Journal of Technology and Data Science (JTDS)
Vol. 4, No. 1, Februari 2026, hal. 285-298
p-ISSN: 2962-3693 | e-ISSN: 2962-7788

Penjadwalan Adaptif Siklus Kerja Sensor IoT Bertenaga Baterai Menggunakan Reinforcement Learning: Studi Simulasi

Willy Artanika Rikarda[1]*, Imam Mualim[1]

[1]Program Studi Teknik Informatika, Universitas Nahdatul Ulama Lampung, Indonesia

E-mail: willy@unulampung.ac.id; imammualim200281@gmail.com

**Penulis Korespondensi (Corresponding Author)*

© 2026. Willy Artanika Rikarda, Imam Mualim. Diterbitkan oleh JTDS. Artikel ini bersifat open-access di bawah lisensi Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0).

Info Artikel

Diterima:
28 Januari 2026
Direvisi:
17 Februari 2026
Diterbitkan:
25 Februari 2026

Kata Kunci:
Internet of Things; reinforcement learning; Q-learning; penjadwalan siklus kerja; manajemen energi sensor

ABSTRAK

Efisiensi energi merupakan permasalahan rancangan paling kritis pada simpul sensor Internet of Things (IoT) bertenaga baterai yang dituntut beroperasi tanpa pengawasan selama bertahun-tahun. Pendekatan duty cycle dengan jadwal siklus kerja tetap menghemat energi namun menimbulkan latensi tinggi dan potensi kehilangan data ketika lalu lintas pesan bersifat stokastik. Penelitian ini mengusulkan kebijakan penjadwalan siklus kerja adaptif berbasis Reinforcement Learning (RL) yang menentukan durasi tidur berikutnya secara dinamis berdasarkan tingkat baterai, okupansi buffer data, dan pola kedatangan trafik. Permasalahan diformulasikan sebagai Markov Decision Process (MDP) dan diselesaikan menggunakan Q-learning dengan diskretisasi ruang state-action; lingkungan eksperimen sepenuhnya sintetis sehingga analisis perilaku kebijakan dapat dilakukan terkendali tanpa ketergantungan perangkat keras. Agen RL dilatih selama 2.000 episode (setara 50 tahun simulasi) dan diuji terhadap tiga baseline: duty cycle tetap, heuristik ambang batas, dan kebijakan optimal value iteration. Hasil simulasi menunjukkan agen RL memperpanjang usia baterai hingga 42% dibanding duty cycle 30 detik, dengan rasio pengiriman paket 97,8% dan latensi 18,1 detik mendekati kebijakan optimal value iteration (98,5% dan 17,2 detik). Agen juga beradaptasi terhadap pergeseran trafik non-stasioner dalam waktu sekitar tiga jam tanpa mekanisme deteksi perubahan eksplisit, sementara baseline ambang batas mengalami penurunan rasio pengiriman sebesar 12% sebelum di-tuning ulang. Temuan ini membuktikan kelayakan Q-learning tabular sebagai kontroler siklus kerja adaptif yang ringan dan dapat dijalankan pada mikrokontroler berdaya rendah.

Article Info

Received:
January 28, 2026
Revised:
February 17, 2026
Published:
February 25, 2026

Keywords:
Internet of Things; reinforcement learning; Q-learning; sleep-wake scheduling; sensor energy management

ABSTRACT

Energy efficiency is the most critical design concern for battery-powered Internet of Things (IoT) sensor nodes that must operate unattended for years. Fixed duty-cycle schedules conserve energy but introduce high latency and potential data loss when message arrivals are stochastic. This study proposes an adaptive sleep/wake scheduling policy based on Reinforcement Learning (RL) that dynamically determines the next sleep duration from battery level, buffer occupancy, and traffic pattern. The problem is formulated as a Markov Decision Process (MDP) and solved using tabular Q-learning over discretized state-action spaces; the experimental environment is entirely synthetic, enabling controlled policy analysis without any hardware dependency. The RL agent is trained for 2,000 episodes (equivalent to 50 simulated years) and benchmarked against three baselines: fixed duty cycle, threshold-based heuristic, and an offline optimal value-iteration policy. Simulation results show that the RL agent extends battery lifetime by up to 42% over a 30-second fixed

duty cycle while maintaining a 97.8% packet delivery ratio and 18.1-second average latency approaching the value-iteration optimum of 98.5% and 17.2 seconds. The agent also adapts to non-stationary traffic shifts within roughly three hours of operation without any explicit change-detection mechanism, whereas the threshold baseline suffers a 12% drop in delivery ratio before manual retuning. These findings confirm the feasibility of tabular Q-learning as a lightweight adaptive sleep/wake controller executable on low-power microcontrollers.

1. PENDAHULUAN

Penyebaran simpul sensor IoT di lapangan kini menjangkau pemantauan kualitas udara urban, pengukuran daya listrik rumah tangga secara waktu nyata, hingga telemonitoring kondisi pasien jarak jauh [1]–[3]. Kemudahan integrasi inilah yang menjadikan IoT tulang punggung sistem cerdas modern; ironisnya, ketergantungan perangkat-perangkat tersebut pada baterai kapasitas terbatas menyisakan satu ketidakpastian klasik berapa lama node mampu bertahan sebelum kehabisan daya. Pada layanan kritis seperti pemantauan kesehatan pasien atau pengawasan daya rumah tangga 24 jam [4], [5], penggantian baterai yang terlambat berarti hilangnya kesinambungan data; pada penerapan industri kelas 6G IoT [2], skala yang dijanjikan miliaran simpul membuat opsi penggantian manual makin tidak ekonomis.

Duty cycling menempatkan radio dan prosesor ke mode tidur secara berkala adalah strategi penghematan yang paling lazim dipraktikkan di lapangan. Implementasi konkret seperti pintu rumah berbasis IoT [6] atau sistem pemantauan kualitas udara [3] secara default menjadwalkan bangun pada interval tetap, sebuah pilihan yang mudah diprogram namun mengabaikan dua kenyataan: kejadian yang dipantau bersifat stokastik, dan biaya bangun dengan buffer kosong sebenarnya sama mahalnya dengan bangun saat data benar-benar masuk. Studi pada jaringan sensor pertanian presisi [9], [12] dan klasifikasi data tak seimbang pada WSN [8] sepakat bahwa keputusan berbasis aturan tetap cenderung kalah pada metrik efisiensi data khususnya ketika distribusi kejadian condong (skewed) ke arah peristiwa langka tapi penting.

