

Implementasi Algoritma Reinforcement Learning dalam Pengelolaan Energi untuk Sistem Smart Grid

Nor Milatul Khusna*, Evanita, Aditya Akbar Riadi

Fakultas Teknik, Teknik Informatika, Universitas Muria Kudus, Kudus, Indonesia

Email: ¹1202151170@std.umk.ac.id, ²evanita@umk.ac.id, ³aditya.akbar@umk.ac.id,

Email Penulis Korespondensi: 202151170@std.umk.ac.id

Submitted: 31/05/2025; Accepted: 23/06/2025; Published: 30/06/2025

Abstrak—Pengelolaan energi pada sistem smart grid merupakan tantangan penting dalam mendukung efisiensi dan keberlanjutan energi, khususnya pada skala rumah tangga. Penelitian ini mengusulkan implementasi *algoritma Proximal Policy Optimization* (PPO) berbasis *Deep Reinforcement Learning* (DRL) untuk mengoptimalkan efisiensi penggunaan energi melalui strategi load shifting. Lingkungan simulasi dibangun untuk merepresentasikan konsumsi energi rumah tangga dalam skenario waktu nyata, di mana agen PPO dilatih untuk mengalihkan beban penggunaan listrik ke waktu dengan tarif lebih rendah atau beban sistem yang lebih ringan. Pengujian dilakukan terhadap tiga skema reward dengan dua mode pelatihan, yaitu cepat dan maksimal. Hasil terbaik diperoleh pada kombinasi reward ketiga dengan mode pelatihan maksimal, menghasilkan rata-rata reward sebesar 41690,53 dan efisiensi biaya hingga 95,83% dibandingkan dengan data konsumsi asli. Temuan ini membuktikan bahwa PPO merupakan pendekatan yang efektif dalam pengelolaan energi pada *smart grid* skala rumah tangga, khususnya dalam mendukung strategi pengalihan beban yang adaptif dan hemat biaya..

Kata Kunci: Smart Grid; Reinforcement Learning; PPO; Load Shifting; Efisiensi Energi

Abstract—Energy management in smart grid systems is a critical challenge in supporting energy efficiency and sustainability, particularly at the household scale. This study proposes the implementation of the Proximal Policy Optimization (PPO) algorithm based on Deep Reinforcement Learning (DRL) to optimize energy usage efficiency through load shifting strategies. A simulation environment was developed to represent household energy consumption in real-time scenarios, where the PPO agent was trained to shift electricity loads to periods with lower tariffs or lighter system loads. Testing was conducted on three reward schemes with two training modes, namely fast and maximal. The best results were achieved using the third reward scheme combined with the maximal training mode, yielding an average reward of 41690,53 and cost efficiency of up to 95,83% compared to the original consumption data. These findings demonstrate that PPO is an effective approach for energy management in household-scale smart grids, particularly in supporting adaptive and cost-efficient load shifting strategies.

Keywords: Smart Grid; Reinforcement Learning; PPO; Load Shifting; Energy Efficiency

1. PENDAHULUAN

Peningkatan konsumsi listrik rumah tangga menjadi tantangan besar dalam mewujudkan sistem energi yang efisien dan berkelanjutan. Pola penggunaan yang tidak terjadwal dengan baik terutama pada jam-jam beban puncak yang menyebabkan lonjakan daya, pemborosan energi, dan kenaikan biaya operasional jaringan listrik. Di sisi lain, kesadaran masyarakat terhadap efisiensi energi dan pengaturan waktu penggunaan perangkat masih tergolong rendah dan belum didukung sistem otomatis yang mampu mengelola konsumsi energi secara dinamis dan hemat biaya. Selain itu diperlukan solusi teknologi cerdas yang mampu mengoptimalkan pemanfaatan energi khususnya melalui strategi *load shifting* yaitu mengalihkan konsumsi energi dari jam sibuk ke jam rendah.

Untuk menjawab tantangan tersebut, berbagai pendekatan teknologi cerdas telah dikembangkan. Sistem smart grid menjadi salah satu solusi menjanjikan dalam hal ini untuk mengelola beban secara adaptif. Algoritma pembelajaran penguatan *Reinforcement Learning*, (RL) khususnya *Proximal Policy Optimization* (PPO) dari kelompok *Deep Reinforcement Learning* (DRL) memiliki potensi besar untuk mengatur penggunaan perangkat rumah tangga secara otonom. PPO memiliki kemampuan belajar dari interaksi dengan lingkungan dan mengoptimalkan kebijakan pengendalian berdasarkan umpan balik secara dinamis.

Penelitian sebelumnya telah membuktikan efektivitas RL dalam manajemen energi. Peirelinck et al.[1] menunjukkan bahwa PPO dengan transfer learning mampu menurunkan biaya energi hingga 14,51%. Chung et al. [2] menerapkan distributed DRL untuk penjadwalan beban rumah tangga dan berhasil menurunkan puncak beban hingga 12%. Sementara itu, Zhou et al. [3] menggunakan Deep Q-Network (DQN) dalam sistem multi-energi rumah tangga, namun belum mengintegrasikan variasi reward dan strategi pelatihan secara bersama. Sedangkan Chen et al.[4] menerapkan *Deep Reinforcement Learning* (DRL) pada sistem manajemen energi gedung pintar berskala besar, namun tidak difokuskan pada konteks rumah tangga individu yang memiliki pola konsumsi lebih dinamis dan variatif, Model Chifu et al.[5] penjadwalan pengisian dan pengosongan EV menggunakan Deep Q-Learning mempertimbangkan mobilitas kendaraan dan perilaku acak pengguna. Model ini mencakup waktu dan kebutuhan pengisian, status baterai, serta lokasi perjalanan dengan fungsi hadiah dinamis untuk mengoptimalkan efisiensi dan menekan biaya. Namun, keakuratan fungsi hadiah menurun jika perilaku pengguna menyimpang dari asumsi model.

