Jurnal Informatika Vol. 15, No. 1, Juni 2026, Page. 227~240
E-ISSN: 2528-0090 https://ejournal.unidayan.ac.id/index.php/JIU
P-ISSN: 2252-9160 DOI: https://doi.org/10.55340/jiu.v15i1.2686

Sistem Akuisisi Data Debit Air Sungai Berbasis IoT Menggunakan Sensor Aliran Air dengan *Random Forest Regression* dan *XGBoost Regression*

IoT-Based River Flow Data Acquisition System Using Water Flow Sensor with Random Forest Regression and XGBoost Regression

Ery Muchyar Hasiri*[1], Helson Hamid[2], Asniati[3]
[1,2,3]Program Studi Teknik Informatika, Universitas Dayanu Ikhsanuddin
E-mail: *[1]erymuchyar82@unidayan.ac.id

Article Info:	Received	Revised	Accepted	Published:
	14 Mei 2026	16 Mei 2026	24 Juni 2026	27 Juni 2026

Abstrak

Pemantauan debit air sungai merupakan komponen kritis dalam sistem peringatan dini banjir dan pengelolaan sumber daya air. Penelitian ini merancang dan mengimplementasikan sistem akuisisi data debit air sungai berbasis Internet of Things (IoT) menggunakan sensor aliran air YF-S201 yang terintegrasi dengan mikrokontroler ESP32, serta menerapkan dua algoritma regresi machine learning, yaitu Random Forest Regression (RF) dan XGBoost Regression (XGB), untuk memprediksi debit satu langkah ke depan. Dataset yang digunakan terdiri dari 1.488 rekaman dengan interval 30 menit sepanjang bulan Juli 2025, mencakup variabel FlowRate, Velocity, WaterLevel, DebitRiver, Rainfall, Temperature, serta fitur temporal dan fitur lag. Data dibagi secara kronologis menjadi tiga subset: pelatihan (70%, n = 1.042), validasi (15%, n = 223), dan uji (15%, n = 222). Hasil evaluasi menunjukkan bahwa XGBoost Regression mencapai kinerja yang lebih unggul dengan nilai MAE = 0,1692 m³/s, RMSE = 0,3429 m³/s, MAPE = 5,15%, dan R^2 = 0,9924 dibandingkan dengan Random Forest dengan MAE = 0,1791 m³/s, RMSE = 0,3498 m³/s, MAPE = 5,58%, dan R^2 = 0,9921. Kedua model menunjukkan akurasi tinggi (R^2 > 0,99), dengan 82,9–85,6% prediksi menghasilkan absolute error di bawah 0,25 m³/s. XGBoost memberikan prediksi yang lebih presisi pada kondisi normal hingga sedang, sementara keduanya mengalami peningkatan error pada debit puncak ekstrem (> 10 m³/s). Sistem ini berpotensi menjadi infrastruktur dasar pemantauan sungai secara cerdas dan terotomasi untuk mendukung sistem peringatan dini banjir.

Kata Kunci: *Internet of Things; Debit Sungai; Random Forest; XGBoost; Sensor YF-S201; ESP32; Prediksi Debit*

Abstract

River discharge monitoring is a critical component of early flood warning systems and water resource management. This study designs and implements an IoT-based river flow data acquisition system using a YF-S201 water flow sensor integrated with an ESP32 microcontroller, and applies two machine learning regression algorithms, namely Random Forest Regression (RF) and XGBoost Regression (XGB), for one-step-ahead discharge prediction. The dataset consists of 1,488 records with 30-minute intervals over July 2025, covering variables such as FlowRate, Velocity, WaterLevel, DebitRiver, Rainfall, Temperature, and temporal/lag features. Data were split chronologically into three subsets: training (70%, n = 1,042), validation (15%, n = 223), and test (15%, n = 222). Evaluation results show that XGBoost Regression achieves superior performance with MAE = 0,1692 m³/s, RMSE = 0,3429 m³/s, MAPE = 5,15%, and R^2 = 0,9924 compared to Random Forest with MAE = 0,1791 m³/s, RMSE = 0,3498 m³/s, MAPE = 5,58%, and R^2 = 0,9921. Both models demonstrate high accuracy (R^2 > 0,99), with 82,9–85,6% of predictions achieving absolute error below 0,25 m³/s. XGBoost provides more

precise predictions across normal to moderate discharge conditions, while both models exhibit higher errors under extreme peak discharge (> 10 m³/s). The system demonstrates strong potential as a foundational infrastructure for intelligent, automated river monitoring to support flood early warning systems.

Keywords: *Internet of Things; River Discharge; Random Forest; XGBoost; YF-S201 Sensor; ESP32; Discharge Prediction*

This is an open access article under the CC BY-SA license.

1. PENDAHULUAN

Banjir merupakan salah satu ancaman hidrometeorologi paling dominan di Indonesia dan terus menimbulkan risiko terhadap keselamatan masyarakat, kerusakan infrastruktur, serta terganggunya aktivitas sosial-ekonomi. Data Bencana Indonesia 2024 menunjukkan bahwa dari 3.472 kejadian bencana yang tercatat sepanjang tahun 2024, banjir menempati urutan tertinggi dengan 1.420 kejadian, disusul kebakaran hutan dan lahan, cuaca ekstrem, tanah longsor, kekeringan, gelombang pasang dan abrasi, gempa bumi, serta erupsi gunung api [1]. Fakta ini menegaskan bahwa penguatan sistem pemantauan sungai secara real-time tidak lagi cukup diposisikan sebagai perangkat pelengkap, tetapi menjadi kebutuhan teknis dalam mitigasi banjir dan pengelolaan risiko bencana hidrometeorologi [2].

Salah satu parameter hidrologis yang paling penting dalam pemantauan sungai adalah debit air atau streamflow, yaitu volume air yang melewati penampang sungai dalam satuan waktu tertentu. Informasi debit diperlukan untuk memahami dinamika aliran, menilai potensi luapan, mendukung pengelolaan sumber daya air, serta menjadi dasar sistem peringatan dini banjir [3], [4]. Literatur hidrologi terbaru menegaskan bahwa prediksi *streamflow* yang akurat berperan penting dalam *flood monitoring, early warning system,* pengelolaan waduk, mitigasi kekeringan, dan perencanaan sumber daya air, tetapi proses prediksinya tetap menantang karena dipengaruhi hubungan nonlinier antara curah hujan, kondisi tangkapan air, penyimpanan air, serta respons hidrologis lokal [5], [6], [7].

Pengukuran debit sungai secara konvensional umumnya bergantung pada pengamatan manual, current meter, papan duga air, atau stasiun hidrometri permanen. Pendekatan tersebut relatif lebih mapan secara hidrologis, tetapi memiliki keterbatasan dalam cakupan spasial, frekuensi akuisisi, biaya instalasi, kebutuhan pemeliharaan, dan kemampuan menyediakan data kontinu pada lokasi-lokasi kecil atau daerah dengan infrastruktur terbatas [8]. Perkembangan *Internet of Things* (IoT) membuka peluang untuk membangun sistem akuisisi data hidrologi yang lebih murah, terdistribusi, dan mampu mengirimkan data secara *real-time* melalui jaringan komunikasi nirkabel [3]. Kajian sistematis menunjukkan bahwa IoT banyak digunakan dalam monitoring air karena mampu mengintegrasikan sensor, mikrokontroler, komunikasi data, penyimpanan cloud, dan visualisasi data secara berkelanjutan [9], [10], [11]. Pada konteks sungai, penelitian berbasis IoT juga telah menunjukkan potensi penggunaan ESP32, protokol *Message Queuing Telemetry Transport* (MQTT), *platform cloud*, dan sensor lingkungan untuk memantau aliran serta kualitas air secara *real-time* [12], [13].

Sensor YF-S201 merupakan salah satu alternatif berbiaya rendah untuk pengukuran laju aliran air yang bekerja berdasarkan prinsip efek *Hall* dengan menghasilkan sinyal pulsa yang dapat dihitung oleh mikrokontroler untuk mengestimasi laju aliran. YF-S201 pada dasarnya dirancang untuk aliran tertutup, sehingga penerapannya pada sungai perlu memperhatikan desain hidraulik, seperti penggunaan pipa pengarah atau *diverted-flow section* agar aliran dapat melewati sensor secara stabil. Kombinasi YF-S201, ESP32, komunikasi nirkabel, *cloud storage*, dan *dashboard* web dapat dikembangkan sebagai *proof-of-concept* untuk pemantauan *streamflow* berbiaya rendah pada aliran kecil [14].