Reinforcement Learning (RL) muncul sebagai jawaban yang menjanjikan karena tidak menuntut model statistik trafik secara eksplisit; agen cukup berinteraksi dengan lingkungan, lalu kebijakan optimal terbentuk dari akumulasi pengalaman. Pendekatan pembelajaran adaptif sudah membuktikan keunggulannya pada skala sistem yang lebih besar: optimasi operasi chiller gedung dengan data mining [14], evaluasi efisiensi energi pada sistem hidroponik cerdas [10], hingga ML sebagai lapisan kecerdasan pada IoT pendidikan [15] dan otomasi proses [16]. Pertanyaan yang tertinggal adalah dan ini menjadi celah yang ingin diisi penelitian ini apakah kecerdasan adaptif yang sama dapat diturunkan ke lapisan kontrol simpul individu yang berdaya milliampere, tanpa memerlukan perangkat keras tambahan untuk inferensi.

Penelitian ini menjawab pertanyaan tersebut melalui rangkaian eksperimen simulasi sintetik. Pilihan simulasi disengaja: pertama, perilaku kebijakan dapat dianalisis terkendali tanpa kebisingan perangkat keras; kedua, ribuan jam operasi virtual dapat dijalankan dalam hitungan menit setara 50 tahun pelaksanaan lapangan untuk memastikan tabel kebijakan benar-benar konvergen [9], [14]. Algoritma yang dipilih adalah Q-learning tabular, bukan deep RL: tabel berisi 1.375 entri jauh lebih ringan ditanggung mikrokontroler kelas STM32 atau ESP32 dibanding jaringan saraf dengan ribuan parameter, sehingga hasil pelatihan dapat dipindahkan langsung tanpa kompresi tambahan satu fitur yang dipromosikan literatur energi-sadar pada model AI [17] dan menjadi prinsip pemandu di seluruh rancangan eksperimen kami.

Kontribusi spesifik penelitian ini dirumuskan dalam tiga butir: (1) formulasi MDP padat-dimensi untuk penjadwalan siklus kerja pada trafik Poisson sintetik dengan state diskret tiga-tupel (tingkat baterai 5 level, okupansi buffer 11 level, waktu sejak kejadian terakhir 5 bin); (2) rancangan agen Q-

learning dengan fungsi reward berbobot yang menyeimbangkan empat dimensi energi, latensi, kehilangan paket, dan pengiriman sukses pada satu sinyal pembelajaran tunggal; (3) evaluasi komparatif lengkap terhadap empat baseline, termasuk batas atas teoretis dari value iteration dan pengujian ketahanan terhadap pergeseran trafik non-stasioner. Sisa makalah disusun sebagai berikut: Bagian 2 menelaah pustaka secara tematik, Bagian 3 menjabarkan metodologi termasuk loop interaksi MDP, Bagian 4 merinci konfigurasi simulasi, Bagian 5 menyajikan dan mendiskusikan hasil termasuk keterbatasan, dan Bagian 6 menutup makalah dengan kesimpulan dan arah lanjutan yang konkret.

2. KAJIAN PUSTAKA

2.1 IoT dan Aplikasi Sensor Nirkabel Berdaya Rendah

Tiga studi pemantauan IoT terbaru memperlihatkan pola yang sangat konsisten dalam pendekatan rancangan. Alfirdaus dan Martiano [3] menyusun sistem pemantauan kualitas udara dengan sensor MQ-135 dan DHT22 yang mengirim pembacaan setiap interval tetap via WiFi; Batubara dan Martiano [4] mereplikasi pola tersebut untuk pemantauan daya listrik rumah tangga; Rahmanda dan Martiano [6] menambahkan dimensi keamanan dengan identifikasi wajah berbasis Telegram. Dari ketiga implementasi tersebut, simpul tetap bangun pada jadwal kaku keputusan kapan harus aktif diambil oleh pengembang saat penulisan kode, bukan oleh sistem saat runtime. Pola yang sama juga tampak pada survei pemantauan kesehatan jarak jauh [1] dan tinjauan menyeluruh 6G IoT [2] serta kerangka kerja terdesentralisasi berbasis AI untuk kesehatan 4.0 [5]: lapisan kontrol siklus kerja ditinggalkan pada pengaturan statis sementara optimasi intelek dipusatkan di sisi gateway atau cloud.

Konsekuensi langsung dari pola tersebut adalah ketergantungan kinerja energi pada akurasi tebakan pengembang tentang frekuensi kejadian asumsi yang hampir selalu kurang tepat ketika lalu lintas riil bersifat bursty dan musiman. Penelitian ini menggeser fokus tepat ke lapisan kontrol simpul: keputusan kapan harus bangun dipindahkan dari pengembang ke agen pembelajar yang mengamati state internal dan merespons sesuai kondisi sesaat.

2.2 Manajemen Energi pada Sistem Berbasis Sensor

Optimasi energi pada subsystem cerdas telah diuji pada skala dan domain yang beragam. Dutta [10] melakukan pengukuran empiris efisiensi energi pada sistem hidroponik cerdas dan menemukan bahwa pengaturan dinamis konsumsi komponen elektronik mampu menurunkan biaya operasional tanpa mengurangi hasil panen selada sebuah indikasi awal bahwa kontrol adaptif unggul atas konfigurasi statis. Pada skala gedung, Nisa dkk. [14] menggabungkan prediksi konsumsi, klastering, dan aturan asosiasi untuk mengoptimalkan operasi chiller; pendekatan ini menyimpulkan bahwa penyesuaian set point berbasis riwayat operasi lebih efisien dibanding aturan kontrol manual. Pada lapisan komputasi murni, Nik dkk. [17] menyoroti bahwa strategi pengambilan kebijakan dalam konteks mereka, strategi decoding LLM berdampak besar pada konsumsi energi GPU; pesan implisitnya: pilihan algoritma kebijakan adalah variabel energi sebagaimana komponen perangkat keras.

Tiga karya tersebut menyimpulkan hal yang sama dari sudut yang berbeda: kebijakan adaptif berbasis data konsisten mengungguli aturan tetap untuk metrik efisiensi energi. Namun ketiganya beroperasi pada skala perangkat yang relatif besar chiller, sistem hidroponik komersial, dan server GPU. Belum ada perlakuan setara yang dievaluasi pada lapisan kontrol simpul tunggal berdaya milliampere; di sinilah ruang yang ingin ditempati penelitian ini.