Beberapa studi lainnya seperti Tang et al. [6] dan Amin et al. [7] telah menggunakan PPO namun belum membandingkan konfigurasi reward dan mode pelatihan secara menyeluruh. GridLearn mengembangkan pendekatan multi-agent reinforcement learning yang memungkinkan koordinasi pengelolaan energi secara terdistribusi dan adaptif

di berbagai unit bangunan dalam jaringan smart grid [8]. Sementara itu, Zulfikar et al. [9] mengusulkan kerangka kerja multi-agen berbasis *Deep Reinforcement Learning* untuk mengelola respons permintaan dan distribusi energi di lingkungan perumahan. Dengan memanfaatkan skema harga real-time dan *algoritma Deep Q-Network (DQN)* sistem ini mampu meningkatkan efisiensi pemanfaatan energi terbarukan serta mengurangi ketergantungan terhadap generator cadangan. Hasil simulasi menunjukkan adanya peningkatan efisiensi operasional jaringan serta keuntungan bagi prosumer dan penyedia layanan energi.

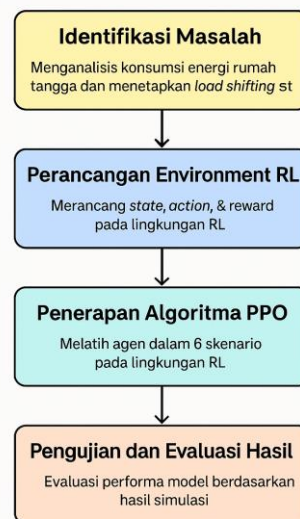
Berdasarkan studi-studi tersebut, dapat disimpulkan bahwa masih terdapat gap penelitian yaitu belum adanya pendekatan yang secara bersamaan mengevaluasi kombinasi fungsi reward dan mode pelatihan dalam konteks *load shifting* rumah tangga menggunakan PPO. Padahal kombinasi tersebut berpotensi menghasilkan strategi kontrol energi yang lebih adaptif dan hemat biaya. Hasilnya diharapkan dapat memberikan kontribusi nyata dalam mengembangkan strategi *load shifting* otomatis berbasis kecerdasan buatan yang efektif untuk sistem smart grid skala rumah tangga. Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi algoritma PPO pada simulasi sistem smart grid skala rumah tangga. Lingkungan simulasi mencakup 11 perangkat rumah tangga dan skema tarif dinamis dengan tiga varian fungsi reward dan dua mode pelatihan. PPO dipilih karena stabilitasnya dalam pembelajaran kebijakan serta efisiensinya dalam menghadapi lingkungan yang kompleks [10][11]. Skema reward yang dirancang mempertimbangkan trade-off antara biaya listrik, kenyamanan, dan jumlah perangkat aktif [12].

Penelitian ini diharapkan memberikan kontribusi signifikan dalam pengembangan sistem otomatisasi pengelolaan energi rumah tangga berbasis kecerdasan buatan. Secara praktis, sistem yang diusulkan dapat membantu rumah tangga mengurangi biaya listrik tanpa mengorbankan kenyamanan pengguna. Secara akademis, penelitian ini memperkaya literatur tentang penerapan PPO dalam domain energi khususnya dalam skenario realistis dengan variasi reward dan mode pelatihan. Selain itu, hasil penelitian ini dapat menjadi dasar pengembangan sistem manajemen energi berbasis RL untuk aplikasi skala komunitas di masa mendatang.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental berbasis simulasi, dengan penerapan algoritma Proximal Policy Optimization (PPO) dalam kerangka Deep Reinforcement Learning (DRL). Tujuan dari metode ini adalah membentuk kebijakan pengoperasian perangkat rumah tangga yang efisien secara energi dan hemat biaya. Berikut ini adalah tahapan pelaksanaan penelitian yang menggambarkan alur penerapan dan pengujian metode. Berikut ini adalah tahapan penelitian yang dilakukan:

Alur Tahapan Penelitian



Gambar 1 Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan utama yang disusun secara sistematis, yaitu:

- a) **Identifikasi Masalah**
Menganalisis permasalahan konsumsi energi tinggi pada rumah tangga dan menetapkan *load shifting* sebagai pendekatan solusi melalui DRL.
- b) **Perancangan Environment**
Membuat environment berbasis OpenAI Gym untuk mensimulasikan kondisi rumah tangga. Komponen utama terdiri dari *state*, *action*, dan *reward* yang dirancang agar agen dapat belajar dari interaksi dengan lingkungan.
- c) **Implementasi PPO**
Melatih agen menggunakan algoritma PPO pada environment yang telah dirancang. Pelatihan dilakukan untuk

enam kombinasi skenario (tiga reward × dua mode pelatihan) guna mengamati respons dan efisiensi hasil yang diperoleh.

d) Analisis dan Pengujian

Melakukan evaluasi performa model PPO berdasarkan hasil simulasi. Penilaian mencakup perbandingan biaya listrik, efektivitas strategi load shifting, dan analisis reward untuk mengukur keberhasilan agen dalam mencapai efisiensi energi..

