Sejarah, Teori Dasar dan Penerapan *Reinforcement Learning*: Sebuah Tinjauan Pustaka

Jeky Andreanus[#1], Ade Kurniawan[#2]

[1,2]*Departement of Informatics Engineering, Universal University*

Kompleks Maha Vihara Duta Maitreya, Sungai Panas, Batam 29456, Kepulauan Riau - Indonesia

[1]jekyandreanus@gmail.com

[2]*ade.kurniawan@uvers.ac.id*

Abstract*— Today's research on the topic of Machine learning has increased sharply. Machine learning is the future of the world, in the future it will be a revolution in every computer-bound science. This paper examines the field of Reinforcement learning from a computer science perspective. Reinforcement learning is part of Machine learning. In general machine learning is divided into three categories, namely supervised learning, unsupervised learning, and reinforcement learning. Supervised learning requires labeled data to analyze data, training and make conclusions, which can be used for mapping new values. Otherwise, Unsupervised learning does not use labeled data, which is more suitable for relatively irregular problems. In contrast to Reinforcement learning that is based on trial and error, by experimenting on the environment then get a response that will improve its ability. This work is summarized based on the history of Reinforcement learning and the selection of current research. This paper discusses central issues in reinforcement learning, from history, Reinforcement learning models, multiinform Reinforcement learning, including comparison of exploration and exploitation. Ends with a Reinforcement learning implementation survey of several systems. Reinforcement learning is the most suitable Machine learning in learning new things from scratch without human intervention in learning, most of Reinforcement learning is used for in-game learning. But the learning may takes a long time and is uncertain.

***Keywords*— reinforcement learning, history, trial and error, RL model, multiagent RL, exploration and exploitation, literature review**

Abstrak*— Dewasa ini penelitian mengenai topik *Machine learning* telah meningkat tajam. *Machine learning* adalah masa depan dunia, kedepannya ini akan menjadi revolusi dalam segalah ilmu yang terikat dengan komputerisasi. Paper ini meneliti bidang *Reinforcement learning* dari perspektif ilmu komputer. *Reinforcement learning* merupakan bagian dari *Machine learning*. Secara umum *machine learning* dibagi menjadi tiga kategori, yaitu *supervised learning, unsupervised learning,* dan *reinforcement learning. Supervised learning* memerlukan data berlabel untuk menganalisis data, pelatihan dan membuat kesimpulan, yang dapat digunakan untuk pemetaan nilai-nilai baru. Sebaliknya, *Unsupervised learning* tidak memenggunakan data berlabel, yang mana lebih cocok untuk masalah yang relatif tidak beraturan. Berbeda dengan *Reinforcement learning* yang berbasis *trial and error,* dengan mencoba-coba pada lingkungannya kemudian mendapatkan respon yang akan meningkatkan kemampuannya. Karya ini dirangkum berdasarkan sejarah bidang *Reinforcement learning* dan pemilihan riset saat ini. Paper ini membahas isu-isu sentral dalam *reinforcement learning,* mulai dari sejarah, model *Reinforcement learning*, *multiagent Reinforcement learning* termasuk melakukan perbandingan dari eksplorasi dan eksploitasi. Diakhiri dengan survei penerapan *Reinforcement Learning* terhadap beberapa sistem. *Reinforcement learning* merupakan *Machine learning* yang paling cocok dalam mempelajari hal baru dari nol tanpa campur tangan manusia dalam pembelajarannya, kebanyakan dari *Reinforcement learning* digunakan untuk belajar dalam game. Namun pembelajaran yang dilakukan membutuhkan waktu lama dan tidak pasti.

***Kata Kunci*— reinforcement learning, sejarah, *trial and error*, model RL, *multiagent* RL, eksplorasi dan eksploitasi, tinjauan pustaka**

I. PENDAHULUAN

Reinforcement learning adalah bagian dari *artificial intelligence* yang melatih algoritma dengan sistem *trial and error*. RL berinteraksi dengan lingkungannya dan mengamati konsekuensi atas tindakannya sebagai tanggapan atas penghargaan maupun hukuman yang diterima. Informasi yang dihasilkan dari setiap interaksi dengan lingkungan, digunakan RL untuk memperbarui pengetahuannya [1].