Integrasi antara sistem IoT dan metode *machine learning* (ML) telah dikembangkan secara ekstensif untuk keperluan analisis dan prediksi parameter hidrologis, di mana kemampuan ML dalam mengekstraksi pola nonlinier dari data deret waktu sensor menjadi landasan utama penerapannya [2], [15] . Dalam domain prediksi *streamflow*, *Random Forest Regression* dan *Extreme Gradient Boosting* (XGBoost) merupakan dua algoritma *ensemble learning* yang secara konsisten diaplikasikan karena kapabilitasnya dalam memodelkan hubungan nonlinier, menangkap interaksi kompleks antarfitur, serta

memproses data tabular hidrometeorologis berdimensi tinggi. *Random Forest* telah diverifikasi memiliki keunggulan kompetitif dalam pemodelan *streamflow* jangka pendek pada berbagai karakteristik daerah aliran sungai [4], [6], sementara *XGBoost* dan varian *boosting* lainnya secara empiris menghasilkan akurasi tinggi dalam prediksi debit harian maupun simulasi limpasan permukaan pada beragam konfigurasi variabel input [16], [17], [18]. Evaluasi komparatif lintas studi mengonfirmasi bahwa kedua algoritma tersebut secara konsisten melampaui model statistik klasik maupun pendekatan *machine learning konvensional* dalam tugas prediksi *streamflow* [5], [19], [20].

Aspek penting lain dalam pemodelan debit adalah pembentukan fitur temporal. Data debit sungai memiliki karakter autoregresif, sehingga informasi debit pada periode sebelumnya, nilai rata-rata bergerak, dan perubahan debit antarperiode dapat meningkatkan kemampuan model dalam membaca memori hidrologis jangka pendek. Studi berbasis *Long Short-Term Memory* (LSTM) menunjukkan bahwa integrasi data *discharge* terbaru dan *lagged inputs* mampu meningkatkan performa prediksi *streamflow* pada skala luas [20]. Kajian tentang input variable selection berbasis *Random Forest* juga menegaskan bahwa pemilihan variabel masukan yang tepat sangat penting dalam pengembangan model *streamflow forecasting* [5], [21]. Selain itu, *time-lag-informed Random Forest* terbukti lebih baik ketika memasukkan *historical flows* dan *time-lag meteorological values* dibanding hanya menggunakan nilai meteorologis aktual [22].

Kajian komparatif terhadap algoritma *XGBoost, LightGBM*, dan *CatBoost* pada pemodelan debit harian menunjukkan bahwa performa masing-masing metode bervariasi tergantung pada karakteristik deret waktu hidrologis yang digunakan [17]. Sejalan dengan temuan tersebut, kapabilitas prediktif *XGBoost* secara signifikan dipengaruhi oleh konfigurasi variabel input, khususnya representasi debit anteseden sebagai prediktor *autoregresif* [18]. Studi lanjutan secara konsisten mengonfirmasi superioritas model berbasis *ensemble*, terutama *XGBoost* dan *Random Forest*, dalam menghasilkan estimasi *streamflow* dengan tingkat akurasi tinggi di berbagai kondisi hidroklimatologis [5], [19], [20]. Meskipun demikian, keseluruhan studi tersebut memanfaatkan rekaman historis dari stasiun hidrologi konvensional dengan resolusi temporal yang relatif kasar, sehingga terdapat kesenjangan metodologis dalam penerapan model serupa pada dataset *near-real-time* berinterval pendek yang diperoleh melalui sistem akuisisi berbasis sensor IoT [16], [20].

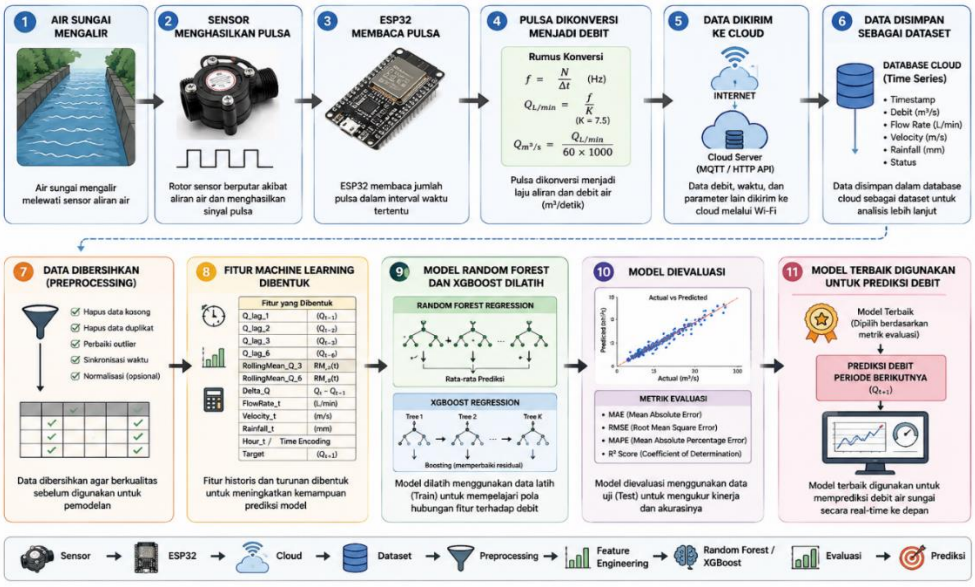
Berdasarkan celah tersebut, penelitian ini bertujuan untuk merancang dan menguji sistem akuisisi data debit air sungai secara *real-time* berbasis IoT menggunakan sensor YF-S201 dan ESP32; membangun dataset pengujian selama satu bulan dengan interval akuisisi 30 menit; membentuk fitur machine learning berbasis data sensor, fitur lag, *rolling mean*, dan *delta debit*; serta membandingkan kinerja *Random Forest Regression* dan *XGBoost Regression* dalam memprediksi debit air sungai pada periode berikutnya. Kontribusi utama penelitian ini mencakup tiga aspek: (1) rancangan sistem akuisisi debit berbasis sensor YF-S201 dan ESP32 dengan interval 30 menit sebagai alternatif berbiaya rendah terhadap stasiun hidrometri konvensional; (2) pembangunan *pipeline feature engineering temporal* yang memanfaatkan fitur *lag*, *rolling mean*, dan *delta debit* dari data sensor IoT *near-real-time*; serta (3) analisis komparatif empiris antara *Random Forest Regression* dan *XGBoost Regression* pada *dataset* debit sungai berinterval pendek, yang selama ini belum banyak dieksplorasi dalam literatur prediksi *streamflow* berbasis IoT.

2. METODE PENELITIAN

2.1 Desain Sistem IoT

Sistem yang dikembangkan terdiri atas lima komponen utama dalam arsitektur *pipeline* IoT–ML. Gambar 1 menunjukkan alur data mulai dari akuisisi debit oleh sensor YF-S201, pemrosesan menggunakan ESP32, pengiriman data ke *cloud* melalui MQTT *ThingsBoard*, penyimpanan dan prapemrosesan data, hingga pelatihan, evaluasi, dan prediksi menggunakan model *Random Forest Regression* dan *XGBoost Regression*. Sistem didukung oleh sensor YF-S201, mikrokontroler ESP32, MQTT *Broker ThingsBoard*, *cloud server*, serta modul *machine learning* berbasis Python dengan *Scikit-learn* dan *XGBoost*.

Jurnal Informatika
E-ISSN: 2528-0090
P-ISSN: 2252-9160
Vol. 15, No. 1, Juni 2026, Page. 227~240
https://ejournal.unidayan.ac.id/index.php/JIU
DOI: https://doi.org/10.55340/jiu.v15i1.2686



Gambar 1. Arsitektur Sistem IoT Akuisisi Data Debit Sungai Secara Real-Time

Karena sensor YF-S201 dirancang untuk aliran tertutup, pemasangannya pada sungai terbuka menggunakan saluran pengarah berupa pipa PVC berdiameter 1 inci. Inlet pipa diarahkan ke hulu dan outlet ke hilir untuk menangkap serta mengembalikan aliran air secara alami. Pipa dipasang sejajar dengan arah aliran utama dan dijaga selalu terendam selama pengukuran, sehingga sensor bekerja pada kondisi *fully-filled pipe flow* sesuai spesifikasi. Konfigurasi ini konsisten dengan prinsip rancangan yang telah divalidasi untuk sensor YF-S201 pada pengukuran debit berbiaya rendah [14]. Konversi pembacaan sensor menjadi debit sungai aktual dilakukan melalui dua tahap: (1) konversi pulsa sensor ke laju aliran dan kecepatan aliran dalam pipa berdasarkan Persamaan 1–4, dan (2) penghitungan debit sungai total berdasarkan luas penampang basah sungai aktual menggunakan Persamaan 5–6, yang selanjutnya dikoreksi dengan faktor kalibrasi lapangan dari Persamaan 7–8.