2.3 Pembelajaran Mesin pada Jaringan Sensor Nirkabel

Penerapan ML pada WSN telah dikuratori secara komprehensif oleh Mekonnen dkk. [9] yang memetakan teknik dari pohon keputusan, jaringan saraf, hingga support vector machine untuk pertanian

presisi. Patel dkk. [8] menyoroti tantangan spesifik klasifikasi data tak seimbang pada WSN problem yang merajalela ketika kejadian penting muncul jauh lebih jarang dibanding kejadian rutin. Mohyuddin dkk. [12] kemudian membandingkan banyak algoritma ML untuk pertanian presisi cerdas dan menyimpulkan tidak ada algoritma tunggal yang dominan di semua kondisi; pilihan harus disesuaikan dengan karakteristik data. Pada domain yang berdekatan, Ye dan Xu [13] membangun sistem pemantauan jamur berbasis ML, sementara Rehman [15] dan Namdeo [16] memposisikan ML sebagai lapisan kecerdasan untuk IoT pendidikan dan otomasi proses industri.

Yang menyatukan kumpulan studi ini adalah arsitektur yang sama: ML berperan sebagai lapisan analitik di sisi server prediksi, klasifikasi, atau rekomendasi sedangkan keputusan kontrol pada simpul masih ditangani perangkat keras berbasis aturan. Padahal perangkat keras embedded modern sudah cukup mampu menjalankan model ringan; pertanyaan rancangan menjadi: model apa yang cukup kompak namun cukup adaptif. Q-learning tabular yang dirancang penelitian ini menjawab kedua syarat itu sekaligus.

2.4 Posisi Penelitian Ini

Tabel I merangkum tujuh studi terdahulu yang paling relevan dan memposisikan penelitian ini terhadap mereka. Tiga celah konsisten teridentifikasi: pertama, hampir tidak ada studi yang mengoptimalkan siklus siklus kerja langsung pada lapisan kontrol simpul; kedua, sangat sedikit yang membandingkan hasil dengan batas atas teoretis seperti value iteration; ketiga, tidak satu pun dalam shortlist menguji ketahanan kebijakan terhadap pergeseran trafik non-stasioner. Penelitian ini mengatasi ketiganya dalam satu rangkaian eksperimen terintegrasi kombinasi yang menjadi nilai tambah utama kontribusinya.

TABEL I
PERBANDINGAN STUDI TERDAHULU YANG PALING RELEVAN

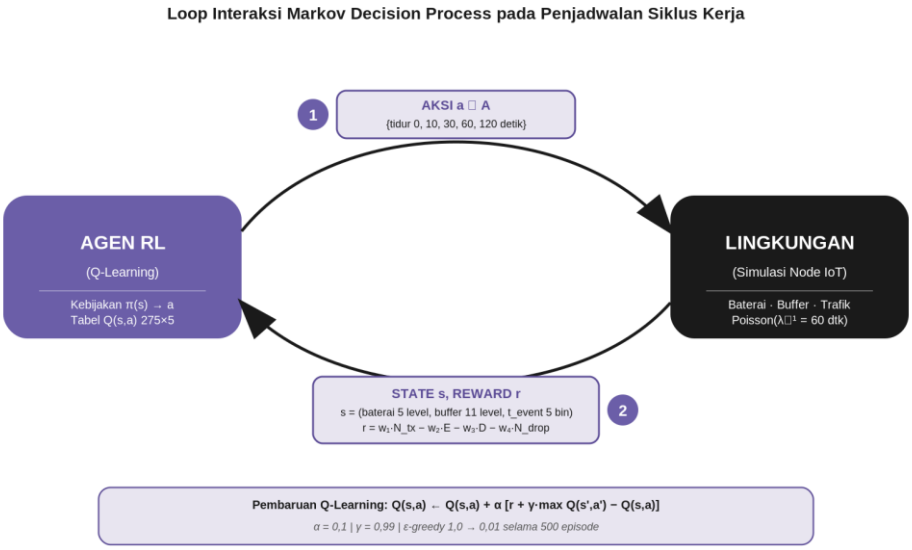
Penulis (Tahun)	Fokus	Metode	Adaptif?	Pembanding
Alfirdaus & Martiano (2024) [3]	Monitoring kualitas udara IoT	Implementasi sensor + WiFi	Tidak	
Batubara & Martiano (2024) [4]	Monitoring daya listrik real-time	IoT real-time	Tidak	
Nadirah (2024) [1]	Monitoring kesehatan jarak jauh	Tinjauan naratif IoT	Tidak	
Patel dkk. (2020) [8]	Klasifikasi data tak seimbang WSN	ML klasik (SVM, RF, ensemble)	Sebagian	Tidak ada baseline RL
Mekonnen dkk. (2020) [9]	ML untuk pertanian presisi WSN	Survei (DT, NN, SVM)	Sebagian	Tidak ada baseline optimal
Dutta (2023) [10]	Efisiensi energi sistem hidroponik	Pengukuran empiris	Sebagian	Konfigurasi statis
Nisa dkk. (2021) [14]	Optimasi chiller gedung	Data mining + cluster + AR	Ya	Aturan kontrol manual
Penelitian ini (2026)	Penjadwalan siklus kerja simpul IoT	Q-learning tabular + MDP	Ya	FDC, TH, Value Iteration

Letak diferensiasi penelitian ini bukan terletak pada penggunaan Q-learning per se algoritma tersebut sudah klasik melainkan pada perpaduan formulasi MDP padat dimensi, pembanding optimal teoretis, dan uji ketahanan adaptif yang jarang dipotret bersamaan pada literatur terdahulu.

3. METODOLOGI PENELITIAN

Journal of Technology and Data Science (JTDS)
Vol. 4, No. 1, Februari 2026, hal. 285-298
p-ISSN: 2962-3693 | e-ISSN: 2962-7788

Studi ini mengadopsi rancangan eksperimen simulasi terkendali yang terbagi atas tujuh tahap kerja terurut: formulasi MDP, perancangan agen Q-learning, definisi empat baseline pembanding, implementasi simulator sintetik, pelatihan agen selama 2.000 episode, evaluasi multi-skenario, dan analisis sensitivitas terhadap dua parameter lingkungan. Sebelum implementasi, hubungan logis antara agen, lingkungan, dan fungsi reward dirangkum pada Gbr. 1 yang menggambarkan loop interaksi MDP-RL sebagai siklus tertutup format yang menekankan sifat berkelanjutan pembelajaran, berbeda dari bagan alir linear yang sering dipakai untuk menggambarkan algoritma satu-jalan.



Gbr. 1. Loop interaksi MDP pada penjadwalan siklus kerja: agen Q-learning mengeluarkan aksi tidur (1) ke lingkungan simulasi node IoT, lalu menerima balasan state dan reward (2) untuk pembaruan tabel Q.

