



Rekomendasi Aktivitas Pembelajaran Anak Usia Dini Berbasis Q-Learning dan Profil Perkembangan

Faiz Ahmad Fauzan*, Joko Aryanto

Fakultas Sains & Teknologi, Program Studi Informatika, Universitas Teknologi Yogyakarta, Yogyakarta, Indonesia

Email: ^{1,*}fauzanfaizahmad@gmail.com, ²joko.aryanto@uty.ac.id

Email Penulis Korespondensi: fauzanfaizahmad@gmail.com

Abstrak—Penelitian ini merancang sistem rekomendasi aktivitas pembelajaran anak usia dini di lingkungan Raudhatul Athfal menggunakan *Q-Learning*. Fokus penelitian adalah membantu guru memilih aktivitas yang sesuai dengan profil perkembangan anak, terutama aspek fisik-motorik dan kognitif. Metode penelitian menggunakan pendekatan rancang bangun dengan formulasi *Markov Decision Process*. *State* memuat usia anak dalam satuan bulan, level fisik-motorik, level kognitif, aktivitas terakhir, dan skor partisipasi terakhir. *Action* berupa dua belas aktivitas pembelajaran, sedangkan *reward* menggabungkan partisipasi, kesesuaian perkembangan, dan variasi aktivitas. Pengujian dilakukan menggunakan data uji berbasis skenario pembelajaran dengan lima *seed* eksperimen. Hasil menunjukkan *Q-Learning* memperoleh *reward* evaluasi rata-rata 24.667 dengan standar deviasi 0.222 dari batas teoritis 30 poin pada skenario evaluasi. Evaluasi peringkat menghasilkan *Precision@1* 0.645, *Recall@5* 0.448, dan *NDCG@5* 0.641. Hasil ini menunjukkan bahwa pendekatan *tabular* layak sebagai *baseline* yang transparan, tetapi belum cukup untuk klaim dampak pedagogis. *Q-Learning* dipilih karena *state* bersifat diskrit, ruang tindakan terbatas, dan nilai Q dapat ditelusuri. Karena belum menggunakan data operasional, klaim penelitian dibatasi pada rancangan sistem, keterlacakan rekomendasi, dan evaluasi komputasional.

Kata Kunci: *Q-Learning*; Sistem Rekomendasi; Aktivitas Pembelajaran; Anak Usia Dini; Profil Perkembangan; *Reinforcement Learning*

Abstract—This study designs a Q-Learning-based recommendation system for early childhood learning activities in a Raudhatul Athfal setting. The system supports teachers in selecting activities aligned with children's developmental profiles, especially physical-motor and cognitive aspects. The recommendation problem is formulated as a Markov Decision Process. The state consists of children's age in months, physical-motor level, cognitive level, previous activity, and previous participation score. The action space contains twelve learning activities, while the reward combines participation, developmental fit, and activity variation. Testing was conducted using scenario-based learning data with five experimental seeds. The results show that Q-Learning achieved an average evaluation reward of 24.667 with a standard deviation of 0.222 from a theoretical scenario bound of 30 points. Ranking evaluation produced *Precision@1* of 0.645, *Recall@5* of 0.448, and *NDCG@5* of 0.641. These results support Q-Learning as a transparent tabular baseline, but they do not prove pedagogical impact. Q-Learning was selected because the state is discrete, the action space is limited, and Q-values are traceable. Since operational data have not been used, the claims are limited to system design, recommendation traceability, and computational evaluation.

Keywords: Q-Learning; Recommender System; Learning Activities; Early Childhood Education; Developmental Profile; Reinforcement Learning

1. PENDAHULUAN

Pendidikan anak usia dini memerlukan aktivitas pembelajaran yang sesuai dengan karakteristik perkembangan anak karena kegiatan bermain, gerak, dan interaksi sosial berperan dalam stimulasi perkembangan kognitif, fisik-motorik, dan regulasi diri anak (Jylänki et al., 2022; Sutapa et al., 2021; Yi et al., 2024). Pada lingkup Raudhatul Athfal, aktivitas pembelajaran juga terkait dengan pembiasaan, nilai keislaman, dan keterbatasan sarana kelas. Profil digital RA mitra dipakai hanya sebagai landasan awal untuk menyusun katalog aktivitas, bukan sebagai bukti empiris bahwa sistem sudah diterapkan pada data operasional anak. Dengan demikian, informasi institusi diposisikan sebagai dasar perumusan kebutuhan dan batas masalah, sedangkan validitas sistem dinilai melalui formulasi komputasional, eksperimen skenario, dan keterlacakan rekomendasi.

Raudhatul Athfal memiliki karakter yang berbeda dari sekolah umum karena pembelajaran anak usia dini disusun dengan perhatian pada perkembangan, pembiasaan, dan nilai keislaman. Pedoman implementasi kurikulum RA menempatkan kegiatan bermain dan pembiasaan sebagai bagian penting dari pengalaman belajar, sedangkan standar kompetensi lulusan PAUD menekankan capaian perkembangan yang holistik, bukan hanya pencapaian akademik sempit. Karena itu, sistem rekomendasi untuk RA tidak cukup hanya menyalin pola rekomendasi materi digital pada pendidikan tinggi. Sistem perlu memuat pembatasan yang jelas: rekomendasi adalah alat bantu pemilihan aktivitas, sedangkan keputusan akhir tetap berada pada guru.

Dalam pendidikan, sistem rekomendasi dipakai untuk membantu personalisasi sumber belajar, jalur belajar, dan aktivitas pembelajaran berdasarkan kebutuhan pengguna, situasi pembelajaran, serta tujuan pedagogis yang ingin dicapai (da Silva et al., 2023; Deschênes, 2020; Ricci et al., 2022; Siafis et al., 2024; Urdaneta-Ponte et al., 2021). Literatur sistem rekomendasi menjelaskan bahwa rekomendasi tidak hanya memilih *item* terbaik pada satu waktu, tetapi juga mengelola relevansi, urutan, dan kebutuhan pengguna. Dalam ranah pendidikan, masalah ini lebih sensitif karena rekomendasi perlu dibaca sebagai dukungan keputusan pedagogis, bukan pengganti guru.

Reinforcement learning relevan untuk masalah ini karena rekomendasi aktivitas merupakan keputusan berurutan. Aktivitas yang dipilih hari ini mempengaruhi riwayat aktivitas dan *reward* evaluasi berikutnya. *Q-Learning* menjadi



pilihan utama karena *state* penelitian ini bersifat diskrit, ruang tindakan terbatas, dan proses pembaharuannya transparan. *Q-Learning* sesuai untuk penelitian ini karena *state* dan *action* masih dapat direpresentasikan secara diskrit dalam tabel nilai, sehingga proses pembaruan nilai *action* pada setiap *state* dapat ditelusuri secara langsung dan mudah diaudit pada rancangan rekomendasi pendidikan berskala kecil (Afsar et al., 2023; Riedmann et al., 2025). Dengan karakter proyek saat ini, penggunaan model yang lebih kompleks berisiko menaikkan beban penjelasan tanpa memberi keuntungan metodologis yang sebanding.