Dunia semakin mengarah ke serba digital dan *autonomous* dengan kehadiran Internet, *artificial intelligence dan machine learning* [2][3]. *Machine Learning* merupakan bidang penelitian dari salah satu cabang ilmu pegetahuan yang menggabungkan gagasan dari beberapa cabang ilmu pengetahuan seperti kecerdasan buatan, statistik, teori informasi, matematika, dll [4]. *Machine learning* secara umum fokus pada teori, kinerja, dan sifat sistem pembelajaran dan algoritma [5]. *Machine learning* digunakan untuk menyelesaikan berbagai masalah, mulai dari robotika, sistem pengenalan, *data mining*, dan sistem control otomatis, hingga informatika [5]. *Machine learning* yang mandiri ini sangat membantu dalam mengatasi masalah-masalah yang selama ini sulit dipecahkan, sehingga masalah yang rumit bisa menjadi hal yang mungkin dilakukan.

Machine learning merupakan bagian dari kecerdasan buatan yang membantu menemukan solusi dari berbagai masalah. Sebelum menjadi mesin pintar, *machine learning* harus mampu belajar dan beradabtasi dengan lingkungan. Jika sistem dapat belajar dan menyesuaikan diri dengan keadaan

tersebut, pengembang tidak perlu meramalkan atau memberi solusi untuk situasi yang mungkin terjadi [6].

Secara umum *machine learning* dibagi menjadi tiga kategori, yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning* [4]. Secara singkat, *supervised learning* membutuhkan pelatihan dengan data berlabel untuk menganalisis data pelatihan dan membuat fungsi yang disimpulkan, yang dapat digunakan untuk pemetaan nilai-nilai baru. Berbeda dengan *unsupervised learning*, tidak memerlukan data pelatihan berlabel dengan mengharapkan lingkungan memberi masukan tanpa target yang diinginkan. Sedangkan *reinforcement learning* memungkinkan pembelajaran dari umpan balik yang diterima melalui interaksi dengan lingkungan eksternal [4][5]. Dari sudut pandang pemorsesan data, *supervised learning* dan *unsupervised learning* lebih mengacuh ke analisa data, sedangkan *reinforcement learning* lebih condong ke pengambilan keputusan untuk menyelesaikan masalah. Dalam artikel ini akan membahas lebih lanjut mengenai *reinforcement learning*

II. METODOLOGI

Tinjauan Pustaka atau *Literature Review* adalah salah satu metode penelitian. Tinjauan pustaka adalah suatu teknik mengkaji kebanyakan kasus di atas, Penilaian tertulis tentang apa yang sudah diketahui dan topik pengetahuan yang sudah ditemukan, tanpa metodologi yang ditentukan, tinjauan akan menjadi bagian dari proyek penelitian dan disertasi [7].

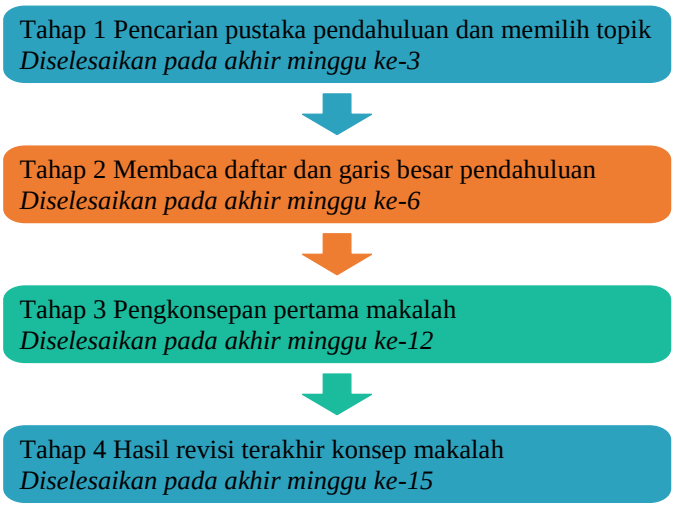
Ada dua model atau pendekatan dari *Literature Review*. Pertama adalah *Traditional Review*, pendekatannya bersifat kritis, menilai teori atau hipotesis dengan memeriksa secara kritis, metode dan hasil studi primer tunggal, dengan penekanan pada latar belakang dan materi kontekstual. Kedua yaitu *Systematic Review* berguna bagi mereka yang ingin mempromosikan pengetahuan dari penelitian dan menerapkannya [7].

