

SHORT-TERM TIME-SERIES FORECASTING OF HYDROPONICS WATER QUALITY PARAMETERS USING XGBOOST BASED IoT SENSOR DATA

I Putu Romyadhy Mahaputra¹, I Ketut Resika Arthana², Bagus Gede Krishna Yudistira³, Ja'far Shiddiq⁴, Putu Zasya Eka Satya Nugraha⁵, Gede Defry Widhi Adnyana⁶

^{1,2,3,6}Jurusan Teknologi Informatika, Fakultas Teknik dan Kejuruan, Universitas Pendidikan Ganesha; Jl. Udayana No.11, Banjar Tegal, Singaraja, Kabupaten Buleleng, Bali 81116

^{4,5}Dago Engineering; Jl. Bukit Dago Selatan No.27-29, Dago, Kecamatan Coblong, Kota Bandung, Jawa Barat; 40135

Keywords:

Hidroponik, IoT, TDS, Turbidity, Water Level, Short-term forecasting, XGBoost, Kualitas Air; Preventif

Correspondent Email:

romyadhy@student.undiksha.ac.id

Abstrak. Pemantauan kualitas air merupakan faktor kritis dalam keberhasilan pertanian hidroponik. Sistem pemantauan konvensional yang bersifat reaktif seringkali terlambat mendeteksi perubahan parameter kualitas air seperti *Total Dissolved Solids* (TDS), kekeruhan (*turbidity*), dan ketinggian air (*water level*), yang dapat berujung pada kegagalan panen. Penelitian ini bertujuan mengembangkan model prediksi jangka pendek (*short-term forecasting*) menggunakan algoritma *Extreme Gradient Boosting* (XGBoost) untuk mendukung sistem peringatan dini. Data time-series dikumpulkan melalui sensor IoT selama lima bulan dengan interval 15 menit. Evaluasi kinerja model menunjukkan akurasi yang tinggi pada seluruh parameter. Model prediksi nutrisi (TDS) mencapai nilai *R-squared* di atas 0,94 dengan MAPE di bawah 2%, menunjukkan presisi yang sangat baik. Pada parameter turbidity, model memperoleh nilai *R-squared* tertinggi sebesar 0,986 dengan MAE hanya 0,72 NTU. Sementara itu, prediksi water level mampu memetakan dinamika fluktuasi air dengan *R-squared* 0,85-0,92. Hasil ini membuktikan bahwa model XGBoost efektif memberikan estimasi kondisi air yang akurat, memungkinkan petani melakukan tindakan preventif sebelum terjadi penurunan kualitas air yang signifikan dalam rentang waktu yang relatif pendek.

Abstract. *Water quality monitoring is a critical factor in the success of hydroponic farming. Conventional reactive monitoring systems are often too slow to detect changes in water quality parameters such as Total Dissolved Solids (TDS), turbidity, and water level, which can lead to crop failure. This study aims to develop a short-term forecasting model using the Extreme Gradient Boosting (XGBoost) algorithm to support an early warning system. Time-series data was collected via IoT sensors over five months at 15-minute intervals. Model performance evaluation showed high accuracy across all parameters. The nutrient prediction model (TDS) achieved an R-squared value above 0.94 with MAPE below 2%, indicating excellent precision. For the turbidity parameter, the model obtained the highest R-squared value of 0.986 with MAE of only 0.72 NTU. Meanwhile, water level predictions were able to map water fluctuation dynamics with R-squared ranging from 0.85-0.92. These results prove that the XGBoost model is effective in providing accurate water condition estimates, enabling farmers to take preventive measures before a significant decline in water quality occurs within a relatively short period of time.*



Copyright © [JITET](http://www.jitet.org) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

1. PENDAHULUAN

Pertanian hidroponik telah menjadi solusi utama dalam mengatasi keterbatasan lahan pertanian konvensional, dengan ketergantungan mutlak pada manajemen kualitas air [1], [2]. Dalam sistem hidroponik, parameter seperti Total Dissolved Solids (TDS), kekeruhan (turbidity), dan ketinggian air (water level) memegang peranan vital karena berhubungan langsung dengan kualitas air, nutrisi dan kesehatan akar tanaman [3], [4]. Perkembangan teknologi Internet of Things (IoT) memungkinkan pemantauan parameter ini secara real-time. Namun, sistem monitoring konvensional umumnya bersifat reaktif, petani atau aktuator hanya bertindak setelah nilai parameter melampaui ambang batas berbahaya, yang seringkali terlambat untuk mencegah stres pada tanaman.

Kualitas air pada sistem hidroponik bersifat sangat dinamis dan dipengaruhi oleh faktor lingkungan seperti siklus penyerapan nutrisi harian tanaman. Suhu lingkungan tinggi yang menyebabkan penurunan ketersediaan air pada tanaman, serta lonjakan kekeruhan akibat pertumbuhan alga, dapat terjadi dalam hitungan jam. Oleh karena itu, pendekatan berbasis prediksi (forecasting) menjadi krusial. Prediksi jangka pendek (short-term forecasting) diperlukan untuk memberikan estimasi kondisi air beberapa jam ke depan, memberikan waktu yang cukup bagi sistem kontrol atau petani untuk melakukan intervensi preventif sebelum kerusakan permanen terjadi [5].

Berbagai metode Machine Learning telah diterapkan untuk prediksi time series pada data pertanian. Beberapa penelitian menggunakan Long Short-Term Memory (LSTM) karena kemampuannya menangani data sekuensial [6], namun metode Deep Learning seringkali membutuhkan sumber daya komputasi yang besar dan data yang sangat massif [7]. Di sisi lain, metode regresi linear sederhana seringkali gagal menangkap pola non-linear yang kompleks pada parameter biologis [8]. Dalam konteks data sensor IoT berbiaya rendah yang rentan terhadap noise dan outliers, diperlukan algoritma yang robust namun efisien secara komputasi [9], [10].