Penelitian sebelumnya tentang *adaptive learning*, *intelligent tutoring*, dan sistem rekomendasi pendidikan banyak menggunakan lingkungan platform digital, jalur materi akademik, atau dukungan keputusan bagi guru (Bassen et al., 2020; Riedmann et al., 2025; Siafis et al., 2024). Sementara itu, penelitian teknologi pendidikan anak usia dini lebih sering membahas pemanfaatan media digital, kesiapan guru, literasi AI, atau rancangan *computing education* daripada sistem rekomendasi aktivitas harian yang terhubung langsung dengan katalog kegiatan kelas (Afsar et al., 2023; J. J. Chen, 2024; J. J. Chen & Lin, 2024; J. J. Chen & Perez, 2023; Liu et al., 2023; Yang, 2022; Yi et al., 2024). Celah penelitian muncul pada kebutuhan memformalkan pemilihan aktivitas RA sebagai keputusan berurutan yang tetap dapat dijelaskan secara sederhana. Pemilihan aktivitas pembelajaran juga perlu mempertimbangkan karakter bermain pada anak usia dini. Aktivitas fisik-motorik dan kognitif tidak berdiri sebagai *item* terpisah seperti kursus daring, tetapi muncul sebagai pengalaman kelas yang melibatkan gerak, alat, instruksi guru, dan respons anak. Kajian recommender pendidikan juga menunjukkan bahwa faktor afektif, preferensi, dan situasi belajar dapat mempengaruhi penerimaan rekomendasi, sehingga model awal perlu menyatakan batas data secara terbuka (Salazar et al., 2021). Dalam penelitian ini, unsur yang dimodelkan dibatasi pada level perkembangan dan riwayat aktivitas, sedangkan emosi, minat, dan kondisi kelas ditempatkan sebagai pengembangan lanjutan.

Kontribusi penelitian ini adalah perumusan aktivitas pembelajaran RA sebagai masalah keputusan berurutan yang memetakan *state*, *action*, *reward*, dan transisi untuk mendukung pemilihan aktivitas secara terukur. Perumusan ini mengikuti cara *reinforcement learning* memandang rekomendasi sebagai interaksi berurutan antara kondisi pengguna, tindakan sistem, dan umpan balik yang diterima (Afsar et al., 2023; J. J. Chen & Perez, 2023; Riedmann et al., 2025). Kontribusi tersebut sengaja dibuat konservatif: sistem menghasilkan daftar rekomendasi beserta alasan singkat agar guru dan penguji dapat menelusuri hubungan antara *state*, *action*, *reward*, dan nilai Q, tetapi naskah tidak mengklaim peningkatan perkembangan anak sebelum data operasional dan penilaian guru tersedia. Tujuan penelitian adalah merancang dan mengevaluasi sistem rekomendasi aktivitas pembelajaran anak usia dini berbasis *Q-Learning* dengan batas pada aspek fisik-motorik dan kognitif, karena kedua aspek tersebut dapat direpresentasikan secara eksplisit melalui katalog aktivitas awal dan didukung literatur tentang hubungan aktivitas gerak, perkembangan motorik, dan fungsi kognitif anak (J. J. Chen, 2024; Jylänki et al., 2022; Sutapa et al., 2021).

2. METODOLOGI PENELITIAN

2.1 Kerangka Dasar Penelitian

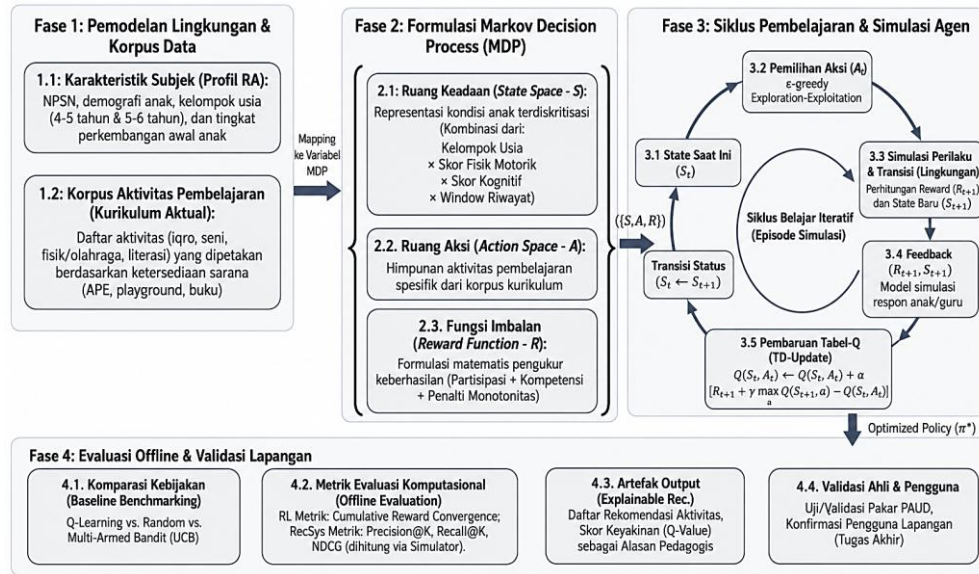
Penelitian ini menggunakan pendekatan rancang bangun sistem dalam bidang rekayasa perangkat lunak dan kecerdasan buatan. Objek penelitian adalah mekanisme rekomendasi aktivitas pembelajaran anak usia dini berbasis *Q-Learning*, sedangkan lingkup penerapan diarahkan pada RA mitra. Artefak yang dibangun meliputi katalog aktivitas, data uji berbasis skenario pembelajaran, lingkungan uji pembelajaran, agen *Q-Learning*, agen pembanding (*baseline*), evaluator peringkat rekomendasi, dan *generator output* artikel. Bukti lingkungan sekolah diambil dari profil digital At-Tayibah yang memuat profil institusi, program pembelajaran, fasilitas, dan alur pendaftaran. Pada tahap ini penelitian belum menggunakan data pribadi anak dari sekolah, sehingga hasil yang diklaim dibatasi pada rancang bangun sistem dan evaluasi komputasional.

Kerangka dasar penelitian menempatkan kebutuhan RA mitra dan literatur *reinforcement learning*, sistem rekomendasi, serta PAUD/RA sebagai dasar perumusan sistem. Masukan sistem berupa *state* anak yang didefinisikan secara diskrit, yaitu usia anak dalam satuan bulan, level fisik-motorik, level kognitif, aktivitas terakhir, dan skor partisipasi terakhir. Proses utama berupa pembelajaran *Q-Learning* untuk memilih aktivitas dari ruang tindakan. Keluaran sistem berupa daftar rekomendasi aktivitas pembelajaran beserta alasan rekomendasi. Dengan struktur tersebut, sistem diposisikan sebagai alat bantu pengambilan keputusan guru, bukan alat diagnosis perkembangan anak.

