



# Pemilihan Parameter Crossover Moving Average Adaptif pada BTC/USDT Menggunakan Proximal Policy Optimization

Anandava Eka Buana Baskara<sup>1,\*</sup>, Joko Ariyanto<sup>2</sup>

<sup>1</sup>Fakultas Sains dan Teknologi, Program Studi Informatika, Universitas Teknologi Yogyakarta, Yogyakarta, Indonesia

<sup>2</sup>Fakultas Sains dan Teknologi, Universitas Teknologi Yogyakarta, Yogyakarta, Indonesia

Email: <sup>1,\*</sup>anandava759@gmail.com, <sup>2</sup>joko.aryanto@uty.ac.id

Email Penulis Korespondensi: anandava759@gmail.com

**Abstrak**—Pasar *cryptocurrency* memiliki volatilitas tinggi, sehingga menentukan kombinasi *Moving Average* (MA *Short* dan MA *Long*) yang optimal merupakan tantangan penting bagi strategi perdagangan berbasis indikator teknikal. Penelitian ini bertujuan untuk mengembangkan strategi *crossover Moving Average* yang adaptif menggunakan algoritma *Proximal Policy Optimization* (PPO) untuk menilai dan merekomendasikan kombinasi MA yang paling efektif. Data harga harian *cryptocurrency* dari 1 Januari 2021 hingga 15 Mei 2026, dengan total 1960 *candle*, diambil melalui API dari platform *exchange*. Data tersebut diproses melalui *preprocessing* untuk membentuk *state* pasar yang mencakup harga, volume, dan indikator volatilitas, yang kemudian digunakan sebagai input bagi agen PPO dalam melakukan *training* dan evaluasi strategi. Hasil pengujian menunjukkan bahwa kombinasi MA 3/50 paling sering dipilih oleh PPO berdasarkan probabilitas rata-rata, sedangkan kombinasi MA 25/40 menghasilkan performa finansial terbaik jika dilihat dari *profit factor*, *net profit*, dan *win rate*. Visualisasi *equity curve*, *drawdown*, serta titik *entry* dan *exit* menegaskan kemampuan strategi dalam menyesuaikan keputusan secara adaptif, menangkap tren pasar, dan menyeimbangkan antara risiko dan profitabilitas. Temuan ini memberikan panduan praktis untuk pemilihan parameter *crossover Moving Average* yang adaptif, memungkinkan strategi perdagangan berbasis indikator teknikal untuk menghadapi dinamika pasar *cryptocurrency* yang kompleks dan cepat berubah.

**Kata Kunci:** Moving Average; Proximal Policy Optimization; Strategi Perdagangan; *Cryptocurrency*; Crossover MA

**Abstract**—*Cryptocurrency* markets exhibit high volatility, making it challenging to determine the optimal combination of Moving Averages (MA *Short* and MA *Long*) for technical indicator-based trading strategies. This study aims to develop an adaptive crossover Moving Average strategy using the Proximal Policy Optimization (PPO) algorithm to evaluate and recommend the most effective MA combinations. Daily *cryptocurrency* price data from January 1, 2021, to May 15, 2026, comprising a total of 1960 candles, were obtained through an exchange platform API. The data were processed through preprocessing to form market states that include price, volume, and volatility indicators, which were then used as input for the PPO agent during training and strategy evaluation. Test results indicate that the MA 3/50 combination was most frequently selected by PPO based on average probability, while the MA 25/40 combination produced the best financial performance in terms of profit factor, net profit, and win rate. Visualizations of the equity curve, drawdown, and entry and exit points confirm the strategy's ability to adaptively adjust decisions, capture market trends, and balance risk and profitability. These findings provide practical guidance for selecting adaptive crossover Moving Average parameters, enabling technical indicator-based trading strategies to navigate the complex and rapidly changing dynamics of *cryptocurrency* markets.

**Keywords:** Moving Average; Proximal Policy Optimization; Trading Strategy; *Cryptocurrency*; Crossover MA

## 1. PENDAHULUAN

Pasar *cryptocurrency* menunjukkan karakteristik yang secara fundamental berbeda dari pasar finansial tradisional. Bitcoin sebagai aset kripto dengan kapitalisasi terbesar, mencatat volatilitas harian rata-rata antara 3% hingga 8%, jauh melampaui ekuitas konvensional yang umumnya berada di bawah 1% (Bandarupalli, 2025). Lebih dari itu, pasar ini beroperasi 24 jam sehari, tujuh hari seminggu, tanpa mekanisme *circuit breaker*, sehingga perubahan regime pasar, dari trending ke *sideways* atau dari *low-volatility* ke *high-volatility*, terjadi secara tiba-tiba dan tidak terprediksi. Kondisi ini secara langsung mengekspos kelemahan mendasar strategi berbasis aturan tetap.

