Penyusunan Fungsi Keanggotaan

Pada tahap ini, setiap variabel *fuzzy* diberikan fungsi keanggotaan untuk mengubah nilai numerik menjadi istilah linguistik. Jenis fungsi keanggotaan yang digunakan adalah fungsi segitiga (*trimf*), karena memiliki bentuk sederhana namun efektif dalam merepresentasikan transisi antar nilai. Implementasi fungsi keanggotaan tersebut dilakukan menggunakan pustaka Scikit-Fuzzy, dengan visualisasi grafik fungsi keanggotaan disajikan pada Gambar 1.

Label linguistik yang digunakan untuk masing-masing variabel adalah sebagai berikut:

- Suhu: Dingin, Normal, Panas.
- pH: Asam, Netral, Basa.
- Kekeruhan: Rendah, Biasa, Tinggi.
- DO: Sedikit, Cukup, Banyak.
- Salinitas: Tawar, Payau, Laut.
- Kualitas Air: Buruk, Sedang, Baik.



Gambar 1. Grafik fungsi keanggotaan.

3.3. Pengembangan Basis Aturan (Rule Base)

Basis aturan (*rule base*) disusun dalam bentuk pernyataan IF-THEN yang merepresentasikan hubungan antara variabel masukan dan variabel keluaran dalam sistem inferensi fuzzy. Pada penelitian ini, dibentuk sebanyak 243 aturan (3^5) untuk mencakup seluruh kombinasi kondisi dari lima variabel masukan.

Sebagai contoh, salah satu aturan yang dirumuskan adalah: "JIKA suhu dingin DAN pH asam DAN kekeruhan rendah DAN DO sedikit DAN salinitas payau MAKA kualitas air buruk". Seluruh aturan tersebut diimplementasikan secara terintegrasi dalam sistem inferensi fuzzy untuk menghasilkan model yang komprehensif dan mampu merepresentasikan kondisi lingkungan perairan secara lebih realistis.

3.4. Generasi Data Sintesis

Proses generasi data sintesis dilakukan secara otomatis melalui simulasi sistem inferensi *fuzzy* yang telah dibangun menggunakan pustaka Scikit-Fuzzy. Proses generasi dilakukan untuk 1500 data sintesis yang diperoleh melalui pengambilan nilai acak untuk setiap variabel masukan. Setiap kombinasi nilai masukan tersebut kemudian diproses dalam sistem simulasi *fuzzy*.

Berdasarkan basis aturan yang telah ditetapkan, sistem melakukan proses inferensi untuk menghasilkan nilai numerik tunggal pada variabel keluaran, yaitu kualitas air. Selanjutnya, pasangan nilai masukan dan keluaran tersebut disimpan sebagai satu entri dalam *dataset*. Data yang dihasilkan mencakup parameter suhu, pH, kekeruhan, DO, salinitas, serta nilai numerik kualitas air.

3.5. Klasifikasi dan Penyimpanan Dataset

Setelah nilai numerik kualitas air dihasilkan, tahap selanjutnya adalah proses klasifikasi dan penyimpanan data. Nilai numerik tersebut dikonversi menjadi label kategoris, yaitu buruk, sedang, dan baik, berdasarkan rentang skor yang telah ditetapkan. Proses konversi ini dilakukan agar *dataset* dapat langsung dimanfaatkan dalam pelatihan model klasifikasi.

Seluruh data masukan dan keluaran, termasuk label kategoris, disusun dalam bentuk *DataFrame* menggunakan pustaka Pandas. Selanjutnya, *DataFrame* tersebut diekspor ke dalam format CSV (.csv) agar dapat digunakan lebih lanjut dalam proses pelatihan model *machine learning*. Dataset sintesis yang dihasilkan terdiri dari 1500 data, dengan distribusi kelas sebagai berikut: 882 data kategori buruk, 342 data kategori sedang, dan 276 data kategori baik.