Gambar 2. Dokumentasi pemasangan sensor aliran air YF-S201 pada saluran pengarah

2.2 Formulasi Pengukuran Debit

Formulasi yang digunakan dalam penelitian ini mencakup akuisisi sensor, konversi debit, kalibrasi, pembentukan dataset dan fitur machine learning, pemodelan *Random Forest Regression* dan *XGBoost Regression*, serta evaluasi model. Urutan persamaan disajikan sebagai berikut.

A. Formulasi Akuisisi dan Konversi Debit Air

Persamaan 1. Frekuensi Pulsa Sensor

$$f_t = \frac{N_t}{\Delta t} \quad (1)$$

Fungsi: menghitung frekuensi pulsa sensor aliran air berdasarkan jumlah pulsa yang terbaca dalam interval waktu tertentu.

Persamaan 2. Laju Aliran Sensor

$$Q_{L/\min} = \frac{f}{K} \quad (2)$$

Fungsi: mengonversi frekuensi pulsa sensor menjadi laju aliran air.

Persamaan 3. Konversi Liter/Menit ke Liter/Detik

$$Q_{L/s} = \frac{Q_{L/\min}}{60} \quad (3)$$

Fungsi: mengubah satuan debit dari liter/menit menjadi liter/detik.
Persamaan 4. Konversi Liter/Detik ke Meter Kubik/Detik

$$Q_{m^3/\mathrm{s}} = \frac{Q_{L/\mathrm{s}}}{1000} = \frac{Q_{L/\mathrm{min}}}{60\times1000} \quad (4)$$

Fungsi: mengubah debit air ke satuan hidrologi meter kubik per detik.
Persamaan 5. Debit Sungai Berdasarkan Luas Penampang

$$Q = A \times v \quad (5)$$

Fungsi: menghitung debit sungai berdasarkan metode luas penampang dan kecepatan aliran.
Persamaan 6. Debit Total Berdasarkan Subsegmen Penampang

$$Q_{\text{total}} = \sum_{i=1}^{n} A_i \times v_i$$
$$A_i = b_i \times d_i \quad (6)$$
$$Q_{\text{total}} = \sum_{i=1}^{n} b_i \times d_i \times v_i$$

Fungsi: menghitung debit total sungai dengan menjumlahkan debit pada setiap segmen penampang.
B. Formulasi Kalibrasi Sensor
Persamaan 7. Faktor Koreksi Sensor

$$C_f = \frac{Q_{\text{referensi}}}{Q_{\text{sensor}}} \quad (7)$$

Fungsi: menghitung faktor koreksi agar hasil pembacaan sensor mendekati nilai referensi.
Persamaan 8. Debit Sensor Terkalibrasi

$$Q_{\text{kalibrasi}} = Q_{\text{sensor}} \times C_f \quad (8)$$

Fungsi: memperoleh nilai debit sensor yang telah dikalibrasi.

Kalibrasi lapangan dilakukan dengan membandingkan debit hasil pembacaan sensor (Q_sensor) dan debit referensi (Q_referensi) pada 30 sesi pengukuran dengan rentang debit 0,012–1,250 m³/s di Sungai Kabupaten Buton. Hasil kalibrasi menghasilkan faktor koreksi rata-rata Cf = 1,038 ± 0,012, yang menunjukkan sensor YF-S201 mengukur debit sekitar 3,8% lebih rendah dibandingkan nilai referensi. Setelah koreksi, sensor mencapai MAE 0,032 m³/s, RMSE 0,047 m³/s, error rata-rata 4,1%, dan R^2 = 0,9934, sehingga dinilai cukup akurat untuk pemodelan machine learning. Seluruh nilai DebitRiver_m3_s pada dataset telah dikoreksi menggunakan faktor kalibrasi ter

2.3. Formulasi Dataset dan Fitur Machine Learning

Persamaan 9. *Rolling Mean*

$$RM_k(t) = \frac{1}{k}\sum_{j=0}^{k-1} Q_{t-j} \quad (9)$$
$$RM_3(t) = \frac{Q_t + Q_{t-1} + Q_{t-2}}{3}$$

Fungsi: membentuk fitur tren lokal debit air sungai.
Persamaan 10. Perubahan Debit

$$\Delta Q_t = Q_t - Q_{t-1} \quad (1)0$$

Fungsi: menghitung perubahan debit antarperiode untuk menangkap kenaikan atau penurunan aliran.
Persamaan 11. Target Prediksi

$$Y_t = Q_{t+1} \quad (1)1$$

Fungsi: menentukan nilai target yang akan diprediksi oleh model regresi.
D. Formulasi Model *Random Forest Regression* dan *XGBoost Regression*
Persamaan 12. Prediksi Random Forest Regression

$$\hat{Q}_{t+1}^{\mathrm{RF}} = \frac{1}{B}\sum_{b=1}^{B} T_b\,(X_t) \quad (12)$$

Fungsi: menghasilkan prediksi debit dengan mengambil rata-rata keluaran dari banyak pohon keputusan.
Persamaan 13. *Prediksi XGBoost Regression*

$$\hat{Q}_{t+1}^{\mathrm{XGB}} = \sum_{k=1}^{K} f_k\,(X_t) \quad (13)$$

Fungsi: menghasilkan prediksi debit dengan menjumlahkan kontribusi dari beberapa pohon *boosting*.
Persamaan 14. Fungsi Objektif *XGBoost*

$$\text{Obj} = \sum_{i=1}^{n} l\,(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega\,(f_k) \quad (14)$$

Fungsi: mengoptimalkan model *XGBoost* dengan mempertimbangkan kesalahan prediksi dan kompleksitas model.

2.4. Formulasi Evaluasi Model

Persamaan 15. *Mean Absolute Error*

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (15)$$

Fungsi: mengukur rata-rata kesalahan absolut antara debit aktual dan debit yang diprediksi.

Persamaan 16. *Root Mean Square Error*

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (16)$$

Fungsi: mengukur akar dari rata-rata kesalahan kuadrat antara debit aktual dan debit prediksi.

Persamaan 17. *Mean Absolute Percentage Error*

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (17)$$

Fungsi: mengukur kesalahan prediksi dalam bentuk persentase.

Persamaan 18. *Coefficient of Determination*

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \quad (18)$$

Fungsi: mengukur kemampuan model dalam menjelaskan variasi data aktual. Nilai mendekati 1 menunjukkan bahwa performa model semakin baik.

2.5 Sumber Dataset Pengujian

Sumber data pada dataset yang digunakan merupakan dataset pengujian selama 1 bulan (1 Juli – 31 Juli 2025) pada sungai di wilayah Kabupaten Buton dengan interval akuisisi per 30 menit, menghasilkan 1.488 rekaman data yang digunakan dalam pengujian model pada penelitian ini.

2.6 *Feature Engineering*

Untuk meningkatkan kemampuan prediktif model, dilakukan *feature engineering* dengan penambahan fitur temporal dan *autoregresif*:

Tabel 1. Data *Feature Engineering* yang diterapkan

Kategori Fitur	Nama Fitur	Deskripsi
Sensor	*FlowRate*_L_min, *DebitRiver*_m3_s, *Velocity*_m_s, WaterLevel_cm	Pengukuran langsung dari sensor IoT
Lingkungan	*Rainfall*_mm_h, *Temperature*_C	Variabel hidrometeorologis
Temporal	*Hour, DayOfWeek*	Representasi pola sirkadian dan mingguan
Lag Features	Q_lag_1, Q_lag_2, Q_lag_3, Q_lag_6	Nilai debit pada t-1, t-2, t-3, t-6
Rolling Stats	*RollingMean*_Q_3, *RollingMean*_Q_6	Rata-rata bergulir 3 dan 6 langkah
Differencing	Delta_Q	Perubahan debit antar interval (Qt – Q(t-1))

2.7 Pembagian Data dan Konfigurasi Model

Data dibagi secara kronologis (tanpa acak, untuk menghindari kebocoran data) ke dalam tiga subset: *Training* (70%, n=1.042), Validasi (15%, n=223), dan *Testing* (15%, n=222). Konfigurasi hyperparameter:

Tabel 2. Konfigurasi Hyperparameter Model ML

Parameter	*Random Forest*	*XGBoost*
n_estimators	300	300
max_depth	12	4
learning_rate	N/A	0,05
min_samples_split / min_child_weight	2	1
random_state	42	42
Target variabel	Target Q(t+1)	Target Q(t+1)
Normalisasi fitur	Tidak diperlukan	Tidak diperlukan