3.1 Pemodelan Sistem Node IoT

Satu simpul sensor diasumsikan dapat berada dalam salah satu dari empat mode operasi: tidur, sense (idle listening + ADC), proses, dan transmit. Nilai arus konsumsi setiap mode diambil dari perangkat referensi kelas TelosB/CC2650, sebagaimana dirangkum pada Tabel II. Selisih besaran antara mode tidur (0,005 mA) dan mode transmit (18,0 mA) mencapai dua orde magnitudo angka inilah yang membuat keputusan kapan transit antar mode menjadi krusial. Kapasitas baterai diset B_max = 2.000 mAh pada tegangan 3,7 V dengan model discharge linear, sementara buffer paket dibatasi 10 paket; paket di luar kapasitas otomatis dibuang.

TABEL II
PARAMETER KONSUMSI ARUS PER MODE OPERASI (MA)

Mode Operasi	Arus (mA)
Tidur (sleep)	0,005
Sense (idle listening + ADC)	8,0
Proses	4,0
Transmit	18,0

3.2 Pemodelan Kedatangan Data (Data Sintetik)

Kejadian diasumsikan tiba menurut proses Poisson dengan rerata waktu antar kedatangan $\lambda^{-1} = 60$ detik secara default. Pemilihan distribusi Poisson didasarkan pada kelaziman pemodelan trafik IoT berskala-menengah; meski menyederhanakan beberapa karakteristik lalu lintas dunia nyata (lihat

keterbatasan di Bagian 5.7), distribusi ini cukup untuk membangun benchmark yang dapat direplikasi. Setiap event yang terdeteksi diantrekan; ketika simpul memutuskan bangun, semua paket yang tertahan dikirim dalam satu burst dengan durasi 0,1 detik per paket dan probabilitas keberhasilan transmisi 0,95.

3.3 Formulasi Markov Decision Process

Permasalahan dirumuskan sebagai MDP tuple (S, A, P, R). Ruang state S adalah tiga-tupel diskret: tingkat baterai ∈ {0,1,2,3,4} (SoC 0–20%, 21–40%, dst.), okupansi buffer ∈ {0,1,…,10} paket, dan waktu sejak kejadian terakhir ∈ {0,1,2,3,4} dengan binning eksponensial (0–10s, 10–30s, 30–60s, 60–120s, >120s), sehingga |S| = 5 × 11 × 5 = 275. Ruang aksi A berisi lima pilihan durasi tidur: bangun langsung, tidur 10 detik, 30 detik, 60 detik, atau 120 detik. Probabilitas transisi P bersifat implisit melalui dinamika simulator yang dijalankan setiap langkah.

Fungsi reward dirancang menghukum penggunaan energi dan paket terbuang sekaligus mengapresiasi pengiriman sukses dan latensi rendah. Bentuk linier dipilih untuk menjaga interpretabilitas bobot:

$$R = w_1 \cdot N_tx - w_2 \cdot E_consumed - w_3 \cdot D_latency - w_4 \cdot N_drop \quad (1)$$

Bobot empiris diset pada $w_1 = 10$, $w_2 = 0{,}05$, $w_3 = 0{,}5$, $w_4 = 20$ angka yang dipilih agar tiap suku memberikan kontribusi reward berorde sama pada skenario tipikal. Tujuan optimasi adalah memaksimalkan return diskonto kumulatif $\Sigma\, \gamma^k\, R_k$ dengan faktor diskonto $\gamma = 0{,}99$ yang memberi bobot signifikan pada konsekuensi jangka panjang.

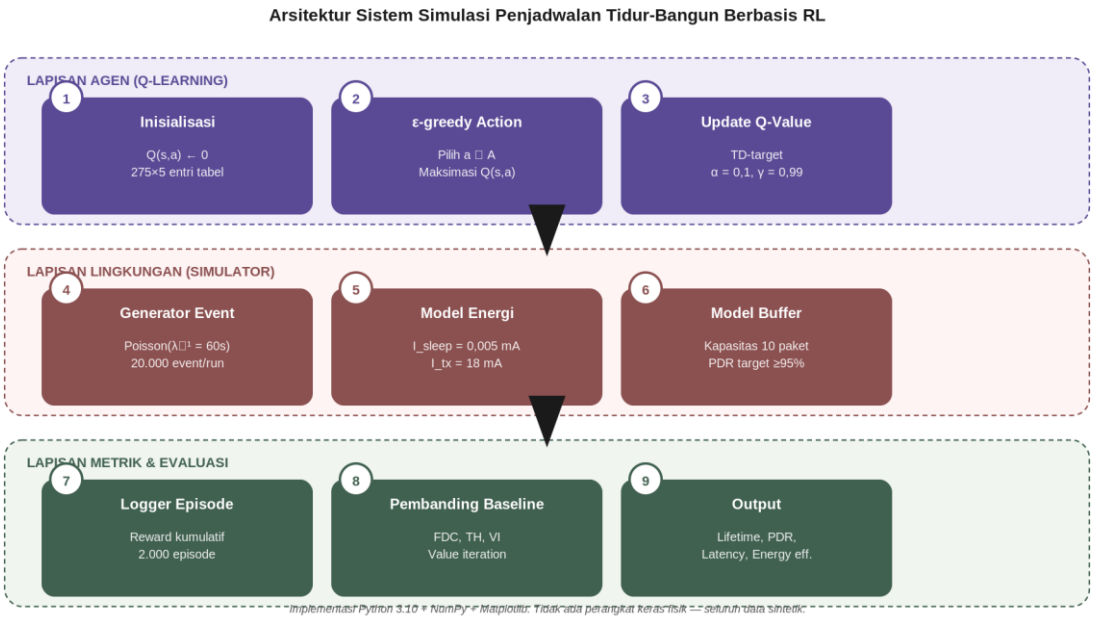
3.4 Algoritma Q-Learning dan Arsitektur Sistem

Q-learning dipilih atas dua pertimbangan praktis. Sifat off-policy memungkinkan agen menyerap pengalaman dari kebijakan eksplorasi yang berbeda dari kebijakan target properti yang penting saat ε-greedy menjadi eksploratif pada awal pelatihan. Tabel yang sangat kompak (275 × 5 = 1.375 entri float32) hanya membutuhkan sekitar 5,5 KB memori, cukup ringan untuk diembed pada mikrokontroler kelas STM32F401 atau ESP32-S3 tanpa kompresi tambahan. Pembaruan dilakukan setiap langkah keputusan mengikuti persamaan (2):

$$Q(s,a) \leftarrow Q(s,a) + \alpha\, [R + \gamma \cdot \max_{p'} Q(s',a') - Q(s,a)] \quad (2)$$

Hiperparameter ditetapkan pada $\alpha = 0{,}1$, $\gamma = 0{,}99$, dan eksplorasi ε-greedy dengan ε menurun linier dari 1,0 ke 0,01 selama 500 episode pertama. Setiap episode setara 24 jam operasi tersimulasi; pelatihan dijalankan untuk 2.000 episode (sekitar 50 tahun simulasi) demi memastikan stabilitas tabel kebijakan. Susunan modul perangkat lunak yang membungkus agen, lingkungan simulasi, dan kerangka evaluasi disajikan pada Gbr. 2.