Menurut [8] penulisan tinjauan pustaka dalam istilah pembuatan *paper* bukanlah hal yang mudah dikerjakan, penulisan menyangkut beberapa tugas seperti, (a) memilih topik pada bidang yang mungkin baru bagi Anda, (b) mengidentifikasi dan menemukan sejumlah artikel penelitian yang sesuai dengan database yang mungkin tidak Anda kenal, (c) menulis dan menyunting *essay* dengan baik, ketiga ini dikerjakan sekitar tiga sampai empat bulan. Harapannya tinjauan pustaka yang dikerjakan bisa diteliti dengan sepenuhnya dan tertulis dengan baik.

Penulisan dipandang banyak orang sebagai hal yang sulit dilakukan, maka dari itu penting bagi Anda seorang penulis untuk membuat rencana yang matang sebelum memulai menulis topik Anda. Pertama-tama Anda harus dipmemastikan bahwa topik yang diambil sudah dimengerti dan diketahui dengan jelas oleh pembimbing pada saat-saat awal untuk memulai menulis. Yang kedua dalam proses penulisan Anda harus menyesuaikan diri dengan baik. Pastikan Anda memiliki waktu yang cukup dalam mengikuti langka-langka seperti di atas, mulai dari pemilihan topik, membaca, dan mengevaluasi artikel-artikel yang bersangkutan dengan topik, mensintesis, mengatur catatan, menulis, konsep ulang, merevisi atau mengeditnya untuk mengoreksi penulisan dan kesesuaian terhadap tata bahasa yang baik. Di bawah ini adalah contoh ‘saran’ untuk penulisan 15 minggu, dibagi menjadi empat bagian [8].

Pengerjaan dengan batas waktu akan memberikan Anda motivasi dalam pembuatan makalah, dalam penentuan topik dan pengerjaan akan lebih singkat. Selain itu juga membantu Anda melihat sebuah bidang baru dengan terperinci [8].

Saran untuk pengerjaan dalam waktu 15 minggu



Saran untuk pengerjaan dalam waktu 15 minggu

Tahap 1 Pencarian pustaka pendahuluan dan memilih topik
Diselesaikan pada akhir minggu ke-3

Tahap 2 Membaca daftar dan garis besar pendahuluan
Diselesaikan pada akhir minggu ke-6

Tahap 3 Pengkonsepan pertama makalah
Diselesaikan pada akhir minggu ke-12

Tahap 4 Hasil revisi terakhir konsep makalah
Diselesaikan pada akhir minggu ke-15

III. HASIL DAN PEMBAHASAN

A. *Sejarah*

Diawali dengan istilah *Optimal Control* yang digunakan pada akhir tahun 1950 dalam menjelaskan desain solusi dari sistem kontrol untuk meminimalkan ukuran perilaku sistem dinamis dari waktu ke waktu. Salah satu pencapaian dari masalah ini dikembangkan oleh Richard Bellman dan teman kuliahnya pada pertengahan tahun 1950. Pencapaian ini menggunakan konsep dari kondisi sistem dinamis dan fungsi nilai, atau yang disebut *optimal return function*, menjelaskan fungsi persamaan yang sekarang dikenal persamaan Bellman.

Pada tahun 1957 Bellman mengenalkan metode penyelesaian *optimal control* dikenal dengan pemrograman

dinamis dan *stochastic* diskrit versi *optimal control* atau dikenal sebagai *Markovian decision processes* (MDPs), dan pada tahun 1960 Ron Howard mengembangkan kebijakan metode perulangan untuk MDPs. Pemrograman dinamis telah dikembangkan panjang lebar selama empat dekade yang lalu [9].

