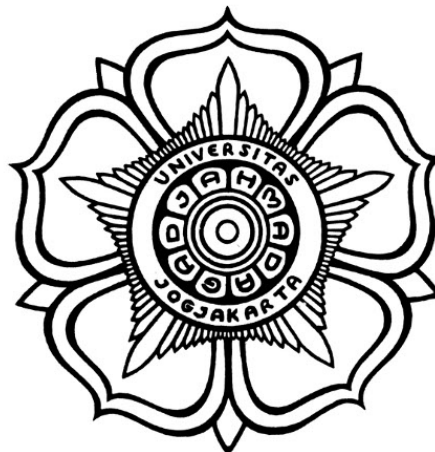


SKRIPSI

**PEMODELAN LAWAN BELAJAR MENGGUNAKAN NONLINEAR
AUTOREGRESSIVE WITH EXOGENOUS INPUT (NARX) PADA
ITERATED PRISONER'S DILEMMA DENGAN HORIZON TERBATAS**

***LEARNING OPPONENT MODELLING USING A NONLINEAR
AUTOREGRESSIVE WITH EXOGENOUS INPUT (NARX) MODEL IN THE
ITERATED PRISONER'S DILEMMA WITH A BOUNDED HORIZON***



FAQIH MAHARDIKA
21/482551/PA/21039

**PROGRAM STUDI ILMU KOMPUTER
DEPARTEMEN ILMU KOMPUTER DAN ELEKTRONIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS GADJAH MADA
YOGYAKARTA**

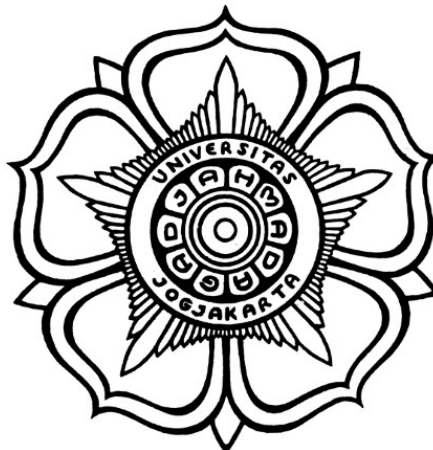
2026

SKRIPSI

**PEMODELAN LAWAN BELAJAR MENGGUNAKAN NONLINEAR
AUTOREGRESSIVE WITH EXOGENOUS INPUT (NARX) PADA
ITERATED PRISONER'S DILEMMA DENGAN HORIZON TERBATAS**

***LEARNING OPPONENT MODELLING USING A NONLINEAR
AUTOREGRESSIVE WITH EXOGENOUS INPUT (NARX) MODEL IN THE
ITERATED PRISONER'S DILEMMA WITH A BOUNDED HORIZON***

Diajukan untuk memenuhi salah satu syarat memperoleh derajat
Sarjana Sains Ilmu Komputer



FAQIH MAHARDIKA
21/482551/PA/21039

**PROGRAM STUDI ILMU KOMPUTER
DEPARTEMEN ILMU KOMPUTER DAN ELEKTRONIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS GADJAH MADA
YOGYAKARTA**

2026

HALAMAN PENGESAHAN

SKRIPSI

PEMODELAN LAWAN BELAJAR MENGGUNAKAN NONLINEAR AUTOREGRESSIVE WITH EXOGENOUS INPUT (NARX) PADA ITERATED PRISONER'S DILEMMA DENGAN HORIZON TERBATAS

Telah dipersiapkan dan disusun oleh

FAQIH MAHARDIKA
21/482551/PA/21039

Telah dipertahankan di depan Tim Penguji
pada tanggal 8 Mei 2026

Susunan Tim Penguji

Dr. Sri Mulyana, M.Kom.
Pembimbing

Bob, M.Sc.
Ketua Penguji

Eve, Ph.D.
Anggota Penguji

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 8 Mei 2026

Faqih Mahardika

DAFTAR ISI

Halaman Judul	ii
Halaman Pengesahan	iii
Halaman Pernyataan	iv
DAFTAR ISI	v
DAFTAR TABEL	vii
DAFTAR GAMBAR	viii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
II TINJAUAN PUSTAKA	6
2.1 Synthesis / Discussion	6
2.1.1 Asumsi perilaku lawan dalam <i>opponent modelling</i> untuk <i>Repeated Games</i>	6
2.1.2 Pendekatan metodologis dalam <i>opponent modelling</i> untuk <i>Repeated Games</i>	8
2.1.3 Bagaimana efektivitas strategi <i>opponent modelling</i> dievaluasi dalam <i>Repeated Games</i> ?	10
2.2 Kesimpulan	13
III LANDASAN TEORI	15
3.1 Dilema Sosial dalam Game Multi-Agen	15
3.1.1 Prisoner's Dilemma	15
3.1.2 Multi-agent system	15
3.2 Game Theory sebagai Model Matematika Interaksi Strategis	15

3.3	Repeated Game dan Ketergantungan Temporal	16
3.3.1	History dependent-strategies dalam Repeated Game	16
3.3.2	Perkembangan lawan	16
3.4	Pemodelan Lawan	16
3.4.1	Pendekatan Reinforcement Learning dalam Repeated Game . .	16
3.5	Opponent Modeling sebagai Pemodelan Time Series	17
3.5.1	Evolusi Model Prediktif: AR, ANN, ARX, dan NARX	17
3.6	Evaluasi Model Prediksi Perilaku Lawan	19
3.6.1	Keterbatasan Prediksi Global pada Horizon Panjang	19
IV	ANALISIS DAN PERANCANGAN	21
V	HASIL DAN PEMBAHASAN	22
VI	KESIMPULAN DAN SARAN	23
	DAFTAR PUSTAKA	24
	LAMPIRAN	26

DAFTAR TABEL

2.1	Asumsi perilaku lawan berdasarkan dependensi perilaku yang dapat diamati.	7
2.2	Perbandingan pendekatan <i>opponent modelling</i> berdasarkan paradigma pemodelan dominan dan mekanisme pembelajaran.	9
2.3	Lingkungan evaluasi dan metrik dalam penelitian	12

DAFTAR GAMBAR

INTISARI

PEMODELAN LAWAN BELAJAR MENGGUNAKAN NONLINEAR AUTOREGRESSIVE WITH EXOGENOUS INPUT (NARX) PADA ITERATED PRISONER'S DILEMMA DENGAN HORIZON TERBATAS

oleh

Faqih Mahardika

21/482551/PA/21039

Kerja sama dan konflik merupakan karakteristik fundamental dari interaksi dalam masyarakat, sistem biologis, dan lingkungan agen-artifisial, di mana agen secara berulang menghadapi pertukaran strategis antara kepentingan diri jangka pendek ataupun hasil kolektif jangka panjang. *Opponent Modelling* memiliki peran dasar dalam konteks tersebut dengan memungkinkan agen untuk menginferensi, mengantisipasi, dan beradaptasi terhadap perilaku pihak lain, dengan pengaplikasiannya mencakup negosiasi, interaksi pasar, hingga sistem kecerdasan buatan multi-agen. Meskipun informatif untuk kinerja jangka panjang, metrik-metrik tersebut umumnya diterapkan pada horizon interaksi yang tidak dibatasi atau cukup panjang, sehingga membatasi pemahaman mengenai efisiensi dan ketepatan waktu dalam proses identifikasi serta adaptasi terhadap lawan secara langsung. Tinjauan ini menemukan adanya kesenjangan struktural dalam praktik evaluasi yang ada dan menekankan perlunya kerangka penilaian yang sadar akan horizon (*horizon-aware*) agar lebih merefleksikan keterbatasan interaksi berulang di dunia nyata.