Tabel 1. Lingkup Aktivitas dan Batas Data Penelitian

Aspek	Temuan dari Profil Digital RA Mitra
Identitas	RA/PAUD swasta di Situraja, Sumedang; dipakai sebagai gambaran lingkungan RA, bukan sebagai sumber data operasional anak.
Batas penggunaan	Identitas personal dan informasi administratif tidak dipakai sebagai variabel model.
Skala lingkup	Katalog aktivitas awal dipakai untuk menyusun ruang tindakan terbatas.
Program utama	Tahfidz & Iqro, Seni & Kreativitas, Bermain & Olahraga, Seni Suara & Nasyid, Calistung, Bahasa Arab & Inggris
Kegiatan rutin	Lingkar literasi, Iqro harian, Hafalan surat pendek, Doa harian
Motorik halus	Mewarnai, Menggunting, Menempel, Meronce, Melipat

Aspek	Temuan dari Profil Digital RA Mitra
Fasilitas	Ruang belajar, <i>playground</i> , APE, dan buku anak sebagai indikasi kelayakan aktivitas.



Gambar 1. Kerangka Penelitian

Kerangka penelitian pada Gambar 1 menunjukkan hubungan antara kebutuhan RA, *state* anak, proses *Q-Learning*, dan keluaran rekomendasi aktivitas. Secara konseptual, penelitian ini memformalkan masalah rekomendasi sebagai *Markov Decision Process*. *State* didefinisikan sebagai kondisi perkembangan dan riwayat aktivitas, *action* adalah dua belas aktivitas pembelajaran, *reward* adalah fungsi penilaian yang menggabungkan partisipasi, kesesuaian perkembangan, dan variasi aktivitas. Relasi antara *state*, *action*, *reward*, dan update *Q-table* menjadi inti kerangka penelitian karena seluruh hasil rekomendasi berasal dari proses pembaruan nilai *Q* pada pasangan *state-action*.

2.2 Tahapan Penelitian

Tahapan penelitian dimulai dari pemetaan kebutuhan sistem, telaah literatur, penyusunan katalog aktivitas, perumusan MDP, penyusunan data uji berbasis skenario pembelajaran, pelatihan agen *Q-Learning*, evaluasi *multi-seed*, analisis peringkat rekomendasi, dan penyusunan naskah sesuai template TIN. Tahap pemetaan kebutuhan dilakukan secara konseptual berdasarkan kebutuhan RA mitra dan literatur, belum menggunakan data pribadi anak. Jika penelitian dilanjutkan ke tugas akhir, tahap berikutnya adalah mengganti skenario pembelajaran dengan data operasional yang telah memperoleh izin sekolah dan persetujuan pihak terkait.

Pada tahap perumusan MDP, *state* didefinisikan sebagai kombinasi usia anak dalam satuan bulan, level fisik-motorik, level kognitif, aktivitas terakhir, dan skor partisipasi terakhir. *State* dibuat diskrit: usia dikelompokkan pada rentang usia RA, level perkembangan memakai kategori rendah, sedang, dan tinggi, aktivitas terakhir berasal dari katalog tertutup, sedangkan partisipasi terakhir dipakai sebagai kategori evaluatif dari skenario pembelajaran. *Action* adalah dua belas aktivitas pembelajaran yang terbagi seimbang antara domain fisik-motorik dan kognitif. Diskritisasi ini dipakai untuk mencegah *state explosion* pada tahap awal dan membuat *Q-table* dapat diaudit. Jika penelitian diperluas ke seluruh aspek perkembangan, riwayat lebih panjang, preferensi anak, dan kondisi kelas, pendekatan *function approximation* atau *DQN* baru perlu dibandingkan sebagai tahap lanjutan. Pembaruan nilai *Q* pada penelitian ini mengikuti Persamaan (1).

$$Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)] \quad (1)$$

Pada Persamaan (1), $Q(s,a)$ menunjukkan nilai tindakan a pada *state* s , α adalah laju pembelajaran, r adalah *reward*, γ adalah faktor diskon, dan $\max_{a'} Q(s',a')$ menunjukkan estimasi nilai terbaik pada *state* berikutnya. *Reward* dihitung dari tiga komponen utama sebagaimana ditunjukkan pada Persamaan (2).

$$R = w_p P + w_f F + w_v V \quad (2)$$

Pada Persamaan (2), P adalah skor partisipasi, F adalah kesesuaian perkembangan, V adalah variasi aktivitas, sedangkan w_p , w_f , dan w_v adalah bobot setiap komponen. Konfigurasi awal memakai $w_p=0,4$, $w_f=0,4$, dan $w_v=0,2$. Partisipasi dan kesesuaian perkembangan diberi bobot sama karena keduanya menentukan apakah aktivitas dapat diikuti dan sesuai dengan kebutuhan perkembangan, sedangkan variasi diberi bobot lebih kecil agar sistem menghindari pengulangan tanpa mengalahkan kesesuaian aktivitas. Bobot tersebut belum diklaim sebagai bobot pedagogis final karena kalibrasi yang kuat memerlukan data observasi guru dan hasil kelas. Prinsip ini mengikuti catatan literatur bahwa evaluasi sistem rekomendasi pendidikan perlu dibedakan dari bukti hasil belajar aktual (Riedmann et al., 2025; Siafis et al., 2024).



Nilai *reward* evaluasi 24.667 dibaca sebagai *reward* kumulatif pada skenario uji dan paling bermakna ketika dibandingkan dengan agen pembanding dalam rancangan yang sama. Dalam konfigurasi evaluasi, *reward* setiap transisi dinormalisasi pada rentang 0 sampai 1 dan evaluasi memakai 30 transisi skenario, sehingga batas maksimum teoritis adalah 30. Dengan demikian, *reward* 24.667 setara dengan sekitar 82,22% dari batas skenario, bukan angka absolut tanpa pembanding rancangan. Karena komponen *reward* dibangun dari partisipasi, kesesuaian perkembangan, dan variasi, nilai tersebut tetap tidak boleh disamakan dengan ukuran pedagogis atau bukti peningkatan perkembangan anak.

Tabel 2. Status Kesiapan Data Penelitian

Komponen	Status	Implikasi Akademik
Profil institusi	Ada	Dipakai sebagai landasan kebutuhan penelitian
Katalog aktivitas sekolah	Ada	Dipakai untuk menyelaraskan ruang tindakan
Fasilitas pendukung aktivitas	Ada	Dipakai untuk kelayakan aktivitas
Data pendaftaran	Ada	Tidak dipakai sebagai data belajar
Riwayat aktivitas anak	Belum ada	Tidak boleh diklaim sebagai data operasional
<i>Reward</i> hasil kegiatan	Belum ada	Masih memakai skenario pembelajaran terkontrol
<i>API</i> rekomendasi	Belum ada	Menjadi target integrasi tugas akhir

Tabel 3. Spesifikasi Metode

Komponen	Nilai
Metode utama	<i>Q-Learning</i>
Lingkup penerapan	RA mitra sebagai dasar kebutuhan, bukan sumber data operasional anak
<i>State</i>	Usia, Level Fisik-Motorik, Level Kognitif, Aktivitas Terakhir, Partisipasi Terakhir
<i>Action</i>	12 aktivitas pembelajaran
<i>Reward</i>	Partisipasi, Kesesuaian Perkembangan, Variasi Aktivitas
<i>Baseline</i>	<i>Random Agent</i> dan <i>Epsilon-Greedy Bandit</i>
Evaluasi	<i>Reward</i> , <i>Precision@K</i> , <i>Recall@K</i> , <i>NDCG@K</i> , Multi-Seed