Strategi *Moving Average Crossover* tetap digunakan secara luas karena sederhana dan mudah diinterpretasikan. Namun, bukti empiris menunjukkan bahwa strategi ini sering menghasilkan kinerja yang tidak stabil ketika diterapkan dengan parameter tetap. Penelitian oleh (Romo et al., 2025) menunjukkan bahwa penggunaan parameter tetap pada *moving average* menghasilkan sinyal yang tertunda dan tidak responsif terhadap perubahan harga yang cepat. Dampaknya terlihat pada penurunan *Sharpe ratio* serta peningkatan *drawdown* pada periode volatilitas tinggi. Secara teoritis, kelemahan ini berasal dari sifat *moving average* sebagai indikator *lagging* serta penggunaan parameter tetap yang tidak mampu mengikuti perubahan distribusi return yang bersifat dinamis (Romo et al., 2025).

Pendekatan optimasi heuristik sering digunakan untuk mengatasi keterbatasan parameter tetap tersebut. Metode seperti *Learning-Based Linear Balancer* yang diusulkan oleh (Romo et al., 2025) mampu menemukan parameter optimal berdasarkan data historis. Namun, pendekatan ini memiliki keterbatasan mendasar karena bersifat *offline* dan tidak adaptif setelah proses optimasi selesai. Parameter yang dihasilkan hanya optimal untuk kondisi tertentu dan tidak mampu menyesuaikan diri terhadap perubahan pasar yang baru. Hal ini menjadi masalah serius pada pasar *cryptocurrency* yang bersifat *non-stationary* dan memiliki distribusi *return* yang terus berubah (Romo et al., 2025).

*Deep Reinforcement Learning* menawarkan pendekatan yang berbeda secara konseptual. Metode ini memodelkan perdagangan sebagai proses pengambilan keputusan sekuensial yang memungkinkan agen belajar langsung dari interaksi dengan lingkungan pasar. *Proximal Policy Optimization* menjadi salah satu algoritma yang relevan karena mampu menjaga stabilitas pembelajaran pada data yang bersifat *noisy* (Edy Prasetyo et al., 2025).

Pendekatan ini tidak hanya mengoptimasi parameter secara statis, tetapi membangun kebijakan yang adaptif terhadap setiap kondisi pasar. Pada penelitian tersebut, *moving average* tidak dihilangkan, tetapi digunakan sebagai representasi state yang tetap informatif dan terstruktur.

Literatur terbaru menunjukkan berbagai upaya untuk mengintegrasikan *Deep Reinforcement Learning* dalam perdagangan *cryptocurrency*, tetapi masih memiliki keterbatasan yang signifikan. Penelitian oleh (Edy Prasetyo et al., 2025) menunjukkan bahwa PPO dan DQN memiliki performa yang berbeda tergantung kondisi pasar, namun tidak memanfaatkan indikator teknikal sebagai struktur informasi awal sehingga meningkatkan kompleksitas pembelajaran. (Kochliaridis et al., 2023) menggabungkan *Deep Reinforcement Learning* dengan analisis teknikal dan mekanisme pengendalian risiko, tetapi tidak mengoptimasi parameter indikator secara adaptif. (Ospina-Holguín & Padilla-Ospina, 2025) mengeksplorasi kombinasi berbagai sinyal teknikal, namun menghadapi masalah korelasi tinggi antar fitur yang dapat menurunkan kualitas pembelajaran. Sementara itu, (Khujamatov et al., 2026) menekankan pentingnya desain *reward function*, tetapi tidak secara langsung menyelesaikan masalah adaptasi parameter pada strategi *crossover*.

Berdasarkan kajian tersebut, terdapat kekosongan penelitian yang jelas dan terukur. Belum ada pendekatan yang secara eksplisit mengintegrasikan adaptasi parameter *moving average crossover* ke dalam kerangka *Deep Reinforcement Learning*. Selain itu, belum terdapat evaluasi empiris yang membandingkan secara langsung kinerja strategi *moving average* adaptif berbasis *Proximal Policy Optimization* dengan strategi statis dan pendekatan lain pada data *cryptocurrency* yang beragam (Edy Prasetyo et al., 2025; Kochliaridis et al., 2023; Romo et al., 2025).