Cooperation and conflict are fundamental features of interaction in human societies, biological systems, and artificial-agent environments, where agents repeatedly face strategic trade-offs between short-term self-interest and long-term collective outcomes. Opponent modelling plays a central role in such settings by enabling agents to infer, anticipate, and adapt to the behavior of others, with applications ranging from negotiation and market interactions to multi-agent artificial intelligence systems. This literature review synthesizes prior work on opponent modelling in repeated strategic games, using the Iterated Prisoner’s Dilemma as a canonical instantiation of repeated social dilemmas rather than as a restrictive domain. The review analyzes existing approaches along three dimensions: assumptions about opponent behavior, opponent-modelling methodologies, and evaluation practices. The surveyed literature exhibits substantial diversity in opponent behavior assumptions, including stationary, reactive, learning-based, and population-mediated opponents, often embedded implicitly within experimental setups. Methodologically, approaches span gradient-based learning, deep reinforcement learning, recursive belief reasoning, evolutionary dynamics, and system identification. Evaluation practices predominantly emphasize outcome-based metrics such as cooperation rate, average payoff, and equilibrium-related measures. While informative for long-horizon performance, these metrics are typically applied under unconstrained interaction horizons, limiting insight into the efficiency and timeliness of opponent identification and adaptation. This review highlights a structural gap in existing evaluations and underscores the need for horizon-aware assessment frameworks that better reflect the constraints of real-world repeated interactions.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan kecerdasan buatan (*Artificial Intelligence/AI*) dalam beberapa dekade terakhir menunjukkan kemajuan yang signifikan, khususnya dalam bidang pembelajaran mesin dan pembelajaran penguatan (*reinforcement learning*). Pendekatan-pendekatan tersebut memungkinkan agen cerdas untuk mempelajari pengambilan keputusan secara otonom melalui interaksi berulang dengan lingkungan yang bersifat dinamis dan tidak pasti. Salah satu kerangka yang banyak digunakan untuk mengkaji pengambilan keputusan strategis dan interaksi antar agen adalah permainan berulang (*iterated games*), seperti *Iterated Prisoner's Dilemma* (IPD). Meskipun memiliki struktur yang sederhana, IPD mampu merepresentasikan dinamika kerja sama, kompetisi, serta adaptasi strategi yang kompleks.

Dalam skenario IPD, kinerja suatu agen tidak hanya ditentukan oleh kebijakan yang dimilikinya, tetapi juga oleh kemampuannya dalam memahami dan mengantisipasi perilaku lawan. Oleh karena itu, pemodelan lawan (*opponent modelling*) menjadi komponen penting dalam perancangan agen cerdas. Pemodelan lawan bertujuan membangun representasi perilaku atau kecenderungan strategi lawan berdasarkan riwayat interaksi, sehingga agen dapat menyesuaikan tindakannya secara adaptif. Berbagai pendekatan pemodelan lawan telah diusulkan, mulai dari strategi berbasis aturan hingga metode pembelajaran yang memanfaatkan ketergantungan temporal antar aksi.

Dalam kerangka pembelajaran penguatan, pemodelan lawan sering kali diintegrasikan secara langsung ke dalam proses pembelajaran kebijakan agen. Pendekatan terintegrasi ini umumnya menggabungkan proses pengumpulan data, pembaruan model lawan, dan pembelajaran kebijakan dalam satu siklus pembelajaran tertutup. Meskipun efektif dalam beberapa kondisi, pendekatan tersebut menyulitkan analisis terhadap kontribusi masing-masing komponen, khususnya ketika jumlah interaksi terbatas atau data observasi bersifat terbatas. Dalam kondisi tersebut, sulit untuk memisahkan apakah peningkatan kinerja agen disebabkan oleh kualitas pemodelan lawan atau oleh mekanisme pembelajaran kebijakan itu sendiri.

Selain itu, pemodelan lawan dalam permainan berulang juga menghadapi tan-

tangan praktis terkait horizon interaksi. Banyak kajian teoretis mengasumsikan horizon interaksi tak terbatas, sementara dalam praktik simulasi dan eksperimen, interaksi sering kali dibatasi oleh jumlah iterasi tertentu atau bersifat stokastik. Horizon interaksi yang terbatas berdampak pada jumlah data yang tersedia untuk mempelajari perilaku lawan, kestabilan estimasi model, serta reliabilitas evaluasi kinerja agen. Kesalahan prediksi pada fase awal interaksi dapat memberikan pengaruh yang tidak proporsional terhadap hasil keseluruhan permainan.

Berdasarkan kondisi tersebut, diperlukan pendekatan pemodelan lawan yang mampu merepresentasikan ketergantungan temporal perilaku lawan secara eksplisit, sekaligus memungkinkan analisis yang terkontrol terhadap pengaruh kualitas pemodelan tersebut terhadap kinerja agen. Salah satu cara pandang yang relevan adalah memformulasikan pemodelan lawan sebagai permasalahan prediksi deret waktu, di mana perilaku lawan dipelajari dari riwayat interaksi tanpa harus terikat secara langsung pada pembaruan kebijakan agen.

Penelitian ini berangkat dari kebutuhan tersebut dengan mengkaji pemodelan lawan dalam permainan *Iterated Prisoner's Dilemma* pada skenario interaksi dengan horizon terbatas. Fokus penelitian diarahkan pada hubungan antara kualitas pemodelan perilaku lawan, keterbatasan interaksi, dan kinerja agen dalam kerangka pembelajaran berbasis evaluasi. Tinjauan pustaka pada bab selanjutnya akan membahas secara lebih rinci pendekatan-pendekatan pemodelan lawan yang telah ada, asumsi yang digunakan, serta keterbatasannya sebagai dasar perumusan metode penelitian.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, permasalahan yang dikaji dalam penelitian ini dirumuskan sebagai berikut:

1. Bagaimana pengaruh pemisahan (*decoupling*) proses pengumpulan data pemodelan lawan dari pembelajaran kebijakan aktor terhadap kualitas pemodelan lawan dan kinerja agen?
2. Bagaimana perbedaan karakteristik horizon terbatas dengan panjang tetap (*fixed horizon*) dan horizon terbatas yang bersifat stokastik (*stochastic horizon*) memengaruhi stabilitas dan akurasi pemodelan lawan berbasis NARX?
3. Sejauh mana kualitas pemodelan lawan yang dihasilkan oleh *critic* NARX berkorelasi dengan kinerja agen aktor dalam kerangka *actor-critic* pada kondisi

interaksi yang terbatas?