Tabel 4. Hyperparameter *Q-Learning*

Parameter	Nilai	Alasan
Laju Pembelajaran (α)	0,1	Laju pembelajaran konservatif untuk <i>Q-table</i>
Faktor Diskon (γ)	0,95	Mempertimbangkan <i>reward</i> berurutan
Eksplorasi Awal (ϵ -start)	1,0	Eksplorasi penuh di awal
Eksplorasi Minimum (ϵ -min)	0,02	Eksplorasi kecil setelah <i>training</i>
Langkah Penurunan	4.000	Penurunan eksplorasi bertahap
<i>Train_episodes</i>	240 per <i>seed</i>	Cukup untuk <i>state</i> diskrit kecil
<i>Seed</i>	7, 13, 21, 42, 99	Uji stabilitas hasil

Agan *Q-Learning* dilatih dengan pembaruan *temporal-difference* pada setiap transisi. Pada awal pelatihan, *epsilon* dibuat tinggi agar agen mengeksplorasi beragam aktivitas. Nilai *epsilon* kemudian diturunkan secara bertahap sampai batas minimum, sehingga agen bergerak dari eksplorasi menuju pemanfaatan nilai *Q* yang telah dipelajari. Lima *seed* digunakan untuk memeriksa apakah hasil hanya muncul pada satu kondisi awal atau relatif konsisten pada beberapa urutan random. *Baseline Random Agent* dipakai sebagai batas bawah, sedangkan *Epsilon-Greedy Bandit* dipakai untuk melihat apakah model yang mengabaikan *state* sudah cukup kuat, karena *bandit* dapat menjadi pembanding sederhana untuk keputusan berulang tanpa representasi *state* yang rinci (Afsar et al., 2023; Ricci et al., 2022). Dengan susunan ini, keunggulan *Q-Learning* harus muncul dari pemanfaatan *state*, bukan semata dari eksplorasi.

Relevansi peringkat rekomendasi didefinisikan dari *state* anak. Aktivitas dianggap relevan bila menargetkan domain perkembangan terlemah dan level target aktivitas tidak berbeda lebih dari satu tingkat. Definisi ini dipakai karena penelitian belum menggunakan label preferensi guru. Dengan membuat aturan relevansi eksplisit, evaluasi *Precision@K* dan *Recall@K* dapat direplikasi dan tidak bergantung pada penilaian tersembunyi. Metrik *NDCG@K* ditambahkan karena sistem rekomendasi tidak hanya perlu mencakup aktivitas relevan, tetapi juga menempatkannya pada posisi atas; praktik ini sejalan dengan evaluasi sistem rekomendasi modern yang membedakan ketepatan, cakupan, dan kualitas urutan rekomendasi (Bauer et al., 2024; Ricci et al., 2022). Evaluasi peringkat rekomendasi dihitung menggunakan *Precision@K*, *Recall@K*, dan *NDCG@K* sebagaimana ditunjukkan pada Persamaan (3), Persamaan (4), dan Persamaan (5).

$$Precision@K = |Rel_K| / K \quad (3)$$

$$Recall@K = |Rel_K| / |Rel| \quad (4)$$

$$NDCG@K = DCG@K / IDC@K \quad (5)$$

Pada Persamaan (3)-(5), *Rel_K* adalah aktivitas relevan dalam *K* rekomendasi teratas, *|Rel|* adalah seluruh aktivitas yang relevan untuk *state* anak, dan *IDCG@K* adalah nilai *DCG* ideal pada *K* rekomendasi teratas.

3. HASIL DAN PEMBAHASAN

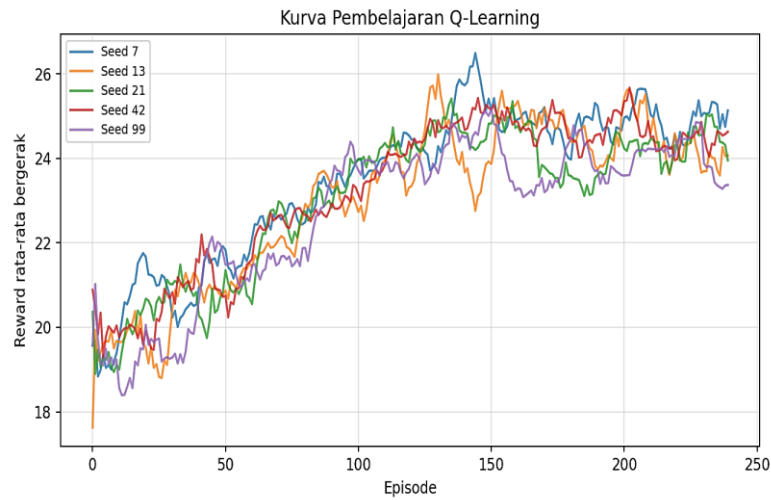
3.1 Hasil

Bagian hasil diawali dengan ringkasan uji *multi-seed* untuk menunjukkan apakah performa agen muncul secara konsisten atau hanya terjadi pada satu kondisi acak. Ringkasan numerik pada Tabel 5 ditempatkan sebelum grafik agar pembaca dapat membaca nilai *reward*, sebaran, dan metrik peringkat terlebih dahulu.

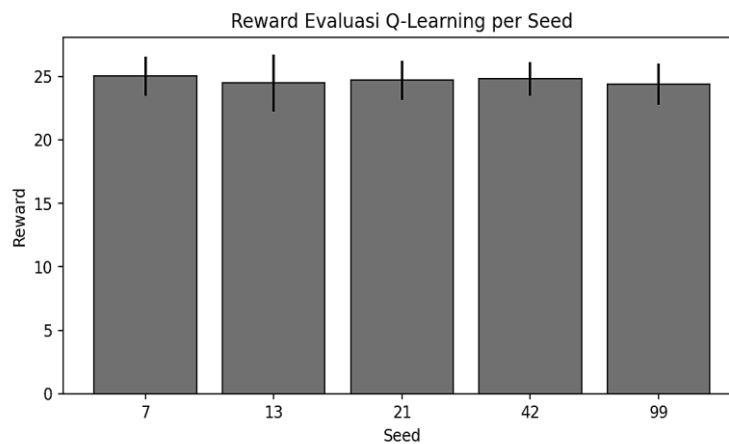
Tabel 5. Hasil Multi-Seed *Q-Learning*

Seed	State	Reward	SD	P@1	R@5	NDCG@5
7.0	117.0	25.011	1.538	0.454	0.383	0.555
13.0	117.0	24.469	2.233	0.726	0.477	0.7
21.0	117.0	24.678	1.536	0.553	0.414	0.593
42.0	117.0	24.785	1.336	0.854	0.491	0.767
99.0	117.0	24.394	1.625	0.636	0.478	0.592

Hasil *multi-seed* pada Tabel 5 memperlihatkan bahwa *reward* evaluasi relatif stabil pada lima *seed*, sedangkan Gambar 2 dan Gambar 3 memperlihatkan pola pembelajaran dan perbandingan *reward* evaluasi antar-*seed*.