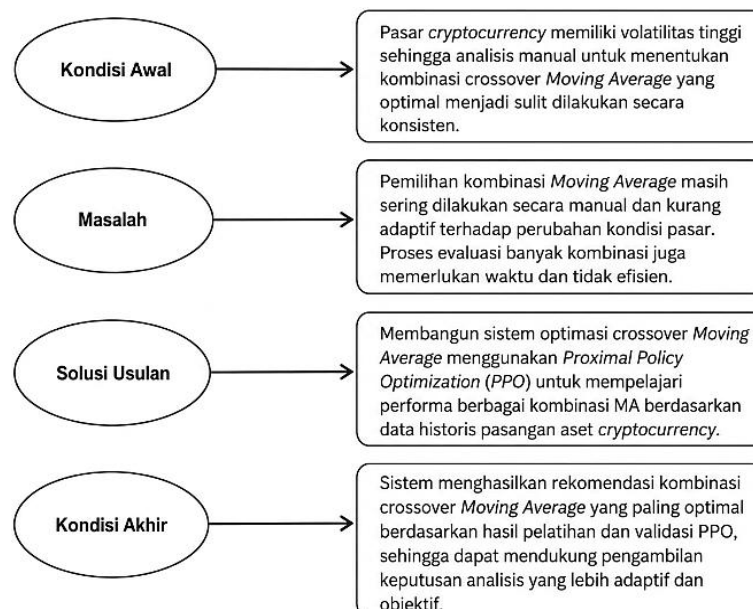
Penelitian ini mengembangkan kerangka kerja yang mengintegrasikan *Moving Average Crossover* dengan *Proximal Policy Optimization* untuk menghasilkan kerangka strategi perdagangan yang adaptif. Parameter *moving average* dimasukkan sebagai bagian dari representasi *state* sehingga dapat disesuaikan secara dinamis oleh agen. Model dilatih dan diuji menggunakan data historis Bitcoin untuk mengevaluasi kinerja secara komprehensif dalam berbagai kondisi pasar yang nyata.

Penelitian ini memberikan tiga kontribusi utama. Pertama, penelitian ini merumuskan pendekatan baru yang mengintegrasikan indikator teknikal dengan pembelajaran kebijakan adaptif dalam satu kerangka yang sistematis. Kedua, penelitian ini menyediakan evaluasi empiris yang membandingkan kinerja strategi adaptif dengan strategi statis dan pendekatan lain berbasis pembelajaran mesin (Edy Prasetyo et al., 2025; Kochliaridis et al., 2023). Ketiga, penelitian ini menghasilkan bukti empiris mengenai bagaimana parameter *moving average* dapat disesuaikan secara dinamis untuk meningkatkan kinerja perdagangan (Khujamatov et al., 2026; Ospina-Holguín & Padilla-Ospina, 2025). Pendekatan ini diharapkan dapat memberikan kontribusi teoritis dan praktis dalam pengembangan sistem perdagangan otomatis yang lebih adaptif dan *robust*.

## 2. METODOLOGI PENELITIAN

### 2.1 Kerangka Dasar Penelitian

Pasar *cryptocurrency* memiliki volatilitas tinggi yang membuat investor sulit menentukan waktu beli atau jual yang tepat secara manual, terutama ketika memantau banyak aset sekaligus. Analisis manual bersifat subjektif, lambat, dan tidak konsisten, sehingga diperlukan sistem berbasis data yang adaptif dan mampu belajar dari pengalaman pasar (Bandarupalli, 2025; Khujamatov et al., 2026).



Gambar 1. Kerangka Penelitian

Untuk mengatasi permasalahan tersebut, penelitian ini mengusulkan sistem yang mengoptimalkan strategi *crossover Moving Average* melalui *Proximal Policy Optimization (PPO)*. Sistem menerima input pasangan aset (*pair*) dan melakukan *training* PPO untuk mempelajari performa berbagai kombinasi MA terhadap data historis. Setelah pembelajaran selesai, sistem divalidasi untuk mengevaluasi efektivitas kombinasi MA yang telah dipelajari, kemudian menghasilkan rekomendasi kombinasi MA yang paling optimal berdasarkan skor yang diperoleh dari agen PPO. Dengan pendekatan ini, sistem dapat memberikan saran yang adaptif, objektif, dan responsif terhadap dinamika pasar *cryptocurrency* yang cepat dan kompleks.

Kerangka penelitian pada Gambar 1 menjelaskan alur dasar sistem yang dikembangkan dalam penelitian ini, dimulai dari kondisi awal pasar *cryptocurrency* yang memiliki volatilitas tinggi sehingga analisis manual untuk menentukan kombinasi *crossover Moving Average* yang optimal menjadi sulit dilakukan secara konsisten. Masalah yang muncul adalah pemilihan kombinasi *Moving Average* yang masih sering dilakukan secara manual dan kurang adaptif terhadap perubahan kondisi pasar, serta proses evaluasi yang memerlukan waktu dan tidak efisien. Untuk mengatasi hal ini, penelitian mengusulkan pembangunan sistem optimasi *crossover Moving Average* menggunakan *Proximal Policy Optimization (PPO)*, di mana agen PPO mempelajari performa berbagai kombinasi MA berdasarkan data historis pasangan aset *cryptocurrency*. Akhirnya, sistem menghasilkan rekomendasi kombinasi MA yang paling optimal berdasarkan hasil pelatihan dan validasi PPO, sehingga mendukung pengambilan keputusan analisis yang lebih adaptif, objektif, dan responsif terhadap dinamika pasar.