1.3 Batasan Masalah

Agar penelitian ini terfokus dan memiliki ruang lingkup yang jelas, maka ditetapkan beberapa batasan masalah sebagai berikut:

1. Permainan yang dikaji dalam penelitian ini dibatasi pada *Iterated Prisoner's Dilemma* (IPD) dengan struktur payoff standar, tanpa mempertimbangkan variasi payoff atau bentuk permainan sosial lainnya.
2. Interaksi dalam permainan dibatasi pada skenario dua agen (*dyadic interaction*), sehingga interaksi multi-agen atau interaksi kelompok tidak menjadi bagian dari kajian.
3. Pemodelan lawan dilakukan berdasarkan riwayat aksi dalam permainan, dengan asumsi observasi sempurna terhadap aksi lawan, tanpa mempertimbangan mekanisme komunikasi eksplisit, negosiasi, atau pertukaran informasi di luar aksi permainan.
4. Penelitian ini memfokuskan kajian pada perilaku lawan yang bersifat adaptif dan bergantung pada riwayat interaksi, tanpa mengasumsikan kemampuan prediksi sempurna atau perilaku rasional global dari lawan.
5. Evaluasi dilakukan pada interaksi dengan horizon terbatas, tanpa mempertimbangkan skenario horizon tak terbatas atau analisis keseimbangan jangka panjang.
6. Fokus evaluasi dibatasi pada kualitas pemodelan perilaku lawan dan dampaknya terhadap kinerja agen aktor dalam permainan, tanpa membahas aspek optimalitas global, efisiensi komputasi, maupun generalisasi ke domain permainan lain.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Menganalisis kinerja model *Nonlinear AutoRegressive with eXogenous input* (NARX) ketika digunakan sebagai *critic* untuk pemodelan lawan pada permainan *Iterated Prisoner's Dilemma* dengan horizon interaksi terbatas.
2. Mengkaji pengaruh pemisahan (*decoupling*) proses pengumpulan data pemodelan lawan dari pembelajaran kebijakan aktor terhadap kualitas pemodelan lawan.
3. Menganalisis dampak perbedaan karakteristik horizon terbatas, yaitu horizon dengan panjang tetap (*fixed horizon*) dan horizon terbatas yang bersifat stokastik (*stochastic horizon*), terhadap stabilitas dan akurasi pemodelan lawan berbasis NARX.
4. Mengidentifikasi hubungan antara kualitas pemodelan lawan yang dihasilkan oleh *critic* NARX dan kinerja agen aktor dalam kerangka *actor-critic* pada kondisi interaksi yang terbatas.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Memberikan pemahaman empiris mengenai penggunaan model deret waktu berbasis NARX sebagai *critic* dalam pemodelan lawan pada permainan berulang dengan horizon interaksi terbatas.
2. Menyediakan kerangka analisis yang lebih terkontrol untuk mengkaji pengaruh kualitas pemodelan lawan terhadap kinerja agen aktor melalui pendekatan pemisahan (*decoupling*) proses pengumpulan data dan pembelajaran kebijakan.
3. Menjadi referensi bagi penelitian selanjutnya yang mengkaji pemodelan lawan atau pembelajaran multi-agen pada kondisi data dan interaksi yang terbatas, khususnya dalam konteks permainan sosial sederhana.
4. Memberikan kontribusi metodologis dalam perancangan dan evaluasi agen berbasis *actor-critic*, khususnya terkait pemanfaatan model eksplisit sebagai komponen evaluasi (*critic*).

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun sebagai berikut:

- Bab I Pendahuluan, berisi latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.
- Bab II Tinjauan Pustaka, membahas penelitian-penelitian terdahulu yang relevan dengan permainan *Iterated Prisoner's Dilemma*, pemodelan lawan (*opponent modelling*), serta pendekatan pemodelan perilaku lawan dalam kerangka pembelajaran multi-agen.
- Bab III Landasan Teori, menjelaskan konsep dan teori yang mendasari penelitian, meliputi permainan *Iterated Prisoner's Dilemma*, pemodelan lawan, kerangka *actor-critic*, serta model *Nonlinear AutoRegressive with eXogenous input* (NARX).
- Bab IV Metodologi Penelitian, memaparkan perancangan metode penelitian, termasuk skema pengumpulan data secara terpisah (*decoupled*), perancangan *critic* berbasis NARX, skenario horizon terbatas, serta prosedur evaluasi yang digunakan.
- Bab V Hasil dan Pembahasan, menyajikan hasil eksperimen yang diperoleh serta pembahasan mengenai kinerja pemodelan lawan dan dampaknya terhadap kinerja agen aktor.
- Bab VI Kesimpulan dan Saran, berisi kesimpulan yang diperoleh dari hasil penelitian serta saran untuk pengembangan penelitian selanjutnya.

BAB II

TINJAUAN PUSTAKA

2.1 Synthesis / Discussion

Bagian ini mensintesis temuan-temuan utama dari studi yang ditinjau dengan tujuan mengidentifikasi pola asumsi perilaku lawan, keterbatasan metodologis yang berulang, serta celah penelitian yang masih terbuka dalam *opponent modelling* untuk *repeated games*. Sintesis difokuskan pada dimensi perilaku yang dapat diamati dari riwayat interaksi, alih-alih pada asumsi internal seperti tujuan optimisasi, struktur pembaruan parameter, atau representasi kebijakan lawan yang sering kali tidak dapat diakses secara langsung oleh agen.

Meskipun berbagai atribut telah diidentifikasi dalam studi-studi yang disertakan, hanya dimensi yang bersifat diskriminatif secara metodologis yang digunakan untuk perbandingan lintas karya. Atribut terkait lingkungan permainan dan protokol evaluasi dibahas secara terpisah untuk menghindari pencampuran antara asumsi perilaku dan pengaturan eksperimental.

2.1.1 Asumsi perilaku lawan dalam *opponent modelling* untuk *Repeated Games*

Untuk mengoperasionalkan asumsi perilaku lawan secara konsisten pada pengaturan permainan yang beragam, tinjauan ini mengabstraksikan properti perilaku yang dapat diinferensi dari *interaction traces*. Secara khusus, perilaku lawan dikarakterisasi berdasarkan empat jenis dependensi yang dapat diamati: apakah aksi lawan (i) bervariasi lintas kondisi lingkungan dalam permainan yang sama, (ii) dimediasi oleh dinamika pada tingkat populasi, (iii) merespons secara langsung aksi agen, dan (iv) menunjukkan divergensi aksi pada riwayat interaksi terkini yang ekuivalen.

Kriteria-kriteria ini dievaluasi sepenuhnya pada tingkat perilaku eksternal dan tidak mengasumsikan adanya pengetahuan mengenai model internal lawan, mekanisme pembelajaran, maupun tujuan strategis yang dioptimalkan. Dengan demikian, kategorisasi yang digunakan tidak dimaksudkan sebagai taksonomi formal dari metode *opponent modelling*, melainkan sebagai kerangka analitis untuk memungkinkan perbandingan lintas studi yang menggunakan paradigma pembelajaran, representasi strategi, dan abstraksi permainan yang berbeda.