2.1 Data dan Simulasi yang Digunakan

Dalam penelitian ini, data konsumsi energi rumah tangga diperoleh melalui simulasi berbasis Python yang merepresentasikan skenario realistis penggunaan perangkat elektronik harian di rumah tangga. Simulasi dilakukan untuk menciptakan dataset sintetik yang dapat digunakan dalam pelatihan model PPO [13].

a. Perangkat Elektronik dan Konsumsi Daya

Sebanyak 11 perangkat rumah tangga disimulasikan, masing-masing dengan nilai konsumsi daya rata-rata (dalam Watt) sebagai berikut:

Tabel 1 Perangkat Rumah Tangga

Nama Perangkat	Daya (Watt)
AC	1200
Mesin Cuci	500
Lampu	100
Pemanas Air	1500
Kulkas	200
Kompas Listrik	1000
TV	150
Oven	1800
Kipas Angin	75
Pompa Air	600
Rice Cooker	400

b. Skema Tarif Listrik per Jam

Harga listrik ditentukan secara dinamis setiap jam selama 24 jam. Nilai tarif berkisar antara Rp 800/kWh hingga Rp 4000/kWh, mencerminkan variasi beban puncak dan non-puncak. Data tarif dimodelkan dalam array.

c. Simulasi Aktivitas Perangkat

Untuk setiap sampel simulasi 1000 rumah tangga, dilakukan iterasi selama 24 jam. Pada setiap jam, perangkat-perangkat diaktifkan secara acak menggunakan fungsi random.sample, sehingga menciptakan variasi kondisi beban harian. Kemudian, total daya dan biaya listrik dihitung berdasarkan status perangkat dan tarif jam terkait. Data simulasi akhirnya disimpan ke dalam format .csv menggunakan pandas, dengan atribut utama:

1. hour: Jam keberapa (0–23)
2. active_devices: Daftar perangkat aktif
3. cost_this_hour: Biaya listrik per jam
4. total_power_consumed: Total konsumsi daya per hari
5. total_cost: Akumulasi biaya listrik harian

Data ini menjadi input utama untuk environment RL di mana sistem akan belajar melakukan pengambilan keputusan (action) berdasarkan pengamatan (state) dan menghasilkan reward sesuai strategi efisiensi yang diinginkan. Model environment dalam penelitian ini dirancang untuk merepresentasikan kondisi rumah tangga dengan 11 perangkat, serta memanfaatkan ruang aksi biner untuk setiap perangkat. [12]

2.2 Model Environment untuk Reinforcement Learning

Reinforcement Learning (RL) merupakan metode pembelajaran yang berfungsi untuk menghubungkan kondisi lingkungan dengan aksi yang diambil oleh agen, dengan tujuan utama untuk memperoleh imbalan kumulatif sebesar mungkin melalui proses interaksi dan eksplorasi terhadap lingkungan tersebut[14]. Digunakan model environment berbasis simulasi yang merepresentasikan kondisi rumah tangga dan konsumsi energinya. Komponen utama dari environment ini meliputi:

a. State

$$S_t = [t, \frac{cost_{thishour_t}}{10^5}, \frac{totalpowerconsumend_t}{P_{max}}, \frac{totalcost}{10^5}, d_t] \quad (1)$$

Di mana :

1. t = jam ke-t,
2. cost_this_hour adalah biaya listrik pada jam itu (distandarisasi dengan 10^5),
3. total_power_consumed distandarisasi terhadap kapasitas maksimum P_{max}

4. $total_cost$ adalah akumulasi biaya hingga saat itu (distandarisasi),
5. $dt \in \{0,1\}^{11}$ adalah vektor status perangkat (1=aktif).

b. Action

$$a_t \in [0,1]^{11} \quad (2)$$

Setiap elemen 0/1 mengendalikan OFF/ON satu perangkat

c. Reward

Untuk mengakomodasi berbagai preferensi pengguna terhadap kenyamanan dan penghematan energi, digunakan tiga fungsi reward berbeda. Pendekatan serupa dalam multi-mode DRL telah terbukti efektif untuk mengadaptasi preferensi pengguna dalam skenario rumah tangga[15]. Reward diberikan berdasarkan seberapa hemat dan efisien keputusan agen. Ada tiga varian fungsi reward yang mengakomodasi trade_off antara penghematan biaya dan kenyamanan :

1. Reward V1 (Hemat Biaya)

$$r_t = -\frac{cost_{this_hour_t}}{10^5} + \epsilon, \epsilon \sim N(0,0.1) \quad (3)$$

2. Reward V2 (Biaya dan Penalti Aktif)

$$r_t = \frac{cost_{this_hour_t}}{10^5} - a \frac{\|a_t\|_1}{11} + n, n \sim N(0,0.05) \quad (4)$$

3. Reward V3 (Biaya, Efisiensi dan Penalti)

$$eff_t = \frac{total_{cost_t} - cost_{this_hour_t}}{total_{cost_t}}, r_t = -\frac{cost_{this_hour_t}}{10^5} + 2eff_t - a \frac{\|a_t\|_1}{11} \quad (5)$$

2.3 Algoritma Proximal Policy Optimization (PPO)

Proximal Policy Optimization (PPO) adalah salah satu algoritma DRL yang stabil dan efisien untuk training agen dalam pengambilan keputusan. PPO mengoptimalkan policy $\pi\theta$ dengan membatasi perubahan terlalu drastis dari policy sebelumnya, menggunakan *clipped surrogate objective*[10]:

$$L^{CLIP}(\theta) = E_t[\min(r_t(\theta)\widehat{A}_t, clip(r_t(\theta), 1 + \epsilon)\widehat{A}_t)] \quad (6)$$

Dimana: $r_t(\theta) = \frac{\pi\theta(a_t|s_t)}{\pi\theta_{old}(a_t|s_t)}$, dan $\widehat{A}_t$ adalah *advantage function*. PPO dilatih menggunakan update bertahap untuk menghindari overfitting pada satu policy.