Gambar 2. Kurva Pembelajaran *Q-Learning*



Gambar 3. Reward Evaluasi *Q-Learning* per *Seed*

Reward evaluasi rata-rata antar-*seed* adalah 24.667 dengan standar deviasi 0.222. Nilai ini berbeda dari kolom SD pada tabel *seed*, karena kolom tersebut menunjukkan variasi *reward* episode evaluasi dalam masing-masing *seed*. Berdasarkan *reward* evaluasi, *seed* terbaik adalah 7. Namun, untuk metrik peringkat rekomendasi, *seed* 42 menghasilkan *Precision@1* dan *NDCG@5* tertinggi. Rata-rata *reward* antar-*seed* tidak banyak berubah. Artinya, hasil tidak bergantung pada satu *seed* saja. Variasi ini wajar karena eksplorasi awal dan urutan *state-action* yang ditemui agen dapat mempengaruhi isi *Q-table*.

Pengujian *multi-seed* diperlukan karena hasil *reinforcement learning* dapat berubah mengikuti urutan eksplorasi. Pada eksperimen ini, *reward* antar-*seed* berada pada rentang sempit, sehingga hasil tidak bergantung pada satu kondisi

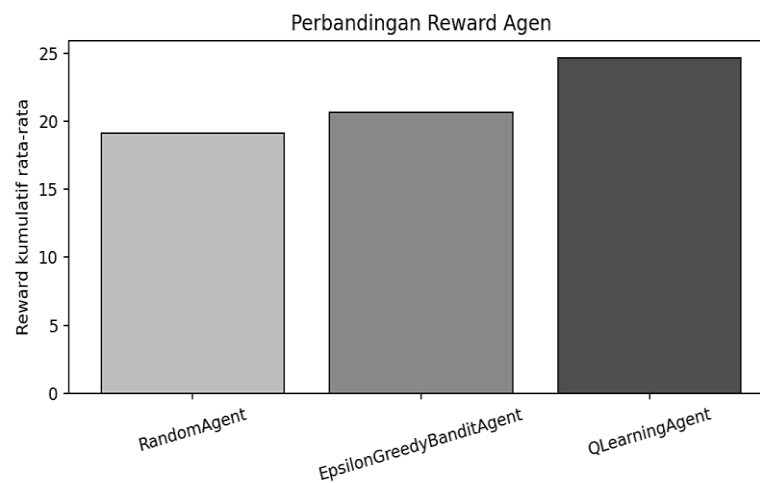
awal. Semua *seed* mengunjungi jumlah *state* yang sama, sehingga perbedaan utama bukan berasal dari kegagalan menjangkau *state*, melainkan dari nilai *Q* yang terbentuk selama pelatihan. Ini memperkuat interpretasi bahwa ruang tindakan saat ini cukup kecil untuk ditangani oleh pendekatan *tabular*.

Setelah stabilitas antar-*seed* dibaca, pembahasan dilanjutkan pada perbandingan agen. Tabel 6 digunakan untuk memperlihatkan posisi *Q-Learning* terhadap *Random Agent* dan *Epsilon-Greedy Bandit*, sedangkan Gambar 4 memudahkan pembacaan selisih *reward* secara visual.

Tabel 6. Perbandingan *Reward* Agen

Agen	Reward	SD	Partisipasi	Kesesuaian	Variasi	Agen
<i>Random</i>	19.086	2.065	0.563	0.567	0.993	<i>Random</i>
<i>Bandit</i>	20.652	2.838	0.666	0.676	0.912	<i>Bandit</i>
<i>Q-Learning</i>	24.667	0.222	-	-	-	<i>Q-Learning</i>

Perbandingan pada Tabel 6 dan Gambar 4 menunjukkan bahwa *Q-Learning* memperoleh *reward* evaluasi lebih tinggi daripada *Random Agent* dan *Epsilon-Greedy Bandit* pada skenario uji yang sama.



Gambar 4. Perbandingan *Reward* Agen

Perbandingan agen menunjukkan bahwa *Q-Learning* lebih proporsional untuk rancangan awal ini daripada *DQN*, dan lebih *state-aware* daripada *rule-based* murni atau *bandit* sederhana. *Rule-based* dapat memilih aktivitas berdasarkan aturan domain, tetapi tidak memperbarui nilai tindakan dari riwayat interaksi. *Epsilon-Greedy Bandit* memperbarui estimasi *reward* tindakan, tetapi tidak menyimpan nilai tindakan secara spesifik untuk setiap *state*. *Q-Learning* berada di tengah: lebih adaptif daripada *rule-based* dan *bandit*, tetapi tetap lebih sederhana dan lebih mudah diaudit daripada *DQN*. Pada *state* diskrit kecil, nilai *Q* dapat ditelusuri langsung, proses pembaruan mudah dijelaskan, dan *output* rekomendasi dapat diberi alasan berdasarkan *state* dan nilai *Q*. Ini penting untuk penelitian RA karena sistem pendukung keputusan guru harus dapat dipertanggungjawabkan, bukan hanya menghasilkan angka performa.

Random Agent memperoleh *reward* yang lebih rendah karena pemilihan aktivitas tidak mempertimbangkan kondisi anak maupun riwayat aktivitas. *Epsilon-Greedy Bandit* lebih baik daripada *Random Agent* karena mampu mengeksplorasi pilihan sekaligus memanfaatkan estimasi *reward* aktivitas, tetapi pendekatan *bandit* dasar tetap menyederhanakan masalah karena menekankan pilihan tindakan dari umpan balik rata-rata dan belum cukup untuk rekomendasi yang berubah mengikuti profil anak serta riwayat aktivitas. Perbedaan ini menjelaskan mengapa hasil *bandit* tetap kompetitif, tetapi belum otomatis menggugurkan *Q-Learning*. Pada penelitian ini, alasan mempertahankan *Q-Learning* bukan hanya *reward* evaluasi, melainkan keterlacakan nilai tindakan pada *state* tertentu.

Hasil ini menjadi dasar untuk menempatkan *DQN* sebagai pembanding lanjutan, bukan model utama pada tahap awal. Pendekatan *deep reinforcement learning* lebih relevan ketika sistem menangani *state* besar, interaksi dinamis, dan representasi fitur yang lebih kompleks (X. Chen et al., 2023). Implementasi penelitian ini masih memakai *state* diskrit kecil dan katalog aktivitas terbatas. Pada kondisi tersebut, *Q-Learning* dipakai sebagai *baseline tabular* yang sengaja dipilih untuk menjaga kesederhanaan, auditabilitas, dan kesesuaian dengan data skenario.

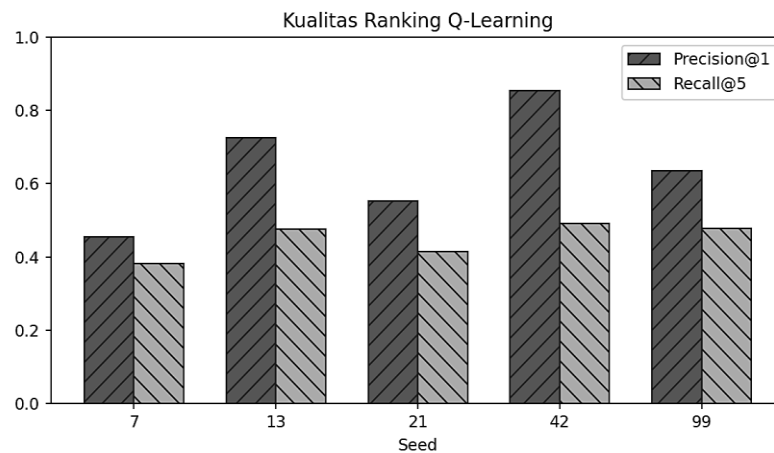
Untuk memperlihatkan bentuk keluaran sistem, Tabel 7 menyajikan contoh rekomendasi pada beberapa kombinasi *state*. Contoh ini penting karena kontribusi artikel bukan hanya nilai *reward*, tetapi juga keterlacakan alasan rekomendasi melalui domain aktivitas, nilai *Q*, dan *reward*.