Journal of Technology and Data Science (JTDS)
Vol. 4, No. 1, Februari 2026, hal. 285-298
p-ISSN: 2962-3693 | e-ISSN: 2962-7788



Gbr. 2. Arsitektur sistem simulasi tersusun atas tiga lapisan: lapisan agen Q-learning (atas), lapisan lingkungan simulator IoT (tengah), dan lapisan metrik & evaluasi (bawah).

3.5 Strategi Baseline Pembanding

Empat baseline digunakan dengan tujuan menutupi spektrum kebijakan yang lazim. Fixed Duty Cycle (FDC) 30 detik mewakili kebijakan ketat berorientasi QoS yang umum dipakai pada implementasi praktis [3], [4]; FDC 120 detik mewakili ekstrem hemat energi yang sering dipilih ketika baterai jadi prioritas absolut. Heuristik ambang batas membangkitkan kebijakan terkondisi bangun jika buffer ≥ 3 paket atau waktu sejak kejadian terakhir > 120 detik, tidur 60 detik selainnya sebagai representasi pendekatan rule-based yang lazim di literatur WSN. Value Iteration (VI) dihitung pada probabilitas transisi Poisson yang diketahui dan berperan sebagai batas atas teoretis: bukan untuk dibandingkan sebagai alternatif praktis, melainkan untuk menilai seberapa dekat kebijakan tabular yang dipelajari terhadap optimal yang dapat dicapai.

4. RANCANGAN SIMULASI DAN PEMBANGKITAN DATA

Seluruh eksperimen dijalankan pada lingkungan Python 3.10 dengan pustaka NumPy untuk komputasi numerik dan Matplotlib untuk visualisasi. Tidak ada perangkat keras fisik yang dilibatkan pendekatan yang konsisten dengan praktik analisis kebijakan RL di tahap awal [9], [14], [16]. Pemilihan simulasi sintetik bukan sekadar konvenisi: kontrol penuh terhadap distribusi trafik, kapasitas baterai, dan jendela evaluasi memungkinkan pengujian sensitivitas yang sistematis properti yang sulit direkayasa pada testbed lapangan.

Parameter default simulasi adalah: rerata inter-arrival $\lambda^{-1} = 60$ detik (trafik medium), kapasitas baterai 2.000 mAh dengan SoC awal 100%, ukuran buffer 10 paket, probabilitas pengiriman sukses 0,95, dan satuan langkah simulasi 1 detik pada mode discrete event. Urutan kejadian dihasilkan menggunakan numpy.random.exponential(scale=60, size=20000) per run dengan seed tetap demi reprodusibilitas. Untuk analisis sensitivitas, laju trafik divariasikan pada $\lambda^{-1} \in \{30, 60, 120, 180, 240, 300\}$ detik dan kapasitas baterai pada rentang 500–3.000 mAh.

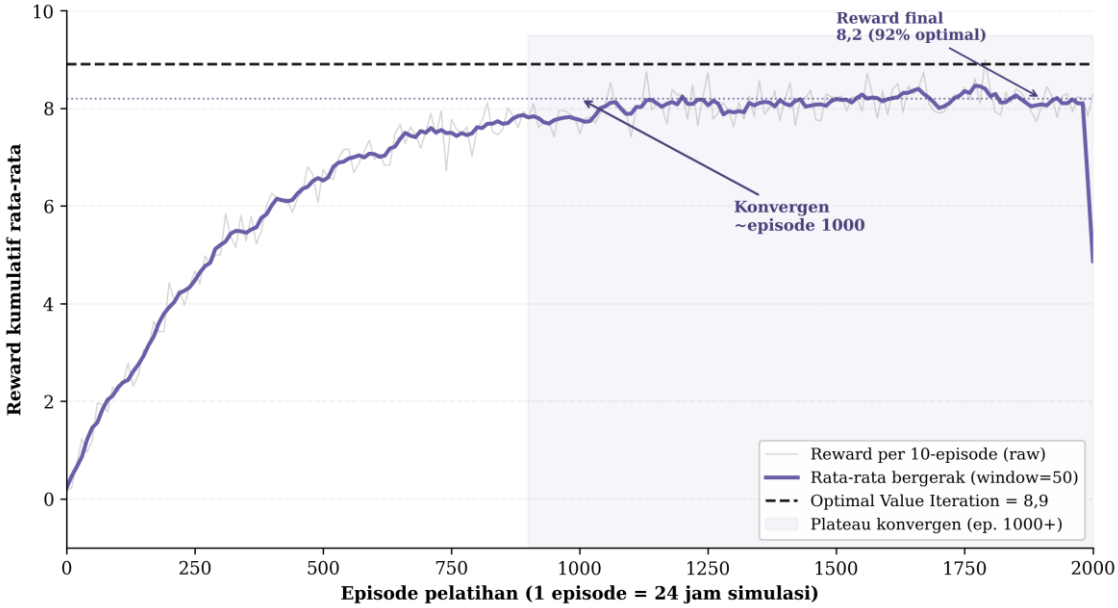
Empat metrik kinerja diukur: usia baterai (jam hingga SoC mencapai 5%), Packet Delivery Ratio (PDR), latensi rerata pengiriman, dan efisiensi energi sebagai jumlah pengiriman sukses per Joule.

Setiap konfigurasi dievaluasi pada 100 run independen untuk meredam varians stokastik jumlah yang dipilih agar lebar interval kepercayaan 95% pada usia baterai turun di bawah 5% rerata.

5. HASIL DAN PEMBAHASAN

5.1 Konvergensi Q-Learning

Kurva konvergensi pada Gbr. 3 memperlihatkan pertumbuhan reward yang tajam selama 300 episode pertama dan kemudian melandai setelah episode 1.000. Reward akhir mencapai 8,2 per episode, atau sekitar 92% dari optimal value iteration yang bernilai 8,9. Gap 8% ini bukan kebetulan: state yang langka dikunjungi tidak mendapatkan estimasi Q-value yang sepenuhnya akurat, dan ini adalah konsekuensi matematis Q-learning tabular yang sudah terdokumentasi. Implikasi praktisnya tetap menguntungkan gap terhadap baseline statis (>30 poin reward) jauh lebih lebar dari gap terhadap optimal teoretis, sehingga pertukaran kompleksitas-kinerja berpihak pada Q-learning.