## 2.2 Cara Mendapatkan Data

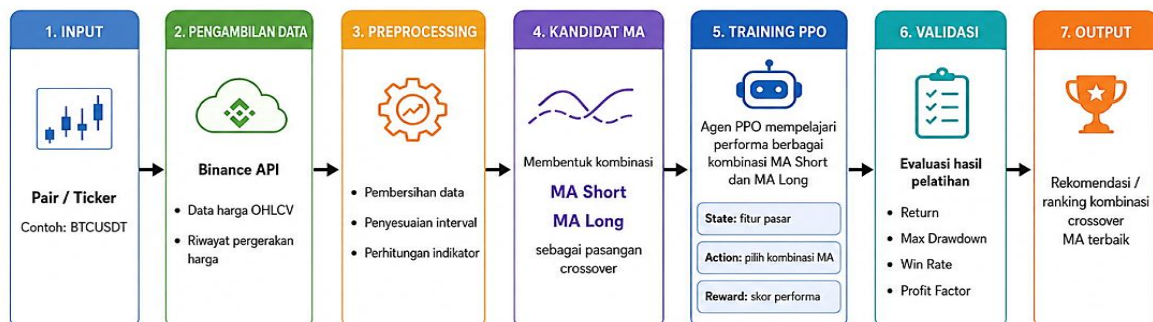
Data dalam penelitian ini diperoleh melalui integrasi dengan Binance API, yang menyediakan akses *real-time* terhadap data pasar *cryptocurrency*. Melalui API ini, sistem mengambil data harga *candlestick (OHLC)*, volume, serta riwayat pergerakan harga yang diperlukan untuk membangun *state* pasar dan melakukan evaluasi berbagai kombinasi *crossover Moving Average* dalam proses pembelajaran agen *Proximal Policy Optimization (PPO)*. Pengambilan data dilakukan secara otomatis menggunakan skrip *Python* dengan library *python-binance*, sehingga proses pengumpulan data berjalan efisien, terstruktur, dan siap digunakan langsung untuk *training* dan *validation* model PPO. Data pasar yang diperoleh kemudian digunakan untuk mengevaluasi performa kombinasi MA dan menghasilkan rekomendasi parameter *crossover* yang optimal berdasarkan pengalaman pembelajaran PPO.

## 2.3 Arsitektur Sistem

Arsitektur sistem dalam penelitian ini dirancang untuk menghasilkan rekomendasi kombinasi *crossover Moving Average* yang optimal berdasarkan pengalaman pembelajaran agen *Proximal Policy Optimization (PPO)*. Sistem ini memungkinkan pemilihan kombinasi *MA Short* dan *MA Long* secara adaptif, sehingga strategi yang dihasilkan mampu menyesuaikan diri dengan dinamika pasar *cryptocurrency* yang cepat dan kompleks. Secara menyeluruh, arsitektur sistem terdiri dari beberapa tahap utama yang saling terintegrasi, dimulai dari input pasangan aset (*pair/ticker*) yang menjadi fokus analisis. Setelah input diterima, data harga historis diambil melalui Binance API, mencakup informasi *candlestick (OHLCV)* dan riwayat pergerakan harga, yang kemudian diproses melalui tahap *preprocessing* untuk pembersihan data, penyesuaian interval, dan perhitungan indikator yang relevan sebagai *state* bagi agen PPO.

# ARSITEKTUR SISTEM

Optimasi Kombinasi Moving Average Crossover dengan PPO



Gambar 2. Arsitektur Sistem

Gambar 2 mengilustrasikan arsitektur sistem secara menyeluruh, yang terdiri dari beberapa tahap utama. Pertama, pengguna memasukkan pasangan aset (*pair/ticker*) yang ingin dianalisis, misalnya BTC/USDT. Selanjutnya, sistem mengambil data harga melalui *Binance API*, meliputi *candlestick (OHLCV)* dan riwayat pergerakan harga. Setelah data diperoleh, dilakukan tahap *preprocessing*, termasuk pembersihan data, penyesuaian interval, dan perhitungan indikator yang relevan.

Tahap berikutnya membentuk kandidat kombinasi *Moving Average (MA Short* dan *MA Long)* yang akan dievaluasi. Agen PPO kemudian melakukan *training* menggunakan *state* pasar yang mencakup harga, *MA Short*, *MA*



*Long*, volatilitas, dan informasi lainnya, dengan *action* berupa pemilihan kombinasi MA dan *reward* berdasarkan skor performa. Setelah proses *training* selesai, sistem melakukan validasi untuk mengevaluasi performa kombinasi MA yang telah dipelajari, menggunakan metrik seperti *total return*, *Sharpe ratio*, *maximum drawdown*, *win rate*, dan *profit factor*.