Tabel 7. Contoh *Output* Rekomendasi *Q-Learning*

Seed	FM	Kog	Aktivitas	Domain	Q	Reward
7	Rendah	Sedang	Berjalan di atas garis lurus	Fisik-motorik	5.061	0.587
7	Sedang	Rendah	Permainan pola bilangan 1-20	Kognitif	7.543	0.986

Seed	FM	Kog	Aktivitas	Domain	Q	Reward
7	Tinggi	Sedang	Permainan pola bilangan 1-20	Kognitif	8.533	0.96
7	Rendah	Rendah	Permainan pola bilangan 1-20	Fisik-motorik	5.982	0.92
7	Sedang	Tinggi	Permainan pola bilangan 1-20	Variasi	6.214	0.88

Contoh keluaran pada Tabel 7 menunjukkan bentuk rekomendasi yang dapat ditelusuri melalui domain aktivitas, nilai Q, dan reward. Setelah contoh keluaran dibaca, kualitas urutan rekomendasi diringkas melalui Gambar 5. Grafik ini dipakai untuk membedakan ketepatan rekomendasi pertama, cakupan aktivitas relevan pada lima rekomendasi, dan kualitas urutan aktivitas relevan.



Gambar 5. Kualitas Peringkat Rekomendasi *Q-Learning*

Kualitas peringkat rekomendasi pada Gambar 5 digunakan untuk membaca hubungan antara *Precision@1*, *Recall@5*, dan *NDCG@5*. Kualitas peringkat rekomendasi menghasilkan *Precision@1* rata-rata 0.645, *Recall@5* rata-rata 0.448, dan *NDCG@5* rata-rata 0.641. *Precision@1* menunjukkan ketepatan rekomendasi utama, sedangkan *Recall@5* menunjukkan seberapa banyak aktivitas relevan yang tercakup dalam lima rekomendasi teratas. *NDCG@5* dipakai untuk melihat apakah aktivitas relevan muncul pada posisi atas. Nilai *Precision@1* tidak dibaca sebagai kegagalan tunggal, karena sistem dirancang untuk memberi daftar prioritas aktivitas, bukan memaksa satu keputusan otomatis. Bagi guru RA, daftar pendek yang dapat ditinjau lebih sesuai daripada satu rekomendasi final.

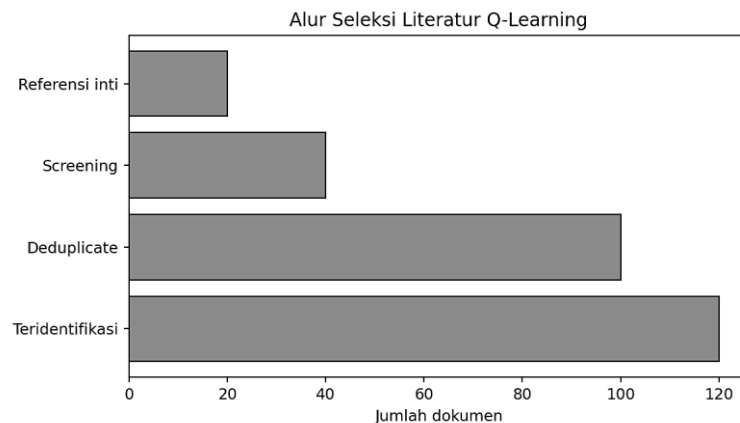
Precision@1 yang belum mendekati satu menunjukkan bahwa rekomendasi utama belum selalu sama dengan aturan relevansi yang dibuat evaluator. Hasil ini menjadi masukan desain, bukan alasan untuk menghapus *Q-Learning*. *Reward* agen menilai partisipasi, kesesuaian perkembangan, dan variasi, sedangkan evaluator peringkat rekomendasi lebih menekankan domain perkembangan terlemah. Perbedaan ini perlu dibuat eksplisit agar pembaca tidak mengira semua metrik mengukur hal yang sama. Jika tahap lanjutan menggunakan label guru, fungsi *reward* dapat dikalibrasi agar lebih dekat dengan preferensi pedagogis dan kebutuhan kelas.

Recall@5 yang lebih tinggi daripada *Recall@1* menunjukkan bahwa daftar pendek rekomendasi lebih berguna daripada satu rekomendasi tunggal. Dalam praktik pembelajaran, sistem rekomendasi lebih tepat diposisikan sebagai pendukung keputusan yang memberi beberapa alternatif terurut, bukan sebagai mekanisme otomatis yang menggantikan pertimbangan guru. Lebih realistis jika sistem menampilkan beberapa aktivitas teratas, lalu guru memilih yang paling sesuai dengan kondisi kelas, alat tersedia, durasi kegiatan, dan suasana anak. Dengan demikian, *output* sistem sebaiknya diposisikan sebagai daftar prioritas aktivitas, bukan keputusan final yang menggantikan guru.

Makna akademik dari hasil komputasional diringkas pada Tabel 8. Ringkasan ini sengaja dipisahkan dari angka mentah agar pembaca dapat melihat mana temuan teknis, mana batas klaim, dan mana implikasi untuk pengembangan sistem. Keterkaitan referensi pendukung dirangkum melalui Gambar 6. Alur ini menunjukkan bahwa literatur dipilih untuk menopang empat kebutuhan utama: RA/PAUD, sistem rekomendasi pendidikan, *reinforcement learning*, dan evaluasi peringkat rekomendasi.

Tabel 8. Interpretasi Akademik Hasil

Temuan	Makna
<i>Reward</i> tinggi	Agen memilih aktivitas yang lebih menguntungkan menurut fungsi <i>reward</i>
<i>Precision@1</i>	Rekomendasi utama sering selaras dengan aturan relevansi
<i>Recall@5</i>	Daftar pendek rekomendasi mencakup lebih banyak aktivitas relevan
<i>NDCG@5</i>	Aktivitas relevan cenderung berada pada posisi atas
<i>Q-table</i>	Keputusan lebih transparan dibanding model <i>deep learning</i>



Gambar 6. Alur Seleksi Literatur

Alur seleksi literatur pada Gambar 6 menjelaskan bagaimana referensi dipilih untuk mendukung landasan PAUD/RA, sistem rekomendasi pendidikan, *reinforcement learning*, dan evaluasi peringkat rekomendasi.