Penelitian ini mengimplementasikan algoritma Extreme Gradient Boosting

(XGBoost) untuk memprediksi parameter kualitas air hidroponik. XGBoost dipilih karena kemampuannya yang unggul dalam menangani hubungan non-linear, kecepatan komputasi yang tinggi, serta ketahanannya terhadap outliers dibandingkan model boosting tradisional [11], [12]. Penelitian oleh Beloev et al. (2025) menunjukkan bahwa XGBoost efektif dalam menangkap dinamika non-linear pada peramalan beban listrik jangka pendek, mengungguli metode konvensional dalam hal akurasi dan efisiensi komputasi [11]. Penelitian ini menggunakan dataset time series yang dikumpulkan selama lima bulan dengan interval pengambilan data setiap 15 menit. Pemilihan interval tersebut didasarkan pada kebutuhan untuk menangkap pola fluktuasi pada parameter air secara efektif tanpa membebani model dengan noise frekuensi tinggi [13].

Tujuan utama dari penelitian ini adalah membangun model prediksi yang mampu mengestimasi nilai TDS, turbidity, dan water level untuk jangka waktu pendek berdasarkan data historis. Kinerja model XGBoost akan dievaluasi menggunakan metrik Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Determinations Koefisien R^2 , dan Mean Absolute Percentage Error (MAPE), untuk mengukur tingkat akurasi prediksi. Hasil penelitian ini diharapkan dapat menjadi dasar pengembangan sistem peringatan dini yang lebih presisi dalam menjaga pertumbuhan tanaman hidroponik [11], [14].

2. TINJAUAN PUSTAKA

2.1. Hidroponik

Hidroponik merupakan metode budidaya tanaman tanpa tanah yang menggunakan larutan nutrisi sebagai media utama pertumbuhan. Sistem ini menjadi solusi pertanian modern terutama pada lahan terbatas, namun sangat bergantung pada kestabilan kualitas air. Parameter seperti Total Dissolved Solids (TDS), kekeruhan (turbidity), dan ketinggian air (water level) memiliki peran penting dalam menjaga kesehatan tanaman dan keberhasilan panen, sehingga monitoring kualitas air menjadi komponen krusial dalam sistem hidroponik [1], [4].

2.2. IoT

Internet of Things (IoT) adalah teknologi yang memungkinkan perangkat fisik saling terhubung dan bertukar data secara real-time melalui jaringan internet. Dalam sistem hidroponik, IoT digunakan untuk mengintegrasikan sensor guna memantau parameter kualitas air secara kontinu. Meskipun demikian, sistem IoT konvensional masih bersifat reaktif, sehingga diperlukan integrasi dengan sistem prediksi untuk meningkatkan kemampuan deteksi dini terhadap perubahan kondisi air [5].

2.3. Data Mining

Data mining merupakan proses ekstraksi informasi dan pola dari dataset dalam jumlah besar menggunakan teknik statistik dan komputasi. Dalam penelitian ini, data mining digunakan untuk mengolah data sensor IoT yang bersifat time-series guna menemukan pola perubahan kualitas air. Proses ini melibatkan tahapan seperti preprocessing, transformasi data, dan pemodelan, yang bertujuan meningkatkan kualitas data sebelum digunakan dalam analisis lebih lanjut [8], [15], [16].

2.4. Machine Learning

Machine Learning adalah cabang kecerdasan buatan yang memungkinkan sistem belajar dari data untuk mengenali pola dan membuat prediksi. Dalam konteks time-series, machine learning digunakan untuk memprediksi nilai masa depan berdasarkan data historis. Dibandingkan metode konvensional, machine learning mampu menangani hubungan non-linear, meskipun beberapa metode seperti deep learning memerlukan sumber daya komputasi yang besar sehingga perlu pemilihan algoritma yang efisien [6], [8].

2.5. Prediksi

Prediksi atau forecasting adalah proses memperkirakan nilai masa depan berdasarkan data historis, khususnya dalam bentuk time-series. Prediksi jangka pendek (short-term forecasting) sangat penting dalam sistem hidroponik karena memungkinkan deteksi dini terhadap perubahan kualitas air dalam waktu dekat. Model prediksi yang baik harus mampu menangkap pola tren, musiman, serta fluktuasi data untuk menghasilkan estimasi yang akurat. [17]

2.6. Algoritma XGBoost

XGBoost merupakan algoritma machine learning berbasis ensemble learning yang menggunakan metode gradient boosting untuk

meningkatkan akurasi prediksi. Algoritma ini mampu menangani hubungan non-linear, tahan terhadap noise, serta efisien secara komputasi. XGBoost telah terbukti efektif dalam berbagai kasus prediksi time-series, sehingga cocok digunakan untuk memodelkan data sensor IoT yang dinamis dan kompleks [11], [12], [18].

2.7. Evaluasi Model

Evaluasi model bertujuan untuk mengukur kinerja prediksi menggunakan metrik kuantitatif seperti Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), dan Koefisien Determinasi (R^2). Kombinasi metrik ini memberikan evaluasi yang komprehensif terhadap akurasi model, di mana nilai error yang rendah dan nilai R^2 yang mendekati 1 menunjukkan performa model yang baik [11], [19], [20].