Tahap akhir menghasilkan output berupa rekomendasi atau ranking kombinasi MA yang paling optimal berdasarkan hasil pembelajaran PPO, sehingga memudahkan pengguna dalam menentukan strategi *crossover* yang adaptif dan responsif terhadap dinamika pasar *cryptocurrency*.

### 2.3.1 Pengambilan Data

Pengambilan data merupakan tahap penting yang memastikan model pembelajaran dapat menangkap pola pasar *cryptocurrency* secara akurat, di mana penelitian ini mengadopsi data harga historis dan *real-time* dari Binance API sebagai sumber utama. Pendekatan serupa telah digunakan dalam studi yang memanfaatkan data pasar *cryptocurrency* secara langsung untuk tugas analisis dan pemodelan, termasuk penerapan skrip Python untuk mengambil data OHLC dari API dan menyusunnya menjadi dataset yang siap (Li, 2025). Selain itu, penggunaan data *real-time* telah diperlihatkan membantu dalam *deep learning* dan *machine learning* terkait model pasar, sehingga memungkinkan pemahaman dinamika harga yang lebih responsif (Tsoku & Makatjane, 2026). Proses pengambilan data dalam penelitian ini dikelola secara otomatis menggunakan skrip Python yang memanggil Binance API berdasarkan *pair/ticker* yang diinginkan, mengambil rangkaian harga yang lengkap dan lengkap untuk disiapkan ke tahap *preprocessing* dan evaluasi lebih lanjut oleh agen PPO, mengingat data yang representatif diperlukan agar metode pembelajaran dapat menghasilkan kombinasi *Moving Average* optimal berdasarkan pengalaman pasar.

### 2.3.2 Preprocessing

Tahap *preprocessing* merupakan langkah penting untuk menyiapkan data pasar *cryptocurrency* agar siap digunakan dalam pelatihan agen *Proximal Policy Optimization (PPO)*. Setelah data harga OHLCV diperoleh dari Binance API, sistem melakukan pembersihan dan penyesuaian data, termasuk penghapusan nilai yang hilang, normalisasi harga, serta transformasi interval waktu agar konsisten. Selanjutnya, indikator *Moving Average* untuk *MA Short* dan *MA Long* dihitung untuk setiap kombinasi kandidat yang akan dievaluasi, sehingga membentuk *state* pasar yang komprehensif dan representatif. Proses ini memastikan bahwa input yang diberikan kepada agen PPO bersih, konsisten, dan relevan untuk proses *training* dan *validation*, sehingga model dapat belajar pola pasar dengan akurat dan menghasilkan rekomendasi kombinasi MA yang optimal. Pendekatan serupa dalam *preprocessing* data *cryptocurrency* telah diterapkan dalam penelitian terkini untuk mendukung *machine learning* dan *deep learning*, menunjukkan pentingnya pembersihan dan transformasi data sebagai fondasi performa model (Tsoku & Makatjane, 2026).

### 2.3.3 Kandidat MA

Tahap Kandidat MA merupakan langkah krusial dalam penelitian ini di mana berbagai kombinasi nilai *Moving Average (MA)* dibentuk dan menjadi elemen yang akan dievaluasi oleh agen *Proximal Policy Optimization (PPO)*. Tujuan dari tahap ini adalah menyediakan sekumpulan pasangan *MA Short* dan *MA Long* yang mungkin berpotensi menghasilkan sinyal *crossover* yang efektif berdasarkan karakteristik pasar *cryptocurrency*. Kombinasi tersebut dibentuk dalam rentang nilai tertentu untuk *MA Short* dan *MA Long*, lalu setiap pasangan akan menjadi kandidat *state* yang dinilai oleh mekanisme pembelajaran PPO dalam proses *training* dan *validation*.

Strategi *Crossover Moving Average* umumnya menggunakan suatu pasangan MA di mana MA dengan periode pendek (*MA Short*) melintasi MA dengan periode panjang (*MA Long*) sebagai sinyal perubahan tren (*trend reversal*). Secara matematis, *Simple Moving Average (SMA)* dihitung sebagai rata-rata bergerak dari harga penutupan selama *n* periode:

$$SMA_t(n) = \frac{1}{n} \sum_{i=0}^{n-1} S_{t-i} \quad (1)$$

Di mana  $P_t$  adalah harga penutupan pada waktu  $t$  dan  $n$  adalah jumlah periode. Setiap pasangan kandidat dipilih dengan syarat  $MA_{short} < MA_{long}$  agar sinyal *crossover* bermakna. Agen PPO kemudian mengevaluasi setiap pasangan berdasarkan performa historisnya melalui *state* dan *reward*, sehingga sistem dapat merekomendasikan kombinasi MA yang paling optimal. Pendekatan ini didukung oleh penelitian (Chiu et al., 2023), yang menunjukkan bahwa eksplorasi berbagai kombinasi parameter *moving average* melalui visualisasi dan analisis data besar dapat membantu mengidentifikasi strategi yang lebih efektif, serta oleh (Abbasi et al., 2024) yang mengevaluasi berbagai aturan teknikal termasuk *moving average*, menunjukkan bahwa variasi parameter teknikal berpengaruh signifikan terhadap performa sinyal di pasar *cryptocurrency* yang volatil (Abbasi et al., 2024; Chiu et al., 2023).