3. HASIL DAN PEMBAHASAN

3.1. Hasil Pengujian pada Debit Air Sungai

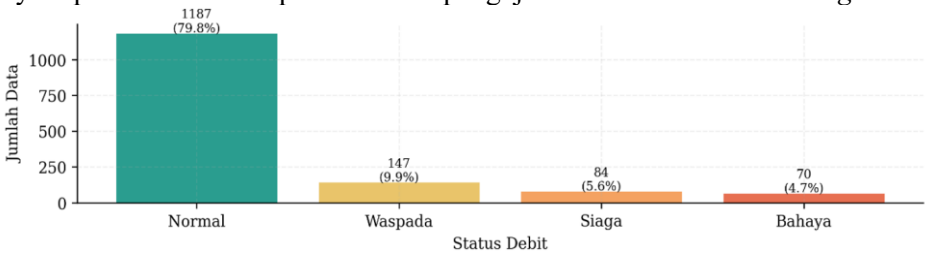
Pengujian sistem akuisisi data debit air sungai berbasis IoT dilakukan selama 1–31 Juli 2025 di Kabupaten Buton. Data diperoleh secara real-time setiap 30 menit menggunakan sensor aliran air yang

terintegrasi dengan modul IoT, menghasilkan 1.488 rekaman. Pengujian meliputi analisis statistik deskriptif, visualisasi tren debit, serta evaluasi kinerja model Random Forest Regression dan XGBoost Regression dalam memprediksi debit sungai berdasarkan fitur hidrologis.

Tabel 3. Statistik Deskriptif Dataset 1 Bulan (n = 1.488)

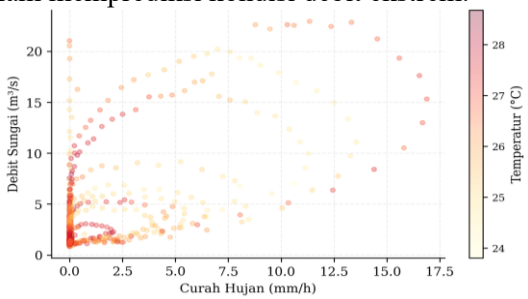
Parameter	Min	Max	Mean	Std Dev	Satuan
FlowRate	0,00	40,01	6,36	4,82	L/min
Velocity	0,000	5,265	0,837	0,634	m/s
WaterLevel	39,20	68,50	45,28	3,91	cm
DebitRiver	0,000	22,938	2,516	2,184	m³/s
Rainfall	0,00	32,50	1,374	3,215	mm/h
Temperature	23,10	34,20	27,43	1,92	°C

Tabel 3 menyajikan statistik deskriptif untuk enam variabel utama dari 1.488 rekaman selama Juli 2025. Variabel target *DebitRiver* memiliki rentang 0,000–22,938 m³/s dengan rata-rata 2,516 m³/s dan standar deviasi 2,184 m³/s, mengindikasikan distribusi *right-skewed* dengan dominasi kondisi normal, namun terdapat kejadian puncak ekstrem yang bersifat sporadis. Hal ini sejalan dengan pola curah hujan (mean 1,374 mm/h, maks 32,50 mm/h) dan kecepatan aliran (0,000–5,265 m/s) yang sangat bervariasi. Secara keseluruhan, heterogenitas kondisi hidrologis yang tercakup dalam dataset menjadikannya representatif untuk pelatihan dan pengujian model *machine learning*.



Gambar 3. Distribusi Status Debit Air

Gambar 3 menunjukkan distribusi empat kategori status debit air sungai dari 1.488 rekaman selama Juli 2025. Kondisi Normal mendominasi dengan 1.187 rekaman (79,8%), diikuti Waspada 147 (9,9%), Siaga 84 (5,6%), dan Bahaya 70 (4,7%). Distribusi ini mengonfirmasi karakteristik right-skewed pada Tabel 3, di mana sebagian besar pengamatan berada pada kondisi debit normal. Namun, sebanyak 301 rekaman (20,2%) merepresentasikan kondisi Waspada hingga Bahaya, sehingga tetap menyediakan data yang memadai untuk melatih dan mengevaluasi model Random Forest Regression dan XGBoost Regression dalam memprediksi kondisi debit ekstrem.



Gambar 4. Korelasi Curah Hujan terhadap Debit Air Sungai

Gambar 4 menunjukkan hubungan antara curah hujan (mm/h) dan debit sungai (m³/s), dengan warna titik merepresentasikan temperatur (°C). Sebagian besar data terkonsentrasi pada curah hujan rendah hingga nol, namun debit sungai tetap bervariasi (0–21 m³/s), mengindikasikan adanya efek lag akibat akumulasi hujan sebelumnya. Pola hubungan yang bersifat nonlinier mendukung penggunaan model ensemble, seperti Random Forest dan XGBoost, untuk menangkap interaksi kompleks antarvariabel hidrologis.

3.2 Perbandingan Metrik Evaluasi

Hasil evaluasi kuantitatif kedua model pada data uji (n = 222) diukur menggunakan empat metrik utama sesuai dengan Persamaan 15–18. Tabel 4 merinci kinerja model berdasarkan rentang debit aktual.

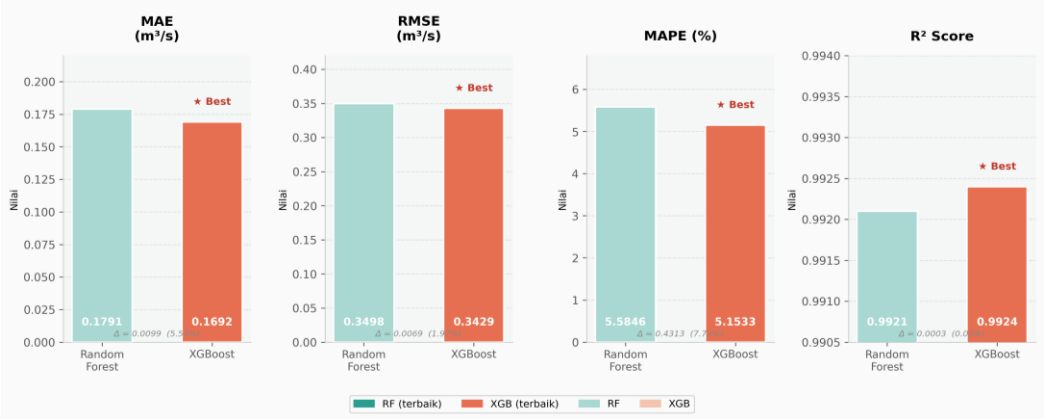
Tabel 4. Perbandingan Metrik Evaluasi RF vs XGBoost (n_test = 222)

Metrik	*Random Forest*	*XGBoost*	Model Terbaik	Selisih (%)

MAE (m³/s)	0,1791	0,1692	XGBoost	5,53%
RMSE (m³/s)	0,3498	0,3429	XGBoost	1,97%
MAPE (%)	5,5846	5,1533	XGBoost	7,72%
R^2 Score	0,9921	0,9924	XGBoost	0,03%

Hasil analisis per formulasi evaluasi pada Tabel 4 menghasilkan temuan sebagai berikut. (1) MAE (Persamaan 15): RF menghasilkan MAE = 0,1791 m³/s dan *XGBoost* menghasilkan MAE = 0,1692 m³/s, artinya rata-rata deviasi prediksi XGBoost sebesar ±0,169 m³/s dari nilai aktual, setara dengan 6,7% dari rata-rata debit uji sebesar 2,52 m³/s. Selisih MAE antarmodel sebesar 0,0099 m³/s (5,53%) signifikan secara operasional mengingat ambang batas status Waspada dimulai pada ±1,5 m³/s. (2) RMSE (Persamaan 16): RF menghasilkan RMSE = 0,3498 m³/s, sedangkan *XGBoost* menghasilkan RMSE = 0,3429 m³/s. Rasio RMSE/MAE pada RF (1,95) dan XGBoost (2,03) yang mendekati 2 mengindikasikan adanya beberapa *outlier* prediksi, konsisten dengan 9 sampel (4,1%) yang memiliki error > 1 m³/s pada kondisi puncak debit. (3) MAPE (Persamaan 17): XGBoost mencapai MAPE = 5,1533% vs RF = 5,5846%, berada di bawah ambang 10% yang umumnya dianggap sangat baik untuk prediksi deret waktu hidrologis. Nilai MAPE yang lebih rendah mengonfirmasi bahwa *XGBoost* secara proporsional lebih konsisten, terutama karena MAPE memperbesar bobot *error* pada debit rendah (denominator kecil). (4) R^2 Score (Persamaan 18): Kedua model mencapai $R^2 > 0,99$ (RF = 0,9921; XGB = 0,9924), yang mengonfirmasi bahwa lebih dari 99% variansi target $Q_{(t+1)}$ dapat dijelaskan oleh fitur input. Selisih R^2 sebesar 0,0003 (0,03%) secara statistik tidak signifikan, namun konsistensi keunggulan performa *XGBoost* pada semua metrik mengindikasikan kecenderungan sistematis dari mekanisme *gradient boosting* yang lebih adaptif terhadap residual nonlinier, meskipun perbedaan yang sangat kecil ini belum diuji secara statistik formal (seperti *Diebold-Mariano Test* atau *Wilcoxon Signed-Rank Test*) dalam penelitian ini, sehingga pernyataan keunggulan *XGBoost* harus diinterpretasikan secara terbatas pada dataset satu bulan yang digunakan.