Kembali mengarah ke bidang pembelajaran penguatan modern, yang berpusat pada gagasan pembelajaran *trial and error*. Hal ini dimulai di bidang psikologi, di mana teori pembelajaran *reinforcement* umum terjadi. Mungkin yang pertama mengekspresikan esensi pembelajaran *trial and error* adalah Edward Thorndike [9].

Thorndike mengenalkan *Law of Effect* yang menjelaskan efek dari penguatan *even-even* pada kecenderungan untuk memilih tindakan. Terkadang ini menjadi perdebatan, *Law of Effect* secara luas dianggap sebagai prinsip dasar yang jelas yang mendasari banyak perilaku. *Law of Effect* mengandung dua aspek paling penting dimana yang dimaksud adalah pembelajaran *trial and error.* Pertama adalah *selectional* yang mana memilih di antara alternatif dengan membandingkan konsekuensi yang diterima. Kedua adalah *associative,* yaitu alternatif diperoleh dengan pemilihan terkait dengan situasi tertentu. Dengan kata lain *Law of Effect* adalah cara dasar menggabungkan pencarian dan penghafalan [9].

Di tahun 1961 dan 1963 Donald Michie menjelaskan pembelajaran *trial and error* sederhana untuk belajar bagaimana bermain *Tic Tac Toe* dikenal dengan *Matchbox Educable Noughts and Crosses Engine* (MENACE). Kemudian pada tahun 1968 Michie dan Chambers mengenalkan *Tic Tac Toe* mesin pembelajar *reinforcement* versi berbeda yang dikenal dengan *Game Learning Expectimaxing Engine* (GLEE) dan kontroler mesin pembelajar *reinforcement* yang dinamakan BOXES. BOXES diterapkan pada tugas belajar menyeimbangkan sebuah tiang yang digantung pada gerobak bergerak berdasarkan sinyal kegagalan yang terjadi hanya saat tiang jatuh atau gerobak mencapai ujung trek. Michie secara menekankan peran *trial and error* dan belajar sebagai aspek penting kecerdasan buatan [9].

Pada tahun 1977 publikasi artikel dari Ian Witten menjelaskan aturan tentang *temporal difference.* Dia mengusulkan metode yang sekarang sebut TD (0) digunakan sebagai bagian dari kontroler adaptif untuk menyelesaikan MDPs [9].

Secara bersamaan *temporal difference* dan *optimal control* pada tahun 1989 digunakan untuk mengembangkan *Q-learning* oleh Chris Watkins. Pada saat Watkins bekerja menghasilkan peningkatan yang luar biasa dalam penelitian pembelajaran *reinforcement*, terutama di bidang pembelajaran mesin kecerdasan buatan, serta pada *neural networks* dan kecerdasan buatan menjadi lebih luas [9].

B. Reinforcement Learning

Reinforcement learning adalah belajar apa yang akan dilakukan pembelajaran dengan, pemetaan situasi dalam menentukan tindakan, dan memaksimalkan angka sinyal penghargaan yang bisa diperoleh dari lingkungannya [9][10]. RL berinteraksi dengan lingkungannya dan mengamati konsekuensi atas tindakannya sehingga dapat belajar mengubah tingkah lakunya sendiri sebagai tanggapan atas penghargaan yang diterima. Informasi yang dihasilkan dari setiap interaksi dengan lingkungan yang kemudian digunakan RL untuk memperbarui pengetahuannya [1].

RL mengaplikasikan pembelajaran *trial and error* untuk mencapai target yang diharapkan [10]. Dalam menghadapi masalah RL akan mempelajari tingkah laku melalui pembelajaran *trial and error* untuk berinteraksi dengan lingkungan yang dinamis [11]. Begitu RL mampu menyesuaikan diri dengan lingkungan, sehingga bisa mendapatkan pengetahuan dan mencapai tujuan.

Dalam menyelesaikan masalah RL memiliki dua strategi utama. Pertama menemukan ruang dari tingkah laku dengan tujuan menemukan performa yang baik di dalam lingkungannya. Pencapaian ini telah diterapkan dalam *genetic algorithms* dan *genetic programming*, serta beberapa teknik pencarian novel. Kedua adalah menggunakan teknik statistik dan metode pemrograman dinamis untuk melengkapi pengambilan keputusan yang ada pada kondisi dunia nyata [11].