2.4 Tools dan Library yang Digunakan

Berikut adalah daftar alat (*tools*) yang digunakan dalam penelitian ini:

- a. Bahasa Pemrograman Python
Python adalah bahasa pemrograman yang fleksibel dan kuat, sangat cocok untuk pengembangan sistem berbasis machine learning dan Reinforcement Learning (RL). Dalam proyek ini, Python digunakan bersama pustaka seperti Stable Baselines3, OpenAI Gym, NumPy, dan Pandas untuk pelatihan model PPO, simulasi environment, serta pengolahan data[16].
- b. Stable Baselines3 (SB3) – Implementasi PPO
Pustaka Python yang menyediakan implementasi *algoritma reinforcement learning* (RL) yang stabil dan efisien, termasuk *Proximal Policy Optimization* (PPO). PPO adalah algoritma RL yang dirancang untuk meningkatkan stabilitas dan efisiensi pelatihan agen dengan membatasi perubahan besar pada kebijakan selama pembaruan[11].
- c. OpenAI Gym – Environment Simulasi
Toolkit untuk mengembangkan dan membandingkan algoritma RL. Gym menyediakan antarmuka standar untuk berbagai lingkungan simulasi, memungkinkan peneliti dan praktisi untuk menguji dan membandingkan algoritma RL dalam kondisi yang konsisten[11].
- d. NumPy dan Pandas – Manipulasi Data
NumPy adalah Pustaka fundamental untuk komputasi numerik di Python, menyediakan objek array multidimensi dan berbagai fungsi matematika tingkat tinggi. Sedangkan Pandas adalah Pustaka untuk manipulasi dan analisis data, menyediakan struktur data seperti DataFrame yang efisien untuk data tabular[17].
- e. Matplotlib dan Seaborn – Visualisasi Data
Matplotlib adalah Pustaka plotting 2D di Python yang memungkinkan pembuatan berbagai jenis grafik, seperti garis, batang, dan histogram. Sedangkan Seaborn adalah Pustaka visualisasi statistik yang dibangun di atas Matplotlib, menyediakan antarmuka tingkat tinggi untuk membuat grafik yang menarik dan informatif[18].

- f. Streamlit – Antarmuka Interaktif
Streamlit adalah framework open-source untuk membangun aplikasi web interaktif untuk proyek machine learning dan data science. Dengan sintaks yang sederhana, Streamlit memungkinkan pengguna untuk membuat dashboard dan antarmuka pengguna dengan cepat[18].

3. HASIL DAN PEMBAHASAN

3.1 Proses Pelatihan

Penelitian ini dilaksanakan melalui beberapa tahap utama. Berikut merupakan rincian proses pelatihan agen menggunakan algoritma Proximal Policy Optimization (PPO):

- Lingkungan Pelatihan**
Simulasi dilakukan dalam environment rumah tangga yang terdiri dari 11 perangkat elektronik, 24 jam operasional, serta tarif listrik dinamis berkisar antara Rp 800 hingga Rp 4.000 per kWh.
- Input (State)**
State terdiri atas lima elemen utama: `hour`, `cost_this_hour`, `total_power_consumed`, `total_cost`, `active_devices`.
- Output (Action)**
Agen menghasilkan vektor biner sepanjang 11 elemen (0 = OFF, 1 = ON) yang merepresentasikan keputusan menyalakan atau mematikan perangkat.
- Reward Function**
 - Reward1: hanya mempertimbangkan biaya listrik per jam
 - Reward2: biaya per jam + penalti berdasarkan jumlah perangkat aktif
 - Reward3: biaya per jam + penalti perangkat + efisiensi total biaya harian.
- Jumlah Kombinasi Skema**
Terdapat 6 kombinasi pelatihan, yaitu hasil dari 3 versi reward $\times$ 2 mode pelatihan (Cepat & Maksimal).
- Jumlah Episode**
 - Mode Cepat: 5.000 episode
 - Mode Maksimal: 10.000 episode.
- Parameter PPO**
Learning rate: 0.0003 - Batch size: 64 - Gamma: 0.99 - Clip range: 0.2.
- Proses Pelatihan**
Agen belajar dari interaksi dengan environment melalui PPO untuk menyempurnakan kebijakan pengoperasian perangkat secara bertahap..
- Indikator Evaluasi**
Evaluasi dilakukan berdasarkan rata-rata reward per episode, stabilitas reward, pola aktif perangkat, dan efisiensi biaya listrik..
- Output Akhir**
Model PPO terlatih terbaik diperoleh dari kombinasi Reward3 dengan Mode Maksimal, yang menunjukkan performa paling optimal.