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan time-series forecasting berbasis machine learning untuk memprediksi parameter kualitas air hidroponik secara jangka pendek (short-term). Tahapan penelitian dapat dilihat pada Gambar 1, dimulai dari pengumpulan dataset melalui perangkat IoT khususnya dari sensor TDS (*Total Dissolved Solids*) sensor, *Turbidity* sensor, dan *water level* sensor. Tahapan selanjutnya melakukan pre-processing dataset, raw data di cleaning untuk menghilangkan noise. Pada tahapan ini juga dilakukan proses resampling dan pembentukan feature (lag history). Tahapan selanjutnya adalah pembangunan model prediksi menggunakan XGBoost, dataset dibagi 80% untuk training, 20% untuk testing. Tahapan berikutnya, evaluasi model menggunakan beberapa metode standar prediksi. Tahapan terakhir dilakukan analisis hasil dan kesimpulan terhadap hasil dari evaluasi model.



Gambar 1. Tahapan Penelitian

3.1. Deskripsi Dataset

Data yang digunakan dalam penelitian ini berasal dari sistem hidroponik berbasis IoT yang merekam parameter kualitas air berupa

total dissolved solids (TDS), turbidity, dan water level. Instrumen pengumpulan data terdiri dari konfigurasi 5 sensor utama yang ditempatkan beberapa titik pada pipa hidroponiknya, untuk merepresentasikan kondisi air pada tanaman. Rincian sensornya yaitu TDS 1, TDS 2, Turbidity, Water Level 1, Water Level 2. Dua unit sensor Total Dissolved Solids digunakan untuk mengukur konsentrasi larutan nutrisi. Penggunaan dua sensor bertujuan untuk memvalidasi distribusi nutrisi dalam sistem. Dengan membandingkan pembacaan pada titik 1 dan titik 2, sehingga sistem dapat memastikan bahwa larutan nutrisi terdistribusi secara merata pada tanaman. Satu unit turbidity sensor, digunakan untuk memantau kejernihan air. Penggunaan sensor IoT seperti sensor TDS, dan water level pada sistem hidroponik juga telah diterapkan pada penelitian sebelumnya untuk mendukung kualitas air secara real-time dan otomatis [21].

Parameter ini krusial untuk mendeteksi indikasi awal pertumbuhan alga, pengendapan partikel nutrisi tak larut, atau kontaminasi air yang dapat menghambat pertumbuhan tanaman. Dua sensor water level ditempatkan pada kedua ujung pipa hidroponik (ujung inlet dan outlet). Konfigurasi posisi ini bertujuan untuk memonitoring ketinggian air, dan memastikan ketersediaan air di sepanjang pipa hidroponik. Data yang digunakan adalah data dengan interval 15 menit, dengan total waktu kurang lebih selama 5 bulan, jumlah data yang digunakan sekitar 16.594 data, untuk setiap parameter. Pemilihan interval 15 menit bertujuan untuk menjaga keseimbangan antara pengurangan noise sensor dan kemampuan menangkap fluktuasi perubahan kualitas air yang relevan secara operasional [13]. Data yang diperoleh bersifat numerik [22], dan time-series [12], sehingga sesuai untuk pendekatan predictions time-series menggunakan XGBoost.

3.2. Pre-Processing Data

Tahapan pra-pemrosesan meliputi pemeriksaan missing values, penanganan data anomali, serta normalisasi data untuk menjaga kestabilan proses pelatihan model [15]. Tahap ini bertujuan meningkatkan kualitas data, sebelum diproses oleh model [16]. Langkah-langkah yang diterapkan sesuai dengan kode program yang dibangun adalah:

- Penyaringan Data (Data Filtering): Menghapus outliers atau kesalahan pembacaan sensor yang tidak logis.
- Agregasi Temporal (Resampling): Data mentah yang mungkin memiliki interval tidak teratur diubah menjadi data per 15 menit dengan menghitung nilai rata-rata (mean) setiap 15 menit [13]. Hal ini dilakukan untuk mengurangi fluktuasi sesaat (noise) dan meringankan beban komputasi.
- Imputasi Data Hilang: Jika terdapat kekosongan data akibat kegagalan transmisi pada waktu tertentu, nilai tersebut diisi menggunakan metode interpolasi linear untuk menjaga kontinuitas tren data.

3.3. Pembentukan Feature

Untuk menangkap pola ketergantungan waktu (temporal dependencies) pada data time-series, penelitian ini tidak hanya menggunakan data mentah, melainkan menciptakan fitur-fitur baru (input features) sebagai variabel predictor selanjutnya. Fitur yang dibangun meliputi:

1. *Lag 1 (t-1)*: Nilai parameter pada 15 menit sebelumnya, fitur ini menangkap kondisi terkini sistem serta menangkap autokorelasi jangka sangat pendek yang umumnya memiliki pengaruh paling kuat terhadap nilai saat ini [23].
2. *Lag 4 (t-4)*: Nilai parameter pada 1 jam sebelumnya, fitur ini bertujuan untuk menangkap tren perubahan jangka pendek yang terjadi dalam satu jam terakhir [23], [24].
3. *Lag 96 (t-96)*: Nilai parameter pada 24 jam sebelumnya, fitur ini berperan penting dalam menangkap pola siklus harian (daily cycle), seperti pengaruh perbedaan kondisi lingkungan antara siang dan malam [23], [25].
4. *Hour (t)*: Informasi jam saat ini (0-23). Fitur temporal ini ditambahkan secara eksplisit agar model dapat membedakan kondisi operasional berdasarkan waktu, misalnya perbedaan karakteristik sistem pada siang hari [25].
5. *Rolling Mean/Window 3H*: Informasi untuk moving average dari 3 jam terakhir 1 (t-1 sampai t-3). Fitur ini digunakan untuk menangkap tren jangka pendek dan memperhalus fluktuasi [23], [24].