Dalam studi yang tidak mendefinisikan asumsi perilaku lawan secara eksplisit,

klasifikasi diturunkan secara konservatif berdasarkan pengaturan eksperimental dan dinamika interaksi yang dilaporkan. Tabel 2.1 merangkum hasil kategorisasi tersebut beserta referensi terkait.

Tabel 2.1 Asumsi perilaku lawan berdasarkan dependensi perilaku yang dapat diamati.

Env.	Pop.	AgentSy	Div.	Kategori Perilaku Lawan	Ref.
—	—	✓	—	Reactive	Jin et al. 2025
—	✓	—	—	Population-Conformist	Gómez et al. 2025
—	✓	✓	—	Contextual Reactive	Elhamer et al. 2020
—	—	✓	✓	Learning Opponent	Qiao et al. 2024; Lv et al. 2023; Li et al. 2025; Freire et al. 2023; Hu et al. 2023; Wang et al. 2019; De Weerd et al. 2022
—	✓	✓	✓	Population-Contextual Strategic	Perera et al. 2025
✓	✓	—	✓	Heterogeneous Collective Behavior	Zhu et al. 2025
✓	✓	✓	✓	Environment-Conditioned Strategic	Di et al. 2023

Catatan:

Env. — perilaku bervariasi lintas lingkungan dalam permainan yang sama;

Pop. — perilaku dimediasi oleh interaksi tingkat populasi;

Agent — perilaku merespons secara langsung aksi agen;

Div. — divergensi aksi terjadi pada riwayat interaksi terkini yang ekuivalen.

Tanda centang menunjukkan adanya dependensi.

Sejumlah pola konsisten muncul dari Tabel 2.1. Pertama, mayoritas studi terkini mengasumsikan lawan yang aksinya responsif terhadap agen dan menunjukkan divergensi perilaku pada riwayat interaksi yang setara Qiao et al. 2024; Lv et al. 2023; Li et al. 2025; Freire et al. 2023; Hu et al. 2023; Wang et al. 2019; De Weerd et al.

2022; Perera et al. 2025; Di et al. 2023. Pola ini mengindikasikan pergeseran fokus riset dari optimisasi terhadap strategi lawan yang tetap menuju ketahanan dan adaptasi terhadap lawan yang belajar atau berperilaku strategis secara dinamis.

Kedua, dependensi perilaku yang dimediasi populasi terutama muncul dalam studi pada lingkungan evolusioner atau berbasis jaringan, di mana aksi individu dipengaruhi secara tidak langsung oleh dinamika agregat populasi Zhu et al. 2025; Gómez et al. 2025; Elhamer et al. 2020; Di et al. 2023; Perera et al. 2025. Dalam pengaturan ini, responsivitas terhadap agen sering kali terpisah dari perubahan strategi pada tingkat populasi, menghasilkan dinamika yang berbeda secara kualitatif dibandingkan skenario pembelajaran dua pemain.

Terakhir, hanya sebagian kecil karya yang secara simultan memodelkan perilaku lawan yang bergantung pada lingkungan, dimediasi populasi, dan responsif terhadap aksi agen Di et al. 2023. Keterbatasan ini menunjukkan bahwa lawan strategis yang sepenuhnya terkondisi oleh konteks interaksi—baik pada tingkat individu maupun kolektif—masih relatif kurang dieksplorasi, terutama dalam pengaturan *repeated games* dengan riwayat interaksi yang panjang dan tidak stasioner.

2.1.2 Pendekatan metodologis dalam *opponent modelling* untuk *Repeated Games*

Distribusi asumsi perilaku lawan pada Tabel 2.1 mencerminkan pergeseran metodologis yang lebih luas dalam riset *opponent modelling*, dari pengaturan dengan asumsi lawan yang tetap dan terkontrol menuju lawan yang adaptif, heterogen, dan berperilaku tidak stasioner. Namun, karakterisasi berbasis asumsi perilaku semata belum menjelaskan bagaimana kompleksitas tersebut dihadapi secara komputasional di sisi agen. Untuk itu, Tabel 2.2 mereorganisasi studi-studi terdahulu berdasarkan paradigma pemodelan dominan dan mekanisme pembelajaran yang digunakan oleh agen.

Pendekatan-pendekatan tersebut dapat dibedakan lebih lanjut berdasarkan apakah adaptasi terhadap lawan ditangani secara eksplisit atau implisit. Paradigma *opponent modelling* yang eksplisit—seperti *gradient-based opponent shaping* Qiao et al. 2024; Hu et al. 2023; Wang et al. 2019, *recursive belief reasoning* Freire et al. 2023; De Weerd et al. 2022, serta *system identification* Li et al. 2025—secara langsung membangun representasi internal perilaku lawan untuk memprediksi atau memengaruhi respons lawan di masa depan. Sebaliknya, pendekatan implisit, termasuk

Tabel 2.2 Perbandingan pendekatan *opponent modelling* berdasarkan paradigma pemodelan dominan dan mekanisme pembelajaran.

Paradigma Pemodelan	Mekanisme Pembelajaran / Pembaruan	Makalah
Reactive reinforcement learning	RL berbasis nilai (DQN)	Lv et al. 2023
Gradient-based opponent shaping	<i>Gradient descent</i> / pembaruan sadar lawan	Qiao et al. 2024; Hu et al. 2023; Wang et al. 2019
Recursive belief reasoning	Pembaruan keyakinan Bayesian / <i>cognitive hierarchy</i>	Freire et al. 2023; De Weerd et al. 2022
Population-based training	<i>Policy-gradient</i> MARL (sampling populasi)	Perera et al. 2025
Evolutionary population dynamics	Imitasi strategi berbasis <i>fitness</i>	Zhu et al. 2025; Di et al. 2023; Elhamer et al. 2020
Bandit-based learning-in-games	<i>Multi-armed bandit</i> / <i>smooth best response</i>	Gómez et al. 2025
System identification	Model autoregresif dengan input eksogen (NARX)	Li et al. 2025
Communication-driven coordination	<i>Policy-gradient</i> dengan objektif komunikasi	Jin et al. 2025

reactive reinforcement learning Lv et al. 2023, *population-based training* Perera et al. 2025, dinamika evolusioner Zhu et al. 2025; Di et al. 2023; Elhamer et al. 2020, dan *bandit-based learning-in-games* Gómez et al. 2025, beradaptasi terhadap lawan tanpa mempertahankan model perilaku lawan yang terpisah.

Meskipun pendekatan berbasis *reinforcement learning* dan *policy-gradient* mendominasi literatur terkini, paradigma tersebut umumnya menggabungkan pemodelan lawan ke dalam proses optimisasi kebijakan agen. Akibatnya, dinamika perilaku lawan sering kali terenkapsulasi secara implisit dalam parameter kebijakan atau fungsi nilai, sehingga sulit untuk mengisolasi, menginterpretasi, atau memanfaatkan prediksi eksplisit terhadap respons lawan, terutama pada horizon interaksi yang panjang dan tidak stasioner.