3.2 Pembahasan

Reward, *Precision@1*, *Recall@5*, dan *NDCG@5* dalam penelitian ini perlu dibaca sebagai bukti bahwa mekanisme rekomendasi berjalan konsisten pada *state* dan *action* yang telah didefinisikan, bukan sebagai bukti langsung bahwa aktivitas tertentu meningkatkan perkembangan anak. *Reward* evaluasi rata-rata 24.667 menunjukkan bahwa agen mampu memilih aktivitas yang menguntungkan menurut fungsi *reward* skenario, sedangkan *Precision@1* 0.645, *Recall@5* 0.448, dan *NDCG@5* 0.641 menunjukkan bahwa daftar rekomendasi sudah cukup selaras dengan aturan relevansi internal. Interpretasi ini penting karena sistem rekomendasi untuk anak usia dini harus dibedakan dari sistem otomatisasi keputusan; keluaran sistem seharusnya membantu guru membaca alternatif aktivitas, bukan menggantikan pertimbangan pedagogis guru.

Hasil penelitian ini sejalan dengan literatur sistem rekomendasi pendidikan yang menekankan pentingnya personalisasi, relevansi, dan dukungan keputusan belajar. Namun, berbeda dengan banyak penelitian *adaptive learning* dan *instructional sequencing* yang berfokus pada urutan materi digital atau jalur belajar akademik, penelitian ini menempatkan objek rekomendasi pada aktivitas pembelajaran RA yang melibatkan gerak, alat, instruksi guru, variasi pengalaman, dan kondisi kelas. Perbedaan objek tersebut membuat *reward* tidak cukup hanya mengejar ketepatan *item*, tetapi juga perlu memperhatikan kesesuaian perkembangan, partisipasi, dan variasi aktivitas agar rekomendasi tetap masuk akal dalam praktik pembelajaran anak usia dini.

Dibandingkan dengan penelitian *reinforcement learning* untuk *recommender systems*, rancangan ini lebih sederhana karena masih memakai *Q-Learning tabular*, *state* diskrit, dan ruang tindakan yang terbatas. Kesederhanaan tersebut bukan kelemahan utama pada tahap awal, karena kebutuhan penelitian saat ini adalah menghasilkan *baseline* yang dapat diaudit. Pada model yang lebih kompleks seperti *DQN* atau pendekatan *deep reinforcement learning*, kemampuan menangani *state* besar memang lebih kuat, tetapi proses penjelasan keputusan menjadi lebih berat. Dalam konteks RA, nilai *Q* pada setiap pasangan *state-action* justru memberi keuntungan akademik karena penguji dan calon pengguna dapat menelusuri mengapa sebuah aktivitas direkomendasikan.

Dibandingkan dengan penelitian teknologi dan AI pada pendidikan anak usia dini, artikel ini juga memiliki posisi yang lebih sempit tetapi lebih operasional. Studi terdahulu banyak membahas peluang AI, literasi guru, media digital, asesmen, atau rancangan kurikulum komputasi untuk anak usia dini. Penelitian ini tidak berusaha membuktikan dampak luas AI terhadap perkembangan anak, melainkan mengisi celah pada perumusan pemilihan aktivitas harian sebagai masalah keputusan berurutan yang dapat dihitung, diuji, dan dijelaskan. Dengan demikian, kontribusi penelitian terletak pada cara menghubungkan profil perkembangan, katalog aktivitas, fungsi *reward*, *baseline* perbandingan, dan metrik peringkat dalam satu artefak rekomendasi yang terbatas tetapi dapat direplikasi.

Dengan posisi tersebut, hasil penelitian ini memperkuat literatur terdahulu pada sisi personalisasi dan dukungan keputusan, tetapi juga membatasi klaimnya secara eksplisit. Temuan komputasional menunjukkan bahwa *Q-Learning* layak dipakai sebagai *baseline* awal untuk rekomendasi aktivitas RA karena stabil pada uji *multi-seed*, lebih baik daripada *Random Agent* dan *Epsilon-Greedy Bandit* dalam skenario yang sama, serta menghasilkan rekomendasi yang dapat diberi alasan. Namun, penelitian ini belum dapat menyatakan bahwa sistem meningkatkan capaian perkembangan anak karena belum memakai data operasional kelas, observasi guru, dan validasi pedagogis langsung. Klaim paling kuat adalah bahwa penelitian ini memformalkan aktivitas pembelajaran RA sebagai masalah *state-action-reward* yang transparan dan siap dikembangkan ke tahap validasi operasional.

Implikasi langsung dari perbandingan tersebut adalah perlunya menjaga sistem sebagai alat bantu guru. Untuk masuk ke penggunaan operasional, sistem perlu diuji pada data RA mitra setelah izin etik, izin sekolah, dan persetujuan wali/guru tersedia. Konfirmasi awal dapat diarahkan pada kegunaan sistem, alur rekomendasi, kewajaran istilah aktivitas, dan keterbacaan alasan rekomendasi, bukan pada klaim asesmen perkembangan anak secara klinis.



Implikasi teknisnya adalah proyek perlu menjaga *Q-Learning* sebagai *baseline* utama sampai *state* diperluas. Jika seluruh aspek perkembangan, riwayat lebih panjang, preferensi anak, ketersediaan alat, dan kondisi kelompok dimasukkan, jumlah *state* dapat bertambah besar. Pada titik itu, *DQN* atau pendekatan *function approximation* dapat dibandingkan kembali. Akan tetapi, perbandingan tersebut sebaiknya muncul sebagai pengembangan, bukan sebagai klaim utama artikel ini. Dengan cara ini, artikel tetap jujur terhadap bentuk data sekarang dan tugas akhir masih memiliki ruang pengembangan yang jelas.

Dari perspektif validitas internal, kekuatan penelitian berada pada keterlacakan antara rumus, kode, tabel, dan grafik. Formula *reward* yang ditulis pada metode dipakai langsung oleh lingkungan uji, sedangkan metrik peringkat rekomendasi dihitung dari aturan relevansi yang didefinisikan terbuka. Hal ini membuat penguji dapat memeriksa apakah angka hasil berasal dari proses yang konsisten. Dari perspektif validitas eksternal, kekuatan tersebut belum cukup untuk mewakili kondisi kelas nyata. Anak di kelas dapat memiliki kondisi emosi, minat, kebutuhan alat bantu, dan respons terhadap instruksi yang belum masuk ke *state*. Karena itu, rencana pengembangan tidak boleh hanya menambah algoritma, tetapi juga perlu meningkatkan kualitas data dan kedekatan katalog dengan praktik RA mitra.

Dari perspektif rekayasa perangkat lunak, hasil penelitian juga menunjukkan bahwa *output* dokumen perlu diperlakukan sebagai bagian dari artefak, bukan pekerjaan akhir yang terpisah. Tabel metrik, contoh rekomendasi, gambar alur, dan daftar referensi dibuat ulang dari *generator* yang sama agar angka di naskah tidak terpisah dari hasil eksperimen. Pendekatan ini penting untuk tugas akhir karena revisi judul, perubahan metode, atau perubahan *seed* dapat langsung diikuti oleh pembaruan dokumen. Dengan demikian, kontribusi proyek bukan hanya model *Q-Learning*, tetapi alur reproduktif untuk menghasilkan bukti, naskah, dan dokumen siap diperiksa.

Batasan penelitian dan rencana tindak lanjut dirangkum pada Tabel 9. Tabel ini diperlukan agar pembaca tidak menafsirkan eksperimen skenario sebagai penerapan operasional penuh, sekaligus melihat arah validasi berikutnya.