3.4. Model XGBoost

Model yang digunakan dalam penelitian ini adalah Extreme Gradient Boosting (XGBoost), yang merupakan metode ensemble learning berbasis gradient boosting decision trees. XGBoost dipilih karena kemampuannya menangani hubungan non-linear, ketahanannya terhadap data sensor yang bersifat fluktuatif, serta efisiensinya pada dataset berukuran menengah [12], [18]. Model dilatih menggunakan fitur lag dari masing-masing parameter kualitas air, dengan konfigurasi parameter utama seperti jumlah pohon (number of estimators), kedalaman pohon (max depth), dan learning rate. Pendekatan ini telah terbukti efektif dalam berbagai studi prediksi lingkungan dan kualitas air [18]. Untuk memprediksi parameter hidroponik, penelitian ini menggunakan model XGBoost yang merupakan metode ensemble berbasis pohon keputusan. Berbeda dengan model tunggal, XGBoost menghasilkan prediksi akhir \hat{y}_t dengan menjumlahkan skor prediksi dari sekumpulan pohon K menggunakan prinsip model aditif, yang sudah didefinisikan dalam persamaan (1):

$$\hat{y}_t = \sum_{k=1}^K f_k(x_t), f_k \in F \quad (1)$$

Dimana f_k merepresentasikan fungsi prediksi dari pohon ke- k . Pendekatan ini memungkinkan model untuk mengakumulasi hasil pembelajaran dari pohon-pohon sebelumnya guna meminimalkan error prediksi secara iterative [18]. Dataset yang telah memiliki fitur tambahan dibagi menggunakan metode time-based split untuk mencegah data leakage. Train dataset yang digunakan untuk melatih model adalah 80%. Testing dataset yang digunakan untuk validasi model yaitu sebanyak 20%. Hyperparameter model ditentukan dari beberapa literature seperti penelitian oleh Kramar & Alchakov (2023) yang terbukti menghasilkan akurasi tinggi ($R^2 > 0.95$) pada prediksi time-series, dengan konfigurasi $n_estimators = 500$ dan $learning_rate = 0.05$ untuk mencapai konvergensi yang optimal [14]. Langkah implementasi secara prosedural ditunjukkan pada Gambar 2. pseudocode algoritma prediksi XGBoost.

Algorithm 1 Proposed XGBoost Forecasting Model

Require: Raw Dataset D_{raw} containing timestamps and sensor values

Ensure: Evaluation Metrics (R^2 , MSE, RMSE, MAE, MAPE)

- 1: **Phase 1: Pre-processing**
- 2: **for** each sensor in D_{raw} **do**
- 3: Filter outliers based on physical limits ($0 \leq val \leq Limit$)
- 4: Resample data to 15-minute intervals $\rightarrow \Delta t = 15m$
- 5: Impute missing values using Linear Interpolation
- 6: **end for**
- 7: **Phase 2: Feature Engineering**
- 8: **for** each timestamp t in Dataset **do**
- 9: $Lag_1 \leftarrow x_{t-1}$
- 10: $Lag_4 \leftarrow x_{t-4}$
- 11: $Lag_{96} \leftarrow x_{t-96}$
- 12: $Hour \leftarrow ExtractHour(t)$
- 13: $Roll_{3h} \leftarrow MovingAvg(x_{t-12} \dots x_{t-1})$
- 14: Remove rows containing NaN values
- 15: **end for**
- 16: **Phase 3: Training and Prediction**
- 17: Split Dataset: D_{train} (80%) and D_{test} (20%)
- 18: Initialize $Model_{XGB}$ ($n_estimators = 500, \eta = 0.05$)
- 19: $Model_{XGB}.fit(D_{train})$
- 20: $\hat{y}_{pred} \leftarrow Model_{XGB}.predict(D_{test})$
- 21: **Phase 4: Evaluation**
- 22: Calculate $R^2, MSE, RMSE, MAE, MAPE$
- 23: **return** Evaluation Metrics and Visualization

Gambar 2. Pseudocode Algoritma XGBoost

3.5. Evaluasi Model

Performa model dievaluasi menggunakan lima metrik kuantitatif standar, yaitu Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), dan Determination Koefisien R^2 [11]. Kombinasi metrik ini dipilih untuk memberikan evaluasi yang komprehensif, MSE dan RMSE sensitif terhadap outliers [19], MAE mengukur rata-rata kesalahan absolut secara linear, MAPE memberikan interpretasi kesalahan relatif dalam bentuk persentase [26], dan R^2 mengukur seberapa baik model menangkap variabilitas data actual [20]. Mean Squared Error (MSE) untuk menghitung rata-rata kuadrat kesalahan antara nilai prediksi dan actual, seperti yang bisa dilihat pada persamaan (2). Metrik ini memberikan penalti yang lebih besar pada kesalahan prediksi yang ekstrem [11].

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

Root Mean Squared Error (RMSE) untuk mengetahui tingkat kesalahan dalam satuan asli parameter (seperti PPM untuk TDS, NTU untuk Turbidity, dan % untuk water level), sehingga lebih mudah diinterpretasikan secara fisik, dapat dilihat pada persamaan (3) [11].

$$RMSE = \sqrt{MSE} \quad (3)$$

Mean Absolute Percentage Error (MAPE), perhitungannya dapat dilihat pada

persamaan (4) evaluasi ini untuk rata-rata persentase kesalahan absolut relative, terhadap nilai aktual. Metrik ini sangat berguna untuk memahami akurasi model dalam perspektif persentase yang agnostik terhadap skala data [11].

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (4)$$

Mean Absolute Error (MAE), mengukur rata-rata selisih absolut antara prediksi dan data aktual tanpa memperhatikan arah kesalahan, perhitungan evaluasi ini dapat dilihat pada persamaan (5). Berbeda dengan RMSE, MAE memberikan bobot yang sama untuk setiap kesalahan, sehingga lebih stabil terhadap outliers [11].

