Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1373−1382
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.5998

Model Prediksi Kualitas Air Untuk Budidaya Ikan Lele Dengan Algoritma Extreme Gradient Boosting

Mokhammad Irvan Maulana*, Fajar Nugraha, Arif Setiawan

Fakultas Teknik, Program Studi Sistem Informasi, Universitas Muria Kudus, Kudus, Indonesia
Email: [1,*]202053011@std.umk.ac.id, [2]fajar.nugraha@umk.ac.id, [3]arif.setiawan@umk.ac.id
Email Penulis Korespondensi: 202053011@std.umk.ac.id
Submitted: **30/09/2024**; Accepted: **07/11/2024**; Published**: 03/12/2024**

Abstrak−Pertumbuhan sektor akuakultur di Indonesia khususnya budidaya ikan lele mengalami peningkatan signifikan, tetapi tantangan utama yang dihadapi adalah kebutuhan untuk prediksi perkembangan ikan yang akurat guna mengoptimalkan produksi dan meminimalkan risiko kerugian. Penelitian ini bertujuan mengembangkan sistem prediksi pertumbuhan ikan lele berbasis *machine learning* menggunakan algoritma XGBoost, yang mempertimbangkan faktor lingkungan kritis seperti kualitas air (suhu, pH, kadar oksigen terlarut, amonia, dan nitrat). Dengan sistem ini, petani lele dapat memantau kualitas air secara *real-time*, memungkinkan mereka untuk mengambil tindakan pencegahan yang tepat waktu dan optimal dalam pemberian pakan, sehingga meningkatkan hasil panen dan menekan biaya operasional. Model XGBoost menunjukkan performa baik dengan nilai *Mean Absolute Error (MAE)* sebesar 0,073 untuk berat ikan dan 14,66 untuk panjang ikan, *Mean Squared Error (MSE)* sebesar 0,123 untuk berat ikan dan 1,278 untuk panjang ikan, serta R^2 mencapai 0,998 untuk kedua variabel yang menandakan akurasi tinggi dalam prediksi pertumbuhan ikan. Diharapkan, penelitian ini tidak hanya meningkatkan produktivitas dan efisiensi budidaya lele, tetapi juga mendukung transformasi digital di sektor perikanan Indonesia, memberikan keuntungan kompetitif bagi petani dalam menghadapi tantangan industri yang semakin kompleks.

Kata Kunci: Akuakultur; Budidaya Ikan Lele; Prediksi Pertumbuhan Ikan ; Machine Learning; XGBoost

Abstract−The growth of the aquaculture sector in Indonesia, particularly in catfish farming, has experienced significant increases. However, a major challenge faced is the need for accurate predictions of fish development to optimize production and minimize the risk of losses. This study aims to develop a growth prediction system for catfish based on machine learning using the XGBoost algorithm, which considers critical environmental factors such as water quality (temperature, pH, dissolved oxygen, ammonia, and nitrate). With this system, catfish farmers can monitor water quality in real-time, allowing them to take timely and optimal preventive actions regarding feed provision, thereby improving harvest yields and reducing operational costs. The XGBoost model demonstrates good performance with a Mean Absolute Error (MAE) of 0.073 for fish weight and 14.66 for fish length, a Mean Squared Error (MSE) of 0.123 for fish weight and 1.278 for fish length, and an R^2 value of 0.998 for both variables, indicating high accuracy in predicting fish growth. It is expected that this research will not only enhance productivity and efficiency in catfish farming but also support digital transformation in Indonesia's fisheries sector, providing a competitive advantage for farmers in facing increasingly complex industry challenges.

Keywords: Aquaculture; Catfish Farming; Fish Growth Prediction; Machine Learning; XGBoost

1. PENDAHULUAN

Pertumbuhan sektor akuakultur di Indonesia [1] mengalami peningkatan signifikan, terutama karena tingginya permintaan pasar terhadap ikan lele, salah satu komoditas dengan nilai ekonomi tinggi. Namun, salah satu tantangan utama dalam budidaya ikan lele adalah memprediksi perkembangan ikan secara akurat untuk mengoptimalkan produksi dan mengurangi risiko kerugian [2]. Faktor-faktor lingkungan seperti kualitas air, suhu, oksigen terlarut, serta kandungan senyawa nitrogen sangat berpengaruh terhadap pertumbuhan ikan lele [3]. Oleh karena itu, teknologi yang mampu memantau dan menganalisis data secara efisien dan real-time sangat diperlukan untuk mendukung pengambilan keputusan yang lebih tepat dalam budidaya ikan lele.

Selama ini, pemantauan kualitas air dalam budidaya ikan lele masih bergantung pada metode manual atau semi-otomatis[4]. Meskipun teknologi berbasis sensor telah diterapkan, kebanyakan sistem tersebut masih belum mampu memberikan data secara real-time atau dalam skala yang memadai untuk mengukur seluruh faktor lingkungan secara menyeluruh. Hal ini menyebabkan ketidakakuratan dalam memprediksi kondisi lingkungan seperti suhu air, kadar oksigen terlarut, dan konsentrasi senyawa nitrogen (amonia, nitrit, nitrat), yang pada gilirannya dapat menurunkan produktivitas serta meningkatkan mortalitas ikan. Sistem pemantauan manual dan sensor sederhana masih terbatas dalam mengatasi kompleksitas dan perubahan dinamika lingkungan, sehingga risiko kerugian bagi petani ikan tetap tinggi [5].

Penelitian sebelumnya umumnya terbatas pada pemantauan parameter tunggal, seperti suhu atau pH yang tidak memperhitungkan interaksi kompleks antara berbagai faktor lingkungan yang dapat memengaruhi pertumbuhan ikan lele secara bersamaan. Ketidakmampuan untuk menganalisis parameter ini dalam konteks yang lebih luas mengakibatkan kurangnya gambaran menyeluruh mengenai kondisi lingkungan, sehingga menyulitkan petani dalam mengambil keputusan yang tepat terkait pemeliharaan ikan. Selain itu sistem yang ada cenderung tidak mampu memberikan prediksi secara *real-time*, yang sangat penting untuk mengambil tindakan preventif yang diperlukan dalam menghadapi fluktuasi kualitas air [6]. Untuk mengatasi masalah ini, diperlukan pendekatan baru yang lebih holistik dan akurat dalam memprediksi kondisi lingkungan, termasuk penggunaan algoritma *machine learning* yang

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1373−1382
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.5998

dapat menganalisis dan mengintegrasikan berbagai faktor secara bersamaan, sehingga petani dapat memanfaatkan data secara lebih efektif untuk meningkatkan hasil panen dan menjaga kesehatan ikan.