### 2.3.4 Training PPO

Tahap *training* PPO merupakan inti dari proses pembelajaran dalam penelitian ini, di mana agen *Proximal Policy Optimization (PPO)* mempelajari strategi kombinasi *Moving Average* optimal berdasarkan *state* pasar yang dihasilkan dari tahap *preprocessing* dan kandidat MA. *State* mencakup harga historis, nilai *MA Short* dan *MA Long*, serta indikator volatilitas, sedangkan *action* berupa pemilihan pasangan kandidat MA yang akan dievaluasi. Algoritma PPO bertujuan untuk memaksimalkan *expected cumulative reward*  $J(\theta)$  dengan memperbarui parameter kebijakan  $\theta$  secara stabil menggunakan *clipped surrogate objective*. Fungsi objektif PPO dapat ditulis sebagai berikut:



$$L^{CLIP}(\theta) = \widehat{E}_t[\min(r_t(\theta)\widehat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\widehat{A}_t)] \quad (2)$$

Di mana  $r_t(\theta) = \frac{\pi_{\theta}(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$  adalah *probability ratio* antara kebijakan baru dan lama,  $\widehat{A}_t$  adalah *advantage estimate*, dan  $\epsilon$  adalah parameter clipping yang membatasi perubahan kebijakan agar tidak terlalu ekstrem. *Advantage estimate* dihitung menggunakan *Generalized Advantage Estimation (GAE)*:

$$\widehat{A}_t = \delta_t + (\gamma\lambda)\delta_{t+1} + \dots + (\gamma\lambda)^{T-t+1}\delta_{T-1} \quad (3)$$

$$\delta_t = r_t + \gamma V(s_{t+1}) - V(s_t) \quad (4)$$

Di mana  $\gamma$  adalah *discount factor*,  $r_t$  adalah *reward* pada langkah  $t$ ,  $V(s_t)$  adalah nilai fungsi (*value function*) pada state  $s_t$ , dan  $\lambda$  adalah parameter *GAE* yang mengontrol bias-variance trade-off.

Selama proses training, agen *Proximal Policy Optimization (PPO)* mempelajari hubungan kompleks antara berbagai kombinasi parameter *Moving Average (MA Short dan MA Long)* dengan performa historis yang tercermin dalam fungsi *reward*, sehingga setelah konvergensi agen mampu mengevaluasi dan merekomendasikan pasangan MA yang paling optimal untuk menghasilkan sinyal *crossover* yang menguntungkan. Pendekatan ini memungkinkan sistem mengeksplorasi ruang parameter secara adaptif melalui interaksi langsung dengan data pasar, menyesuaikan kebijakan berdasarkan pengalaman yang diperoleh, serta menghasilkan strategi trading yang lebih *robust* di pasar *cryptocurrency* yang sangat volatil. Penelitian terbaru menunjukkan bahwa *deep reinforcement learning* berbasis PPO memberikan performa yang lebih responsif terhadap dinamika pasar yang cepat berubah, seperti penerapan PPO pada prediksi nilai BNB *cryptocurrency* yang berhasil mengidentifikasi pola tersembunyi antara volume trading dan harga aset (Firsov et al., 2023), studi konstruksi strategi transaksi Bitcoin berbasis PPO dengan LSTM yang menghasilkan profit rate signifikan melebihi benchmark (Liu et al., 2021), serta strategi trading saham berbasis PPO yang berhasil melampaui indeks pasar baik dari segi return maupun Sharpe Ratio (Xia et al., 2023).

### 3. HASIL DAN PEMBAHASAN

Pengujian dilakukan untuk mengevaluasi kinerja sistem dalam merekomendasikan kombinasi *crossover Moving Average (MA Short dan MA Long)* yang optimal menggunakan agen *Proximal Policy Optimization (PPO)*. Data historis pasangan aset *cryptocurrency* diolah melalui proses *preprocessing*, kemudian setiap kandidat MA dievaluasi selama *training* dan validasi PPO. Hasil pengujian diekstraksi dalam beberapa bentuk ringkasan dataset yang menunjukkan kondisi harga dan volume, data OHLCV terbaru untuk mengilustrasikan pergerakan pasar, ranking kombinasi MA terbaik berdasarkan skor PPO, dan formula *score* yang digunakan untuk mengevaluasi performa masing-masing kombinasi. Analisis ini memungkinkan penilaian objektif terhadap efektivitas setiap kombinasi MA, serta membantu mengidentifikasi strategi *crossover* yang adaptif dan paling optimal di pasar *cryptocurrency* yang sangat volatil.