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

Koefisien Determinasi R^2 , evaluasi ini menunjukkan proporsi varians pada variabel dependen (*target*) yang dapat diprediksi dari variabel independen (*fitur*) [20]. Nilai R^2 mendekati 1 menunjukkan bahwa model sangat sesuai (*goodness of fit*) dengan data observasi [11]. Keterangan perhitungan pada persamaan (6) yaitu, \hat{y}_i : Nilai aktual data ke- i , \hat{y}_i : Nilai hasil prediksi ke- i , dan \bar{y}_i : Nilai rata-rata dari data aktual.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (6)$$

4. HASIL DAN PEMBAHASAN

Pada bagian ini akan dipaparkan mengenai hasil dan pembahasan terkait model prediksi menggunakan XGBoost. Pembahasan dimulai dari karakteristik data sensor IoT, efektivitas pra-pemrosesan, visualisasi hasil prediksi time-series, hingga evaluasi statistik model XGBoost.

4.1. Deskripsi Dataset

Data penelitian dikumpulkan menggunakan sistem akuisisi data berbasis ESP32 yang terintegrasi dengan sensor TDS 1&2 (Total Dissolved Solids), sensor kekeruhan (Turbidity), dan sensor Water Level 1&2. Pengambilan data berlangsung selama periode 12 Juli 2025 - 5 Januari 2026, menghasilkan

total data mentah sebanyak 207.340 pada setiap sensor. Analisis statistik deskriptif pada Tabel 1, menunjukkan adanya noise yang signifikan dan nilai outliers ekstrem yang tidak mungkin terjadi secara fisik.

Sebagai contoh, sensor TDS1 mencatat nilai maksimum hingga 483.835 PPM dan minimum negatif -684.593 PPM, yang mengindikasikan kesalahan pembacaan sensor (sensor error) atau gangguan elektrik pada mikrokontroler. Hal serupa juga terjadi pada sensor Turbidity dengan nilai maksimum mencapai 573.911 NTU. Hal ini mengonfirmasi bahwa data mentah dari sensor, tidak dapat langsung digunakan untuk pemodelan tanpa melalui proses pembersihan yang ketat [27]. Oleh karena itu agar model bisa memprediksi lebih baik dan presisi, diperlukan pre-processing dataset untuk menjaga kondisi ideal dari rentang dataset yang akan digunakan [15].

Tabel 1. Raw Dataset

Sensor	Total Data	Min Value	Max Value
TDS1	207.340	-684.593	483.835
TDS2	207.340	-67.997	6.221
Turbidity	207.340	-431	573.911
WLevel 1	207.340	0	100
WLevel 2	207.340	0	100

4.2. Hasil Pre-processing Data

Untuk menangani anomali yang ditunjukkan pada Tabel 1, dilakukan tahapan pra-pemrosesan yang meliputi filtering batas fisik (misalnya membatasi TDS pada 0-5000 PPM) dan resampling temporal. Proses ini mereduksi jumlah data dari 207.340 baris (data mentah) menjadi 16.889 baris data bersih dengan interval satu jam (hourly). Pengurangan volume data ini tidak hanya menghilangkan noise frekuensi tinggi, tetapi juga menetapkan dataset agar model tidak memprediksi diluar rentang nilai optimal [28]. Data bersih ini dibagi dengan rasio 80:20, menghasilkan 13.511 training data dan 3.378 testing data. Hasil data pada pre-processing ini selanjutnya akan digunakan model XGBoost untuk melakukan prediksi terhadap nilai numerik pada setiap parameter seperti TDS (1&2), Turbidity, dan Water Level (1&2).

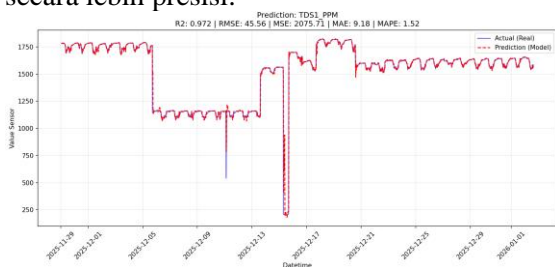
4.3. Hasil Train Model

Model XGBoost dilatih menggunakan dataset hasil pre-processing dengan konfigurasi hyperparameter $n_estimators=500$ dan

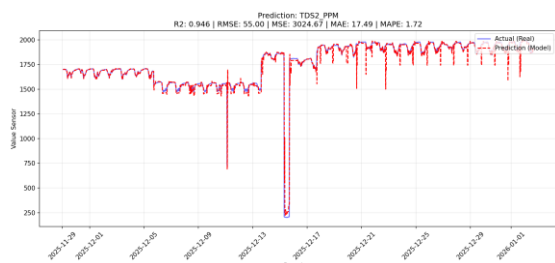
learning_rate=0.05 [14]. Kemampuan model dalam mempelajari pola data divalidasi melalui perbandingan visual antara nilai aktual (Actual) dan nilai prediksi (Prediction) pada testing dataset. Berikut adalah hasil training pada setiap parameter yang di visualisasikan dalam sebuah grafik, untuk garis biru merupakan grafik data asli, kemudian garis putus – putus merah merupakan grafik prediksi model.

4.3.1. Prediksi Parameter TDS 1&2

Pada hasil train parameter TDS, Gambar 3, dan Gambar 4, model menunjukkan akurasi yang kuat antara hasil model XGBoost dengan data aktual. Pada hasil training prediksi tersebut, memperlihatkan pola gelombang harian yang konsisten, di mana model berhasil menangkap fluktuasi ritmik akibat siklus penyerapan nutrisi tanaman dan pengaruh suhu lingkungan. Selain itu, model menunjukkan responsivitas tinggi terhadap perubahan fisik yang ekstrem saat terjadi penurunan konsentrasi nutrisi secara drastis akibat intervensi operasional, seperti air habis atau maintenance, garis prediksi mampu menyesuaikan diri mengikuti pola yang signifikan. Hal ini mengindikasikan bahwa fitur historis yang diterapkan efektif memberikan konteks adaptif, sehingga memungkinkan model memprediksi secara lebih presisi.