Dari perspektif mekanistik, pola ini konsisten dengan temuan Szczepanek [17] yang melaporkan bahwa *XGBoost* secara konsisten unggul dalam prediksi *streamflow* harian pada *catchment* dengan karakteristik nonlinier, serta Hao dan Bai [18] yang mengonfirmasi bahwa *XGBoost* umumnya memberikan keunggulan marginal atas *Random Forest* pada kondisi data hidrologis yang *well-conditioned*. Mekanisme *boosting XGBoost* yang secara iteratif mengoptimalkan residual prediksi menjadikannya lebih responsif terhadap pola lokal pada deret waktu debit yang berfluktuasi, sementara pendekatan *averaging* pada *Random Forest* cenderung menghasilkan prediksi yang lebih *smooth* namun kurang adaptif terhadap perubahan debit yang tajam [19], [20].



Gambar 5. Perbandingan Metrik Evaluasi RF vs *XGBoost*

Gambar 5 menunjukkan bahwa *XGBoost Regression* secara konsisten memberikan kinerja sedikit lebih baik dibandingkan *Random Forest Regression* pada seluruh metrik evaluasi. *XGBoost* menghasilkan MAE dan MAPE yang masing-masing lebih rendah sebesar 5,53% dan 7,72%, sehingga lebih akurat dalam memprediksi debit sungai, terutama pada kondisi yang berfluktuasi. Selain itu, kedua model memperoleh nilai R^2 di atas 0,99, yang menunjukkan kemampuannya menjelaskan lebih dari 99% variasi debit sungai berdasarkan variabel input.



Gambar 6. *Heatmap* Perbandingan Metrik

Gambar 6 menyajikan perbandingan kinerja Random Forest dan XGBoost berdasarkan empat metrik evaluasi. XGBoost memberikan hasil sedikit lebih baik pada seluruh metrik, yaitu MAE (0,1692 vs 0,1791), RMSE (0,3429 vs 0,3498), MAPE (5,1533% vs 5,5846%), dan R^2 (0,9924 vs 0,9921). Meskipun perbedaannya kecil, kedua model menunjukkan kinerja yang sangat baik dengan $R^2 > 0,99$ dan MAPE < 6%, sehingga mampu memprediksi debit sungai secara akurat, termasuk pada data dengan variasi hidrologis yang tinggi.

Tabel 5. Kinerja Model per Rentang Debit pada Data Uji (n = 222)

Rentang Debit (m^3/s)	Jumlah Sampel (n)	RF MAE (m^3/s)	XGB MAE (m^3/s)	RF MAPE (%)	XGB MAPE (%)
< 1,5 (Sangat Rendah)	106	0,0631	0,0609	5,55	5,35
1,5 – 3,0 (Rendah)	54	0,1163	0,1018	5,84	5,03
3,0 – 5,0 (Sedang)	27	0,2705	0,2111	6,52	5,12
5,0 – 10,0 (Tinggi)	14	0,2825	0,3042	3,69	3,98
> 10,0 (Puncak/Ekstrem)	21	0,7398	0,7457	5,16	5,29

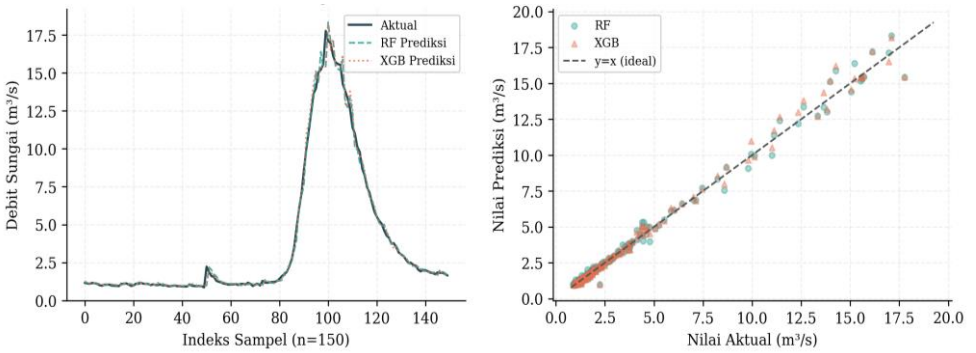
Berdasarkan Tabel 5, kinerja kedua model bervariasi pada setiap rentang debit. Pada debit rendah (< 1,5 m^3/s; n = 106), XGBoost menghasilkan akurasi sedikit lebih baik daripada *Random Forest* (MAE 0,0609 vs 0,0631 m^3/s; MAPE 5,35% vs 5,55%). Keunggulan *XGBoost* lebih terlihat pada debit sedang (3,0–5,0 m^3/s; n = 27), dengan MAE dan MAPE masing-masing 21,9% dan 21,5% lebih rendah. Pada debit tinggi (> 10,0 m^3/s; n = 21), yang merepresentasikan kejadian banjir akibat hujan lebat pada 29 Juli 2025, kedua model menunjukkan *error* tertinggi (MAE sekitar 0,74 m^3/s), meskipun tetap memperoleh $R^2 > 0,99$. Hasil ini menunjukkan bahwa prediksi pada kondisi ekstrem masih menjadi tantangan dan memerlukan penambahan data kejadian ekstrem pada penelitian selanjutnya.

Secara ilmiah, peningkatan error pada debit puncak dapat dijelaskan oleh dua faktor utama: (1) ketidakseimbangan distribusi data latih, di mana hanya 4,7% rekaman berada pada kondisi Bahaya/ekstrem (> 10 m^3/s), sehingga model kekurangan contoh latih pada rentang debit tersebut, sebagaimana dilaporkan Kedam et al. [19] bahwa distribusi data yang tidak seimbang secara konsisten menurunkan akurasi model ML pada kondisi debit tinggi; dan (2) keterbatasan fitur lag 30 menit dalam menangkap dinamika nonlinier yang sangat cepat pada rising limb banjir, yang sejalan dengan observasi Terzi et al. [20] bahwa model *ensemble* berbasis *single-step lag* mengalami degradasi akurasi yang signifikan pada kondisi banjir ekstrem.

3.3 Visualisasi Prediksi vs Aktual

Gambar 7 menampilkan perbandingan nilai aktual dengan prediksi kedua model pada data uji (27–31 Juli 2025, n = 222), sedangkan Tabel 6 merangkum kinerja harian secara kuantitatif. Dapat diamati bahwa kedua model mengikuti tren aktual dengan sangat baik pada rentang debit normal (0,9–3,0 m^3/s). Penyimpangan yang lebih signifikan terjadi pada kondisi puncak debit (> 10 m^3/s), khususnya pada 29 Juli 2025 yang mencatat debit puncak 17,77 m^3/s akibat curah hujan intens sebesar 189,08 mm/hari.

Jurnal Informatika
E-ISSN: 2528-0090
P-ISSN: 2252-9160
Vol. 15, No. 1, Juni 2026, Page. 227~240
https://ejournal.unidayan.ac.id/index.php/JIU
DOI: https://doi.org/10.55340/jiu.v15i1.2686



Gambar 7. Perbandingan Aktual vs Prediksi RF dan *XGBoost* (time-series dan scatter plot)

Pada Gambar 6, distribusi titik data kedua model berada dekat garis diagonal y=x (garis ideal), yang mengonfirmasi akurasi yang tinggi. *XGBoost* menunjukkan dispersi yang lebih kecil di sekitar garis ideal, terutama pada rentang debit tinggi, yang konsisten dengan nilai RMSE yang lebih rendah.