#### 3.1 Parameter Pengujian PPO

Pengujian dilakukan menggunakan *Proximal Policy Optimization (PPO)* pada pasangan BTC/USDT dengan data harian (1d) dari 1 Januari 2021 hingga 15 Mei 2026, dengan jumlah timesteps sebesar 5.000 dan proporsi data training 70 persen agar agen memiliki cukup data historis untuk mempelajari pola pasar yang relevan (Gort et al., 2023; Purnawa et al., 2022; Yang et al., 2025). Konfigurasi manajemen risiko ditetapkan melalui modal awal 100 USD, risiko per transaksi 2 USD, dan rasio risiko-untung 1,5 untuk menjaga keseimbangan antara potensi keuntungan dan pengendalian drawdown, sejalan dengan praktik *deep reinforcement learning* untuk trading yang menekankan stabilitas portofolio dan evaluasi strategi yang reliabel (Gort et al., 2023; Yang et al., 2025).

State lingkungan dibentuk menggunakan jendela lookback sepanjang 120 hari dan chunk size 240, sedangkan volatilitas diukur dengan indikator Average True Range (ATR) periode 14 dan multiplier 1,5 untuk menetapkan batas stop-loss yang adaptif terhadap dinamika pasar, sebagaimana digunakan dalam perancangan agen trading berbasis *deep reinforcement learning* pada aset kripto (Gort et al., 2023; Purnawa et al., 2022). Strategi exit memprioritaskan eksekusi stop-loss sehingga konfigurasi parameter ini mendukung proses training dan evaluasi PPO yang konsisten, serta memungkinkan agen mempelajari hubungan antara kombinasi *Moving Average (MA Short dan MA Long)*, volatilitas, dan reward historis secara lebih terstruktur pada pasar kripto (Gort et al., 2023; Purnawa et al., 2022; Yang et al., 2025).

#### 3.2 Dataset Yang Digunakan

Dataset yang digunakan dalam pengujian mencakup seluruh harga harian BTC/USDT dari 1 Januari 2021 hingga 15 Mei 2026. Data ini diambil melalui *Binance API* dan mencakup informasi OHLCV (*open, high, low, close, volume*) untuk setiap hari. Total data yang diperoleh mencakup 1960 *candle* dalam periode tersebut, sehingga memberikan basis data yang cukup untuk analisis kombinasi *Moving Average* dan *training PPO*. Data terakhir yang digunakan dalam pengujian menampilkan harga penutupan terbaru, volume, serta informasi volatilitas yang relevan untuk pembentukan



state dan evaluasi kombinasi *Moving Average*. Data ini selanjutnya digunakan untuk *preprocessing* dan sebagai input ke agen PPO dalam melakukan *training* dan validasi strategi. Berikut adalah contoh tabel ringkasan OHLCV dari beberapa hari terakhir:

**Tabel 1.** Dataset BTC/USDT

Tanggal	Pembukaan	Tertinggi	Terendah	Penutupan	Volume
05-05-2026	79826.1	81745.4	79761.2	80861.9	170921.528
06-05-2026	80861.9	82828.7	80651.2	81390.1	158200.508
07-05-2026	81390	81680.6	79461.5	79969.4	134727.652
08-05-2026	79969.5	80478	79137.4	80155.2	106762.913
09-05-2026	80155.2	81043.2	80080	80634	55957.334
10-05-2026	80634	82460.5	80225.1	82177.7	94950.803
11-05-2026	82177.7	82350	80394	81699.1	121990.421
12-05-2026	81699.1	81757.3	79801	80461	121313.956
13-05-2026	80461	81270.2	78713.8	79288.3	139700.645
14-05-2026	79288.2	81999	78869.4	81403.7	171736.243