C. Model dari Reinforcement Learning

Dasar dari Model RL adalah RL terhubung dengan lingkungannya melalui presepsi dan tindakan seperti yang ada pada Gambar 1.

Setiap langkah dari interaksi RL diterima sebagai masukan (*i*), dengan beberapa indikasi dari kondisi mutakhir (*s*) pada lingkungan. RL akan memilih tindakan (*a*) sebagai keluaran. Setiap tindakan bisa mempengaruhi kondisi lingkungan dan nilai dari transisi kondisi terhubung dengan RL melalui skalar *reinforcement signal (r).* Tingkah laku RL (*B*), akan memilih tindakan yang akan meningkatkan jumlah nilai jangka panjang dari *reinforcement signal* [11].

Umumnya model ini berisi:

- Seset diskrit kondisi lingkungan (*S*);
- Seset diskrit tindakan RL (*A*); dan
- Seset skalar *reinforcement signal*, khususnya {0,1}, atau angka real.

Pada gambar juga berisikan masukan fungsi *I*, dimana menentukan pandangan RL terhadap kondisi lingkungan. asusmsikan ini adalah identitas fungsi (yaitu, RL melihat kondisi yang tepat dari lingkungan) [11]. Untuk memahami bagaimana RL dengan lingkungan perhatikan contoh berikut:

Lingkungan: Anda pPada kondisi ke 65 memiliki 4 pilihan tindakan.

RL: ASaya akan mengambil tindakan 2.

Lingkungan: MAnda menerima 7 unit *reinforcement* dan sekarang di kondisi 15, memiliki 2 pilihan tindakan.

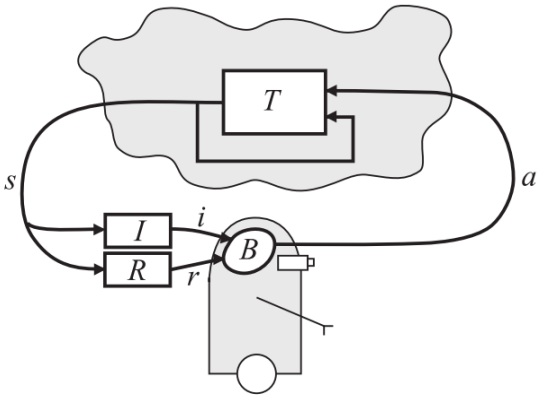
RL: Saya Aakan mengambil tindakan 1.

Lingkungan: Anda mMenerima (-4) unit *reinforcement* dan sekarang di kondisi 65, memiliki 4 pilihan tindakan.

RL: Saya aAkan mengambil tindakan 2.

Lingkungan: Anda mMenerima 5 unit *reinforcement* dan sekarang di kondisi 44, memiliki 5 pilihan tindakan.

… **…**



Gambar 1 Dasar dari Model RL[11]

Tugas RL adalah mencari kebijakan (π), pemetaan kondisi dalam menentukan tindakan-tindakan, ini memaksimalkan perkiraan jangka panjang dari *reinforcement.* Harapannya lingkungan tidak akan bersifat non deterministik, yaitu mengambil tindakan yang sama dalam kondisi yang sama pada dua kesempatan yang berbeda, mungkin menghasilkan perbedaan pada kondisi berikutnya dan/atau nilai *reinforcement* yang berbeda. Seperti yang terjadi pada contoh di atas, dari kondisi 65, menerapkan tindakan 2 yang memperoleh perbedaan *reinforcement* dan perbedaan kondisi dalam dua kali kesempatan. Namun, kita asumsikan bahwa lingkungan ini seimbang, yaitu kemungkinan membuat transisi kondisi atau menerima *reinforcement signal* yang spesifik tidak berubah sepanjang waktu [11].

D. Eksplorasi versus Eksploitasi

Salah satu kesulitan terbesar dalam RL adalah dilema mendasar eksplorasi versus eksploitasi. Kapan sebaiknya agen mencoba (menduga) tindakan yang tidak optimal untuk mengeksplorasi lingkungan (dan berpotensi memperbaiki model), dan kapan sebaiknya mengeksploitasi tindakan optimal untuk menghasilkan kemajuan yang bermanfaat? Meskipun menambahkan bunyi independen untuk eksplorasi dapat digunakan dalam masalah pengendalian terus menerus, strategi yang lebih canggih menyuntikkan suara yang berkorelasi sepanjang waktu untuk mempertahankan momentum dengan lebih baik [1].