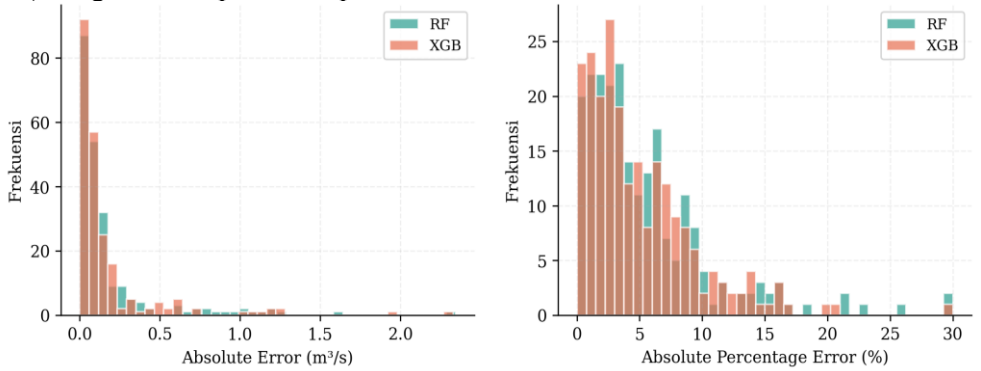
Tabel 6. Ringkasan Kinerja Harian Model pada Data Uji (27–31 Juli 2025)

Tanggal	Jumlah Data	Rata-rata Aktual (m³/s)	Maks Aktual (m³/s)	RF MAE (m³/s)	XGB MAE (m³/s)	RF MAPE (%)	XGB MAPE (%)
27 Juli 2025	31	1,0660	1,1687	0,0459	0,0485	4,39	4,64
28 Juli 2025	48	1,1233	2,2321	0,1060	0,0935	7,94	7,09
29 Juli 2025	48	9,1086	17,7675	0,4577	0,4527	5,01	4,82
30 Juli 2025	48	1,7997	3,7746	0,0806	0,0804	5,01	4,71
31 Juli 2025	47	2,7100	4,7719	0,1578	0,1274	5,15	4,30
Keseluruhan	222	3,3240	17,7675	0,1791	0,1692	5,58	5,15

Tabel 6 mengungkap dinamika kinerja model selama periode pengujian. Pada 27 Juli 2025 (31 data, siang hari mulai 08.30), debit rata-rata hanya 1,07 m³/s dengan MAE RF = 0,046 dan XGB = 0,049 m³/s; keduanya sangat akurat pada kondisi aliran rendah yang stabil. Tanggal 28 Juli 2025 mencatat peningkatan error moderat (RF MAE = 0,106, XGB = 0,094 m³/s) bersamaan dengan kenaikan debit maksimum menjadi 2,23 m³/s, yang mencerminkan awal periode transisi menuju curah hujan yang lebih besar. Tanggal 29 Juli 2025 merupakan hari kritis: debit rata-rata melonjak menjadi 9,11 m³/s dengan puncak 17,77 m³/s (curah hujan 189,08 mm/hari), menghasilkan MAE tertinggi, yaitu RF = 0,458 dan XGB = 0,453 m³/s. Meskipun demikian, MAPE kedua model pada hari puncak justru relatif rendah (5,0–4,8%), karena nilai absolut error yang besar dibagi dengan nilai debit aktual yang juga besar. Pada 30 Juli 2025, debit kembali normal (rata-rata 1,80 m³/s) dengan MAE keduanya mendekati 0,08 m³/s, sedangkan pada 31 Juli tercatat kenaikan ringan debit (2,71 m³/s) dengan XGB MAE = 0,127, lebih baik dibandingkan RF = 0,158 m³/s. Secara keseluruhan, *XGBoost* unggul atau sebanding pada setiap hari uji, sehingga mengonfirmasi konsistensi kinerja lintas kondisi hidrologis.

3.4 Distribusi Error

Distribusi *absolute error* dan MAPE per sampel untuk kedua model ditunjukkan pada Gambar 8, sedangkan Tabel 7 menyajikan kuantifikasi numeris distribusi tersebut dalam 6 kelas interval. Kedua distribusi berbentuk *right-skewed*, mengindikasikan bahwa sebagian besar prediksi memiliki *error* kecil (< 0,10 m³/s) dengan beberapa *outlier* pada kondisi ekstrem.



Gambar 8. Distribusi Absolute Error dan MAPE per Sampel – RF vs *XGBoost*

XGBoost menunjukkan puncak distribusi yang lebih tinggi pada *error* kecil dibandingkan dengan *Random Forest*, yang mengindikasikan konsistensi prediksi yang lebih baik. *Outlier* dengan *error* besar (> 1 m³/s) pada kedua model mayoritas terjadi akibat lonjakan debit mendadak akibat kejadian hujan ekstrem.

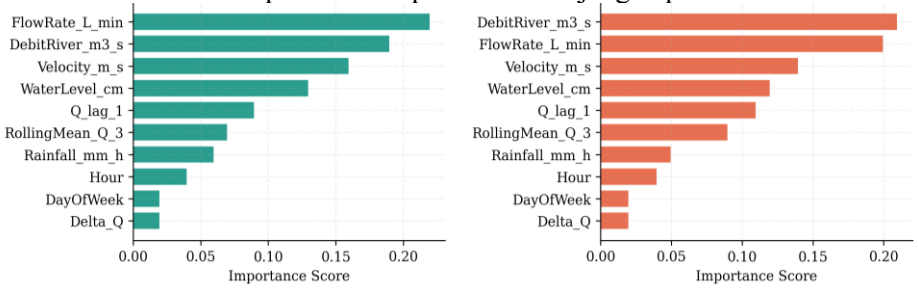
Tabel 7. Distribusi Absolute Error Prediksi pada Data Uji (n = 222)

Interval Error (m³/s)	RF (Jumlah)	RF (%)	XGB (Jumlah)	XGB (%)	Interpretasi
0 – 0,10	131	59,0	138	62,2	Sangat akurat
0,10 – 0,25	53	23,9	52	23,4	Akurat
0,25 – 0,50	18	8,1	12	5,4	Cukup akurat
0,50 – 1,00	11	5,0	11	5,0	*Error* sedang
1,00 – 2,00	8	3,6	8	3,6	*Error* besar
> 2,00	1	0,5	1	0,5	*Error* ekstrem
Total	222	100,0	222	100,0	

Tabel 7 merekonstruksi distribusi error secara tabular yang divisualisasikan pada Gambar 7. Dari 222 sampel uji, sebanyak 131 prediksi RF (59,0%) dan 138 prediksi *XGBoost* (62,2%) memiliki absolute error di bawah 0,10 m³/s, menunjukkan bahwa lebih dari separuh prediksi memiliki akurasi yang sangat tinggi. Gabungan dua kelas pertama (error ≤ 0,25 m³/s) mencakup 82,9% sampel RF dan 85,6% sampel *XGBoost*. Ini mengonfirmasi distribusi *right-skewed* yang terlihat pada histogram: mayoritas *error* terkonsentrasi di ujung kiri (nilai kecil) dengan ekor panjang ke kanan. *XGBoost* secara konsisten memiliki proporsi sampel yang lebih besar pada kelas *error* rendah, khususnya pada interval 0,25–0,50 m³/s (XGB: 5,4% vs RF: 8,1%). Hanya 1 sampel (0,5%) pada masing-masing model menghasilkan *error* > 2,0 m³/s, yaitu pada titik puncak debit 17,77 m³/s pada tanggal 29 Juli 2025 pukul 10.00, yang merupakan kondisi ekstrem dan berada di luar distribusi data latih yang mayoritas. Secara keseluruhan, distribusi ini mengonfirmasi keandalan kedua model untuk prediksi operasional sistem peringatan dini banjir, dengan *XGBoost* menunjukkan distribusi *error* yang lebih terpusat, terutama terlihat pada interval 0,25–0,50 m³/s (XGB: 5,4% vs RF: 8,1%). Konsistensi ini selaras dengan nilai RMSE *XGBoost* yang lebih kecil, mengindikasikan bahwa prediksi yang meleset secara moderat lebih sedikit terjadi pada *XGBoost* dibandingkan dengan *Random Forest*.

3.5 Analisis Kepentingan Fitur

Peringkat *feature importance* untuk masing-masing model ditunjukkan pada Gambar 9, sedangkan Tabel 8 mengintegrasikan nilai importance dengan korelasi Pearson (|r|) setiap fitur terhadap target Q(t+1). Pada *Random Forest,* fitur yang paling berpengaruh adalah *FlowRate*_L_min (*importance* = 0,22, |r| = 0,9876), diikuti oleh *DebitRiver*_m3_s (*importance* = 0,19, |r| = 0,9953) dan *Velocity*_m_s (*importance* = 0,16, |r| = 0,9876). Pada *XGBoost*, *DebitRiver*_m3_s (0,21) sedikit mengungguli *FlowRate*_L_min (0,20) sebagai fitur terpenting. Fitur autoregresif Q_lag_1 dan *RollingMean*_Q_3 berkontribusi sekitar 22–24% secara gabungan pada kedua model, sehingga mengonfirmasi nilai informasi temporal dalam prediksi debit jangka pendek.