3.6. Pengembangan Model Supervised Machine Learning

Dataset sintesis yang telah dihasilkan digunakan untuk melatih model klasifikasi berbasis *supervised machine learning*. Proses pengembangan model meliputi beberapa tahapan sebagai berikut:

- *Encoding* Label Kualitas. Tahapan ini dilakukan dengan mengonversi kategori kualitas air non-numerik (buruk, sedang, baik) menjadi format numerik agar dapat diproses oleh algoritma *machine learning*. Proses *encoding* ini memastikan model mampu mengenali pola data berdasarkan representasi numerik yang konsisten.
- *Oversampling* dengan SMOTE. Untuk mengatasi ketidakseimbangan jumlah data antar kelas, digunakan metode *Synthetic Minority Oversampling Technique* (SMOTE). Teknik ini menghasilkan sampel sintesis baru berdasarkan data kelas minoritas dengan memanfaatkan algoritma *k-nearest neighbors* sehingga distribusi kelas menjadi lebih seimbang [13]. Pendekatan ini meningkatkan kemampuan model dalam mengenali pola pada kelas minoritas, yang sangat penting dalam klasifikasi kualitas air.
- Pembagian Data *Training dan Testing*. Dataset dibagi menjadi dua bagian, yaitu data pelatihan (80%) dan data pengujian (20%). Data pelatihan digunakan untuk membangun model, sedangkan data pengujian digunakan untuk mengevaluasi kinerja model terhadap data yang belum pernah dilihat sebelumnya.
- Normalisasi Data. Proses normalisasi dilakukan untuk menyeragamkan skala antar variabel sehingga seluruh fitur berada dalam rentang nilai yang sebanding. Langkah ini mempercepat konvergensi model, mengurangi pengaruh *outlier*, serta meningkatkan akurasi hasil klasifikasi.
- Pembuatan Model Pembelajaran Mesin. Beberapa algoritma klasifikasi diterapkan untuk menguji performa *dataset* sintesis, yaitu *Naïve Bayes*, *Support Vector Machine* (SVM), dan *Random Forest* [14], [15]. Algoritma *Naïve Bayes* didasarkan pada Teorema Bayes dengan asumsi independensi antarfitur. Meskipun sederhana, algoritma ini efisien dan mampu memberikan kinerja yang baik dengan kebutuhan data pelatihan yang relatif sedikit [16]. *Support Vector Machine* (SVM) digunakan untuk mencari *hyperplane* optimal yang memisahkan kelas dengan margin maksimum [16]. Sementara itu, *Random Forest* menggabungkan sejumlah pohon keputusan untuk meningkatkan akurasi dan mengurangi risiko *overfitting* [17].
- Pelatihan dan Pengujian Model. Model yang dipilih dilatih menggunakan data pelatihan untuk

mempelajari hubungan antara variabel masukan (suhu, pH, kekeruhan, DO, salinitas) dengan label keluaran (kualitas_label). Setelah proses pelatihan selesai, model diuji menggunakan data pengujian untuk menilai kemampuan generalisasi terhadap data baru. Evaluasi kinerja dilakukan menggunakan

metrik akurasi (*accuracy*), yang menggambarkan tingkat keberhasilan model dalam mengklasifikasikan kualitas air secara tepat. Hasil pelatihan dan pengujian masing-masing model ditampilkan pada Gambar 2.

