

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam *machine learning* (ML), dikenal dua kategori variabel parameter: *model parameter* dan *hyperparameter*. *model parameter* merupakan variabel internal yang nilainya dipelajari secara otomatis dari *training data*. Sedangkan *hyperparameter* adalah variabel konfigurasi yang ditentukan sebelum proses pembelajaran.

Hyperparameter seringkali mempengaruhi secara signifikan aspek-aspek model seperti kompleksitas, perilaku, kecepatan, dan sebagainya. Oleh karena itu, nilai suatu *hyperparameter* harus dipilih dengan bijak untuk menghasilkan performa yang optimal.

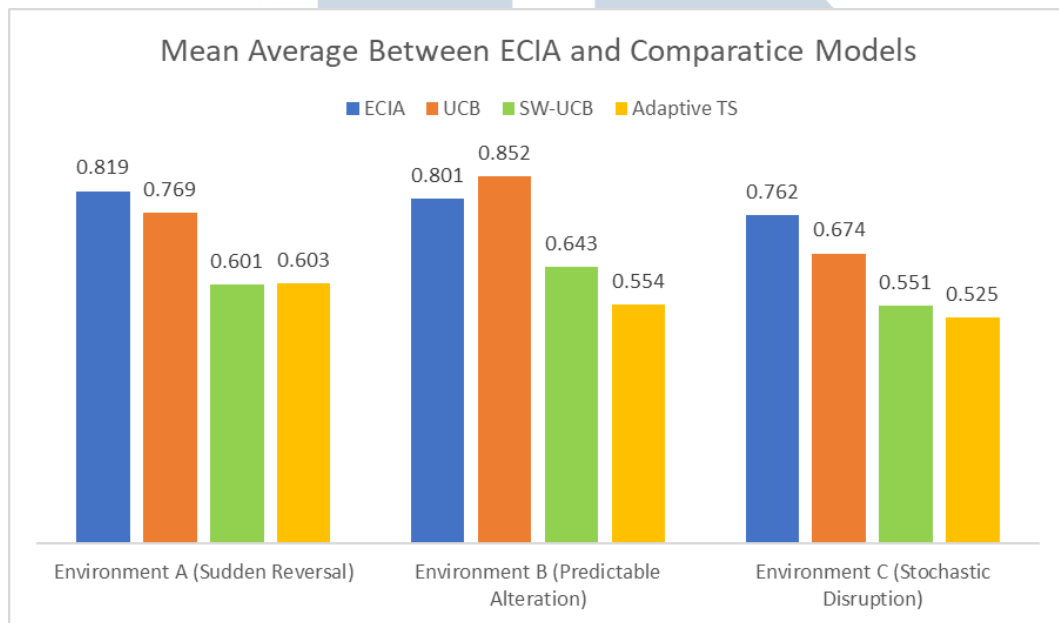
Cara klasik memilih *hyperparameter* dilakukan dengan pemilihan manual oleh pengembang model dengan metode *trial-and-error*. Metode ini memerlukan studi ablatasi dalam skala besar dan seringkali menghabiskan waktu, rentan terhadap bias dan kesalahan manusia, serta sulit direproduksi secara komputasi. Oleh karena itu, *hyperparameter optimization* (HPO) menjadi salah satu cabang keilmuan yang paling sering diteliti dalam dunia AutoML [1].

Penerapan HPO dalam bidang *reinforcement learning* RL cenderung masih lebih jarang ditemui dibandingkan cabang lainnya dalam ML. Studi oleh [2] mengagaskan bahwa fenomena ini disebabkan oleh beberapa faktor. Diantaranya adalah pandangan sebelah mata kepada *hyperparameter* yang dianggap kurang penting, serta keabu-abuan *cost-and-benefit* dari teknik HPO terhadap performa algoritma RL. Menurut temuan [2], *hyperparameter* yang sering dianggap kurang penting ternyata dapat menjadi penentu kesuksesan sebuah algoritma dalam sebuah lingkungan pembelajaran. Dalam banyak kasus, *hyperparameter* juga dapat memiliki efek yang signifikan terhadap stabilitas suatu algoritma saat pembelajaran berlangsung.