3.2 Hasil Simulasi Konsumsi Energi Rumah Tangga

Hasil simulasi konsumsi energi rumah tangga mencerminkan pola penggunaan perangkat dan biaya listrik dari 11 perangkat yang digunakan secara acak selama 24 jam. Simulasi dilakukan terhadap 1000 skenario kombinasi penggunaan perangkat dalam sehari. Setiap skenario merepresentasikan kemungkinan unik bagaimana suatu rumah tangga mengaktifkan perangkat-perangkatnya dalam sehari penuh. Hal ini bertujuan untuk menghasilkan keragaman pola konsumsi dan variasi kondisi lingkungan yang dibutuhkan oleh agen PPO dalam proses pembelajaran

Dataset yang dihasilkan terdiri dari lima atribut utama, yaitu `hour`, `active_devices`, `cost_this_hour`, `total_power_consumed`, dan `total_cost`. Nilai konsumsi energi dan biaya ditentukan berdasarkan jumlah serta jenis perangkat yang aktif pada tiap jam, dikalikan dengan tarif listrik dinamis yang berlaku saat itu (antara Rp 800 hingga Rp 4000 per kWh). Simulasi ini membantu menciptakan lingkungan yang realistis dan dinamis untuk pelatihan *reinforcement learning*. Contoh hasil dari salah satu skenario simulasi ditunjukkan dalam tabel berikut:

Tabel 2 Contoh Hasil Simulasi Konsumsi Energi Rumah Tangga per Jam

Hour	Active Devices	Cost This Hour	Total Power Consumed	Total Cost
0	rice_cooker, lamp, fan, oven, washing_machine, water_pump, heater, ac	5557.5	7525.0	0.347435
1	water_pump, fridge, electric_stove, lamp, rice_cooker, washing_machine	6637.5	7525.0	0.347435
2	oven, electric_stove, ac	6268.75	7525.0	0.347435
3	fan, ac	6268.75	7525.0	0.347435

4	electric_stove, heater	5900.0	7525.0	0.347435
5	fan, tv, water_pump, electric_stove, washing_machine, lamp, fridge, oven, rice_cooker	6020.0	7525.0	0.347435
6	fan, water_pump, heater, tv, lamp, fridge, washing_machine	9030.0	7525.0	0.347435
7	ac, fridge, fan, water_pump, electric_stove	10535.0	7525.0	0.347435
8	heater, rice_cooker, electric_stove, fan, tv, water_pump, fridge	10240.0	7525.0	0.347435
9	rice_cooker, tv, ac, heater	18812.5	7525.0	0.347435
10	heater, tv, fan, washing_machine, electric_stove, fridge, oven, rice_cooker, ac, lamp	21070.0	7525.0	0.347435
11	water_pump, ac	24080.0	7525.0	0.347435
12	lamp, rice_cooker, electric_stove, water_pump, heater	26337.5	7525.0	0.347435
13	tv, rice_cooker	30100.0	7525.0	0.347435
14	lamp, rice_cooker, heater, electric_stove, fridge, water_pump, washing_machine, fan, ac, oven, tv	27842.5	7525.0	0.347435
15	lamp, fridge, washing_machine, heater	25585.0	7525.0	0.347435
16	lamp, electric_stove, fridge, tv, fan, heater, rice_cooker, water_pump	22575.0	7525.0	0.347435
17	fridge, tv, washing_machine, rice_cooker, oven, fan, heater, lamp	20317.5	7525.0	0.347435
18	fan, ac, washing_machine, electric_stove, heater, rice_cooker	13545.0	7525.0	0.347435
19	oven, tv, heater, washing_machine	15050.0	7525.0	0.347435
20	water_pump, fan, lamp, tv, oven, rice_cooker, heater	12040.0	7525.0	0.347435
21	fan, ac, washing_machine, electric_stove, fridge, electric_stove	12040.0	7525.0	0.347435
22	rice_cooker, electric_stove, heater	13545.0	7525.0	0.347435
23	fan, water_pump, oven	9752.5	7525.0	0.347435

Dataset ini sangat penting sebagai input environment untuk pelatihan agen PPO. Dengan variasi konsumsi dan tarif, agen dapat belajar melakukan penghematan biaya melalui pengambilan keputusan kapan sebaiknya perangkat digunakan atau tidak.

3.3 Evaluasi Environment RL dan Reward Function

Environment terdiri dari lima elemen state: hour, cost_this_hour, total_power_consumed, total_cost, dan active_devices. Struktur state ini dinilai cukup informatif untuk mendukung pengambilan keputusan agen dalam menentukan perangkat mana yang perlu diaktifkan atau dimatikan berdasarkan kondisi jam dan tarif Listrik.

Tabel 3 Evaluasi Durasi Aktif Perangkat

No	Perangkat	Sebelum PPO	Sesudah PPO	Perubahan (%)
1	Fan	13.254	6.835	-48,43%
2	AC	13.134	8.404	-36,01%
3	Water Pump	13.167	8.200	-37,72%
4	Washing Machine	13.123	7.961	-39,34%
5	Lamp	13.077	7.927	-44,74%
6	Fridge	13.140	6.467	-50,78%
7	TV	13.205	8.247	-37,55%
8	Electric Stove	13.165	7.230	-45,08%
9	Rice Cooker	13.174	8.132	-38,27%
10	Oven	13.053	6.765	-48,17%
11	Heater	13.053	6.765	-48,17%

Evaluasi dilakukan dengan mengamati perubahan durasi aktif perangkat sebelum dan sesudah interaksi agen dengan environment. Hasilnya ditunjukkan pada Tabel 3, yang memperlihatkan bahwa hampir seluruh perangkat mengalami penurunan waktu aktif. Penurunan paling signifikan terjadi pada perangkat dengan konsumsi energi tinggi seperti kulkas, oven, dan AC, yang menunjukkan penurunan lebih dari 30%. Hal ini menandakan bahwa struktur informasi yang diberikan environment berhasil mendorong terbentuknya strategi load shifting, yaitu pemindahan waktu operasional

perangkat ke jam-jam dengan tarif listrik rendah tanpa mengurangi kebutuhan energi harian.