<p>Naive Bayes - Training Accuracy: 60.63% Naive Bayes - Testing Accuracy: 61.32%</p> <p>Classification report:</p> <table border="1"> <thead> <tr> <th></th> <th>precision</th> <th>recall</th> <th>f1-score</th> <th>support</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0.62</td> <td>0.66</td> <td>0.64</td> <td>172</td> </tr> <tr> <td>1</td> <td>0.49</td> <td>0.40</td> <td>0.44</td> <td>179</td> </tr> <tr> <td>2</td> <td>0.69</td> <td>0.78</td> <td>0.73</td> <td>179</td> </tr> <tr> <td>accuracy</td> <td></td> <td></td> <td>0.61</td> <td>530</td> </tr> <tr> <td>macro avg</td> <td>0.60</td> <td>0.61</td> <td>0.61</td> <td>530</td> </tr> <tr> <td>weighted avg</td> <td>0.60</td> <td>0.61</td> <td>0.61</td> <td>530</td> </tr> </tbody> </table>		precision	recall	f1-score	support	0	0.62	0.66	0.64	172	1	0.49	0.40	0.44	179	2	0.69	0.78	0.73	179	accuracy			0.61	530	macro avg	0.60	0.61	0.61	530	weighted avg	0.60	0.61	0.61	530	<p>SVM - Training Accuracy: 80.29% SVM - Testing Accuracy: 76.60%</p> <p>Classification report:</p> <table border="1"> <thead> <tr> <th></th> <th>precision</th> <th>recall</th> <th>f1-score</th> <th>support</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0.83</td> <td>0.73</td> <td>0.78</td> <td>172</td> </tr> <tr> <td>1</td> <td>0.68</td> <td>0.70</td> <td>0.69</td> <td>179</td> </tr> <tr> <td>2</td> <td>0.79</td> <td>0.87</td> <td>0.83</td> <td>179</td> </tr> <tr> <td>accuracy</td> <td></td> <td></td> <td>0.77</td> <td>530</td> </tr> <tr> <td>macro avg</td> <td>0.77</td> <td>0.77</td> <td>0.77</td> <td>530</td> </tr> <tr> <td>weighted avg</td> <td>0.77</td> <td>0.77</td> <td>0.77</td> <td>530</td> </tr> </tbody> </table>		precision	recall	f1-score	support	0	0.83	0.73	0.78	172	1	0.68	0.70	0.69	179	2	0.79	0.87	0.83	179	accuracy			0.77	530	macro avg	0.77	0.77	0.77	530	weighted avg	0.77	0.77	0.77	530	<p>Random Forest - Training Accuracy: 100.00% Random Forest - Testing Accuracy: 89.06%</p> <p>Classification report:</p> <table border="1"> <thead> <tr> <th></th> <th>precision</th> <th>recall</th> <th>f1-score</th> <th>support</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0.93</td> <td>0.87</td> <td>0.90</td> <td>172</td> </tr> <tr> <td>1</td> <td>0.83</td> <td>0.88</td> <td>0.86</td> <td>179</td> </tr> <tr> <td>2</td> <td>0.92</td> <td>0.92</td> <td>0.92</td> <td>179</td> </tr> <tr> <td>accuracy</td> <td></td> <td></td> <td>0.89</td> <td>530</td> </tr> <tr> <td>macro avg</td> <td>0.89</td> <td>0.89</td> <td>0.89</td> <td>530</td> </tr> <tr> <td>weighted avg</td> <td>0.89</td> <td>0.89</td> <td>0.89</td> <td>530</td> </tr> </tbody> </table>		precision	recall	f1-score	support	0	0.93	0.87	0.90	172	1	0.83	0.88	0.86	179	2	0.92	0.92	0.92	179	accuracy			0.89	530	macro avg	0.89	0.89	0.89	530	weighted avg	0.89	0.89	0.89	530
	precision	recall	f1-score	support																																																																																																							
0	0.62	0.66	0.64	172																																																																																																							
1	0.49	0.40	0.44	179																																																																																																							
2	0.69	0.78	0.73	179																																																																																																							
accuracy			0.61	530																																																																																																							
macro avg	0.60	0.61	0.61	530																																																																																																							
weighted avg	0.60	0.61	0.61	530																																																																																																							
	precision	recall	f1-score	support																																																																																																							
0	0.83	0.73	0.78	172																																																																																																							
1	0.68	0.70	0.69	179																																																																																																							
2	0.79	0.87	0.83	179																																																																																																							
accuracy			0.77	530																																																																																																							
macro avg	0.77	0.77	0.77	530																																																																																																							
weighted avg	0.77	0.77	0.77	530																																																																																																							
	precision	recall	f1-score	support																																																																																																							
0	0.93	0.87	0.90	172																																																																																																							
1	0.83	0.88	0.86	179																																																																																																							
2	0.92	0.92	0.92	179																																																																																																							
accuracy			0.89	530																																																																																																							
macro avg	0.89	0.89	0.89	530																																																																																																							
weighted avg	0.89	0.89	0.89	530																																																																																																							

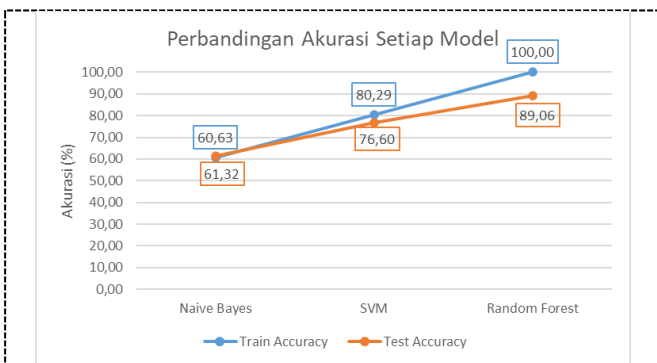
Gambar 2. Akurasi training dan testing setiap model