Gambar 3. Hasil Prediksi TDS 1

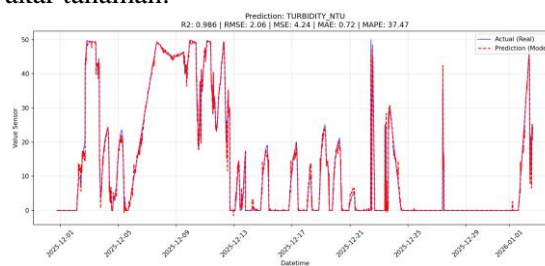


Gambar 4. Hasil Prediksi TDS 2

4.3.2. Prediksi Parameter Turbidity

Pada grafik parameter Turbidity, Gambar 5, terlihat jelas bahwa model mampu mengikuti pola dengan baik pada data aktual, bahkan saat

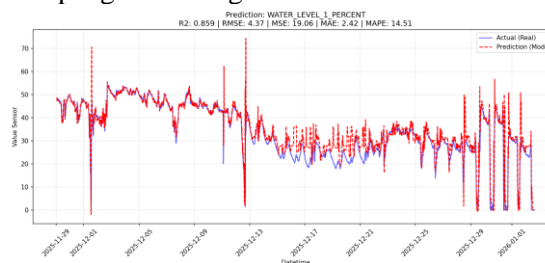
terjadi perubahan nilai yang sangat cepat. Model berhasil menangkap setiap lonjakan yang ada mendadak. Kemampuan respons cepat ini menunjukkan bahwa model sangat sensitif terhadap perubahan kondisi air, sehingga sangat berguna untuk mendeteksi masalah kekeruhan sedini mungkin sebelum berdampak buruk pada akar tanaman.



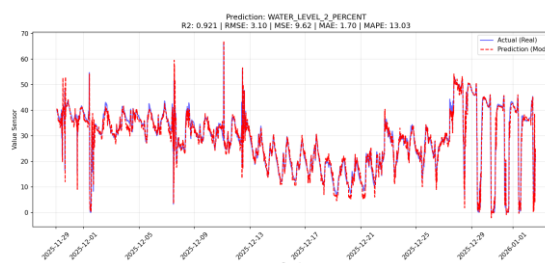
Gambar 5. Hasil Prediksi Turbidity

4.3.3. Prediksi Water Level 1&2

Pada parameter ketinggian air Gambar 6, dan Gambar 7, grafik menunjukkan fluktuasi yang lebih dinamis. Hal ini masih cukup wajar, karena permukaan air dalam pipa hidroponik sering mengalami fluktuasi akibat diserap oleh tanaman dan juga penguapan. Meskipun demikian, model mampu memetakan tren rata-rata penurunan level air, baik akibat konsumsi tanaman atau penguapan dan kenaikan level air saat pengisian ulang.



Gambar 6. Hasil Prediksi Water Level 1



Gambar 7. Hasil Prediksi Water Level 2

4.4. Evaluasi Model

Evaluasi statistik secara menyeluruh terhadap seluruh sensor disajikan pada Tabel 2. Evaluasi ini dihitung berdasarkan perbandingan nilai prediksi model terhadap data aktual pada

testing data. Hasil evaluasi model tersebut berupa Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), dan Determination Koefisien (R^2).

Tabel 2. Hasil Evaluasi Model

Sensor	MAE	R2	MSE	RMSE	MAPE
TD1	9.17	0.97	2075.7	45.55	1.52
TD2	17.4	0.94	3024.6	54.99	1.71
Tur	0.72	0.98	4.24	2.06	37.46
WL1	2.41	0.85	19.05	4.36	14.50
WL2	1.69	0.92	9.61	3.10	13.02

Hasil evaluasi statistik mengindikasikan bahwa model prediksi yang diusulkan mampu mencapai akurasi tinggi pada mayoritas parameter. Analisis mendalam terhadap Tabel 2 adalah sebagai berikut:

1. Evaluasi TDS (Total Dissolved Solid), sensor TDS menunjukkan kinerja yang akurat dengan nilai Koefisien Determinasi (R^2) di atas 0,94 dan nilai MAPE yang sangat rendah, yaitu di bawah 2%. Hal ini mengonfirmasi bahwa model mampu memprediksi fluktuasi konsentrasi nutrisi tanaman dengan tingkat presisi yang tinggi. Kesalahan rata-rata absolut (MAE) yang berkisar 9-17 PPM, tergolong galat (*error*) kecil, jika dibandingkan rentang pengukuran TDS yang ribuan PPM.
2. Evaluasi Turbidity, parameter Turbidity mencatatkan skor (R^2) tertinggi mencapai 0,986, yang menunjukkan model sangat presisi mengikuti pola data. Meskipun nilai MAPE terlihat tinggi (37,47%), angka ini terinflasi akibat sifat matematis rumus MAPE yang sensitif terhadap pembagian nilai aktual yang sangat kecil (mendekati nol) pada kondisi air jernih. Indikator akurasi yang lebih valid untuk parameter ini adalah MAE, yang menunjukkan kesalahan prediksi rata-rata hanya sebesar 0,72 NTU. Angka kesalahan absolut yang sangat kecil ini membuktikan model sangat akurat dalam memprediksi.
3. Evaluasi Water Level, prediksi Water Level menunjukkan performa yang bervariasi (R^2 antara 0,86 hingga 0,92). Sedikit penurunan

akurasi pada Water Level 1 (MAPE 14,5%) mencerminkan tingginya dinamika fisik dalam pipa, seperti gangguan mekanis akibat fluktuasi ketersediaan air pada pipa hidroponik yang menciptakan noise pada sensor. Namun, RMSE yang berada di kisaran 3-4% menunjukkan selisih prediksi masih cukup baik.