Dalam konteks ini, pendekatan *system identification*, khususnya model autoregresif dengan input eksogen (NARX), menawarkan alternatif metodologis yang berbeda. Alih-alih mengasumsikan struktur pembelajaran atau tujuan optimisasi lawan, NARX memodelkan perilaku lawan sebagai proses dinamis yang dapat diinferensi langsung dari riwayat interaksi. Dengan memanfaatkan dependensi temporal dan aksi agen sebagai sinyal eksogen, pendekatan ini secara alami selaras dengan asumsi perilaku lawan yang responsif dan menunjukkan divergensi aksi pada riwayat interaksi yang ekuivalen, sebagaimana diidentifikasi pada Tabel 2.1.

Selain itu, NARX memungkinkan pemisahan yang jelas antara proses prediksi perilaku lawan dan mekanisme pengambilan keputusan agen. Pemisahan ini memberikan fleksibilitas metodologis untuk menganalisis kualitas prediksi lawan secara independen dari kebijakan agen, serta memungkinkan integrasi dengan berbagai skema pengambilan keputusan tanpa memerlukan pelatihan ulang berbasis interaksi penuh seperti pada *reinforcement learning*. Karakteristik ini menjadikan pendekatan berbasis NARX secara khusus menarik pada pengaturan dengan keterbatasan data, sumber daya komputasi, atau horizon waktu penelitian, sekaligus tetap mempertahankan kemampuan untuk menangkap dinamika perilaku lawan yang tidak stasioner dalam *repeated games*.

2.1.3 Bagaimana efektivitas strategi *opponent modelling* dievaluasi dalam *Repeated Games*?

Praktik evaluasi secara implisit mendefinisikan apa yang dianggap sebagai keberhasilan dalam interaksi multi-agent yang adaptif, baik dalam bentuk hasil kerja sa-

ma, ketahanan terhadap eksploitasi, stabilitas perilaku, maupun akurasi prediksi. Tabel 2.3 merangkum lingkungan evaluasi dan metrik yang digunakan dalam penelitian-penelitian terdahulu, dengan tujuan mengidentifikasi pola evaluasi yang berulang serta aspek-aspek yang relatif terabaikan, alih-alih menetapkan standar normatif atau pemeringkatan kinerja.

Seperti terlihat pada Tabel 2.3, sebagian besar studi mengevaluasi efektivitas *opponent modelling* melalui metrik kinerja agregat jangka panjang, khususnya tingkat kerja sama Di et al. 2023; Elhamer et al. 2020; Wang et al. 2019; Jin et al. 2025, payoff rata-rata Lv et al. 2023; Perera et al. 2025; Li et al. 2025; Wang et al. 2019; De Weerd et al. 2022, serta konvergensi menuju equilibrium atau solusi stabil Gómez et al. 2025; Hu et al. 2023; Jin et al. 2025. Metrik-metrik ini secara inheren mengasumsikan interaksi berulang dengan horizon panjang, di mana kerugian eksplorasi pada tahap awal dapat dikompensasikan oleh perbaikan kinerja pada fase selanjutnya.

Namun demikian, asumsi horizon panjang ini membatasi daya representasi evaluasi terhadap skenario di mana interaksi bersifat terbatas, biaya eksplorasi signifikan, atau kesalahan awal sulit dipulihkan. Bahkan dalam studi yang mempertimbangkan lawan adaptif atau tidak stasioner, evaluasi umumnya dilakukan setelah fase pembelajaran mencapai stabilitas atau konvergensi Qiao et al. 2024, sehingga kinerja selama fase identifikasi lawan secara *online* relatif kurang diperhatikan.

Keterbatasan ini menjadi semakin relevan dalam pengaturan *repeated games* dengan horizon tetap yang pendek atau tidak pasti, di mana agen tidak dapat mengandalkan eksplorasi agresif tanpa risiko penurunan kinerja yang substansial. Dalam konteks tersebut, strategi eksplorasi yang terlalu invasif dapat menyebabkan salah koordinasi permanen, eksploitasi oleh lawan, atau kegagalan mencapai kerja sama sebelum interaksi berakhir. Oleh karena itu, evaluasi berbasis horizon panjang cenderung melebihkan keuntungan metode yang mengandalkan eksplorasi mendalam, sementara meremehkan pendekatan yang menekankan kehati-hatian dan efisiensi identifikasi perilaku lawan.

Sebagai respons terhadap celah ini, penggunaan horizon tetap yang pendek atau horizon stokastik dapat dipandang sebagai pilihan evaluasi yang lebih konservatif dan informatif. Horizon semacam ini secara eksplisit membatasi anggaran eksplorasi dan memaksa agen untuk menyeimbangkan antara identifikasi perilaku lawan dan kinerja langsung sejak tahap awal interaksi. Selain itu, horizon stokastik mengurangi insentif bagi strategi yang bergantung pada eksploitasi fase akhir permainan, sehingga mendorong perilaku yang lebih stabil dan berorientasi jangka pendek.

Tabel 2.3 Lingkungan evaluasi dan metrik dalam penelitian

Ref.	Lingkungan Evaluasi	Metrik
Qiao et al. 2024	Self-play simetris	MSE selama pelatihan offline; akurasi memori laten
Zhu et al. 2025	Jaringan scale-free dengan strategi zero-determinant	Frekuensi kerja sama (C) dan eksploitasi (E)
Lv et al. 2023	Opponent adaptif (dirata-ratakan pada beberapa opponent)	Nilai reward
Gómez et al. 2025	Simulator teamwork-game khusus (aggregative public good games); eksperimen sintetis	Produktivitas tim agregat; uji kecocokan χ^2 terhadap equilibrium; konvergensi ke Nash equilibrium; kontribusi individu
Di et al. 2023	Simulator evolutionary game pada jaringan terstruktur	Tingkat kerja sama; fraksi kooperator; ambang fase transisi
Perera et al. 2025	Repeated matrix games dengan populasi opponent sintetis	Payoff rata-rata; tingkat kerja sama; robustness terhadap himpunan opponent; generalisasi
Li et al. 2025	Repeated zero-sum games melawan Hedge, OMD, dan Regret Matching	Galat prediksi; payoff kumulatif; robustness terhadap non-stationarity
Freire et al. 2023	Repeated matrix games; simulasi robotik embodied waktu-kontinu	Efektivitas; stabilitas; akurasi prediksi
Hu et al. 2023	Simulasi repeated matrix game	Payoff rata-rata; kecepatan konvergensi; pemilihan equilibrium
Elhamer et al. 2020	Simulasi continuous-space skala besar (FLAME GPU)	Tingkat kerja sama; ukuran dan jumlah kluster kooperatif; kecepatan agen; stabilitas kluster
Wang et al. 2019	Lingkungan SPD 2D khusus (Fruit Gathering; Apple—Pear games)	Reward individu rata-rata; total kesejahteraan sosial; akurasi deteksi derajat kerja sama
Jin et al. 2025	Benchmark MARL (Cleanup, Harvest, Sequential PD, Tragedy of the Commons) 12	Return ternormalisasi; tingkat kerja sama; kecepatan konvergensi; perbedaan kebijakan (MSE)
De Weerd et al. 2022	Simulasi Colored Trails dengan peningkatan ketidakpastian lingkungan	Skor allocator; skor responder; total kesejahteraan sosial

Dalam konteks *opponent modelling* berbasis prediksi eksplisit, evaluasi dengan horizon terbatas juga memungkinkan analisis yang lebih tajam terhadap kegunaan prediksi perilaku lawan. Alih-alih menilai keberhasilan hanya berdasarkan hasil agregat jangka panjang, pengaturan ini menyoroti seberapa cepat dan seberapa akurat model lawan dapat memberikan informasi yang berguna untuk pengambilan keputusan, serta sejauh mana agen mampu memanfaatkan prediksi tersebut tanpa melakukan eksplorasi yang berlebihan. Dengan demikian, praktik evaluasi ini memberikan perspektif pelengkap terhadap literatur yang ada, khususnya dalam menilai efisiensi dan kehati-hatian strategi *opponent modelling* pada pengaturan interaksi yang terbatas.