4. Hasil dan pembahasan

Pada penelitian ini, *dataset* sintesis dihasilkan dengan memanfaatkan pustaka *open-source* Python Scikit-Fuzzy, yang menyediakan fitur lengkap untuk membangun sistem inferensi *fuzzy*. *Dataset* yang dihasilkan mencakup lima parameter utama kualitas air, yaitu suhu, pH, kekeruhan, dissolved oxygen (DO), salinitas, serta nilai numerik kualitas air. *Dataset* sintesis tersebut kemudian digunakan untuk mengembangkan model *machine learning* dalam klasifikasi kualitas air. Tiga algoritma klasifikasi yang diterapkan adalah *Naive Bayes*, *Support Vector Machine* (SVM), dan *Random Forest*. Hasil proses pelatihan dan pengujian dari ketiga model tersebut disajikan pada Tabel 1 dan divisualisasikan pada Gambar 3

Tabel 1. Hasil akurasi training dan testing setiap model

Model	Train Accuracy	Test Accuracy
Naive Bayes	60,63%	61,32%
SVM	80,29%	76,60%
Random Forest	100,00%	89,06%



Gambar 3. Hasil akurasi training dan testing setiap model

Hasil pengujian menunjukkan adanya variasi performa yang signifikan di antara ketiga algoritma yang diuji. Model *Naive Bayes* menghasilkan akurasi pelatihan sebesar 60,63% dan akurasi pengujian sebesar 61,32%. Nilai akurasi yang relatif rendah dan perbedaan yang kecil antara data latih dan data uji menunjukkan bahwa model ini memiliki keterbatasan dalam mengenali pola kompleks antarp parameter kualitas air. Model *Support Vector Machine* (SVM) menunjukkan kinerja yang lebih baik dengan akurasi pelatihan sebesar 80,29% dan akurasi pengujian sebesar 76,60%. Selisih akurasi yang relatif kecil mengindikasikan kemampuan generalisasi yang baik serta stabilitas model dalam menghadapi variasi data. Sementara itu, model *Random Forest* menunjukkan performa tertinggi dengan akurasi pelatihan mencapai (100,00%) dan akurasi pengujian sebesar 89,06%. Meskipun demikian, perbedaan yang cukup besar antara hasil pelatihan dan pengujian mengindikasikan potensi *overfitting*, di mana model terlalu menyesuaikan diri pada data pelatihan sehingga kinerjanya menurun pada data uji.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa *dataset* sintesis berbasis logika fuzzy dapat dimanfaatkan secara efektif dalam pengembangan model klasifikasi kualitas air tambak lobster. Model SVM dan *Random Forest* menunjukkan tingkat akurasi yang tinggi, meskipun diperlukan perhatian terhadap risiko *overfitting* pada *Random Forest* serta keterbatasan performa pada *Naive Bayes*. Untuk penelitian selanjutnya disarankan untuk penerapan metode validasi yang lebih komprehensif, seperti *k-fold cross-validation*, serta penggunaan metrik evaluasi tambahan seperti *precision*, *recall*, dan *F1-score*. Selain itu, kombinasi antara dataset sintesis dan data empiris lapangan juga direkomendasikan guna meningkatkan reliabilitas serta relevansi model dalam aplikasi pemantauan kualitas air secara nyata

5. Kesimpulan

Penelitian ini berhasil merancang pendekatan sistematis untuk menghasilkan dataset sintesis parameter kualitas air tambak lobster dengan memanfaatkan logika fuzzy melalui pustaka *open-source* Python Scikit-Fuzzy. Dataset yang dihasilkan mencakup lima parameter utama, yaitu suhu, pH, kekeruhan, *dissolved oxygen* (DO), salinitas, serta nilai numerik kualitas air dengan total 1.500 data. Distribusi awal *dataset* menunjukkan adanya ketidakseimbangan kelas, di mana kategori kualitas buruk mendominasi sebanyak 882 data (58,8%), kualitas sedang 342 data (22,8%), dan kualitas baik 276 data (18,4%). Ketidakseimbangan tersebut berhasil diatasi dengan penerapan teknik *Synthetic Minority Over-sampling Technique* (SMOTE), sehingga distribusi data menjadi lebih seimbang dan representatif untuk proses pelatihan model.

Hasil pengujian terhadap tiga algoritma *machine learning* menunjukkan bahwa *Random Forest* memberikan performa terbaik dengan akurasi pelatihan sebesar 100,00% dan akurasi pengujian 89,06%. Model SVM menempati posisi kedua dengan akurasi pelatihan 80,29% dan pengujian 76,60%, sedangkan *Naive Bayes* menunjukkan kinerja terendah dengan akurasi pelatihan 60,63% dan pengujian 61,32%. Temuan ini mengindikasikan bahwa model berbasis *ensemble* seperti *Random Forest* lebih efektif dalam menangkap kompleksitas data sintesis, meskipun memiliki potensi risiko *overfitting*.