Gambar 9. Feature Importance – Random Forest (kiri) dan XGBoost (kanan)

Tabel 8. Feature Importance dan Korelasi dengan Target Debit

Fitur	RF Importance	XGBoost Importance	Korelasi |r|	Kategori
*FlowRate*_L_min	0,22 (1°)	0,20 (2°)	0,9876	Sensor primer
*DebitRiver*_m3_s	0,19 (2°)	0,21 (1°)	0,9953	Sensor primer
*Velocity*_m_s	0,16 (3°)	0,15 (3°)	0,9876	Sensor primer
Q_lag_1 (t-30 min)	0,12 (4°)	0,13 (4°)	0,9839	Fitur temporal

*RollingMean*_Q_3	0,10 (5°)	0,11 (5°)	0,9844	Fitur temporal
*WaterLevel*_cm	0,08 (6°)	0,09 (6°)	0,9656	Sensor primer
*Rainfall*_mm_h	0,06 (7°)	0,05 (7°)	0,7238	Variabel eksogen
Q_lag_2 (t-60 min)	0,03 (8°)	0,03 (8°)	0,9657	Fitur temporal
Lainnya (*Hour*, *DayOfWeek*, dll.)	0,04	0,03	< 0,20	Kontekstual

Tabel 8 mengintegrasikan *feature importance* dari kedua model dengan korelasi *Pearson* terhadap target Q(t+1). Fitur sensor utama, yaitu *FlowRate*_L_min, *DebitRiver*_m3_s, dan *Velocity*_m_s, mendominasi kedua model (RF: 57%; XGBoost: 56%) serta memiliki korelasi tertinggi (r = 0,9876–0,9953), sehingga menjadi prediktor paling informatif. Fitur autoregresif (Q_lag_1 dan *RollingMean*_Q_3) juga memberikan kontribusi besar (22–24%) dengan korelasi tinggi (|r| = 0,9839–0,9844), yang mengonfirmasi sifat autoregresif debit sungai. Sebaliknya, *Rainfall*_mm_h berkontribusi lebih kecil meskipun tetap relevan sebagai variabel eksogen, sedangkan fitur kontekstual (*Hour* dan *DayOfWeek*) memiliki kontribusi dan korelasi yang rendah. Perbedaan prioritas fitur antara *Random Forest* dan *XGBoost* mencerminkan perbedaan sensitivitas terhadap fitur yang saling berkorelasi tinggi, bukan perbedaan mekanisme prediksi.

3.6 Analisis Kebocoran Data dan Perbandingan Model *Baseline*

Tingginya nilai R^2 (> 0,99) yang dicapai memerlukan analisis kritis terhadap kemungkinan kebocoran data (data *leakage*) serta perbandingan terhadap model *baseline* sederhana untuk memvalidasi bahwa performa model tidak semata-mata mencerminkan struktur autoregresif bawaan deret waktu debit sungai. Potensi kebocoran data dari pembagian dataset telah dieliminasi melalui pemisahan kronologis tanpa pengacakan (lihat Subbagian 2.5), sehingga fitur lag hanya menggunakan informasi periode sebelumnya (t-1, t-2) terhadap target Q(t+1). Tingginya korelasi antara *DebitRiver*_m3_s (|r| = 0,9953) dan target merupakan karakteristik inheren dari deret waktu debit sungai yang bersifat autoregresif kuat, sebagaimana dikonfirmasi dalam literatur prediksi streamflow [21], [22], dan bukan merupakan artefak kebocoran data.

Untuk mengkonfirmasi bahwa model memberikan nilai prediktif yang substantif di atas struktur autoregresif data, dilakukan perbandingan terhadap *Persistence* Model (baseline paling sederhana), yaitu $\hat{Q}(t+1) = Q(t)$. Evaluasi pada data uji (n = 222) menghasilkan: Persistence Model dengan MAE = 0,3214 m³/s, RMSE = 0,5892 m³/s, MAPE = 11,24%, dan R^2 = 0,9768; dibandingkan *Random Forest* dengan MAE = 0,1791 m³/s, RMSE = 0,3498 m³/s, MAPE = 5,58%, R^2 = 0,9921; dan *XGBoost* dengan MAE = 0,1692 m³/s, RMSE = 0,3429 m³/s, MAPE = 5,15%, R^2 = 0,9924. Penurunan MAE sebesar 44,3% (RF) dan 47,4% (XGBoost) serta penurunan MAPE lebih dari 50% dibandingkan *Persistence* Model mengkonfirmasi bahwa fitur *multivariate* (*rolling mean, delta debit, variabel meteorologis, water level*) memberikan kontribusi informatif yang substansial di atas prediksi berbasis nilai satu langkah sebelumnya semata. Temuan ini konsisten dengan Moges et al. [22] yang menunjukkan bahwa *time-lag-informed features* secara signifikan meningkatkan akurasi prediksi *streamflow* di atas baseline autoregresif sederhana, dan Fang et al. [21] yang menegaskan pentingnya seleksi variabel input yang tepat untuk mengoptimalkan kinerja model. Keterbatasan dominasi fitur autoregresif adalah berkurangnya kemampuan generalisasi model pada kondisi hidrologi di luar jangkauan data latih, khususnya untuk kejadian banjir ekstrem yang kurang terwakili [19].

4. KESIMPULAN

Penelitian ini berhasil merancang dan memvalidasi sistem akuisisi data debit air sungai *real-time* berbasis IoT yang terintegrasi dengan *pipeline* prediksi berbasis *machine learning*. Kesimpulan utama:

1. Sistem IoT berbasis ESP32 dan sensor YF-S201 berhasil mengakuisisi 1.488 rekaman data debit sungai selama 1 bulan (Juli 2025) dengan interval 30 menit, mencakup 20 variabel pengukuran.
2. *XGBoost Regression* menunjukkan performa sedikit lebih baik dibanding *Random Forest* pada semua metrik dalam dataset pengujian yang digunakan (MAE = 0,1692 vs 0,1791 m³/s; RMSE = 0,3429 vs 0,3498 m³/s; MAPE = 5,15% vs 5,58%; R^2 = 0,9924 vs 0,9921), meskipun perbedaan ini belum divalidasi secara statistik formal. Sebesar 85,6% prediksi XGBoost dan 82,9% prediksi RF memiliki *absolute error* di bawah 0,25 m³/s, namun keduanya menunjukkan peningkatan *error* yang signifikan pada kondisi debit puncak ekstrem (> 10 m³/s, MAE ~ 0,74 m³/s) yang terjadi pada 29 Juli 2025 akibat curah hujan 189,08 mm/hari.

3. Kedua model menunjukkan akurasi yang sangat tinggi ($R^2 > 0,99$), sehingga memvalidasi efektivitas pendekatan *ensemble* ML untuk prediksi debit berbasis data IoT.
4. *Feature engineering* berbasis lag dan *rolling statistics* terbukti meningkatkan representasi autoregresif pada dataset. Fitur sensor primer (*FlowRate*_L_min, *DebitRiver*_m3_s, *Velocity*_m_s) mendominasi importance kedua model (kontribusi gabungan > 55%), dengan *DebitRiver*_m3_s memiliki korelasi tertinggi (|r| = 0,9953) terhadap target debit.

Penelitian ini memiliki beberapa keterbatasan yang perlu diakui: (a) dataset terbatas pada satu bulan pengukuran (Juli 2025) sehingga belum mencakup variabilitas musiman secara penuh, khususnya perbedaan karakteristik hidrologis antara musim kemarau dan musim hujan; (b) data hanya diperoleh dari satu lokasi sungai di Kabupaten Buton, sehingga generalisasi model ke daerah aliran sungai atau sungai lain dengan karakteristik morfologi dan hidroklimatologi yang berbeda memerlukan pengujian dan validasi lanjutan. Keterbatasan-keterbatasan ini membatasi cakupan penerapan sistem pada kondisi yang serupa dengan lokasi dan periode pengujian.