2.2 Kesimpulan

Bab ini mensintesis literatur *opponent modelling* dalam interaksi strategis berulang dengan menelaah tiga dimensi utama: asumsi perilaku lawan yang mendasari, paradigma pemodelan yang digunakan oleh agen, serta praktik evaluasi yang menentukan kriteria keberhasilan. Tinjauan difokuskan pada pengaturan *repeated games*, khususnya dilema sosial, di mana ketergantungan pada riwayat interaksi dan adaptasi perilaku lawan menjadi faktor sentral.

Analisis terhadap asumsi perilaku menunjukkan bahwa literatur terkini semakin menekankan lawan yang adaptif, responsif terhadap aksi agen, dan menunjukkan divergensi perilaku meskipun berada pada riwayat interaksi yang ekuivalen. Namun demikian, asumsi-asumsi tersebut sering kali tidak dinyatakan secara eksplisit dan tertanam secara implisit dalam desain eksperimen atau mekanisme pembelajaran. Akibatnya, pendekatan pemodelan yang tampak serupa secara metodologis dapat beroperasi di bawah asumsi perilaku lawan yang berbeda secara fundamental, sehingga menyulitkan perbandingan lintas studi dan interpretasi hasil secara konsisten.

Dari sisi metodologis, literatur menunjukkan keberagaman paradigma *opponent modelling*, mulai dari pendekatan implisit berbasis optimisasi kebijakan dan dinamika populasi hingga pendekatan eksplisit yang mempertahankan representasi internal perilaku lawan. Meskipun pendekatan berbasis *reinforcement learning* dan *policy-gradient* mendominasi, dinamika perilaku lawan pada paradigma tersebut sering kali terenkapsulasi secara implisit dalam parameter kebijakan agen. Hal ini membatasi kemampuan untuk mengisolasi, mengevaluasi, dan memanfaatkan prediksi perilaku lawan secara langsung, terutama dalam pengaturan interaksi yang tidak stasioner.

Praktik evaluasi yang ada sebagian besar berfokus pada metrik kinerja agregat jangka panjang, seperti tingkat kerja sama, payoff rata-rata, dan konvergensi menuju equilibrium. Meskipun metrik-metrik ini relevan untuk menilai stabilitas pada horizon interaksi yang panjang, penerapannya secara luas mengasumsikan bahwa biaya eksplorasi dapat diabaikan dan kesalahan awal dapat dipulihkan seiring waktu. Akibatnya, skenario dengan horizon interaksi terbatas, eksplorasi yang mahal, serta sensitivitas tinggi terhadap keputusan awal masih relatif kurang terwakili dalam evaluasi yang ada.

Berdasarkan sintesis tersebut, dapat diidentifikasi celah penelitian pada *opponent modelling* dalam pengaturan *repeated games* yang bersifat *online*, tidak stasioner, dan dibatasi oleh horizon interaksi yang pendek atau tidak pasti. Secara khusus, literatur masih kurang mengeksplorasi pendekatan yang memungkinkan identifikasi perilaku lawan dalam horizon yang pendek dan eksplisit dari riwayat interaksi, serta evaluasi yang menyoroti trade-off antara eksplorasi dan kinerja langsung pada fase awal interaksi. Celah ini membuka ruang bagi pendekatan metodologis yang berfokus pada pemodelan perilaku lawan berbasis observasi historis dan pengujian yang sensitif terhadap keterbatasan interaksi.

BAB III

LANDASAN TEORI

3.1 Dilema Sosial dalam Game Multi-Agen

Dilema sosial merupakan situasi di mana kepentingan individu tidak selalu sejalan dengan kepentingan kolektif. Dalam konteks game multi-agen, setiap agen mengambil keputusan yang dapat memengaruhi hasil yang diterima oleh agen lain.

Fenomena dilema sosial banyak ditemukan dalam game berulang, di mana interaksi antar agen bersifat strategis dan bergantung pada perilaku lawan.

3.1.1 Prisoner's Dilemma

Prisoners' dilemma merupakan contoh klasik dari dilema sosial dalam teori permainan. Dalam permainan ini, dua tahanan dihadapkan pada pilihan untuk bekerja sama atau mengkhianati satu sama lain. Jika kedua tahanan bekerja sama, mereka menerima hukuman ringan. Namun, jika salah satu mengkhianati yang lain sementara yang lain tetap diam, pengkhianat akan dibebaskan sementara yang diam menerima hukuman berat. Jika keduanya mengkhianati, keduanya menerima hukuman sedang. Meskipun kerjasama memberikan hasil terbaik secara kolektif, rasionalitas individu sering kali mendorong mereka untuk mengkhianati, menciptakan dilema sosial.

3.1.2 Multi-agent system

Dalam sistem multi-agen, sejumlah agen otonom berinteraksi dalam lingkungan yang sama. Setiap agen memiliki tujuan dan strategi sendiri, yang dapat saling mempengaruhi. Interaksi ini sering kali melibatkan dilema sosial, di mana keputusan satu agen dapat berdampak pada hasil yang diterima oleh agen lain. Studi tentang sistem multi-agen mencakup analisis perilaku kolektif, koordinasi, dan kompetisi antar agen dalam konteks game teori.

3.2 Game Theory sebagai Model Matematika Interaksi Strategis

Oleh karena itu, dilema sosial dalam game dimodelkan menggunakan kerangka matematika *game theory*. Game theory menyediakan representasi formal terhadap agen, ruang aksi, serta struktur payoff yang mendefinisikan hasil interaksi strategis.

Pendekatan ini memungkinkan analisis perilaku agen secara sistematis berdasarkan aturan permainan yang didefinisikan.

3.3 Repeated Game dan Ketergantungan Temporal

Namun, banyak game dengan dilema sosial tidak berlangsung sebagai interaksi satu kali, melainkan sebagai interaksi berulang. Dalam repeated game, keputusan agen pada satu iterasi dipengaruhi oleh hasil interaksi sebelumnya.

Oleh karena itu, konsep *repeated game* digunakan untuk merepresentasikan dinamika strategi yang berkembang seiring waktu dan bergantung pada riwayat interaksi.

3.3.1 History dependent-strategies dalam Repeated Game

Dalam repeated game, strategi yang diadopsi oleh agen sering kali bergantung pada riwayat tindakan sebelumnya. Strategi semacam ini dikenal sebagai *history-dependent strategies*. Contohnya termasuk strategi *tit-for-tat*, *Grim*, dll.