5. KESIMPULAN

- a. Penelitian ini berhasil mengembangkan model prediksi time-series berbasis algoritma XGBoost untuk pemantauan kualitas air pada sistem hidroponik berbasis IoT. Model yang dibangun mampu mengolah data historis sensor dan menghasilkan prediksi jangka pendek (short-term forecasting) terhadap parameter utama, yaitu TDS, turbidity, dan water level secara efektif.
- b. Hasil evaluasi menunjukkan bahwa model memiliki kinerja yang sangat baik dengan tingkat akurasi tinggi, khususnya pada parameter kimia air. Parameter turbidity dan TDS 1 mencatatkan performa terbaik dengan nilai koefisien determinasi (R^2) masing-masing sebesar 0,986 dan 0,972 serta nilai Mean Absolute Error (MAE) yang rendah, yang mengindikasikan kemampuan model dalam menangkap pola data secara akurat.
- c. Model juga mampu mengidentifikasi pola siklus harian (*diurnal*) serta merespons perubahan nilai yang signifikan, sehingga dapat digunakan untuk mendeteksi potensi anomali kualitas air secara dini. Kemampuan ini menunjukkan bahwa fitur time-series yang digunakan, seperti lag dan rolling window, berkontribusi terhadap peningkatan performa model.
- d. Pada parameter water level, meskipun memiliki dinamika fluktuasi yang lebih tinggi akibat pengaruh faktor fisik seperti penguapan dan konsumsi air oleh tanaman, model tetap mampu memprediksi tren ketersediaan air dengan cukup baik, yang ditunjukkan oleh nilai R^2 sebesar 0,921, sehingga masih relevan untuk kebutuhan operasional.
- e. Secara keseluruhan, integrasi antara sensor IoT dan algoritma XGBoost dapat diandalkan sebagai dasar pengembangan sistem peringatan dini (early warning

system). Untuk penelitian selanjutnya, disarankan pengembangan ke arah prediksi jangka panjang (long-term forecasting), eksplorasi model hybrid untuk meningkatkan generalisasi, serta peningkatan kualitas data melalui proses preprocessing dan data cleaning yang lebih optimal, mengingat kualitas data sangat berpengaruh terhadap akurasi model.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Living Lab yang telah memfasilitasi infrastruktur perangkat IoT dan menjadi sumber data utama dalam penelitian ini. Kemudian apresiasi juga disampaikan kepada Fakultas Teknik dan Kejuruan, Universitas Pendidikan Ganesha atas dukungan lingkungan akademik yang kondusif. Penulis menyampaikan rasa hormat dan terima kasih kepada para Dosen Pembimbing serta para mentor atas bimbingan, arahan, dan diskusi mendalam yang sangat berharga. Terima kasih juga ditujukan kepada rekan-rekan seperjuangan yang telah memberikan bantuan dan dukungan dalam penyelesaian penelitian ini. Terakhir, terima kasih sebesar – besarnya kepada Bapak dan Ibu serta keluarga tercinta, yang senantiasa memberikan doa, dan dukungan tanpa henti hingga penelitian ini dapat diselesaikan dengan maksimal.

DAFTAR PUSTAKA

- [1] C. R. Ramadhani, R. A. Hamzah, and A. A. Wahyudi, "Penerapan Teknologi Hidroponik Sebagai Solusi Pertanian Berkelanjutan di Lingkungan Perkotaan," vol. 1, no. 1, 2025.
- [2] T. Alam, Z.-U.- Haq, M. A. Ahmed, and M. Ikram, "Hydroponics as an advanced vegetable production technique: an overview," *Zoo Bot.*, vol. 1, no. 1, pp. 29–42, Aug. 2023, doi: 10.55627/zoobotanica.001.01.0630.
- [3] W. H. Sugiharto, H. Susanto, and A. B. Prasetyo, "Real-Time Water Quality Assessment via IoT: Monitoring pH, TDS, Temperature, and Turbidity," *ISI*, vol. 28, no. 4, pp. 823–831, Aug. 2023, doi: 10.18280/isi.280403.
- [4] R. Anjini, J. Jenifer, and Mrs. A. M. C. Blessy, "IoT Based Automated Hydroponics Greenhouse Monitoring," *IJARST*, pp. 671–681, Apr. 2021, doi: 10.48175/IJARST-960.
- [5] A. Tzounis, N. Katsoulas, T. Bartzanas, and C. Kittas, "Internet of Things in agriculture, recent advances and future challenges," *Biosystems Engineering*, vol. 164, pp. 31–48, Dec. 2017, doi: 10.1016/j.biosystemseng.2017.09.007.
- [6] Y. Hua, Z. Zhao, R. Li, X. Chen, Z. Liu, and H. Zhang, "Deep Learning with Long Short-Term Memory for Time Series Prediction," Oct. 24, 2018, *arXiv*: arXiv:1810.10161. doi: 10.48550/arXiv.1810.10161.
- [7] N. Elsayed, Z. ElSayed, and A. S. Maida, "LiteLSTM Architecture for Deep Recurrent Neural Networks," Oct. 25, 2022, *arXiv*: arXiv:2201.11624. doi: 10.48550/arXiv.2201.11624.
- [8] S. A. Frimpong, M. Han, W. Zheng, X. Li, E. Akpaku, and A. P. Obeng, "Machine and Deep Learning in Agricultural Engineering: A Comprehensive Survey and Meta-Analysis of Techniques, Applications, and Challenges," *Computers*, vol. 14, no. 10, p. 438, Oct. 2025, doi: 10.3390/computers14100438.
- [9] M.Tech, Department of Computer Science & Engineering, SRM Institute of Science & Technology, Chennai, India., R. Ravi*, Dr. B. Baranidharan, and Associate Professor, Department of Computer Science & Engineering, SRM Institute of Science & Technology, Chennai, India., "Crop Yield Prediction using XG Boost Algorithm," *IJRTE*, vol. 8, no. 5, pp. 3516–3520, Jan. 2020, doi: 10.35940/ijrte.D9547.018520.
- [10] M. Saritha, R. M. Irshath, S. Sanjay, A. V. Selvan, and M. S. Abdullah, "Real-Time Plant Disease Prediction Using XGBoost and IoT-Enabled Smart Agriculture System," in *Proceedings of International Conference on Computer Science and Communication Engineering (ICCSCE 2025)*, vol. 124, J. K. Katiyar, D. S. R. Yellampalli, D. Chandra Mohan, K. K. Singh, B. Venkata Ramana, and N. Dinesh Kumar, Eds., in *Advances in Computer Science Research*, vol. 124, Dordrecht: Atlantis Press International BV, 2025, pp. 120–131. doi: 10.2991/978-94-6463-858-5_12.
- [11] H. I. Beloev, S. R. Saitov, A. A. Filimonova, N. D. Chichirova, O. E. Babikov, and I. K. Iliev, "Short-Term Electrical Load Forecasting Based on XGBoost Model," *Energies*, vol. 18, no. 19, p. 5144, Sep. 2025, doi: 10.3390/en18195144.
- [12] Z. Fang, S. Yang, C. Lv, S. An, and W. Wu, "Application of a data-driven XGBoost model for the prediction of COVID-19 in the USA: a time-series study," *BMJ Open*, vol. 12, no. 7, p. e056685, Jul. 2022, doi: 10.1136/bmjopen-2021-056685.
- [13] L. Zhang, W. Bian, W. Qu, L. Tuo, and Y. Wang, "Time series forecast of sales volume based on XGBoost," *J. Phys.: Conf. Ser.*, vol.