Salah satu prinsip utama strategi eksplorasi adalah algoritma *upper confidence bound* (UCB), berdasarkan prinsip "optimisme dalam menghadapi ketidakpastian"[1]. Gagasan di balik UCB adalah memilih tindakan yang memaksimalkan eksplorasi di daerah dengan ketidakpastian tinggi dan perkiraan pengembalian yang moderat. Jika ini terjadi cukup sering, maka agen akan belajar apa hasil sebenarnya dari tindakan ini dan tidak memilihnya di masa depan. UCB juga dapat dianggap sebagai salah satu cara untuk menerapkan motivasi intrinsik, yang merupakan konsep umum yang menganjurkan penurunan ketidakpastian atau kemajuan dalam belajar tentang lingkungan [12].

E. Multiagent RL

Seringkali pembelajaran yang dilakukan dalam sebuah lingkungan dilakukan oleh satu agen RL. Sebaliknya dengan *multiagent RL* (MARL) memperhatikan beberapa agen belajar melalui RL dan seringkali agen lain bertindak secara tidak menentu yang mengubah prilaku kedua agen saat mereka belajar [13]. Hal ini berfokus untuk membuat perbedaan diantara agen, yang diharapkan bisa menjadi hal yang membuat mereka bekerja sama dengan baik. Sudah dilakukan beberapa pendekatan untuk menciptakan hal ini, yaitu menyampaikan pesan ke agen secara berurutan, dengan menggunakan saluran dua arah (menyediakan pemesanan dengan sedikit kehilangan sinyal), dan saluran *all-to-all* [1]. Penambahan saluran komunikasi adalah strategi alami yang diterapkan pada MARL dalam skenario yang kompleks dan tidak menghalangi praktik pemodelan kolaborasi atau perlombaan agen yang biasa diterapkan di tempat lain dalam literatur MARL[1].

MARL adalah bidang penelitian yang baru, namun aktif dan berkembang pesat. Ini mengintegrasikan hasil dari penguatan penguatan agen tunggal, teori permainan, dan pencarian langsung di ruang perilaku [13]. Dalam pandangan [13], kemajuan signifikan dalam bidang multi agent learning dapat dicapai dengan fertilisasi silang yang lebih intensif antara bidang pembelajaran mesin, teori permainan, dan teori kontrol.

F. Penerapan Reinforcement Learning

1) Bermain Game

Bermain game menjadi salah satu pengaplikasian RL, RL akan mempelajari bagaimana cara bermain game dengan menemukan solusi terbaik dari pengetahuan yang dimilikinya dan akan terus bertambah seiring RL mempelajari game tersebut. RL dengan basis *trial and error* dalam menentukan tindakan yang bereaksi pada lingkungan sehingga menerima penghargaan, untuk terus menambah pengetahuan dan informasi yang dimiliki mesin. Sangat cocok untuk menemukan cara terbaik dalam memainkan game. Misalnya RL yang bernama AlphaGo suskes mengalahkan manusia tingkat dunia dalam permainan Go [1]. Kemudian sebuah penelitian dari E. Amadou dkk, meneliti RL dalam permainan *Doubles Pong* bahwa dua agen dari RL dapat bermain dengan lancar meskipun lingkungannya tidak stabil [14]. Contoh lain game-game yang telah di terapkan menggunakan RL, seperti *Three classic Atari 2600 video games, Enduro, Freeway,* dan *Seaquest,* dari *Arcade Learning Environment* (ALE), beberapa game ini menjadi standar pengujian algoritma RL[1]. Untuk menjadi pemain profesional RL butuh belajar bermain berkali-kali tergantung tingkat kesulitanya, hingga puluhan, ratusan, bahkan ribuan percobaan.