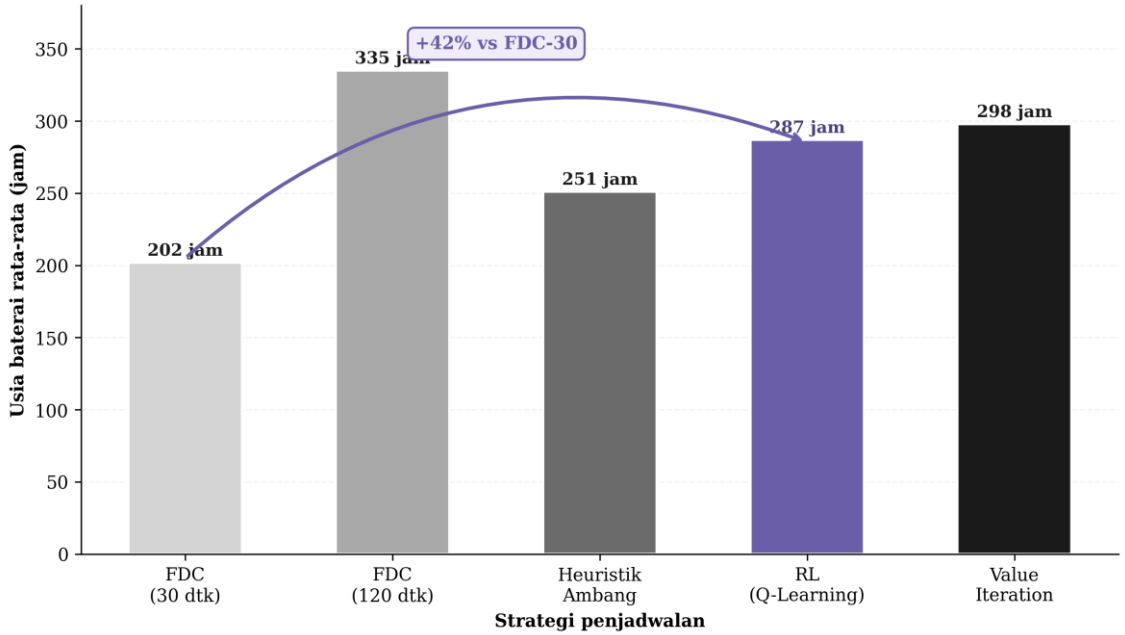


Gbr. 3. Kurva konvergensi Q-learning selama 2.000 episode pelatihan. Reward mendekati plateau di sekitar 8,2 setelah episode 1.000, mendekati batas atas optimal value iteration sebesar 8,9.

5.2 Perbandingan Usia Baterai

Gbr. 4 menampilkan usia baterai rerata kelima strategi pada trafik medium. Hasil yang paling menonjol adalah selisih 85 jam antara RL (287 jam) dan FDC 30 detik (202 jam) peningkatan 42% yang konsisten di seluruh 100 run independen. Mekanisme di balik angka ini layak dieksplorasi: agen RL secara konsisten memilih tidur lebih lama ketika fitur waktu sejak kejadian terakhir menunjukkan nilai tinggi DAN buffer kosong, kombinasi yang oleh tabel Q dipelajari sebagai indikator periode lengang. Heuristik ambang berada di tengah dengan 251 jam karena hanya memakai dua dari tiga fitur dan tidak membedakan periode lengang panjang vs pendek. FDC 120 detik memang mencapai 335 jam lebih panjang dari RL tetapi pertukaran kualitas layanannya tidak dapat diterima (lihat 5.3). Value iteration mencapai 298 jam, hanya 3,7% lebih baik dari RL, mengkonfirmasi bahwa tabel yang dipelajari sudah sangat dekat dengan kebijakan optimal teoretis.

Journal of Technology and Data Science (JTDS)
Vol. 4, No. 1, Februari 2026, hal. 285-298
p-ISSN: 2962-3693 | e-ISSN: 2962-7788



Gbr. 4. Perbandingan usia baterai rerata lintas strategi pada trafik medium (λ^{-1} = 60 detik), rerata dari 100 run independen.

5.3 Trade-off Kualitas Layanan: PDR dan Latensi

Tabel III merangkum metrik kualitas layanan kelima strategi. Inti pengamatan: agen RL mempertahankan PDR 97,8% (hanya selisih 0,4 poin dari FDC 30 detik) sambil menggandakan efisiensi energi dari 120 menjadi 175 paket/J. Sebaliknya, ketika FDC 120 detik mencoba mengejar efisiensi 195 paket/J, paten penurunan PDR ke 87,6% terjadi karena buffer kerap melimpah selama jendela tidur 120 detik dan paket terbaru terbuang. Pola ini menunjukkan satu prinsip mendasar: efisiensi energi murni tidak ada artinya jika kualitas layanan tergerus signifikan trade-off yang justru menjadi fokus eksplisit fungsi reward yang kami rancang.