Emotional-Cognition Integration Architecture (ECIA) merupakan arsitektur pengambilan keputusan adaptif berbasis *reinforcement learning* (RL) yang dikembangkan oleh [3] untuk menjawab tantangan lingkungan non-stasioner.

Arsitektur ini mengambil inspirasi dari sistem biologis pada sistem emosional makhluk hidup yang mampu memproses keputusan secara paralel dan

sinergis. Hal ini dilakukan dengan mengintegrasikan variabel sinyal yang memiliki properti tertentu untuk mensimulasikan delapan emosi. Sinyal-sinyal ini kemudian menjadi indikator bagi agen ECIA untuk menyesuaikan parameter eksplorasi dan eksploitasi.



Gambar 1.1. Perbandingan overall mean reward antara ECIA dan metode baseline pada tiga lingkungan non-stasioner

Dalam penelitian [3], hasil evaluasi yang ditunjukkan pada Gambar 1.1 menunjukkan bahwa model ECIA mendominasi performa dalam dalam lingkungan-lingkungan dengan tingkat ketidakpastian tinggi (Environment A dan Environment C) sedangkan performanya kalah dengan *Upper Confidence Bound* (UCB) dalam lingkungan yang lebih terprediksi (Environment B).

Hasil ini mengindikasikan bahwa ECIA efektif untuk kasus-kasus dengan ketidakpastian lingkungan yang tinggi, seperti pasar keuangan dengan pergeseran rezim, atau penasihat pengambilan keputusan klinis di mana kondisi pasien berubah secara tiba-tiba. ECIA juga bagus untuk kasus dimana ada beberapa opsi strategis yang berkompetisi, membutuhkan evaluasi cepat saat situasi berubah.

Dalam pengimplementasiannya, ECIA dirancang dan dievaluasi melalui kerangka permasalahan *Multi-Armed Bandits* (MAB) dengan lima lengan. ECIA memiliki beberapa hyperparameter yang memerlukan konfigurasi nilai sebelum proses pelatihan dijalankan. Salah satu *hyperparameter* tersebut adalah epsilon (ϵ). Dalam penelitian terdahulu, nilai *hyperparameter* tersebut diperoleh melalui

preliminary experimentation dan *manual tuning*, tanpa menggunakan pendekatan pencarian otomatis seperti *Bayesian Optimization*.

Epsilon (ϵ) adalah *hyperparameter* yang berfungsi sebagai *exploration coefficient* yang nilainya mempengaruhi mekanisme *adaptive exploration rate* pada ECIA. *Exploration rate* umumnya merupakan probabilitas agen melakukan eksplorasi *action* baru daripada memilih berdasarkan data historis (*exploitation*).

Melihat kembali performa ECIA pada Environment B, yang dirancang dengan dinamika perubahan yang teratur (*regular oscillation*) dan lebih dapat diprediksi dibandingkan lingkungan lainnya, mekanisme eksplorasi adaptif pada ECIA diduga memerlukan konfigurasi yang berbeda. Secara teoritis, lingkungan yang memiliki pola transisi yang konsisten cenderung lebih banyak memanfaatkan informasi yang telah dipelajari (*exploitation*) dibandingkan eksplorasi yang berlebihan (*exploration*). Dengan demikian, nilai *exploration coefficient* (ϵ) sebagai pengendali *adaptive exploration rate* menjadi salah satu *hyperparameter* yang layak dievaluasi lebih lanjut terhadap pengaruhnya pada performa ECIA di lingkungan tersebut.

Pemilihan *hyperparameter* epsilon sebagai objek studi didasarkan pada perannya dalam mekanisme eksplorasi adaptif ECIA, sehingga perubahan nilainya secara langsung memengaruhi keseimbangan antara eksplorasi dan eksploitasi yang menjadi karakteristik utama pada Environment B.