Salah satu solusi yang ditawarkan adalah penerapan teknologi *machine learning*, yang memiliki kemampuan untuk menganalisis data historis tanpa membutuhkan pemrograman eksplisit [7]. Algoritma *machine learning* mampu memproses berbagai faktor lingkungan seperti kualitas air, suhu, oksigen terlarut, dan pakan untuk memprediksi pertumbuhan ikan dengan lebih akurat [8]. Prediksi ini memberikan nilai tambah bagi petani dalam membuat keputusan yang tepat, dalam hal pengelolaan kualitas air, yang pada akhirnya dapat meningkatkan produktivitas dan efisiensi budidaya.

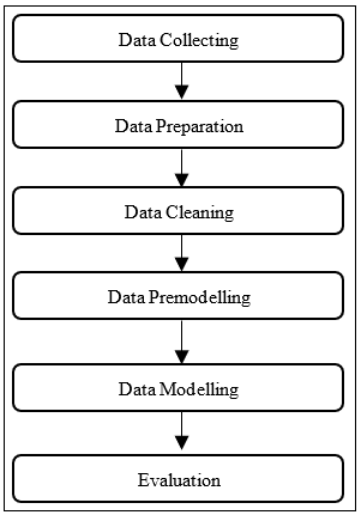
Algoritma XGBoos**t** *(Extreme Gradient Boosting)* adalah salah satu metode *machine learning* yang terbukti efektif dalam meningkatkan akurasi prediksi, terutama saat dihadapkan dengan data yang kompleks dan tidak seimbang [9]. XGBoost bekerja dengan menggabungkan beberapa pohon keputusan dan menggunakan teknik boosting secara bertahap untuk mengurangi kesalahan prediksi [10]. Dibandingkan dengan algoritma lain, XGBoost unggul dalam hal kecepatan, efisiensi, dan kemampuan menangani data dengan distribusi yang tidak merata, yang umum terjadi dalam pengelolaan data lingkungan budidaya ikan [11]. Dalam konteks budidaya ikan lele, XGBoost dapat digunakan untuk memprediksi kondisi optimal dengan mempertimbangkan berbagai parameter lingkungan secara bersamaan, seperti suhu air, pH, kadar oksigen terlarut, nitrit dan nitrat. Penerapan algoritma ini diharapkan mampu meningkatkan ketepatan prediksi serta mendukung pengambilan keputusan yang lebih cepat dan tepat, khususnya dalam mengelola kualitas air dalam budidaya ikan lele.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan Sistem Prediksi Kualitas Air untuk Budidaya Ikan Lele dengan menggunakan algoritma XGBoost. Implementasi algoritma ini diharapkan tidak hanya mampu meningkatkan akurasi prediksi, tetapi juga membantu petani ikan dalam mengoptimalkan pengelolaan kualitas air. Dengan demikian, penelitian ini dapat berkontribusi dalam meningkatkan efisiensi dan produktivitas sektor akuakultur di Indonesia serta mendukung transformasi digital dalam budidaya ikan lele.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Melalui penelitian tetang prediksi kualitas air untuk budidaya ikan lele, pada Gambar 1 menunjukan tahapan tahapan penting dalam proses Data Mining, yang akan membantu menghasilkan prediksi yang lebih akurat untuk memprediksi kualitas air untuk budidaya ikan lele.



Gambar 1. Alur Tahapan Penelitian

Gambar 1 menunjukkan Alur Tahapan Penelitian yang terdiri dari enam langkah utama. Proses dimulai dengan Pengumpulan Data dari berbagai sumber yang relevan. Setelah data terkumpul, dilakukan Persiapan Data untuk menata data agar siap dianalisis. Langkah selanjutnya adalah Pembersihan Data untuk menghilangkan *missing values*, duplikasi, dan kesalahan lainnya. Setelah itu, data dieksplorasi lebih lanjut pada tahap Pra-Pemodelan, di mana variabel penting dipilih untuk pemodelan. Pada tahap Pemodelan Data, algoritma XGBoost diterapkan untuk menganalisis data. Terakhir, hasil dievaluasi pada tahap Finalisasi Data, di mana performa model diukur dan hasil akhir disiapkan untuk digunakan.

a. *Data Collecting*
   Langkah pertama adalah pengumpulan data, di mana data yang relevan diambil dari Kaggle sebagai sumber dataset publik [12]. Tahap ini memastikan bahwa peneliti memiliki data yang cukup untuk dianalisis dalam penelitian.
b. *Data Preparation*

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1373−1382
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.5998

Setelah data dikumpulkan tidak dapat langsung di proses [13], tahap selanjutnya adalah mempersiapkan data tersebut. Data preparation melibatkan penataan data ke dalam format yang dapat dianalisis, termasuk penyelarasan struktur, pengkodean ulang variabel, dan penanganan data yang tidak terstruktur. Langkah ini bertujuan untuk memastikan bahwa data siap untuk diproses lebih lanjut [14].

c. *Data Cleaning*
Pada tahap ini, data yang dikumpulkan dibersihkan dari *noise, outliers, dan anomali* lainnya yang dapat mengganggu analisis. Data cleaning juga mencakup penanganan *missing values* (nilai yang hilang), penghapusan duplikasi, dan koreksi kesalahan pada dataset [15]. Tahap ini penting untuk meningkatkan kualitas data dan keakuratan hasil analisis.

d. *Data Premodelling*
Sebelum masuk ke proses pemodelan, dilakukan analisis pra-pemodelan yang melibatkan eksplorasi data lebih lanjut, pemilihan variabel penting, dan pembentukan variabel baru jika diperlukan [16]. Tujuannya adalah untuk memahami karakteristik data dan mempersiapkannya untuk model pembelajaran mesin atau statistik yang akan digunakan [17].

e. *Data Modelling*
Pada tahap ini data yang telah dibersihkan dan dipersiapkan diimplementasikan ke dalam model yang sesuai dengan menggunakan algoritma XGBoost. Algoritma ini diterapkan untuk menghasilkan prediksi berdasarkan data yang telah diproses. Dalam proses ini, model dilatih untuk mengenali pola dan hubungan yang ada dalam data. Selama pelatihan dilakukan pengoptimalan model dengan mengatur parameter-parameter seperti tingkat pembelajaran dan kedalaman pohon, guna meningkatkan akurasi prediksi. Dengan demikian, penerapan algoritma XGBoost tidak hanya menghasilkan model yang efisien, tetapi juga meningkatkan kemampuan prediksi yang akurat dalam konteks analisis data yang sedang dilakukan. [18].

f. *Evaluation*
Tahap evaluasi model menggunakan tiga metrik utama *Mean Absolute Error (MAE), Mean Squared Error (MSE),* dan R^2. MAE mengukur rata-rata selisih absolut antara nilai prediksi dan aktual, memberikan indikasi seberapa besar kesalahan prediksi dalam satuan asli data, dengan nilai yang lebih rendah menunjukkan prediksi yang lebih akurat [19]. MSE yang menghitung rata-rata kuadrat dari selisih antara prediksi dan aktual lebih sensitif terhadap *outliers* karena memberikan penalti lebih besar pada kesalahan yang signifikan [20]. Oleh karena itu, MSE yang rendah menunjukkan bahwa model memiliki kesalahan prediksi yang kecil dan tidak terpengaruh oleh *outliers*. Sementara itu, R^2 menunjukkan seberapa baik model menjelaskan variasi data target dengan nilai mendekati 1 menunjukkan performa yang baik [21].