Tabel 9. Batasan dan Rencana Pengembangan

Batasan	Dampak	Tindak Lanjut
Data uji berbasis skenario	Belum mewakili data kelas nyata	Gunakan data operasional RA mitra setelah izin etik, izin sekolah, dan persetujuan wali/guru tersedia
Dua aspek perkembangan	Belum mencakup seluruh STPPA	Perluas <i>state</i> dan katalog aktivitas bertahap
Aturan relevansi internal	Belum memakai label guru	Tambahkan <i>checklist</i> kebutuhan pengguna
<i>Q-table</i> diskrit	Kurang cocok jika <i>state</i> sangat besar	Bandingkan lagi dengan <i>DQN</i> setelah <i>state</i> diperluas

4. KESIMPULAN

Penelitian ini membangun sistem rekomendasi aktivitas pembelajaran anak usia dini berbasis *Q-Learning* di lingkungan Raudhatul Athfal. *State* sistem memuat usia anak dalam satuan bulan, level fisik-motorik, level kognitif, aktivitas terakhir, dan partisipasi terakhir, sedangkan *action* berupa dua belas aktivitas pembelajaran. Hasil pengujian *multi-seed* menunjukkan *reward* evaluasi rata-rata 24.667 dengan standar deviasi 0.222. Evaluasi peringkat rekomendasi menghasilkan *Precision@1* 0.645, *Recall@5* 0.448, dan *NDCG@5* 0.641. Berdasarkan hasil tersebut, kontribusi utama penelitian bukan klaim dampak pedagogis, melainkan formulasi aktivitas RA sebagai masalah keputusan berurutan yang dapat diaudit melalui *state*, *action*, *reward*, *baseline*, dan nilai *Q*. *Q-Learning* dipertahankan sebagai *baseline tabular* karena *state* masih diskrit, ruang tindakan kecil, dan nilai *Q* lebih mudah diaudit pada tahap data skenario daripada model yang lebih kompleks. Keterbatasan penelitian tetap jelas: data operasional anak belum dipakai, *reward* belum dikalibrasi dengan label guru, dan aspek perkembangan baru dibatasi pada fisik-motorik dan kognitif. Dengan batas ini, sistem menyediakan dasar komputasional yang dapat dipertanggungjawabkan untuk tahap tugas akhir, terutama jika dilanjutkan dengan izin data operasional, validasi guru, kalibrasi *reward*, dan perbandingan ulang dengan pendekatan yang lebih kompleks setelah *state* diperluas. Revisi ini juga menegaskan bahwa kontribusi metodologis harus dinilai dari kejelasan formulasi dan keterbatasan klaim, bukan dari perluasan model yang belum didukung data.

REFERENCES

- Afsar, M. M., Crump, T., & Far, B. (2023). Reinforcement learning based recommender systems: A survey. *ACM Computing Surveys*, 55(7), 1–38. <https://doi.org/10.1145/3543846>
- Bassen, J., Balaji, B., Schaarschmidt, M., Thille, C., Painter, J., Zimmaro, D., Games, A., Fast, E., & Mitchell, J. C. (2020). Reinforcement learning for the adaptive scheduling of educational activities. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12. <https://doi.org/10.1145/3313831.3376518>
- Bauer, C., Zangerle, E., & Said, A. (2024). Exploring the landscape of recommender systems evaluation: Practices and perspectives. *ACM Transactions on Recommender Systems*, 2(1), 1–31. <https://doi.org/10.1145/3629170>
- Chen, J. J. (2024). A scoping study on AI affordances in early childhood education: Mapping the global landscape, identifying research gaps, and charting future research directions. *Journal of Artificial Intelligence Research*, 81, 701–740. <https://doi.org/10.1613/jair.1.16882>



- Chen, J. J., & Lin, J. C. (2024). Artificial intelligence as a double-edged sword: Wielding the POWER principles to maximize its positive effects and minimize its negative effects. *Contemporary Issues in Early Childhood*, 25(1), 146–153. <https://doi.org/10.1177/14639491231169813>
- Chen, J. J., & Perez, C. (2023). Enhancing assessment and personalized learning through artificial intelligence. *Childhood Education*, 99(6), 72–79. <https://doi.org/10.1080/00094056.2023.2282903>
- Chen, X., Yao, L., McAuley, J., Zhou, G., & Wang, X. (2023). Deep reinforcement learning in recommender systems: A survey and new perspectives. *Knowledge-Based Systems*, 264, 110335. <https://doi.org/10.1016/j.knosys.2023.110335>
- da Silva, F. L., Slodkowski, B. K., Da Silva, K. K. A., & Cazella, S. C. (2023). A systematic literature review on educational recommender systems for teaching and learning: Research trends, limitations and opportunities. *Education and Information Technologies*, 28(3), 3289–3328. <https://doi.org/10.1007/s10639-022-11341-9>
- Deschênes, M. (2020). Recommender systems to support learners' agency in a learning context: A systematic review. *International Journal of Educational Technology in Higher Education*, 17(1), 50. <https://doi.org/10.1186/s41239-020-00219-w>
- Jylänki, P., Mbay, T., Hakkarainen, A., Sääkslahti, A., & Aunio, P. (2022). The effects of motor skill and physical activity interventions on preschoolers' cognitive and academic skills: A systematic review. *Preventive Medicine*, 155, 106948. <https://doi.org/10.1016/j.ypmed.2021.106948>
- Liu, R., Luo, F., & Israel, M. (2023). Technology-integrated computing education in early childhood: A systematic literature review. *Journal of Educational Computing Research*, 61(6), 1275–1311. <https://doi.org/10.1177/07356331231170383>
- Ricci, F., Rokach, L., & Shapira, B. (Eds.). (2022). *Recommender systems handbook*. Springer US. <https://doi.org/10.1007/978-1-0716-2197-4>
- Riedmann, A., Schaper, P., & Lugin, B. (2025). Reinforcement learning in education: A systematic literature review. *International Journal of Artificial Intelligence in Education*, 35(5), 2669–2723. <https://doi.org/10.1007/s40593-025-00494-6>
- Siafis, V., Rangoussi, M., & Psaromiligkos, Y. (2024). Recommender systems for teachers: A systematic literature review of recent (2011–2023) research. *Education Sciences*, 14(7), 723. <https://doi.org/10.3390/educsci14070723>
- Sutapa, P., Pratama, K. W., Rosly, M. M., Ali, S. K. S., & Karakauki, M. (2021). Improving motor skills in early childhood through goal-oriented play activity. *Children*, 8(11), 994. <https://doi.org/10.3390/children8110994>
- Urdaneta-Ponte, M. C., Mendez-Zorrilla, A., & Oleagordia-Ruiz, I. (2021). Recommendation systems for education: Systematic review. *Electronics*, 10(14), 1611. <https://doi.org/10.3390/electronics10141611>
- Yang, W. (2022). Artificial intelligence education for young children: Why, what, and how in curriculum design and implementation. *Computers and Education: Artificial Intelligence*, 3, 100061. <https://doi.org/10.1016/j.caeai.2022.100061>
- Yi, H., Liu, T., & Lan, G. (2024). The key artificial intelligence technologies in early childhood education: A review. *Artificial Intelligence Review*, 57(1), 12. <https://doi.org/10.1007/s10462-023-10637-7>