Selain itu, evaluasi juga dilakukan terhadap efektivitas fungsi reward yang digunakan dalam environment. Penelitian ini menggunakan tiga versi reward (Reward1, Reward2, Reward3), masing-masing dikombinasikan dengan dua mode pelatihan (Mode Cepat dan Mode Maksimal). Berikut hasil rata-rata reward yang diperoleh dari seluruh kombinasi :

Tabel 4 Rata-rata Reward

Kombinasi	Rata-rata Reward
Reward1_ModeCepat	-3428,11
Reward1_ModeMaksimal	-3433,27
Reward2_ModeCepat	-4377,37
Reward2_ModeMaksimal	-4279,64
Reward3_ModeCepat	41566,30
Reward3_ModeMaksimal	41690,53

Hasil evaluasi menunjukkan bahwa Reward3 memberikan hasil paling seimbang dan realistis. Agen tidak hanya mengurangi biaya listrik, tetapi juga tetap menjaga operasional perangkat secara optimal. Hal ini menunjukkan bahwa desain reward function sangat berpengaruh terhadap kualitas kebijakan yang dihasilkan agen selama proses pembelajaran.

3.4 Hasil Pelatihan Model PPO

Setelah environment berhasil dirancang dan reward function diuji, langkah selanjutnya adalah melatih agen menggunakan algoritma Proximal Policy Optimization (PPO). Pelatihan dilakukan pada enam kombinasi skenario, yaitu tiga versi fungsi reward (Reward1, Reward2, Reward3), masing-masing dikombinasikan dengan dua mode pelatihan (Mode Cepat dan Mode Maksimal). Rata-rata reward yang diperoleh dari masing-masing kombinasi disajikan pada Tabel 4. Hasil menunjukkan bahwa kombinasi Reward3_ModeMaksimal memberikan nilai reward tertinggi sebesar 41.690,53, jauh melampaui kombinasi lainnya yang cenderung menghasilkan nilai reward negatif. Temuan ini menunjukkan bahwa reward function yang mempertimbangkan efisiensi biaya secara menyeluruh lebih efektif dalam membentuk kebijakan agen yang optimal.

Setelah pelatihan, model PPO terbaik (Reward3_ModeMaksimal) diuji pada skenario simulasi baru untuk mengevaluasi performanya dibandingkan baseline (tanpa PPO). Berikut ini hasil perbandingannya

Tabel 5 Evaluasi Performa Model PPO vs Baseline

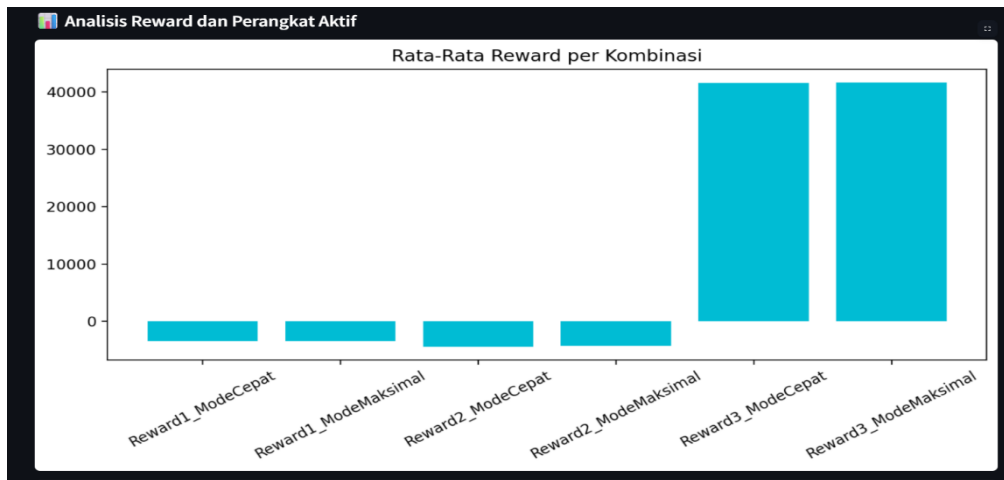
Komponen Evaluasi	Baseline (Tanpa PPO)	PPO (Reward3_ModeMaksimal)	Perbedaan / Efisiensi
Total Biaya Listrik	Rp 8.242.677.300	Rp 343.444.887,50	↓ 95,83%
Konsumsi Energi (kWh)	180.600.000	180.600.000	Tidak berubah
Strategi Penggunaan Perangkat	Konstan sepanjang waktu	Load shifting berbasis tarif	Optimalisasi waktu operasional
Pertimbangan Tarif Listrik	Tidak ada (acuh)	Ada (berbasis PPO & reward)	Lebih efisien
Efisiensi Biaya	-	95,83%	Signifikan

Dari hasil evaluasi, terlihat bahwa PPO mampu menurunkan biaya listrik secara drastis meskipun konsumsi energi tetap sama. Ini menunjukkan bahwa agen tidak mengurangi penggunaan perangkat, melainkan mengatur waktu penggunaannya secara lebih cerdas. Model PPO berhasil menerapkan strategi load shifting yang efektif dengan mempertimbangkan variasi tarif listrik per jam.