Berdasarkan kondisi tersebut, penelitian ini menerapkan metode *Bayesian Optimization* untuk mencari konfigurasi nilai *hyperparameter* epsilon yang menghasilkan *overall mean reward* lebih tinggi pada Environment B. Fokus pencarian konfigurasi diarahkan pada *exploration coefficient* (ϵ) untuk mengevaluasi sejauh mana penyesuaian mekanisme eksplorasi adaptif memengaruhi kinerja ECIA pada lingkungan dengan pola perubahan yang relatif terprediksi. Pendekatan ini diharapkan dapat memberikan pemahaman yang lebih baik mengenai sensitivitas ECIA terhadap konfigurasi ϵ . Pendekatan ini juga menunjukkan bagaimana proses pencarian konfigurasi *hyperparameter* dapat dilakukan secara sistematis dibandingkan pemilihan nilai secara *manual* pada penelitian sebelumnya.

Sejumlah penelitian terdahulu telah mengeksplorasi penerapan *Hyperparameter Optimization* (HPO) dalam *Reinforcement Learning* (RL) dari berbagai pendekatan. [4] memperkenalkan metodologi empiris melalui metrik sensitivitas dan dimensi efektif *hyperparameter* pada varian PPO, yang mengungkapkan bahwa peningkatan kinerja dari teknik normalisasi umum sering

kali lebih didorong oleh ketergantungan pada penyesuaian khusus lingkungan ketimbang proses optimasi itu sendiri.

Di sisi lain, [5] membuktikan bahwa HPO berbasis model seperti *Gaussian Processes*, *Random Forests*, dan *Gradient Boosted Random Trees* pada algoritma DDPG mampu melipatgandakan akumulasi imbalan agen dalam manajemen energi kendaraan listrik *hybrid* sembari tetap mematuhi batasan waktu pelatihan. Untuk mengoptimalkan *hyperparameter* secara dinamis, [6] mengusulkan *Generalized Population-Based Training* (GPBT) yang dipadukan dengan *Pairwise Learning* (PL) guna mengatasi konvergensi dini pada PBT tradisional dengan memanfaatkan perbedaan kinerja berpasangan serta *pseudo-gradient* untuk menjaga keragaman populasi. Memperkuat efektivitas pendekatan otomatisasi ini, [7] turut menunjukkan bahwa HPO otomatis secara signifikan mampu mengungguli konfigurasi manual yang dirancang oleh para ahli pada *Model-based Reinforcement Learning* (MBRL).

Pada *Bayesian Optimization*, [8] menerapkan model Gaussian Process untuk menyetel *hyperparameter* secara sistematis pada Deep Q-Network untuk tugas otomasi navigasi. Pendekatan tersebut mencapai tingkat keberhasilan tugas sebesar 83%, serta menunjukkan efisiensi sampel dan kecepatan konvergensi yang lebih baik dibandingkan grid search maupun random search.

Berangkat dari keberhasilan pendekatan tersebut, penelitian ini mengadopsi *Bayesian Optimization* dengan menggunakan *Tree-structured Parzen Estimator* (TPE). Dibandingkan pendekatan berbasis Gaussian Process, TPE memodelkan distribusi probabilitas dari konfigurasi *hyperparameter* yang menghasilkan nilai objektif baik maupun buruk, sehingga mampu mengarahkan proses pencarian secara lebih efisien tanpa mengasumsikan bentuk fungsi objektif tertentu. Karakteristik ini menjadikan TPE sesuai untuk permasalahan optimasi *hyperparameter* yang memiliki ruang pencarian kompleks dan biaya evaluasi yang relatif mahal, seperti pada pelatihan *reinforcement learning*.

1.2 Rumusan Masalah

1. Bagaimana pengaruh variasi nilai *hyperparameter* epsilon (ϵ) terhadap *overall mean reward* agen ECIA pada Environment B yang bersifat non-stasioner?
2. Apakah penyesuaian nilai *hyperparameter* epsilon (ϵ) menggunakan *Bayesian Optimization* mampu meningkatkan *overall mean reward* agen

ECIA dibandingkan dengan konfigurasi baseline?