2) Stock Price Prediction

Pengambilan keputusan pada tempat yang tidak pasti dan beresiko adalah lokasi penelitian yang menonjol [15]. Prediksi menjadi hal penting dalam pengambilan keputusan untuk menentukan masa depan. Ekonomi merupakan bidang yang penting dalam prediksi untuk menentukan masa depan dari perekonomian. Salah satu contoh pengaplikasian RL pada

bidang ekonimi adalah prediksi *stock-price.* Banyak yang telah mencoba RL untuk mempelajari *stock-price* dan membuat keputusan yang semakin baik. Seperti yang diteliti oleh A. Pastore dkk, mereka berasumsi RL terbukti signifikan secara statistik, meski kecil, karena hanya berlaku untuk subset pemain [15]. Kemudian J. Won menjelaskan, hasil percobaan menunjukkan bahwa metode yang diusulkan dapat digunakan sebagai indikator yang lebih berguna untuk perdagangan saham, yaitu Algoritma TD, dapat langsung beralih dari pengalaman mentah tanpa dinamika lingkungan [16]. Nanum, saat ini prediksi menggunakan RL masih belum mendekati angka sempurna.

3) Robotic

Robot merupakan mesin berbasis komputer yang mampu melakukan berbagai pekerjaan kompeks secara otomatis, akhir-akhir ini telah banyak robot AI yang diciptakan. Robot yang diharapkan bisa melakukan pekerjaan seperti manusia, hingga nanti akan menggantikan berbagai pekerjaan manusia. Namun tantangannya saat ini robot belum mampu melakukan pekerjaan tertentu secara sempurna, oleh karena itu robot AI memerlukan banya pengetahuan untuk bisa mencapai kemampuan dan bersifat seperti manusia.

Penerapan RL pada robot agar robot bisa belajar dengan basis *trial and error* dalam belajar dan bertindak, agar robot bisa menemukan caranya sendiri dalam menyelesaikan masalah. Ini bisa berguna untuk robot bisa belajari dari mengamati kondisi sekitarnya, mengetahui pola dari objek penelitian, agar bisa berpikir, bertindak, dan berkelakuan seperti manusia.

Sebuah penelitian yang dilakukan oleh [17] yaitu penerapan RL pada robot untuk melakukan *Path Planning*, yang tujuannya agar robot bisa melakukan tindakan optimal langsung dari pandangan visual asli tanpa campur tangan manusia. Contoh lain dari penelitian yang dilakukan oleh [18] untuk memungkinkan kebijakan gerakan dengan pembelajaran aman dengan menggunakan *collision avoidance system* yang diterapkan pada RL. Menurut D. Schwung dkk, hasil yang disajikan menunjukkan bahwa kedua robot dapat mempelajari tugas penanganan yang diberikan dengan menerapkan perilaku koperasi yang khas saat belajar menggunakan algoritma *Q-learning* yang terkenal [18]. Dengan mengamati, menerima penghargaan, dan bertindak dengan mempelajari lingkungan, diharapkan robot bisa bertindak dan berkelakuan layaknya manusia

IV. KESIMPULAN

Perkembangan RL dari beberapa dekade dahulu telah menunjukan peningkatan pesat. Banyak penelitian-penelitian pengembangan dan pemanfaatan teori RL ini dalam melakukan berbagai hal. Baik itu bidang teknologi informasi, pendidikan, ekonomi dan, psikologi manusia dan hewan.

RL merupakan sarana yang paling cocok dalam mempelajari hal baru dari nol tanpa campur tangan manusia dalam pembelajarannya. Seperti bermain *game* merupakan sarana paling cocok pengujian RL yang berbasis *trial and error*. Bermain game akan memberikan penghargaan baik itu nilai tambah atau kurang, sehingga membuat RL mengambil tindakan dari penghargaan yang diterima, sampai RL mampu bermain dengan sempurna. Adapun penerapan agen ganda RL dalam bermain, kedua agen bertindak saling mempengaruhi, seringkali agen lain bertindak secara tidak menentu yang mengubah prilaku kedua agen saat mereka belajar, perbedaan kedua agen diharapkan bisa menjadi hal yang membuat mereka bekerja sama dengan baikyang kedepannya mungkin bisa berkolaborasi dengan baik, mampu belajar dengan baik bersama-sama, sehingga tidak memerlukan manusia untuk membantu interaksi mereka.