3.3.2 Perkembangan lawan

lawan sederhana sampai kompleks, menentukan strategi tidak lagi hanya dengan memperhatikan tindakan sebelumnya, tetapi juga dengan memahami perkembangan perilaku lawan seiring waktu. contoh hmd, HDO, memberikan kompleksitas lebih tinggi dalam permainan berulang.

3.4 Pemodelan Lawan

Dalam konteks repeated game, pemodelan lawan (opponent modeling) merupakan aspek penting untuk memahami dan memprediksi perilaku agen lain. Pendekatan ini melibatkan analisis strategi lawan berdasarkan observasi tindakan mereka dalam interaksi berulang. Dengan demikian, agen dapat menyesuaikan strateginya untuk mencapai hasil yang lebih baik.

3.4.1 Pendekatan Reinforcement Learning dalam Repeated Game

Meskipun repeated game menjelaskan dinamika interaksi, pendekatan analitis klasik sering kali tidak memadai untuk menangani agen yang adaptif.

Oleh karena itu, *reinforcement learning* (RL) dan *deep reinforcement learning* (DRL) banyak digunakan untuk mempelajari strategi optimal dalam repeated game.

Pendekatan ini memungkinkan agen menyesuaikan kebijakannya berdasarkan pengalaman interaksi.

Keterbatasan Reinforcement Learning untuk Opponent Modeling Namun, reinforcement learning pada umumnya berfokus pada optimasi kebijakan agen berdasarkan sinyal reward, bukan pada pemodelan eksplisit perilaku lawan.

Selain itu, pendekatan ini sering memerlukan horizon jangka panjang, jumlah episode yang besar, serta sumber daya komputasi yang signifikan. Kondisi ini menjadi keterbatasan ketika tujuan utama penelitian adalah prediksi perilaku lawan.

3.5 Opponent Modeling sebagai Pemodelan Time Series

Dengan demikian, perilaku lawan dalam repeated game lebih tepat dipandang sebagai proses sekuensial. Tindakan lawan pada suatu waktu dipengaruhi oleh riwayat tindakan sebelumnya serta tindakan agen lain.

Pendekatan *time series* memungkinkan pemodelan ketergantungan temporal tersebut secara eksplisit.

3.5.1 Evolusi Model Prediktif: AR, ANN, ARX, dan NARX

Pemodelan perilaku lawan dalam permainan berulang dapat dipandang sebagai permasalahan prediksi deret waktu, di mana aksi lawan pada waktu tertentu bergantung pada riwayat interaksi sebelumnya. Pendekatan paling dasar untuk memodelkan ketergantungan temporal tersebut adalah model autoregresi (Autoregressive Model, AR), yang merepresentasikan keluaran sistem sebagai fungsi linier dari keluaran masa lalu.

Dalam konteks opponent modelling, model AR digunakan untuk memprediksi aksi lawan berdasarkan riwayat aksinya sendiri. Model ini memiliki keunggulan dari sisi kesederhanaan dan efisiensi komputasi, serta mampu menangkap pola perilaku yang relatif stabil. Namun, asumsi linearitas membatasi kemampuan model AR dalam merepresentasikan perilaku lawan yang kompleks dan adaptif, khususnya pada permainan dengan interaksi strategis yang dinamis.

Untuk meningkatkan kapasitas representasi, pendekatan berbasis jaringan saraf tiruan (Artificial Neural Network, ANN) diperkenalkan sebagai model prediktif

nonlinier. Jaringan saraf mampu memodelkan hubungan kompleks antara input dan output tanpa asumsi linearitas yang ketat. Dalam penerapannya pada opponent modelling, riwayat aksi lawan dan agen dapat digunakan sebagai input untuk memprediksi aksi lawan berikutnya.

Meskipun ANN meningkatkan fleksibilitas model, pendekatan ini tidak secara eksplisit membedakan antara ketergantungan temporal internal sistem dan pengaruh input eksternal. Ketergantungan waktu biasanya direpresentasikan dengan memperluas vektor input menggunakan beberapa observasi masa lalu, yang dapat meningkatkan dimensi input secara signifikan dan menyulitkan proses pelatihan serta interpretasi model.

Model AutoRegressive with eXogenous inputs (ARX) diperkenalkan untuk mengatasi keterbatasan tersebut dengan memisahkan secara eksplisit pengaruh keluaran masa lalu dan input eksternal. Dalam model ARX, keluaran pada waktu tertentu dimodelkan sebagai fungsi linier dari keluaran masa lalu dan variabel eksogen. Dalam konteks permainan berulang, variabel eksogen tersebut dapat berupa aksi agen atau faktor kontekstual lain yang memengaruhi keputusan lawan.

Model ARX memungkinkan analisis yang lebih terstruktur terhadap dinamika interaksi antara agen dan lawan. Namun, seperti halnya model AR, ARX masih dibatasi oleh asumsi linearitas, sehingga kurang mampu menangkap dinamika nonlinier yang umum muncul dalam perilaku strategis dan adaptif.

Untuk menggabungkan pemodelan dinamika temporal yang eksplisit dengan kemampuan representasi nonlinier, model Nonlinear AutoRegressive with eXogenous inputs (NARX) diperkenalkan. Model NARX merepresentasikan keluaran sistem sebagai fungsi nonlinier dari keluaran masa lalu dan input eksogen masa lalu, yang umumnya direalisasikan menggunakan jaringan saraf.

Dalam opponent modelling, NARX memungkinkan pemodelan perilaku lawan sebagai respons terhadap riwayat aksinya sendiri dan tindakan agen secara simultan. Struktur ini memberikan keseimbangan antara kapasitas representasi dan eksplisitnya pemodelan dinamika temporal. Selain itu, penggunaan jumlah lag yang terbatas memungkinkan pengendalian kompleksitas model, sehingga tetap feasible secara komputasi untuk diterapkan pada permainan berulang dengan horizon menengah.

Meskipun demikian, model NARX tetap memiliki keterbatasan. Kinerja prediksi sangat dipengaruhi oleh pemilihan jumlah lag dan kualitas data pelatihan. Selain itu, seperti model prediktif lainnya, NARX berfokus pada estimasi perilaku law-

an pada horizon terbatas dan tidak secara langsung menjamin optimalitas keputusan strategis jangka panjang. Oleh karena itu, model ini umumnya digunakan sebagai komponen prediktif yang dikombinasikan dengan mekanisme evaluasi lanjutan.

Secara keseluruhan, perkembangan dari model AR menuju ANN, ARX, dan akhirnya NARX mencerminkan kebutuhan untuk memodelkan dinamika temporal dan interaksi strategis secara lebih akurat tanpa menuntut kompleksitas komputasi yang tidak feasible. Berdasarkan pertimbangan tersebut, penelitian ini memilih NARX sebagai model prediktif perilaku lawan dalam permainan berulang.

3.6 Evaluasi Model Prediksi Perilaku Lawan

Oleh karena itu, diperlukan metode evaluasi untuk menilai sejauh mana model mampu memprediksi perilaku lawan secara akurat dan konsisten.