Secara keseluruhan, hasil penelitian ini menegaskan bahwa *dataset* sintesis berbasis logika *fuzzy* dapat berfungsi sebagai alternatif maupun pelengkap data empiris dalam pengembangan model klasifikasi kualitas air tambak lobster. Pendekatan ini berpotensi mengurangi ketergantungan terhadap data lapangan yang terbatas dan memerlukan biaya tinggi, sekaligus mempercepat proses penelitian dan pengembangan sistem cerdas di bidang akuakultur.

Untuk penelitian selanjutnya, disarankan penerapan metode validasi yang lebih mendalam, seperti *k-fold cross-validation*, serta penggunaan metrik evaluasi tambahan seperti *precision*, *recall*, dan *F1-score*. Selain itu, integrasi antara data sintesis dan data nyata perlu dilakukan guna meningkatkan akurasi, reliabilitas, serta relevansi hasil terhadap kondisi budidaya di lapangan.

Ucapan terima kasih

Penulis mengucapkan terima kasih kepada Universitas Lampung dan Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Lampung atas pendanaan Hibah BLU Tahun 2025 Skema Penelitian Dasar dengan Nomor Kontrak 658/UN26.21/PN/2025, serta dukungan baik dalam

bentuk fasilitas, maupun arahan, sehingga penelitian ini dapat terlaksana dengan baik.

Daftar Pustaka

- Arigo, M.A.A.H., Sofi Berliana Rizky, Zainu Rafsanjani, Isa Rachman, Rini Indarti, Noorman Rinanto, Agus Khumaidi, 2024. Optimasi Penghematan Energi Listrik Menggunakan Metode Fuzzy Logic Pada Sistem Pendingin Udara Berbasis IoT. *J. Elektron. dan Otomasi Ind.* 11, 363–375. <https://doi.org/10.33795/elkolind.v11i2.5467>
- Bolón-Canedo, V., Sánchez-Maróño, N., Alonso-Betanzos, A., 2013. A review of feature selection methods on synthetic data. *Knowl. Inf. Syst.* 34, 483–519. <https://doi.org/10.1007/s10115-012-0487-8>
- Erlania, Radiarta, I.N., Sugama, K., 2014. Dynamics of lobster (*Panulirus* spp.) seeds abundance in Gerupuk Bay, West Nusa Tenggara: A challenge for lobster aquaculture technology development. *J. Ris Akuakultur* 9, 475–486.
- Jordan, M.I., Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. *Science* (80-). 349, 255–260. <https://doi.org/10.1126/science.aaa8415>
- Junaidi, M., Cokrowati, N., Abidin, Z., 2011. Tingkah Laku Induk Betina Selama Proses Pengeraman Telur Dan Perkembangan Larva Lobster Pasir. *J. Akuatika* 2, 1–10.
- Lu, Y., Chen, L., Zhang, Y., Shen, M., Wang, H., Wang, X., van Rechem, C., Fu, T., Wei, W., 2025. Machine Learning for Synthetic Data Generation: A Review. *J. LATEX Cl. FILES* 14, 1–18.
- Lucini, F., 2021. The real deal about synthetic data. *MIT Sloan Manag. Rev.* 63, 1–4.
- Pedrycz, W., Gomide, F., 1998. *An Introduction to Fuzzy Sets: Analysis and Design*. MIT Press.
- Prabowo, E., 2020. Peraturan Menteri Kelautan dan Perikanan Republik Indonesia Nomor 12/PERMEN-KP/2020 Tentang Pengelolaan Lobster (*Panulirus* Spp.), Kepiting (*Scylla* Spp.), dan Rajungan (*Portunus* Spp.) Di Wilayah Negara Republik Indonesia. Indonesia.
- Ross, T.J., 2016. *Fuzzy logic with engineering applications*, 4th ed. Wiley.
- Scikit-fuzzy, n.d. *skfuzzy 0.3 docs* [WWW Document]. URL <https://scikit-fuzzy.readthedocs.io/en/latest/>
- Wang, Q., Gao, J., Lin, W., Yuan, Y., 2019. Learning From Synthetic Data for Crowd Instance Segmentation in the Wild, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). <https://doi.org/10.1109/ICIP46576.2022.9897547>