2.2 XGBoost

XGBoost merupakan salah satu algoritma machine learning yang terbukti efektif dalam tugas prediksi dan klasifikasi, dengan keunggulan dalam akurasi dan kecepatan pemrosesan. Keberhasilan XGBoost dapat dikaitkan dengan model pohon keputusannya yang efisien dan kemampuannya mengatasi ketidakseimbangan data, yang sering muncul dalam konteks pengambilan keputusan [22]. Algoritma ini mengintegrasikan pendekatan berbasis pohon keputusan, baik untuk regresi maupun klasifikasi, tergantung pada tipe variabel dependen yang dianalisis [23]. Variabel kontinu menghasilkan *regression tree*, sedangkan variabel kategorikal membentuk *classification tree*. Dengan kemampuan tersebut, XGBoost menunjukkan kinerja yang unggul dalam pemrosesan data yang kompleks dan dalam jumlah besar [24]. Tahapan perhitungan dalam XGBoost melibatkan fungsi objektif, yang mencakup fungsi kerugian *(loss function)* dan fungsi regularisasi, dirumuskan sebagai berikut:

$$obj(\theta) = L(\theta) + \Omega(\theta) \quad (1)$$

Dimana L adalah fungsi pelatihan yang hilang, dan Ω adalah fungsi regularisasi, dan θ adalah sebagai parameter model terkait. Fungsi pelatihan yang hilang L umumnya dinyatakan sebagai:

$$L(\theta) = \sum_{i=1}^{n} l\,(y_i\,, \hat{y}_i) \quad (2)$$

Dimana yi adalah nilai sebenarnya dan $\hat{y}\,i$ adalah hasil nilai prediksi dari model, sedangkan n adalah jumlah iterasi nilai dari model.

2.3 Evaluasi

MAE, MSE, dan R^2 adalah metrik yang digunakan untuk mengevaluasi model yang dibangun. Mengevaluasi keakuratan model adalah langkah penting dalam setiap model pembelajaran mesin.
Mean Squared Error (MSE) mencerminkan rata-rata perbedaan absolut antara nilai aktual dan prediksi dalam kumpulan data, mengukur rata-rata residu dalam kumpulan data tersebut [25]

$$MAE = \sum_{i=1}^{n} \frac{|y - \hat{y}|}{n} \quad (3)$$

Dalam rumus tersebut, simbol ∑ melambangkan jumlah keseluruhan nilai, sementara y_i merupakan nilai aktual yang diamati. Selain itu y_i menunjukkan nilai hasil prediksi yang dihasilkan oleh model, dan n adalah banyaknya data yang digunakan dalam analisis.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1373−1382
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.5998

Mean Squared Error (MSE) mewakili rata-rata dari perbedaan kuadrat antara nilai asli dan prediksi. Metrik ini mengukur varian dari residual, memberikan penalti yang lebih besar untuk kesalahan yang signifikan. MSE dihitung dengan rumus:

$$MSE = \sum_{i=1}^{N} \frac{|y - \hat{y}|^2}{n} \quad (4)$$

Nilai MAE, MSE, dan RMSE yang lebih rendah menandakan akurasi model regresi yang lebih tinggi. RMSE menunjukkan seberapa baik suatu model regresi dapat memprediksi nilai suatu variabel respon secara absolut. Berikut adalah contoh evaluasi keakuratan prediksi [25]

R^2 Score atau koefisien determinasi (R^2) merupakan metrik yang digunakan untuk menilai seberapa besar kontribusi variabel independen (X) terhadap variabel dependen (Y) secara keseluruhan. Nilai R^2 berada dalam rentang 0 hingga 1 ($0 \leq R^2 \leq 1$), yang mencerminkan proporsi variasi total yang dapat dijelaskan oleh model yang dihasilkan. Apabila nilai R^2 yang diperoleh semakin rendah, hal ini menunjukkan bahwa pengaruh variabel independen terhadap variabel dependen semakin lemah. Sebaliknya, semakin tinggi nilai R^2 yang mendekati angka 1, semakin kuat pengaruh variabel independen tersebut. Secara umum, nilai R^2 di atas 0,9 mengindikasikan bahwa model yang dibangun memiliki kinerja yang baik dan mampu menjelaskan variabilitas data dengan sangat efektif [26].

3. HASIL DAN PEMBAHASAN

Di bagian ini, akan dijelaskan hasil analisis dari setiap langkah penelitian berjudul "Prediksi Kualitas Air untuk Pertumbuhan Ikan Lele Menggunakan Metode XGBoost". Penjelasan ini mencakup cara pengumpulan data, langkah-langkah pra-pemrosesan, proses prediksi, serta pengujian dan evaluasi kinerja metode yang digunakan.

3.1 *Data Collecting*

Dalam penelitian ini menggunakan *Sensor-Based Aquaponics Fish Pond Dataset* yang didapatkan dari platform Kaggle oleh Dr. Collins N. Udanor. Dataset ini memiliki 10 atribut yang memiliki peran penting dalam menentukan pertumbuhan ikan lele di sistem akuaponik. Tabel 2 berikut adalah deskripsi dari masing-masing atribut yang terdapat dalam dataset.

Tabel 1. Deskripsi Atribut Data

Atribute	Deskripsi
Date/Time	Waktu pengambilan data
Temperature	Suhu air di kolam ikan.
Turbidity	Tingkat kejernihan atau kekeruhan air.
Dissolved Oxygen	Jumlah oksigen yang terlarut dalam air.
pH	Tingkat keasaman atau kebasaan air.
Ammonia	Kandungan amonia dalam air, yang berasal dari limbah ikan.
Nitrate	Kandungan nitrat dalam air, hasil dari proses penguraian amonia.
Population of fish in the pond	Jumlah ikan yang ada di kolam.
Length Of Fish	Ukuran panjang ikan dalam kolam.
Weight Of Fish	Bobot ikan dalam kolam.