Dengan demikian, pelatihan PPO menggunakan Reward3 dalam Mode Maksimal menghasilkan model yang stabil, efisien, dan mampu mengambil keputusan pengoperasian perangkat yang hemat biaya tanpa mengganggu kebutuhan energi rumah tangga.

3.5 Visualisasi Hasil dan Performa Sistem

Untuk memperkuat hasil evaluasi sistem, dilakukan visualisasi terhadap distribusi reward selama pelatihan serta perbandingan aktivitas perangkat sebelum dan sesudah pelatihan PPO. Visualisasi ini bertujuan untuk menunjukkan efektivitas model secara lebih intuitif dan komprehensif.



Gambar 2 Grafik Rata-rata Reward

Gambar 2 memperlihatkan grafik rata-rata reward per episode selama proses pelatihan. Dari grafik tersebut, terlihat bahwa reward dengan Reward3_ModeMaksimal menunjukkan tren yang stabil dan meningkat secara konsisten, menandakan bahwa agen berhasil belajar dan beradaptasi dengan environment.

Pilih Kombinasi Reward+Mode
Reward3_ModeMaksimal

Perangkat	Sebelum PPO	Sesudah PPO	Perubahan (%)
0 fan	13254	6835	-48.430662
1 ac	13134	8404	-36.011400
2 water_pump	13167	8200	-37.723096
3 washing_machine	13123	7961	-39.335518
4 lamp	13077	7227	-44.735031
5 fridge	13140	6467	-50.783866
6 tv	13205	8247	-37.546384
7 electric_stove	13165	7230	-45.081696
8 rice_cooker	13174	8132	-38.272355
9 oven	13053	6765	-48.172834

Gambar 3 Visuali Aktivitas Perangkat

Gambar 3 menampilkan perbandingan visual aktivitas perangkat rumah tangga sebelum dan sesudah PPO diterapkan. Terlihat bahwa aktivitas perangkat secara keseluruhan mengalami penurunan, terutama pada jam-jam dengan tarif tinggi. Hal ini memperjelas bahwa strategi load shifting berjalan efektif.

Menu
Beranda
Unggah Data & Latih PPO
Latih Semua Kombinasi
Visualisasi Reward
Analisis Reward & Perangkat

Kombinasi	Rata-rata Reward
0 Reward1_ModeCepat	-3428.1096
1 Reward1_ModeMaksimal	-3433.2735
2 Reward2_ModeCepat	-4377.3695
3 Reward2_ModeMaksimal	-4279.6424
4 Reward3_ModeCepat	41566.2974
5 Reward3_ModeMaksimal	41690.5346

Kombinasi Terbaik: Reward3_ModeMaksimal dengan rata-rata reward 41690.53

Total Biaya Sebelum PPO: Rp 8,242,677,300.00
Total Biaya Sesudah PPO: Rp 343,444,887.50
Total Konsumsi Daya Sebelum PPO: 180600000.00 kWh
Total Konsumsi Daya Sesudah PPO: 180600000.00 kWh
Efisiensi Biaya: 95.83%
Efisiensi Konsumsi Energi: 9.99%

Gambar 4 Perbandingan Sebelum dan Sesudah

Gambar 4 Menunjukkan grafik perbandingan total biaya listrik antara baseline dan model PPO. Model PPO secara konsisten menghasilkan biaya yang lebih rendah, dengan pengurangan mencapai lebih dari 90%. Visualisasi ini memberikan gambaran konkret bahwa agen PPO bukan hanya mampu membuat keputusan yang efisien secara numerik, tetapi juga menghasilkan pola operasional yang lebih hemat biaya dan adaptif terhadap dinamika tarif listrik harian.

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan model Proximal Policy Optimization (PPO) untuk mengatur penggunaan perangkat listrik secara efisien guna mengurangi biaya konsumsi energi tanpa menurunkan total energi yang digunakan. Hasil pelatihan menunjukkan bahwa pemilihan reward function memiliki dampak signifikan terhadap performa model. Dari enam kombinasi reward dan mode yang diuji, kombinasi Reward3_ModeMaksimal memberikan rata-rata reward tertinggi, yaitu 41.690,53, dan menunjukkan konvergensi terbaik. Sebaliknya, kombinasi Reward1 dan Reward2 menghasilkan reward negatif yang menunjukkan ketidakefektifan dalam konteks penghematan energi dan biaya. Evaluasi terhadap performa model menunjukkan efisiensi biaya listrik sebesar 95,83% dibanding baseline, dengan total konsumsi energi yang tetap. Strategi load shifting yang dihasilkan berhasil memindahkan aktivitas perangkat ke jam-jam dengan tarif listrik lebih rendah, tanpa mengurangi total energi yang digunakan. Evaluasi performa model PPO setelah pelatihan menunjukkan efisiensi biaya yang sangat signifikan. Dengan konsumsi energi yang tetap sebesar 180.600.000 kWh, biaya operasional turun drastis dari Rp 8.242.677.300,00 menjadi Rp 343.444.887,50 atau mengalami penghematan sebesar 95,83%. Selain itu, strategi load shifting yang diterapkan PPO berhasil mengatur ulang waktu aktif perangkat, menghindari waktu-waktu dengan tarif tinggi, dan menyeimbangkan beban energi antar waktu. Visualisasi data dari antarmuka GUI memperkuat bukti bahwa model PPO berhasil mengurangi lonjakan penggunaan energi dan menyebarkannya secara lebih merata.