Evaluasi dalam opponent modeling dapat mencakup akurasi prediksi tindakan, pengukuran error time series, serta evaluasi berbasis interaksi dalam skenario game.

Landasan teori evaluasi ini menjadi dasar bagi perancangan eksperimen dan analisis hasil pada bab selanjutnya.

3.6.1 Keterbatasan Prediksi Global pada Horizon Panjang

Dalam permainan berulang dua agen (*dyadic repeated games*), pengambilan keputusan yang optimal secara teoritis mensyaratkan agen untuk mempertimbangkan konsekuensi jangka panjang dari setiap aksi yang diambil. Secara ideal, agen memilih aksi yang memaksimalkan nilai ekspektasi dari akumulasi reward sepanjang horizon permainan, dengan memperhitungkan respons lawan terhadap aksi tersebut pada setiap langkah selanjutnya.

Pendekatan ini berkaitan erat dengan kerangka optimisasi dinamis dan *dynamic programming*, di mana solusi optimal diperoleh dengan mengevaluasi seluruh kemungkinan lintasan interaksi antara agen dan lawannya. Namun, kompleksitas perhitungan solusi tersebut meningkat secara eksponensial terhadap panjang horizon. Jika setiap agen memiliki ruang aksi diskret berukuran $|\mathcal{A}|$ dan permainan berlangsung selama H langkah, maka jumlah kemungkinan urutan aksi bersama berada pada orde $|\mathcal{A}|^{2H}$.

Selain kompleksitas ruang aksi, tantangan tambahan muncul akibat sifat adaptif lawan. Dalam opponent modelling, perilaku lawan tidak dapat diasumsikan statis, melainkan dapat berubah sebagai respons terhadap strategi yang diterapkan agen.

Hal ini menyebabkan agen harus mempertimbangkan tidak hanya prediksi aksi lawan, tetapi juga bagaimana kebijakan lawan berevolusi sepanjang interaksi. Akibatnya, perhitungan kebijakan optimal memerlukan estimasi terhadap dinamika kebijakan lawan, yang semakin memperbesar beban komputasi.

Keterbatasan ini menjadikan pendekatan prediksi global-optimal pada horizon panjang sulit diterapkan dalam praktik, khususnya pada skenario dengan keterbatasan sumber daya komputasi dan waktu. Bahkan pada permainan dengan struktur sederhana, evaluasi seluruh lintasan interaksi menjadi tidak tractable untuk horizon menengah hingga panjang. Oleh karena itu, sebagian besar pendekatan praktis memilih untuk membatasi horizon perencanaan atau menggunakan aproksimasi nilai.

Pembatasan horizon dan pendekatan prediksi jangka pendek memang dapat mengurangi kompleksitas komputasi secara signifikan, namun konsekuensi strategis jangka panjang tidak sepenuhnya tertangkap. Dalam konteks dilema sosial berulang, keputusan yang optimal secara instan dapat memicu perubahan perilaku lawan pada langkah-langkah berikutnya, sehingga menurunkan reward kumulatif jangka panjang. Hal ini menunjukkan bahwa terdapat trade-off mendasar antara efisiensi komputasi dan kualitas keputusan strategis.

Sebagai respons terhadap trade-off tersebut, berbagai pendekatan telah dikembangkan untuk mengaproksimasi dampak jangka panjang tanpa melakukan prediksi global secara eksplisit. Pendekatan-pendekatan ini mencakup penggunaan metrik evaluasi berbasis regret, simulasi terbatas melalui *Monte Carlo rollout*, serta model prediktif dinamis yang memfokuskan pada estimasi konsekuensi jangka menengah. Dengan demikian, kualitas keputusan dapat ditingkatkan tanpa menuntut perhitungan yang tidak feasible secara komputasi.

BAB IV

ANALISIS DAN PERANCANGAN

BAB V
HASIL DAN PEMBAHASAN

BAB VI
KESIMPULAN DAN SARAN

DAFTAR PUSTAKA

- De Weerd, H., Verbrugge, R., dan Verheij, B. (Oct. 2022), Higher-order theory of mind is especially useful in unpredictable negotiations, en, *Autonomous Agents and Multi-Agent Systems* 36.2, p. 30.
- Di, C., Zhou, Q., Shen, J., Wang, J., Zhou, R., dan Wang, T. (2023), The coupling effect between the environment and strategies drives the emergence of group cooperation, *Chaos, Solitons & Fractals* 176, p. 114138.
- Elhamer, Z., Suzuki, R., dan Arita, T. (2020), The effects of population size and information update rates on the emergent patterns of cooperative clusters in a large-scale social particle swarm model, *Artificial Life and Robotics* 25.1, Type: Article, pp. 149–158.
- Freire, I. T., Arsiwalla, X. D., Puigbò, J. Y., dan Verschure, P. (2023), Modeling Theory of Mind in Dyadic Games Using Adaptive Feedback Control, *Information (Switzerland)* 14.8, Type: Article.
- Gómez, A. L. d. A., Sierra, C., dan Sabater-Mir, J. (2025), Grounded predictions of teamwork as a one-shot game: A multiagent multi-armed bandits approach, *Artificial Intelligence* 341, p. 104307.
- Hu, Y., Han, C., Li, H., dan Guo, T. (2023), Modeling opponent learning in multiagent repeated games, *Applied Intelligence* 53.13, Type: Article, pp. 17194–17210.
- Jin, Y., Wei, S., dan Montana, G. (Aug. 2025), Achieving collective welfare in multi-agent reinforcement learning via suggestion sharing, en, *Machine Learning* 114.8, p. 190.
- Li, K., Huang, W., Li, C., dan Deng, X. (2025), Exploiting a No-Regret Opponent in Repeated Zero-Sum Games, *Journal of Shanghai Jiaotong University (Science)* 30.2, Type: Article, pp. 385–398.
- Lv, M., Liu, J., Guo, B., Ding, Y., Zhang, Y., dan Yu, Z. (Sept. 2023), Inducing Coordination in Multi-Agent Repeated Game through Hierarchical Gifting Policies, *2023 IEEE 20th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*, ISSN: 2155-6814, pp. 279–287.

- Perera, I., Nijs, F. de, dan Garcia, J. (2025), Learning to cooperate against ensembles of diverse opponents, *Neural Computing and Applications* 37.23, Type: Article, pp. 18835–18849.
- Qiao, X., Han, C., dan Guo, T. (Dec. 2024), O2M: Online Opponent Modeling in Online General-Sum Matrix Games, *2024 4th International Conference on Artificial Intelligence, Robotics, and Communication (ICAIRC)*, pp. 358–361.
- Wang, W., Wang, Y., Hao, J., dan Taylor, M. E. (2019), Achieving cooperation through deep multiagent reinforcement learning in sequential prisoner’s dilemmas, *ACM International Conference Proceeding Series*, Type: Conference paper.
- Zhu, L., Zhu, Y., dan Xia, C. (May 2025), Evolutionary Dynamics of Cooperation and Extortion on Networks With Fitness-Dependent Rules, *2025 Joint International Conference on Automation-Intelligence-Safety (ICAIS) & International Symposium on Autonomous Systems (ISAS)*, ISSN: 2996-3850, pp. 1–6.

LAMPIRAN