Tabel 1 memberikan pemahaman mengenai variabel-variabel yang terdapat dalam dataset serta peran masing-masing dalam memengaruhi pertumbuhan ikan lele pada sistem akuaponik. Dengan dataset ini, analisis mendetail terhadap faktor-faktor lingkungan yang mempengaruhi kondisi kolam dan kesehatan ikan dapat dilakukan. Tabel 2 berikut menyajikan dataset yang digunakan dalam penelitian ini.

Tabel 2. Data Kolam Ikan Lele

Created_at	Entry_id	Temperature	…	Fish Length	Fish_Weight
0	2021-06-19 00:00:05 CET	24.875	…	7.11	2.91
1	2021-06-19 00:01:02 CET	24.9375	…	7.11	2.91
2	2021-06-19 00:01:22 CET	24.875	…	7.11	2.91
3	2021-06-19 00:01:44 CET	24.9375	…	7.11	2.91
4	2021-06-19 00:02:07 CET	24.9375	…	7.11	2.91
…	…	…	…	…	…
83121	2021-10-13 02:48:31 CET	26.5625	…	33.45	318.64
83122	2021-10-13 03:17:36 CET	26.5625	…	33.45	318.64
83123	2021-10-13 03:46:49 CET	26.5	…	33.45	318.64
83124	2021-10-13 04:13:23 CET	26.375	…	33.45	318.64
83125	2021-10-13 04:14:22 CET	26.375	…	33.45	318.64

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1373−1382
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.5998

Pada Tabel 2 Data Kolam Ikan Lele disajikan informasi penting mengenai kondisi lingkungan dan pertumbuhan ikan lele selama periode tertentu. Tabel ini mencakup kolom *Created_at* yang menunjukkan tanggal dan waktu pengukuran dalam format CET, *Entry_id* yang berfungsi sebagai pengidentifikasi unik untuk setiap entri, *Temperature* yang mencatat suhu air dalam derajat *Celsius* dan merupakan faktor krusial yang mempengaruhi kesehatan ikan, *Fish Length* yang menunjukkan panjang ikan lele dalam sentimeter, serta *Fish Weight* yang mencatat berat ikan dalam gram. Setiap baris tabel mewakili pengukuran yang dilakukan pada waktu tertentu, memberikan data yang berguna untuk analisis kondisi kolam dan pertumbuhan ikan. Dengan memantau data ini, petani dapat mengevaluasi kesehatan ikan dan membuat keputusan yang lebih baik dalam pengelolaan budidaya, sehingga mendukung upaya peningkatan efisiensi dan produktivitas.

3.2 *Data Preparation*

Tahap pra-pemrosesan bertujuan untuk mempersiapkan data sebelum dilakukan proses prediksi. Pra-pemrosesan melibatkan 3 langkah utama, yaitu penggabungan data, analisis data dan pembersihan data.

a. *Data Merged*

Langkah pertama dalam pra-pemrosesan data adalah penggabungan data (*data merging*). Pada tahap ini fokus utama adalah mengkombinasikan berbagai sumber data menjadi satu dataset yang utuh. Metode yang digunakan untuk penggabungan ini meliputi penggabungan berdasarkan kunci atau atribut yang sama di setiap dataset. Seperti yang ditunjukkan pada Gambar 2 pendekatan ini dipilih untuk memastikan bahwa semua informasi relevan tersedia dalam satu tempat sehingga memudahkan analisis dan prediksi lebih lanjut.