DAFTAR REFERENSI

[1] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Process. Mag.*, vol. 34, no. 5, 2017.

[2] A. Kurniawan, I. Riadi, and A. Luthfi, "Forensic Analysis and Prevent of Cross Site Scripting in Single Victim Attack Using Open Web Application Security Project (Owasp) Framework," *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 6, pp. 1363–1371, 2017.

[3] Cristina and A. Kurniawan, "Sejarah , Penerapan , dan Analisis Resiko dari Neural Network : Sebuah Tinjauan Pustaka," vol. 03, no. 02, pp. 259–270, 2018.

[4] Q. He, N. Li, W. J. Luo, and Z. Z. Shi, "A survey of machine learning algorithms for big data," *Moshi Shibie yu Rengong Zhineng/Pattern Recognit. Artif. Intell.*, vol. 27, no. 4, pp. 327–336, 2014.

[5] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, "A survey of machine learning for big data processing," *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 1, 2016.

[6] E. Alpaydin, *Introduction to Machine Learning*. London: The MIT Press, 2014.

[7] J. Jesson, L. Matheson, and F. M. Lacey, *Doing Your Literature Review :Traditional and Sistematic Techniques*. 2011.

[8] J. L. Galvan and M. C. Galvan, *Writing literature reviews: A guide for students of social and behavioral sciences*. 2017.

[9] R. S. Sutton and a G. Barto, "Reinforcement learning: an introduction.," *IEEE Trans. Neural Netw.*, vol. 9, no. 5, pp. 127–144, 1998.

[10] N. D. Nguyen, T. Nguyen, and S. Nahavandi, "System Design Perspective for Human-Level Agents Using Deep Reinforcement Learning: A Survey," *IEEE Access*, vol. PP, no. 99, p. 1, 2017.

[11] R. Giryes and M. Elad, "Reinforcement Learning: A Survey," *Eur. Signal Process. Conf.*, pp. 1–2, 2011.

[12] J. Schmidhuber, "A possibility for implementing curiosity and boredom in model-building neural controllers," *Proc. Int. Conf. Simul. Adapt. Behav.*, pp. 222–227, 1991.

[13] L. Busoniu, R. Babuska, B. De Schutter, and B. De Schutter, "A comprehensive survey of multiagent reinforcement learning," *Syst. Man, Cybern. Part C Appl. Rev.*, vol. 38, no. 2, pp. 156–172, 2008.

[14] E. Amadou, O. Diallo, A. Sugiyama, and T. Sugawara, "Learning to Coordinate with Deep Reinforcement Learning in Doubles Pong Game," 2017.

[15] A. Pastore, U. Esposito, and E. Vasilaki, "Modelling Stock-market Investors as Reinforcement Learning Agents," 2015.

[16] J. Won, "Stock price prediction using reinforcement learning," pp. 690–695, 2001.

[17] J. Xin, H. Zhao, and D. Liu, "Application of Deep Reinforcement Learning in Mobile Robot Path Planning," no. 16.

[18] D. Schwung, F. Csaplar, A. Schwung, and S. X. Ding, "An application of reinforcement learning algorithms to industrial multi-robot stations for cooperative handling operation," *Proc. - 2017 IEEE 15th Int. Conf. Ind. Informatics, INDIN 2017*, pp. 194–199, 2017.

Jeky Andreanus, berkelahiran di kota Tanjungpandan. Saat ini penulis sedang melaksanakan perkuliahan sarjana di Universitas Universal Batam. Kegiatan saat ini adalah pengembangan situs dan mempelajari teori-teori *machine learning*. Penulis memiliki kesukaan dibidang pengembangan situs yang nantinya akan digunakan untuk pengaplikasian *machine learning*.

Ade Kurniawan, Lahir di Sumbawa Besar, NTB. Saat ini bekerja sebagai staf pengejar di Teknik Informatika, Universitas Universal, Batam. Minta riset pada bidang *Cyber Security*, *Deep Learning*, *Digital Forensics*, *Machine Learning*, *Network Forensics*.