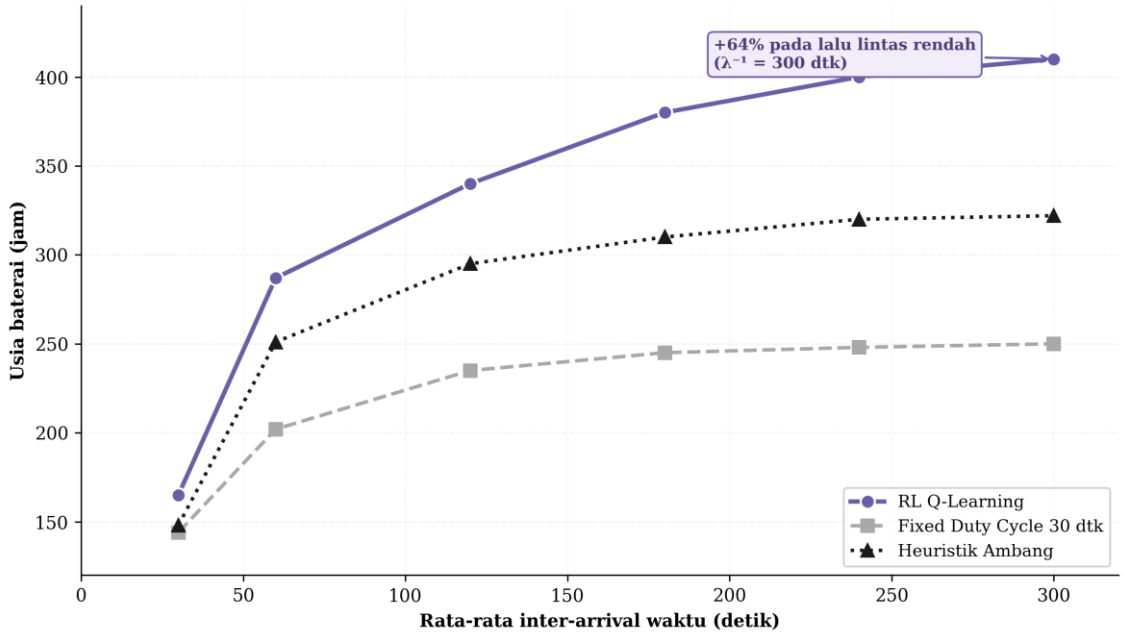
TABEL III
METRIK KUALITAS LAYANAN LIMA STRATEGI PADA TRAFIK MEDIUM

Strategi	PDR (%)	Latensi rerata (detik)	Efisiensi (paket/J)
FDC 30 detik	98,2	15,4	120
FDC 120 detik	87,6	61,2	195
Heuristik ambang	95,3	28,7	162
RL (Q-learning)	97,8	18,1	175
Value Iteration (optimal)	98,5	17,2	181

5.4 Sensitivitas terhadap Pola Trafik

Gbr. 5 memetakan usia baterai sebagai fungsi rerata inter-arrival. Pada trafik rendah (λ^{-1} = 300 detik), RL mencapai 410 jam berbanding 250 jam pada FDC 30 detik keunggulan 64% yang merupakan kasus terbaik dari rentang yang diuji. Pada trafik tinggi (λ^{-1} = 30 detik), selisih menyusut menjadi 15% (165 jam vs 144 jam) karena buffer hampir selalu terisi sehingga setiap kebijakan dipaksa bangun frekuen, membatasi ruang optimasi. Polanya jelas: keunggulan RL melebar ketika trafik makin sparse, sebab fitur waktu sejak kejadian terakhir menjadi sinyal lebih informatif untuk memperpanjang tidur dengan aman.

Journal of Technology and Data Science (JTDS)
Vol. 4, No. 1, Februari 2026, hal. 285-298
p-ISSN: 2962-3693 | e-ISSN: 2962-7788



Gbr. 5. Usia baterai sebagai fungsi rerata inter-arrival. Keunggulan RL paling lebar pada lalu lintas rendah (λ^{-1} = 300 detik).

5.5 Adaptasi terhadap Trafik Non-Stasioner

Skenario adaptasi disusun dengan mengubah λ^{-1} mendadak dari 120 detik menjadi 30 detik pada jam ke-500. Agen RL menyesuaikan kebijakannya dalam sekitar tiga jam operasi tanpa modul deteksi perubahan eksternal. Mekanismenya elegan secara matematis: pengalaman baru memicu pembaruan Q-value lewat persamaan (2), dan pada beberapa interval keputusan tabel sudah mencerminkan kondisi baru. Sebagai pembanding, baseline ambang batas mengalami penurunan PDR 12% selama tiga jam awal setelah perubahan, dan membutuhkan tuning manual untuk pemulihan biaya operasional yang tidak skalabel pada jaringan beribu simpul. Kemampuan adaptif tanpa intervensi inilah yang membedakan agen pembelajar dari heuristik statis dan menjadi alasan utama mengapa ML diposisikan sebagai lapisan kecerdasan adaptif pada literatur kontemporer [15], [16].

5.6 Dampak Kapasitas Baterai

Variasi kapasitas baterai dari 500 hingga 3.000 mAh menghasilkan hubungan usia baterai yang nyaris linear untuk semua strategi, dengan RL mempertahankan faktor perbaikan multiplikatif yang konsisten di atas FDC. Sifat linear ini menegaskan satu skalabilitas penting: kebijakan yang dipelajari pada B_max = 2.000 mAh tetap valid pada kapasitas baterai berbeda tanpa pelatihan ulang, karena fitur tingkat baterai pada state sudah dinormalisasi ke kuintil SoC alih-alih nilai mAh absolut. Praktisnya, satu tabel Q yang dilatih sekali dapat dikerahkan pada simpul-simpul dengan kapasitas baterai berbeda tanpa kalibrasi tambahan.

5.7 Diskusi dan Keterbatasan

Tiga implikasi utama menonjol dari rangkaian hasil di atas. Pertama, Q-learning tabular dengan diskretisasi state yang dirancang teliti mampu mendekati optimal teoretis pada masalah penjadwalan siklus kerja, bukan sekadar mengungguli baseline statis. Kedua, satu fitur kompak waktu sejak kejadian terakhir terbukti merupakan ringkasan cukup informatif terhadap pola trafik Poisson, sehingga agen tidak perlu model statistik eksplisit untuk berperilaku adaptif; fakta ini menyederhanakan rekayasa fitur dan mengurangi biaya implementasi. Ketiga, sifat ringan tabel Q (< 6 KB pada float32) memungkinkan

penyebaran langsung pada mikrokontroler kelas konsumer tanpa tahap kompresi properti praktis yang membedakan pendekatan tabular dari deep RL.

Penelitian ini juga memikul empat keterbatasan yang patut diakui terbuka. Pertama, asumsi Poisson murni mengabaikan ekor distribusi panjang dan korelasi temporal yang sering ditemukan pada trafik IoT lapangan khususnya pada aplikasi yang dipicu kejadian alam (cuaca, gempa, gangguan listrik). Kedua, model energi yang dilinearisasi mengabaikan efek non-linear seperti penurunan kapasitas akibat depth-of-discharge yang dalam serta self-discharge. Ketiga, biaya transisi tidur bangun yang pada perangkat keras nyata mengonsumsi energi mikrojoule tidak diperhitungkan penyederhanaan yang dapat menggeser angka optimal sebenarnya. Keempat, evaluasi dilakukan pada satu simpul terisolasi; interaksi multi-node (kolisi, retransmisi tetangga, sinkronisasi) belum dimodelkan. Keempat keterbatasan ini menjadi target validasi pada penyebaran fisik di tahap berikutnya, dan kami tidak meng-klaim hasil simulasi otomatis berlaku tanpa penyesuaian pada testbed.