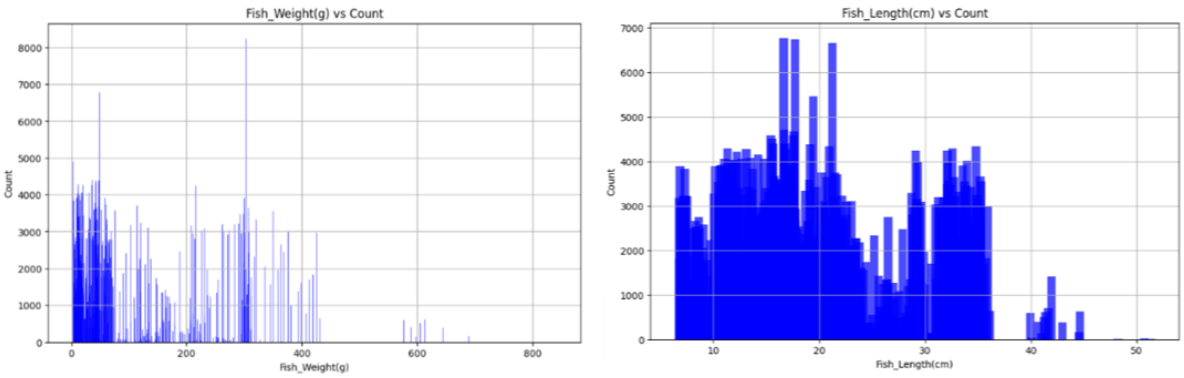
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1114974 entries, 0 to 1114973
Data columns (total 11 columns):
 #   Column                  Non-Null Count    Dtype
---  ------                  --------------    -----
 0   created_at              1114973 non-null  object
 1   entry_id                1114892 non-null  object
 2   Temperature (C)         1114892 non-null  object
 3   Turbidity(NTU)          1114892 non-null  object
 4   Dissolved Oxygen(g/ml)  1114892 non-null  object
 5   PH                      1114892 non-null  object
 6   Ammonia(g/ml)           1020199 non-null  object
 7   Nitrate(g/ml)           1114892 non-null  object
 8   Population              1114974 non-null  object
 9   Fish_Length(cm)         1114972 non-null  object
 10  Fish_Weight(g)          1114972 non-null  object
dtypes: object(11)
memory usage: 93.6+ MB
```

Gambar 2. Hasil Pemprosesan Penggabungan Data Kolam

Pada Gambar 2 menunjukkan total dataset keseluruhan berjumlah 1.114.974 data. Namun, data tersebut belum melalui tahap pembersihan (cleaning), sehingga masih terdapat nilai-nilai yang tidak valid atau hilang yang perlu ditangani sebelum dilakukan proses prediksi untuk memastikan hasil yang akurat.

b. *Data Analysis*

Proses analisis data melibatkan pengkajian terhadap berbagai variabel dalam dataset seperti *fish_length, fish_weight, population, time, temperature, turbidity, Dissolved Oxygen, Nitrate, pH, dan Ammonia*. Setiap variabel dianalisis secara terpisah untuk memahami distribusi, karakteristik, serta pola hubungan antarvariabel. Pada gambar 3 menampilkan hasil analisis data terkait berat dan panjang ikan.



Gambar 3. Hasil Analysis Data Berat dan Panjang Ikan

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1373−1382
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.5998

Pada Gambar 3 merupakan hasil analisis yang menunjukkan bahwa variabel kualitas air, seperti pH, suhu, dan kadar amonia, berpengaruh signifikan terhadap pertumbuhan dan kesehatan ikan. Sebagian besar ikan memiliki berat sekitar 300 gram dan panjang antara 15 hingga 25 cm, dengan konsentrasi tertinggi di sekitar 20 cm. Namun, terdapat beberapa outlier dalam distribusi data, yaitu ikan dengan berat 600-800 gram dan panjang lebih dari 40 cm, yang jauh berbeda dari mayoritas. Data outlier ini dapat mengindikasikan adanya ikan yang tumbuh secara signifikan lebih besar karena faktor lingkungan tertentu atau kesalahan pengukuran. Analisis lebih lanjut diperlukan untuk memahami penyebab di balik outlier ini dan apakah kondisi lingkungan ekstrem berperan dalam menghasilkan variasi pertumbuhan yang signifikan.

c. *Data Cleaning*

Setelah data berhasil dianalisis, tahap berikutnya adalah *data cleaning*. Pada tahap ini, fokus utama adalah membersihkan dataset dari nilai yang hilang, duplikasi, atau data yang tidak valid. Data yang kotor dapat mempengaruhi kualitas analisis dan prediksi, sehingga perlu dilakukan beberapa langkah seperti menghapus entri yang tidak relevan, mengisi nilai yang hilang menggunakan metode tertentu, serta memperbaiki kesalahan format atau inkonsistensi dalam atribut data. Tabel 4 berikut adalah hasil dari pembersihan data.

Tabel 3. Hasil Data Cleaning

	Temperature	Turbidity	Dissolved oxygen	pH	Ammonia	Nitrate	Fish length	Fish Weight
43065	24.125	73	12.729	7.2806	0.064668	14.702842	18.49	59.3600
43066	24.125	75	14.064	7.2806	0.163824	12.48062	18.49	59.3600
43067	24.125	78	13.16	7.2715	0.197164	10.904393	18.49	59.3600
43069	24.125	79	7.945	7.267	0.194727	9.741602	18.49	59.3600
43070	24.125	79	10.514	7.2715	0.184536	9.664083	18.49	59.3600
...	...	...	...	...	...	...	...	…
1020186	27	51	12.88	6.3637	0.345679	26.692506	41.66213	620.9285
1020187	27	51	11.498	6.3546	0.347083	26.124031	41.66213	620.9285
1020188	27	51	12.323	6.3727	0.324023	26.124031	41.66213	620.9285
1020189	27	51	12.916	6.3591	0.303816	26.046512	41.66213	620.9285
1020191	27	51	11.368	6.3591	0.309982	25.994832	41.66213	620.9285

206746 rows × 8 columns

Tabel 3 menunjukkan hasil dari proses *data cleaning*, di mana dataset telah dibersihkan dari nilai-nilai yang hilang dan duplikasi. Terlihat bahwa beberapa nilai pada kolom *Dissolved Oxygen* dan *Fish Length* telah diperbaiki untuk memastikan kelengkapan data. Jumlah total data yang digunakan setelah proses pembersihan adalah 206.746 baris.

3.3 Data Premodelling

Setelah proses *data cleaning* selesai, langkah berikutnya adalah tahap *premodelling.* Pada tahap ini, data yang telah dibersihkan dipersiapkan untuk masuk ke dalam model prediksi. Beberapa langkah yang dilakukan dalam premodelling termasuk pemilihan fitur *(feature selection)* yang relevan, normalisasi atau standarisasi data untuk memastikan setiap fitur memiliki skala yang sebanding, serta pembagian dataset menjadi data pelatihan *(training set)* dan data pengujian *(testing set)*. Proses ini bertujuan untuk memastikan bahwa data yang digunakan dalam pemodelan memiliki kualitas yang optimal sehingga dapat meningkatkan akurasi hasil prediksi. Dapat dilihat pada tabel 5 merupakan hasil dari *premodelling* kualitas air.

Tabel 4. Hasil Kondisi Kualitas Air

	Temperature	…	Fish_health_condition_factor	Water_quality_score	Water_quality_category
0	24.125	…	0.939038	0.716105	Good
1	24.125	…	0.939038	0.716105	Good
2	24.125	…	0.939038	0.716105	Good
3	24.125	…	0.939038	0.716105	Good
4	24.125	…	0.939038	0.716105	Good
…	…	…	…	…	…
206741	27.000	…	0.858652	0.654803	Good
206742	27.000	…	0.858652	0.654803	Good