1.3 Batasan Permasalahan

1. Penelitian menggunakan implementasi *Emotional-Cognition Integration Architecture* (ECIA) sesuai dengan implementasi yang disediakan oleh penelitian terdahulu tanpa melakukan modifikasi terhadap mekanisme internal arsitektur, melainkan hanya melakukan penyesuaian terhadap nilai *hyperparameter* epsilon (ϵ).
2. Pencarian nilai *hyperparameter* dilakukan dengan pendekatan *Bayesian Optimization* menggunakan algoritma *Tree-structured Parzen Estimator* (TPE) yang diimplementasikan sebagai *sampler* pada pustaka Optuna.
3. Optimasi dilakukan terhadap satu *hyperparameter*, yaitu epsilon (ϵ), tanpa mempertimbangkan interaksi dengan *hyperparameter* ECIA lainnya.
4. Kinerja agen dievaluasi berdasarkan *overall mean reward*, yaitu rata-rata dari *episode mean reward* yang diperoleh pada 3.600 *independent runs*, mengikuti protokol evaluasi pada penelitian terdahulu.
5. Implementasi agen ECIA yang digunakan mengacu pada *source code* resmi dari penelitian sebelumnya tanpa melakukan perubahan terhadap algoritma inti.
6. Penelitian berfokus pada peningkatan performa melalui penyesuaian *hyperparameter* dan tidak membahas optimasi parameter internal atau modifikasi *fungsi reward* pada arsitektur ECIA.

1.4 Tujuan Penelitian

1. Menganalisis pengaruh variasi nilai *hyperparameter* epsilon (ϵ) terhadap *overall mean reward* agen ECIA pada Environment B.
2. Mengevaluasi kemampuan *Bayesian Optimization* dalam memperoleh konfigurasi nilai epsilon (ϵ) yang menghasilkan *overall mean reward* lebih tinggi dibandingkan konfigurasi *baseline* ECIA.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi, baik secara teoretis maupun praktis, sebagai berikut:

1. Manfaat Teoretis

- (a) Memberikan pemahaman mengenai pengaruh konfigurasi nilai *hyperparameter* epsilon (ϵ) terhadap performa agen ECIA pada lingkungan non-stasioner dengan pola perubahan yang relatif terprediksi.
- (b) Menambah referensi mengenai penerapan *Bayesian Optimization* untuk pencarian konfigurasi *hyperparameter* pada algoritma *reinforcement learning*, khususnya pada arsitektur *Emotional-Cognition Integration Architecture* (ECIA).

2. Manfaat Praktis

- (a) Menghasilkan konfigurasi nilai *hyperparameter* epsilon yang mampu meningkatkan *overall mean reward* agen ECIA pada Environment B dibandingkan konfigurasi awal yang digunakan pada penelitian terdahulu.
- (b) Menyediakan contoh penerapan *Bayesian Optimization* sebagai pendekatan pencarian konfigurasi *hyperparameter* secara sistematis sehingga dapat menjadi acuan bagi penelitian lanjutan yang menggunakan arsitektur ECIA maupun algoritma *reinforcement learning* sejenis.

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian ini disusun sebagai berikut:

1. Bab 1 Pendahuluan

Bab ini memuat latar belakang penelitian, rumusan masalah, tujuan penelitian, batasan permasalahan, manfaat penelitian, serta sistematika penulisan.

2. Bab 2 Tinjauan Pustaka

Bab ini membahas landasan teori yang mendukung penelitian, meliputi

konsep *reinforcement learning*, *multi-armed bandit*, *Emotional-Cognition Integration Architecture* (ECIA), *hyperparameter optimization*, serta *Bayesian Optimization* menggunakan *Tree-structured Parzen Estimator* (TPE).

3. Bab 3 Metodologi Penelitian

Bab ini menjelaskan tahapan penelitian yang meliputi perancangan eksperimen, lingkungan pengujian, implementasi *Bayesian Optimization*, konfigurasi eksperimen, serta metode evaluasi yang digunakan.

4. Bab 4 Hasil dan Pembahasan

Bab ini menyajikan hasil eksperimen, analisis pengaruh konfigurasi nilai *hyperparameter* epsilon terhadap performa ECIA, serta pembahasan hasil optimasi pada Environment B.

5. Bab 5 Kesimpulan dan Saran

Bab ini berisi kesimpulan yang diperoleh dari hasil penelitian serta saran untuk pengembangan penelitian selanjutnya.