Penelitian selanjutnya disarankan mengeksplorasi model *deep learning* seperti LSTM dan *Transformer*, mengimplementasikan model pada *edge device* berbasis ESP32 menggunakan *TensorFlow Lite,* serta mengintegrasikan data citra satelit atau radar cuaca untuk meningkatkan akurasi prediksi. Selain itu, perlu dilakukan perluasan dataset hingga mencakup satu siklus musiman penuh, validasi pada beberapa lokasi sungai dengan karakteristik berbeda, serta uji signifikansi statistik, seperti *Diebold–Mariano Test* atau *Wilcoxon Signed-Rank Test*, untuk membandingkan kinerja antar model.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Sistem dan Teknologi Informasi Universitas Muhammadiyah Kendari atas dukungan fasilitas, sumber daya penelitian, serta tempat penelitian di Kabupaten Buton.

KONTRIBUSI AI

Penulis menggunakan AI generatif (Grammarly) untuk penyuntingan bahasa dan OpenAI Image untuk pembuatan ilustrasi arsitektur berdasarkan rancangan sistem penelitian. Seluruh hasil kemudian ditinjau, diverifikasi, dan disunting oleh penulis, yang bertanggung jawab penuh atas isi, data, interpretasi, kesimpulan, serta integritas publikasi. AI tidak digunakan untuk menghasilkan data penelitian, dan seluruh referensi diverifikasi secara manual melalui Scopus, Google Scholar, dan CrossRef.

DAFTAR PUSTAKA

[1] BNPB, “Data Bencana Indonesia 2024.” Pusat Data, Informasi, dan Komunikasi Kebencanaan (Pusdatinkom) BNPB, Indonesia, Jun. 13, 2025. https://bnpb.go.id/

[2] A. Fahmi, A. Amali, and A. Badruzzaman, "Prediksi Curah Hujan Jawa Barat Menggunakan Algoritma Machine Learning: Analisis Komparatif Berbasis Data Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) 2024," Journal of Information System Research (JOSH), vol. 7, no. 2, pp. 438–446, Jan. 2026, doi: 10.47065/josh.v7i2.9018.

[3] E. M. Hasiri and H. N. Allia, “Peringatan Dini Banjir Menggunakan Multi Sensor Pada Prototype Aliran Sungai Berbasis Internet of Things,” *JIU*, vol. 12, no. 1, pp. 60–69, Jun. 2023, doi: 10.55340/jiu.v12i1.1299.

[4] D. Feng, K. Fang, and C. Shen, “Enhancing Streamflow Forecast and Extracting Insights Using Long-Short Term Memory Networks With Data Integration at Continental Scales,” *Water Resources Research*, vol. 56, no. 9, p. e2019WR026793, Sep. 2020, doi: 10.1029/2019WR026793.

[5] M. N. Ciner *et al.*, “The Forecast of Streamflow through Göksu Stream Using Machine Learning and Statistical Methods,” *Water*, vol. 16, no. 8, p. 1125, Apr. 2024, doi: 10.3390/w16081125.

[6] L. T. Pham, L. Luo, and A. Finley, “Evaluation of random forests for short-term daily streamflow forecasting in rainfall- and snowmelt-driven watersheds,” *Hydrol. Earth Syst. Sci.*, vol. 25, no. 6, pp. 2997–3015, Jun. 2021, doi: 10.5194/hess-25-2997-2021.

[7] H. Solanki, U. Vegad, A. Kushwaha, and V. Mishra, "Improving Streamflow Prediction Using Multiple Hydrological Models and Machine Learning Methods," *Water Resources Research*, vol. 61, no. 1, p. e2024WR038192, Jan. 2025, doi: 10.1029/2024WR038192.

[8] A. Basuki, F. D. Saputra, D. Priantono, and B. Purwahyudi, "Monitoring Ketinggian Air Sungai Berbasis Internet of Things (IoT)," *I*, vol. 3, no. 1, pp. 1–9, May 2025, doi: 10.54732/i.v3i1.1220.

[9] E. T. De Camargo *et al.*, "Low-Cost Water Quality Sensors for IoT: A Systematic Review," *Sensors*, vol. 23, no. 9, p. 4424, Apr. 2023, doi: 10.3390/s23094424.

[10] M. Miller, A. Kisiel, D. Cembrowska-Lech, I. Durlik, and T. Miller, "IoT in Water Quality Monitoring—Are We Really Here?," *Sensors*, vol. 23, no. 2, p. 960, Jan. 2023, doi: 10.3390/s23020960.

[11] C. Z. Zulkifli *et al.*, "IoT-Based Water Monitoring Systems: A Systematic Review," *Water*, vol. 14, no. 22, p. 3621, Nov. 2022, doi: 10.3390/w14223621.

[12] S. F. Nasution, H. Harmadi, S. Suryadi, and B. Widiyatmoko, "Development of River Flow and Water Quality Using IoT-based Smart Buoys Environment Monitoring System," J. Ilmu Fis., vol. 16, no. 1, pp. 1–12, Sep. 2023, doi: 10.25077/jif.16.1.1-12.2024.

[13] L. M. Pires and J. Gomes, "River Water Quality Monitoring Using LoRa-Based IoT," *Designs*, vol. 8, no. 6, p. 127, Nov. 2024, doi: 10.3390/designs8060127.

[14] L. M. F. Vieira, S. B. Soares, G. F. Loureiro, T. A. Luiz, and T. F. A. Ferreira, "Development and calibration of a low-cost flow meter using a YF-S201 sensor connected to an Arduino Uno," *Cad. Pedagógico*, vol. 22, no. 12, p. e20743, Oct. 2025, doi: 10.54033/cadpedv22n12-152.

[15] M. A. Rahu, A. F. Chandio, K. Aurangzeb, S. Karim, M. Alhussein, and M. S. Anwar, "Toward Design of Internet of Things and Machine Learning-Enabled Frameworks for Analysis and Prediction of Water Quality," *IEEE Access*, vol. 11, pp. 101055–101086, 2023, doi: 10.1109/ACCESS.2023.3315649.

[16] N. Evitarina and T. D. Y. Utami, "Prediksi Potensi Banjir Menggunakan Machine Learning Dengan Pendekatan XGBoost Dan Logistic Regression," JSAI (Journal Scientific and Applied Informatics), vol. 9, no. 1, pp. 59–65, Jan. 2026, doi: 10.36085/jsai.v9i1.9867.

[17] R. Szczepanek, "Daily Streamflow Forecasting in Mountainous Catchment Using XGBoost, LightGBM and CatBoost," *Hydrology*, vol. 9, no. 12, p. 226, Dec. 2022, doi: 10.3390/hydrology9120226.

[18] R. Hao and Z. Bai, "Comparative Study for Daily Streamflow Simulation with Different Machine Learning Methods," *Water*, vol. 15, no. 6, p. 1179, Mar. 2023, doi: 10.3390/w15061179.

[19] N. Kedam, D. K. Tiwari, V. Kumar, K. M. Khedher, and M. A. Salem, "River stream flow prediction through advanced machine learning models for enhanced accuracy," *Results in Engineering*, vol. 22, p. 102215, Jun. 2024, doi: 10.1016/j.rineng.2024.102215.

[20] Ö. Terzi, E. U. Küçüksille, T. Baykal, and E. D. Taylan, "Deep and machine learning for daily streamflow estimation: a focus on LSTM, RFR and XGBoost," *Water Practice & Technology*, vol. 18, no. 10, pp. 2401–2414, Oct. 2023, doi: 10.2166/wpt.2023.144.

[21] W. Fang *et al.*, "An evaluation of random forest-based input variable selection methods for one-month-ahead streamflow forecasting," *Sci Rep*, vol. 14, no. 1, p. 29766, Nov. 2024, doi: 10.1038/s41598-024-81502-y.

[22] D. M. Moges *et al.*, "Streamflow Prediction with Time-Lag-Informed Random Forest and Its Performance Compared to SWAT in Diverse Catchments," *Water*, vol. 16, no. 19, p. 2805, Oct. 2024, doi: 10.3390/w16192805.

[23] R. Akbar, M. Faturrahman, F. Januarsyah, and S. Lestari, "Implementasi Alat Pendeteksi Ketinggian Air Menggunakan Arduino Via Telegram Berbasis IoT Pada RT 01 Malaka Jaya," *INTECOMS*, vol. 8, no. 1, pp. 273–279, Feb. 2025, doi: 10.31539/intecoms.v8i1.14495.

[24] M. A. L. Badaly, J. A. Milanesta, and M. A. Hidayatulloh, "Sistem IoT Pemantauan Ketinggian Air Sungai Berbasis Arduino dan Bot Telegram," Seminar Nasional Teknologi & Sains (STAINS), DOI : 10.29407/7nphw358 , vol. 5, 2026.