206743	27.000	…	0.858652	0.654803	Good
206744	27.000	…	0.858652	0.654803	Good
206745	27.000	…	0.858652	0.654803	Good

Pada Tabel 4, hasil kondisi kualitas air ditentukan dengan menggunakan skor kualitas air (*Water Quality Score*) yang diklasifikasikan ke dalam beberapa kategori. Untuk menentukan kategori ini, dilakukan pembagian skor berdasarkan beberapa batasan nilai menggunakan metode *binning*. Batasan yang digunakan adalah: 0 hingga 0.2 untuk

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1373−1382
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.5998

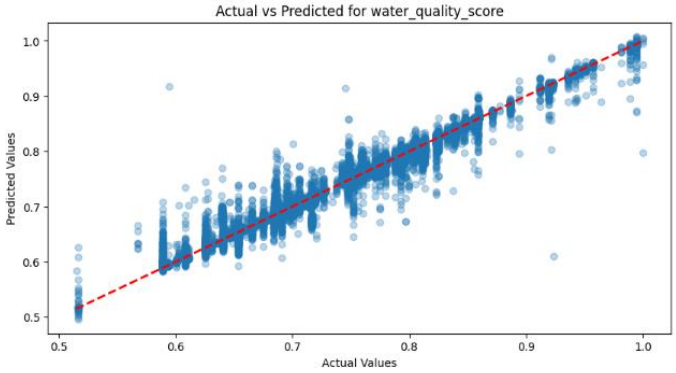
kategori *'poor'*, 0.2 hingga 0.4 untuk kategori *'fair'*, 0.4 hingga 0.6 untuk kategori *'normal'*, 0.6 hingga 0.8 untuk kategori *'good'*, dan 0.8 hingga 1.0 untuk kategori 'excellent'. Hasilnya, data kualitas air di setiap baris diberikan label kategori yang sesuai, seperti yang terlihat pada kolom *Water Quality Category*.

3.4 *Data Modelling*

Setelah tahap premodelling selesai, langkah selanjutnya dalam pengolahan data adalah data modelling, di mana model prediksi dibangun menggunakan data yang telah dibersihkan dan dipersiapkan. Pada tahap ini, algoritma *machine learning* XGBoost diterapkan untuk memprediksi variabel target berdasarkan fitur-fitur relevan, mengingat keunggulannya dalam menangani data kompleks dan mengatasi masalah *overfitting*, serta menghasilkan prediksi akurat. Model XGBoost dilatih menggunakan data training dan dievaluasi dengan data testing, bertujuan untuk membuat model belajar dari pola dalam data agar mampu memberikan prediksi akurat pada data baru. Evaluasi kinerja model dilakukan dengan metrik *mean squared error* (MSE), *mean absolute error* (MAE), dan *R-squared* (R^2), yang membantu menilai kemampuan model dalam memprediksi variabel target. Setelah evaluasi awal, proses optimasi model dilakukan melalui hyperparameter tuning untuk menemukan kombinasi parameter terbaik, seperti kedalaman pohon keputusan dan *learning rate*, yang dapat meningkatkan akurasi prediksi dan meminimalkan kesalahan. Setelah model optimal diperoleh, model ini siap digunakan untuk memprediksi kualitas air atau faktor lain yang mempengaruhi pertumbuhan ikan dalam skala yang lebih luas.

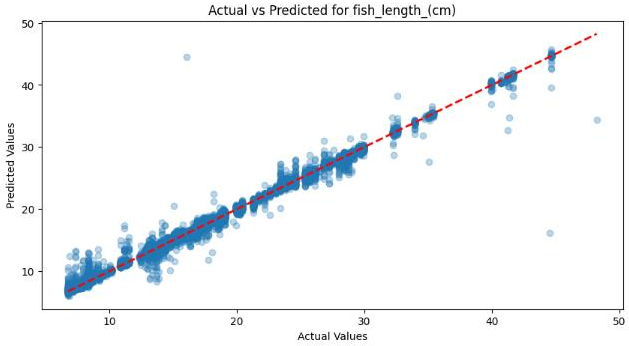
a. *Training Data*

Training data menunjukkan proses pelatihan model menggunakan algoritma XGBoost untuk memprediksi variabel *water quality score*, *fish length* (cm), dan *fish weight* (g) berdasarkan fitur lingkungan seperti suhu, kekeruhan, oksigen terlarut, pH, amonia, dan nitrat. Dataset dibagi menjadi fitur (X) dan target (Y), kemudian dipisah menjadi data latih (80%) dan data uji (20%) menggunakan fungsi *train_test_split* dengan *random_state 42* untuk menjaga konsistensi. Selanjutnya, GridSearchCV digunakan untuk mengoptimalkan hiperparameter seperti *n_estimators*, *learning_rate*, *max_depth*, *subsample*, *dan colsample_bytree* guna menemukan kombinasi parameter terbaik yang memaksimalkan kinerja model. Setiap variabel target dilatih secara terpisah dengan tujuan regresi, dan hasil terbaik diperoleh melalui GridSearchCV berdasarkan metrik R^2, yang digunakan untuk mengevaluasi akurasi prediksi model pada data uji. Gambar 4 menunjukkan hasil prediksi menggunakan XGBoost, yang menggambarkan performa model dalam menghasilkan prediksi yang akurat.



Gambar 4. *Actual vs Prediuct for Water Quality Score*

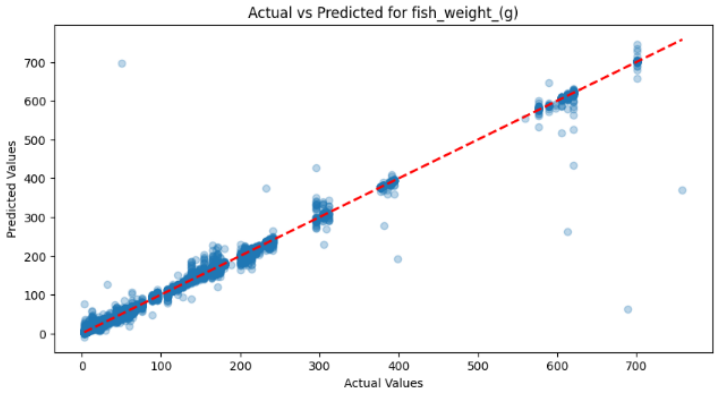
Gambar 4 memperlihatkan perbandingan antara nilai aktual dan prediksi untuk skor kualitas air, di mana sumbu horizontal merepresentasikan nilai aktual dan sumbu vertikal menunjukkan nilai prediksi. Garis merah putus-putus di tengah menggambarkan garis ideal di mana nilai prediksi sama dengan nilai aktual. Sebagian besar titik data berada di sekitar garis ini, menunjukkan bahwa model memiliki akurasi yang baik dalam memprediksi kualitas air. Namun, beberapa titik yang berada jauh dari garis mengindikasikan adanya deviasi atau kesalahan prediksi.



Gambar 5. *Actual vs Predict for Fish Length*

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1373−1382
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.5998

Gambar 5 merupakan hasil perbandingan antara nilai aktual dan prediksi panjang ikan (dalam cm) menggunakan algoritma XGBoost. Sumbu horizontal (x-axis) menggambarkan nilai aktual, sedangkan sumbu vertikal (y-axis) menunjukkan nilai prediksi yang dihasilkan oleh model. Garis putus-putus berwarna merah merepresentasikan garis ideal $y = x$ di mana jika prediksi model tepat, seluruh titik data akan terletak pada garis tersebut. Titik-titik biru yang ada menunjukkan pasangan nilai aktual dan prediksi untuk setiap sampel. Kedekatan titik-titik ini dengan garis referensi menandakan akurasi prediksi model semakin dekat titik-titik dengan garis, semakin baik performa model dalam memprediksi panjang ikan. Secara umum, sebagian besar titik data mendekati garis referensi, yang menunjukkan bahwa model XGBoost memberikan prediksi yang akurat pada sebagian besar data. Namun, terdapat beberapa nilai outlier yang jauh dari garis, menandakan prediksi yang kurang akurat pada sampel-sampel tertentu. Hal ini menunjukkan kinerja model yang baik secara keseluruhan, dengan beberapa penyimpangan yang mungkin dapat diperbaiki melalui pengoptimalan lebih lanjut.



Gambar 6. *Actual vs Predict for Fish Weight*