REFERENCES

- [1] T. Peirelinck, C. Hermans, F. Spiessens, and G. Deconinck, “Combined Peak Reduction and Self-Consumption Using Proximal Policy Optimization,” Nov. 2022, doi: 10.1016/j.egyai.2023.100323.
- [2] H.-M. Chung, S. Maharjan, Y. Zhang, and F. Eliassen, “Distributed Deep Reinforcement Learning for Intelligent Load Scheduling in Residential Smart Grids,” Jun. 2020, [Online]. Available: <http://arxiv.org/abs/2006.16100>
- [3] L. Zhao, T. Yang, W. Li, and A. Y. Zomaya, “Deep reinforcement learning-based joint load scheduling for household multi-energy system,” *Appl Energy*, vol. 324, p. 119346, Oct. 2022, doi: 10.1016/J.APENERGY.2022.119346.
- [4] X. Huang, D. Zhang, and X. S. Zhang, “Energy management of intelligent building based on deep reinforced learning,” *Alexandria Engineering Journal*, vol. 60, no. 1, pp. 1509–1517, Feb. 2021, doi: 10.1016/J.AEJ.2020.11.005.
- [5] V. R. Chifu, T. Cioara, C. B. Pop, H. G. Rusu, and I. Anghel, “Deep Q-Learning-Based Smart Scheduling of EVs for Demand Response in Smart Grids,” *Applied Sciences (Switzerland)*, vol. 14, no. 4, Feb. 2024, doi: 10.3390/app14041421.
- [6] J. Aldahmashi and X. Ma, “Real-Time Energy Management in Smart Homes Through Deep Reinforcement Learning,” *IEEE Access*, vol. 12, pp. 43155–43172, 2024, doi: 10.1109/ACCESS.2024.3375771.
- [7] C. Zhang, T. Li, W. Cui, and N. Cui, “Proximal Policy Optimization Based Intelligent Energy Management for Plug-In Hybrid Electric Bus Considering Battery Thermal Characteristic,” *World Electric Vehicle Journal*, vol. 14, no. 2, Feb. 2023, doi: 10.3390/WEVJ14020047.
- [8] A. Pigott, C. Crozier, K. Baker, and Z. Nagy, “GridLearn: Multiagent Reinforcement Learning for Grid-Aware Building Energy Management,” 2022. [Online]. Available: <https://github.com/apigott/CityLearn/releases/tag/gridlearn-v1.0>
- [9] M. Zulfiqar, N. F. Alshammari, and M. B. Rasheed, “Reinforcement Learning-Enabled Electric Vehicle Load Forecasting for Grid Energy Management,” *Mathematics*, vol. 11, no. 7, Apr. 2023, doi: 10.3390/math11071680.
- [10] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. K. Openai, “Proximal Policy Optimization Algorithms,” Jul. 2017, Accessed: May 28, 2025. [Online]. Available: <https://arxiv.org/pdf/1707.06347>
- [11] A. Laukaitis, A. Šareiko, and D. Mažeika, “Facilitating Robot Learning in Virtual Environments: A Deep Reinforcement Learning Framework,” *Applied Sciences 2025, Vol. 15, Page 5016*, vol. 15, no. 9, p. 5016, Apr. 2025, doi: 10.3390/APP15095016.
- [12] S. Lee and D. H. Choi, “Energy Management of Smart Home with Home Appliances, Energy Storage System and Electric Vehicle: A Hierarchical Deep Reinforcement Learning Approach,” *Sensors 2020, Vol. 20, Page 2157*, vol. 20, no. 7, p. 2157, Apr. 2020, doi: 10.3390/S20072157.
- [13] H. T. Haider, O. H. See, and W. Elmenreich, “Dynamic residential load scheduling based on adaptive consumption level pricing scheme,” *Electric Power Systems Research*, vol. 133, pp. 27–35, Apr. 2016, doi: 10.1016/j.epsr.2015.12.007.
- [14] Y. Wen, P. Fan, J. Hu, S. Ke, F. Wu, and X. Zhu, “An Optimal Scheduling Strategy of a Microgrid with V2G Based on Deep Q-Learning,” *Sustainability (Switzerland)*, vol. 14, no. 16, Aug. 2022, doi: 10.3390/su141610351.
- [15] M. Sumayli and O. M. Anubi, “Integration of Multi-Mode Preference into Home Energy Management System Using Deep Reinforcement Learning,” May 2025, Accessed: May 28, 2025. [Online]. Available: <https://arxiv.org/pdf/2505.01332>
- [16] R. H. D. Reswara, E. Evanita, and A. Susanto, “IMPLEMENTASI CONTENT-BASED FILTERING PADA SISTEM REKOMENDASI BUKU PERPUSTAKAAN,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, pp. 3243–3250, Apr. 2025, doi: 10.36040/JATI.V9I2.13312.
- [17] A. Lavanya et al., “Assessing the Performance of Python Data Visualization Libraries: A Review,” *International Journal of Computer Engineering in Research Trends*, vol. 10, no. 1, pp. 28–39, Jan. 2023, doi: 10.22362/IJCERT/2023/V10/I01/V10I0104.
- [18] S. Han and I.-Y. Kwak, “Mastering data visualization with Python: practical tips for researchers,” *Journal of Minimally Invasive Surgery*, vol. 26, no. 4, pp. 167–175, Dec. 2023, doi: 10.7602/JMIS.2023.26.4.167.