- 1873, no. 1, p. 012067, Apr. 2021, doi: 10.1088/1742-6596/1873/1/012067.
- [14] V. Kramar and V. Alchakov, "Time-Series Forecasting of Seasonal Data Using Machine Learning Methods," *Algorithms*, vol. 16, no. 5, p. 248, May 2023, doi: 10.3390/a16050248.
- [15] A. Tawakuli, B. Havers, V. Gulisano, D. Kaiser, and T. Engel, "Survey: Time-series data preprocessing: A survey and an empirical analysis," *Journal of Engineering Research*, vol. 13, no. 2, pp. 674–711, Jun. 2025, doi: 10.1016/j.jer.2024.02.018.
- [16] B. L. Ortiz *et al.*, "Data Preprocessing Techniques for AI and Machine Learning Readiness: Scoping Review of Wearable Sensor Data in Cancer Care," *JMIR Mhealth Uhealth*, vol. 12, p. e59587, Sep. 2024, doi: 10.2196/59587.
- [17] F. H. Syhadah, R. T. Subagio, and P. Rizqiyah, "PENERAPAN XGBOOST DALAM PREDIKSI PENDAFTARAN SISWA BARU BIMBINGAN BELAJAR QSC DI KOTA CIREBON," *JITET*, vol. 13, no. 3S1, Oct. 2025, doi: 10.23960/jitet.v13i3S1.7998.
- [18] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- [19] K. Maciejowska, A. Lipiecki, and B. Uniejewski, "Statistical and economic evaluation of forecasts in electricity markets: beyond RMSE and MAE," 2025, *arXiv*. doi: 10.48550/ARXIV.2511.13616.
- [20] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Computer Science*, vol. 7, p. e623, Jul. 2021, doi: 10.7717/peerj-cs.623.
- [21] B. G. K. Yudistira, C. Hapsari, G. D. W. Adnyana, W. Nath, I. Putu, and R. M. Putra, "Smart Fisheries: Real-Time Water Quality Management and Automated Feeding System Design for Tilapia Farming using ESP32 Micro Controller," vol. 5, no. 2.
- [22] Z. Qinghe, X. Wen, H. Boyan, W. Jong, and F. Junlong, "Optimised extreme gradient boosting model for short term electric load demand forecasting of regional grid system," *Sci Rep*, vol. 12, no. 1, p. 19282, Nov. 2022, doi: 10.1038/s41598-022-22024-3.
- [23] S. Karmaker, "Feature Engineering in Time Series Forecasting".
- [24] V. Cerqueira, N. Moniz, and C. Soares, "VEST: Automatic Feature Engineering for Forecasting," Oct. 14, 2020, *arXiv*: arXiv:2010.07137. doi: 10.48550/arXiv.2010.07137.
- [25] C. S. Bojer, "Understanding machine learning-based forecasting methods: A decomposition framework and research opportunities," *International Journal of Forecasting*, vol. 38, no. 4, pp. 1555–1561, Oct. 2022, doi: 10.1016/j.ijforecast.2021.11.003.
- [26] A. De Myttenaere, B. Golden, B. Le Grand, and F. Rossi, "Mean Absolute Percentage Error for regression models," *Neurocomputing*, vol. 192, pp. 38–48, Jun. 2016, doi: 10.1016/j.neucom.2015.12.114.
- [27] H. Y. Teh, A. W. Kempa-Liehr, and K. I.-K. Wang, "Sensor data quality: a systematic review," *J Big Data*, vol. 7, no. 1, p. 11, Dec. 2020, doi: 10.1186/s40537-020-0285-1.
- [28] C. Tamilselvi, M. Yeasin, R. K. Paul, and A. K. Paul, "Can Denoising Enhance Prediction Accuracy of Learning Models? A Case of Wavelet Decomposition Approach," *Forecasting*, vol. 6, no. 1, pp. 81–99, Jan. 2024, doi: 10.3390/forecast6010005.