Pada Gambar 6 merupakan plot *Actual vs Predicte for fish_weight* yang menunjukkan hubungan antara nilai aktual dan nilai prediksi untuk berat ikan dalam gram. Garis merah putus-putus adalah garis referensi yang mewakili prediksi sempurna, di mana semua titik seharusnya berada jika model membuat prediksi yang benar-benar akurat. Sebagian besar titik data berada dekat dengan garis ini, menunjukkan bahwa model memiliki kinerja prediksi yang cukup baik. Namun, terdapat beberapa titik yang jauh dari garis *(outlier)*, menunjukkan prediksi yang kurang akurat pada nilai-nilai tertentu, terutama pada ikan dengan berat di atas 600 gram, yang mungkin memerlukan analisis lebih lanjut atau penyempurnaan model.

b. *Predict Data*

Proses *predict data* melibatkan penerapan model yang telah dilatih menggunakan algoritma XGBoost untuk menghasilkan prediksi terhadap variabel seperti *water quality score*, *fish length* (cm), dan *fish weight* (g) berdasarkan fitur lingkungan yang sama, yaitu suhu, kekeruhan, oksigen terlarut, pH, amonia, dan nitrat. Setelah model selesai dilatih dan dioptimalkan, data baru yang berisi fitur-fitur tersebut dipersiapkan untuk tahap prediksi. Model kemudian menggunakan data ini untuk menghitung estimasi nilai target. Gambar berikut memperlihatkan perbandingan antara nilai prediksi dan nilai aktual, memberikan gambaran tentang seberapa baik model dapat melakukan prediksi pada data yang belum pernah dilihat sebelumnya.

Tabel 5. Hasil Prediksi Data Kualitas Air dan Kesehatan Ikan

Fish Length (cm)	**:**	**28.74197**
Fish Weight (g)	:	215.9633
Fish Health Condition Factor	:	0.909558
Fish Health Condition Category	:	poor
Water Quality Score	:	0.693623
Water Quality Category	:	good

Dari tabel 5 tersebut hasil prediksi dari beberapa parameter kualitas air suhu 20°C, kekeruhan 10 NTU, oksigen terlarut 10 mg/L, pH 1, amonia 0.5 mg/L, dan nitrat 25 mg/L. Model memprediksi panjang ikan sebesar 28.74 cm dan beratnya 215.96 gram. Faktor kondisi kesehatan ikan dihitung sebesar 0.99, namun kategorinya dinilai *"Poor"* (buruk). Skor kualitas air yang dihasilkan adalah 0.69, yang termasuk dalam kategori *"Good"* (baik), meskipun kondisi kesehatan ikan tetap buruk, kemungkinan disebabkan oleh pH yang sangat rendah (sangat asam).

3.5 Evaluasi

Hasil evaluasi setelah proses data modeling menggunakan metrik *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan R^2 memberikan gambaran komprehensif tentang kinerja model XGBoost. MAE mengukur rata-rata kesalahan absolut antara nilai prediksi dan nilai aktual, memberikan indikasi seberapa besar kesalahan rata-rata dalam satuan yang sama dengan variabel target. MSE, di sisi lain, memberikan penalti yang lebih besar untuk kesalahan

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1373−1382
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.5998

yang lebih besar, sehingga lebih sensitif terhadap *outlier*. Nilai R^2 digunakan untuk mengevaluasi proporsi variansi dalam data target yang dapat dijelaskan oleh model. Nilai yang mendekati 1 menunjukkan bahwa model sangat baik dalam memprediksi variabel target. Hasil evaluasi menunjukkan nilai MAE dan MSE yang rendah, serta R2 yang tinggi, menandakan bahwa model memiliki akurasi yang baik dalam *memprediksi water quality score, fish length, dan fish weight.*

Tabel 6. Hasil Evaluasi

	Fish_weight_(g):	Fish_length_(cm):
Mean Squared Error (MSE)	0.073323532	14.66617244
Mean Absolute Error (MAE)	0.123376185	1.278238282
R-Squared (R2)	0.998249992	0.998077246

Tabel 6 menunjukkan hasil evaluasi model prediksi berat dan panjang ikan menggunakan tiga metrik *Mean Squared Error (MSE), Mean Absolute Error (MAE), dan R-Squared (R²).* MSE untuk berat ikan adalah 0.0733 dan untuk panjang ikan 14.6661, yang berarti kesalahan prediksi lebih kecil pada berat ikan. MAE, yang mengukur rata-rata kesalahan prediksi, adalah 0.1234 untuk berat ikan dan 1.2782 untuk panjang ikan, menunjukkan prediksi berat lebih akurat. Nilai R^2 untuk berat ikan adalah 0.9982 dan untuk panjang ikan 0.9981, menunjukkan bahwa model sangat baik dalam memprediksi keduanya dengan lebih dari 99% variabilitas data dijelaskan oleh model.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma XGBoost dapat digunakan secara efektif untuk memprediksi kualitas air dan pertumbuhan ikan lele berdasarkan berbagai faktor lingkungan seperti suhu, kekeruhan, oksigen terlarut, pH, amonia, dan nitrat. Dengan dataset dari *Sensor-Based Aquaponics Fish Pond*, model XGBoost terbukti mampu memberikan prediksi akurat untuk panjang ikan dan berat ikan, dengan nilai evaluasi R^2 mencapai 0.9982 untuk berat ikan dan 0.9981 untuk panjang ikan, yang menunjukkan bahwa lebih dari 99% variabilitas dalam data dapat dijelaskan oleh model. Evaluasi kinerja model juga menunjukkan nilai *Mean Absolute Error (MAE)* sebesar 0.1234 untuk berat ikan dan 1.2782 untuk panjang ikan, serta *Mean Squared Error (MSE)* sebesar 0.0733 dan 14.6661, mengindikasikan tingkat akurasi prediksi yang tinggi, terutama untuk berat ikan. Implementasi XGBoost dalam budidaya ikan lele dapat membantu petani memantau kualitas air dan pertumbuhan ikan secara *real-time*, meminimalkan kesalahan dalam pemeliharaan, dan meningkatkan produktivitas, sehingga mendukung adopsi teknologi *machine learning* dalam akuakultur yang lebih efisien di masa depan. Penelitian selanjutnya dapat mengintegrasikan sistem *Internet of Things* (IoT) dengan algoritma XGBoost untuk menghasilkan pemantauan kualitas air secara otomatis dan prediksi pertumbuhan ikan secara *real-time*, guna meningkatkan efektivitas dan efisiensi budidaya ikan lele.

REFERENCES