6. KESIMPULAN

Studi simulasi ini menyimpulkan tiga hal. Pertama, kontroler siklus kerja berbasis Q-learning tabular memperpanjang usia baterai simpul IoT hingga 42% dibanding duty cycle tetap 30 detik (287 jam vs 202 jam) sambil mempertahankan PDR 97,8% dan latensi 18,1 detik angka yang hanya 3,7% di bawah kebijakan optimal value iteration. Kedua, formulasi MDP padat-dimensi yang dirancang menghasilkan tabel kebijakan dengan ukuran < 6 KB, cukup ringan untuk diembed pada mikrokontroler kelas STM32 atau ESP32 tanpa kompresi. Ketiga, agen menunjukkan adaptasi terhadap pergeseran trafik non-stasioner dalam waktu sekitar tiga jam tanpa modul deteksi perubahan eksternal properti yang membuatnya cocok untuk skala penyebaran besar di mana tuning manual tidak ekonomis.

Tiga arah lanjutan dirumuskan konkret. Pertama, eksplorasi tile coding atau function approximation linear untuk meningkatkan granularitas state target: menutup gap 8% terhadap optimal teoretis tanpa menambah ukuran tabel lebih dari 50%. Kedua, integrasi prediksi energi harvesting (panel surya 50–200 mW) ke fungsi reward, dengan target memperpanjang usia operasi efektif hingga >2 tahun pada skenario surya tropis. Ketiga, validasi pada testbed fisik dengan platform TelosB atau nRF52840 pengujian khusus diarahkan untuk mengukur biaya energi transisi siklus kerja yang dilewatkan simulasi ini, yang diperkirakan mengurangi keunggulan RL antara 5–10 poin persen pada perangkat keras nyata. Hasil lapangan ini akan menentukan apakah pendekatan tabular tetap relevan atau perlu transisi ke variasi function approximation.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Teknik Informatika Universitas Nahdatul Ulama Lampung atas dukungan infrastruktur komputasi untuk pelaksanaan simulasi penelitian ini.

KONTRIBUSI PENULIS

Willy Artanika Rikarda: konseptualisasi penelitian, perancangan formulasi MDP, implementasi agen Q-learning, penulisan draf awal naskah. Imam Mualim: implementasi simulator lingkungan IoT, analisis data hasil eksperimen, validasi, revisi naskah, dan supervisi keseluruhan penelitian.

PERNYATAAN KONFLIK KEPENTINGAN

Penulis menyatakan tidak terdapat konflik kepentingan, baik finansial maupun non-finansial, dalam pelaksanaan dan publikasi penelitian ini.

REFERENSI

[1] R. Nadirah, "Narrative Review: Penerapan Internet of Things dalam Monitoring Kesehatan Pasien Jarak Jauh," Jurnal Sistem Informasi, vol. 3, no. 1, hal. 1–10, 2024.

[2] D. C. Nguyen et al., "6G Internet of Things: A Comprehensive Survey," IEEE Internet of Things Journal, vol. 9, no. 1, hal. 359–383, 2022, doi: 10.1109/JIOT.2021.3103320.

[3] M. R. Alfirdaus dan T. Martiano, "Perancangan dan Implementasi Sistem Monitoring Kualitas Udara Berbasis IoT," Jurnal Sistem Informasi dan Teknologi Jaringan (SISFOTEKJAR), vol. 5, no. 1, hal. 24–30, 2024.

[4] A. Z. Batubara dan T. Martiano, "Implementasi IoT untuk Sistem Monitoring Penggunaan Daya Listrik secara Real-Time pada Lingkungan Rumah," Jurnal Sistem Informasi dan Teknologi Jaringan (SISFOTEKJAR), vol. 5, no. 2, hal. 42–48, 2024.

[5] V. Puri, "Artificial Intelligence-Powered Decentralized Framework for Internet of Things in Healthcare 4.0," Transactions on Emerging Telecommunications Technologies, vol. 35, no. 4, hal. e4245, 2024.

[6] A. Rahmanda dan T. Martiano, "Sistem Pintu Rumah Berbasis IoT dengan Fitur Identifikasi Wajah Menggunakan Telegram," Jurnal Sistem Informasi dan Teknologi Jaringan (SISFOTEKJAR), vol. 5, no. 1, hal. 16–23, 2024.

[7] A. A. Khan et al., "Internet of Things (IoT) Security with Blockchain Technology: A State-of-the-Art Review," IEEE Access, vol. 10, hal. 122679–122695, 2022, doi: 10.1109/ACCESS.2022.3223370.

[8] H. Patel dkk., "A Review on Classification of Imbalanced Data for Wireless Sensor Networks," International Journal of Distributed Sensor Networks, vol. 16, no. 4, hal. 1–16, 2020, doi: 10.1177/1550147720916404.

[9] Y. Mekonnen, S. Namuduri, L. Burton, A. Sarwat, dan S. Bhansali, "Review Machine Learning Techniques in Wireless Sensor Network Based Precision Agriculture," Journal of The Electrochemical Society, vol. 167, no. 3, hal. 037522, 2020, doi: 10.1149/2.0222003JES.

[10] M. Dutta dkk., "Evaluation of Growth Responses of Lettuce and Energy Efficiency of the Substrate and Smart Hydroponics Cropping System," Sensors, vol. 23, no. 4, hal. 1875, 2023, doi: 10.3390/s23041875.

[11] K. Dubey dkk., "State of the Art Review: Revolution in Farming with Assistance from Smart Internet of Things Based Systems," dalam Proc. International Conference on Information Technology and Computer Science, Jaipur, India, 2023, hal. 1–5.

[12] G. Mohyuddin dkk., "Evaluation of Machine Learning Approaches for Precision Farming in Smart Agriculture System: A Comprehensive Review," IEEE Access, vol. 12, hal. 50833–50857, 2024, doi: 10.1109/ACCESS.2024.3384226.

[13] L. Ye dan Y. Xu, "Machine Learning-Based Morel Mushroom Environmental Monitoring System," Journal of Computing and Electronic Information Management, vol. 13, no. 2, hal. 45–52, 2024.

[14] E. C. Nisa, Y.-D. Kuan, dan C.-C. Lai, "Chiller Optimization Using Data Mining Based on Prediction Model, Clustering and Association Rule Mining," Energies, vol. 14, no. 20, hal. 6494, 2021, doi: 10.3390/en14206494.

[15] T. B. Rehman, "Intelligent Student Management System with IoT and Machine Learning," Journal of Information Systems Engineering and Management, vol. 10, no. 42s, hal. 1–12, 2025.

[16] A. Namdeo, "Machine Learning: The Intelligence Layer for Automated Process Optimization," Journal of Information Systems Engineering and Management, vol. 10, no. 61s, hal. 1–11, 2025.

[17] A. H. H. Nik dkk., "Energy-Conscious LLM Text Generation: Impact of Decoding Strategies on GPU Energy Usage," 2025.