[1] D. Mariyono, A. N. A. Hidayatullah, And A. N. A. Kamila, *Inovasi Akuakultur Menyatukan Teknologi Dan Budaya Dalam Bisnis Korporasi Di Indonesia*. Cipta Media Nusantara, 2024. [Online]. Available: Https://Books.Google.Co.Id/Books?Id=Dtcdeqaaqbaj

[2] F. Ali, D. Hidayat, N. Hoerniasih, And Others, "Pelatihan Program Kecakapan Hidup Budidaya Ikan Lele Sebagai Upaya Pemberdayaan Masyarakat Di Pkbm Linggih Sinau Banyusari," *Comm-Edu (Community Education Journal)*, Vol. 7, No. 1, Pp. 98–113, 2024.

[3] Rudiyant, *Ensiklopedia Ikan Lele*. Lembar Langit Indonesia, 2023. [Online]. Available: Https://Books.Google.Co.Id/Books?Id=09xmeaaaqbaj

[4] M. Tasya Aulia, N. Anisah, E. Sulistyo, P. M. Negeri, And B. Belitung, "Prosiding Seminar Nasional Inovasi Teknologi Terapan 2022 Sistem Kontrol Dan Monitoring Kualitas Air Pada Budidaya Ikan Lele Dengan Media Kolam Berbasis Iot," 2022.

[5] A. Deni, *Manajemen Strategi Di Era Industri 4.0*. Cendikia Mulia Mandiri, 2023.

[6] N. Fahmi And S. Natalia, "Sistem Pemantauan Kualitas Air Budidaya Ikan Lele Menggunakan Teknologi Iot," *Jurnal Media Informatika Budidarma*, Vol. 4, No. 4, Pp. 1243–1248, 2020.

[7] E. Retnoningsih And R. Pramudita, "Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python," *Bina Insani Ict Journal*, Vol. 7, No. 2, Pp. 156–165, 2020.

[8] S. Jesika, S. Ramadhani, And Y. P. Putri, "Implementasi Model Machine Learning Dalam Mengklasifikasi Kualitas Air," *Jurnal Ilmiah Dan Karya Mahasiswa*, Vol. 1, No. 6, Pp. 382–396, 2023.

[9] N. N. Amiroh *Et Al.*, *Prosiding Seminar Nasional Inovasi Teknologi Informasi & Komunikasi: "Optimalisasi Teknologi Kecerdasan Artifisial Untuk Mendukung Transformasi Digital Dan Masa Depan Otomasi."* Sanata Dharma University Press, 2024. [Online]. Available: Https://Books.Google.Co.Id/Books?Id=N30yeqaaqbaj

[10] E. Sakina And A. H. Mirza, "Prediksi Hasil Produksi Ikan Lele Menggunakan Machine Learning (Studi Kasus Dinas Perikanan Kabupaten Muara Enim)," *Jurnal Instek (Informatika Sains Dan Teknologi)*, Vol. 9, No. 1, Pp. 55–64, 2024.

[11] J. M. A. S. Dachi And P. Sitompul, "Analisis Perbandingan Algoritma Xgboost Dan Algoritma Random Forest Ensemble Learning Pada Klasifikasi Keputusan Kredit," *Jurnal Riset Rumpun Matematika Dan Ilmu Pengetahuan Alam*, Vol. 2, No. 2, Pp. 87–103, 2023.

[12] C. Thomas And A. Engelbrecht, *Data Mining: Concepts And Applications*. In Artificial Intelligence, Volume 8. Intechopen, 2022. [Online]. Available: Https://Books.Google.Co.Id/Books?Id=Btjueaaaqbaj

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1373−1382
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.5998

[13] A. Maghfiroh, Y. Findawati, And U. Indahyanti, "Klasifikasi Penipuan Pada Rekening Bank Menggunakan Pendekatan Ensemble Learning," *Building Of Informatics, Technology And Science (Bits)*, Vol. 4, No. 4, Mar. 2023, Doi: 10.47065/Bits.V4i4.3212.
[14] T. Revathi, K. Muneeswaran, And M. Blessa Binolin Pepsi, *Big Data Processing With Hadoop*. In Advances In Data Mining And Database Management (2327-1981). Igi Global, 2018. [Online]. Available: Https://Books.Google.Co.Id/Books?Id=Qsj0dwaaqbaj
[15] P. W. Rahayu *Et Al.*, *Buku Ajar Data Mining*. Pt. Sonpedia Publishing Indonesia, 2024. [Online]. Available: Https://Books.Google.Co.Id/Books?Id=Vcrueaaaqbaj
[16] D. A. Prasetya, A. Muhaimin, And Others, "Analisis Klaster Partitioning Around Medoids Dengan Gower Distance Untuk Rekomendasi Indekos (Studi Kasus: Indekos Di Sekitar Kampus Upnvjt)," *G-Tech: Jurnal Teknologi Terapan*, Vol. 8, No. 3, Pp. 2060–2069, 2024.
[17] A. H. Pratama, *Belajar Mudah Dan Singkat Machine Learning: Panduan Praktis Dengan Studi Kasus, Kode Program, Dan Dataset*. Penerbit Andi, 2024. [Online]. Available: Https://Books.Google.Co.Id/Books?Id=Pfmoeqaaqbaj
[18] H. Wijaya, D. P. Hostiadi, And E. Triandini, "Meningkatkan Prediksi Penjualan Retail Xyz Dengan Teknik Optimasi Random Search Pada Model Xgboost," In *Seminar Hasil Penelitian Informatika Dan Komputer (Spinter)| Institut Teknologi Dan Bisnis Stikom Bali*, 2024, Pp. 829–833.
[19] I. Amansyah, J. Indra, E. Nurlaelasari, And A. R. Juwita, "Prediksi Penjualan Kendaraan Menggunakan Regresi Linear: Studi Kasus Pada Industri Otomotif Di Indonesia," *Innovative: Journal Of Social Science Research*, Vol. 4, No. 4, Pp. 1199–1216, 2024.
[20] T. D. R. Octavia, N. Rosmawarni, A. Zaidiah, J. R. S. F. Raya, And P. Labu, "Implementasi Algoritma Multiple Linear Regression Untuk Memprediksi Temperatur Udara Berdasarkan Kadar Zat Polutan Di Kota Tangerang Selatan," *Jrsf Raya*, 2024.
[21] H. Hermansyah, A. Abdullah, And P. Y. Utami, "Penerapan Metode Regresi Linier Berganda Untuk Memprediksi Panen Kelapa Sawit," *Progresif: Jurnal Ilmiah Komputer*, Vol. 20, No. 1, Pp. 540–554, 2024.
[22] Iskandar, A. Faisal, B. Utomo, And M. H. Ridho, "Formula_Yang_Bikin_Datamu_Bunyi," 2022.
[23] P. S. Rizky, R. H. Hirzi, And U. Hidayaturrohman, "Perbandingan Metode Lightgbm Dan Xgboost Dalam Menangani Data Dengan Kelas Tidak Seimbang," *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, Vol. 15, No. 2, Pp. 228–236, 2022.
[24] A. Samih, A. Ghadi, And A. Fennan, "Enhanced Sentiment Analysis Based On Improved Word Embeddings And Xgboost.," *International Journal Of Electrical & Computer Engineering (2088-8708)*, Vol. 13, No. 2, 2023.
[25] M. Arifin, "Model Educational Data Mining Berbasis Gradient Boosted Trees Untuk Prediksi Performa Akademik Mahasiswa," 2024.
[26] F. Septia Nugraha And H. Ferdinandus Pardede, "Autoencoder Untuk Sistem Prediksi Berat Lahir Bayi," 2022, Doi: 10.25126/Jtiik.202293868.