Berdasarkan Tabel 1, harga harian BTC/USDT dalam sepuluh hari terakhir menunjukkan variasi yang konsisten dengan karakteristik volatilitas tinggi pada pasar kripto, di mana perubahan return dan volume perdagangan berkontribusi pada dinamika volatilitas jangka pendek (Kuncara & Anugrah, 2023; Rahmat & Iswardhani, 2025). Temuan ini sejalan dengan studi yang menunjukkan bahwa return dan variabel mikrostruktur pasar kripto berperan penting dalam membentuk volatilitas indeks maupun aset kripto utama (Albana et al., 2026; Kuncara & Anugrah, 2023). Pergerakan harga dan volume pada rentang sepuluh hari tersebut menjadi dasar pembentukan state dalam pelatihan PPO, karena setiap observasi harga, volume, dan indikator volatilitas dimasukkan ke dalam sistem sebagai input yang memungkinkan agen belajar mengenali pola tren dan respon pasar (Rahmat & Iswardhani, 2025). Selain itu, data ini memberikan landasan empiris bagi pengujian kandidat kombinasi *Moving Average* pada berbagai kondisi pasar, sehingga evaluasi tidak hanya bertumpu pada performa rata-rata, tetapi juga pada sensitivitas strategi terhadap fluktuasi volatilitas dan perubahan volume, sebagaimana ditekankan dalam penelitian mengenai hubungan volatilitas, volume, dan imbal hasil Bitcoin (Rahmat & Iswardhani, 2025). Dengan demikian, tabel ini memperkuat peran data historis harga harian, volume perdagangan, dan dinamika volatilitas sebagai landasan analisis crossover *Moving Average* sekaligus sebagai basis empiris bagi interpretasi hasil pengujian strategi trading berbasis deep reinforcement learning pada aset kripto (Albana et al., 2026; Kuncara & Anugrah, 2023).

### 3.3 Ranking PPO dari Crossover *Moving Average*

Pengujian dilakukan untuk mengevaluasi kinerja agen *Proximal Policy Optimization (PPO)* dalam memilih kombinasi *Moving Average (MA Short dan MA Long)* yang paling optimal. Hasil pengujian ditampilkan dalam bentuk *ranking PPO*, yang diurutkan berdasarkan nilai *mean PPO probability* tertinggi dari setiap pasangan MA. Artinya, PPO mengevaluasi setiap segmen pasar dan memberikan probabilitas pada semua kandidat kombinasi MA, seperti MA 9/21, MA 20/50, dan seterusnya. Probabilitas ini kemudian dirata-ratakan untuk setiap kandidat, sehingga rank 1 menunjukkan pasangan MA yang paling “dipercaya” oleh policy PPO dalam segmen evaluasi.

**Tabel 2.** Ranking MA PPO

Rank	Pair MA	Short MA	Long MA	Mean Probability	Max Probability	Min Probability
1	MA 3/50	3	50	20.9291	53.2187	4.1566
2	MA 25/40	25	40	10.7432	17.4419	1.795
3	MA 20/30	20	30	8.2816	10.2311	5.2475
4	MA 5/50	5	50	4.6245	10.9988	1.387
5	MA 18/30	18	30	4.348	6.9949	0.5293

Berdasarkan Tabel 2, dapat dilihat bahwa pasangan MA 3/50 memperoleh *mean PPO probability* tertinggi, menunjukkan bahwa kombinasi ini paling dipercaya oleh agen PPO dalam evaluasi setiap segmen pasar. Ranking berikutnya, seperti MA 25/40 dan MA 20/30, menunjukkan performa yang lebih rendah, namun tetap memiliki probabilitas signifikan sehingga menjadi kandidat alternatif yang potensial. Interpretasi ini memperlihatkan bagaimana PPO menilai setiap kombinasi MA secara adaptif berdasarkan pengalaman historis, dengan *mean probability* sebagai indikator utama preferensi agen. Selain itu, perbedaan antara *max* dan *min PPO probability* pada setiap pasangan MA memberikan wawasan mengenai konsistensi pemilihan kombinasi dalam berbagai segmen pasar: pasangan dengan rentang probabilitas yang lebih sempit cenderung dipilih lebih konsisten, sementara pasangan dengan rentang lebar menunjukkan fluktuasi preferensi agen di berbagai kondisi pasar. Analisis ini menjadi dasar untuk memilih kombinasi MA yang tidak hanya optimal secara rata-rata tetapi juga stabil dalam performanya di pasar *cryptocurrency* yang volatil.

### 3.4 Analisis Kinerja Strategi

Analisis kinerja strategi dilakukan dengan meninjau kombinasi MA berdasarkan ranking PPO. Fokus evaluasi menggunakan metrik yang mencerminkan performa strategi:

1. *Win Rate (%)*: persentase transaksi yang menghasilkan keuntungan; indikator konsistensi strategi.
2. *Maximum Drawdown (%)*: penurunan modal terbesar selama periode pengujian; menggambarkan risiko terbesar yang dihadapi.
3. *Profit Factor*: rasio total keuntungan terhadap total kerugian; indikator efisiensi strategi.
4. *Total Trades*: jumlah transaksi yang dieksekusi selama periode pengujian; menunjukkan frekuensi strategi dalam memanfaatkan sinyal.
5. *Net Profit (USD)*: keuntungan bersih akhir dari seluruh transaksi; ukuran hasil finansial strategi.

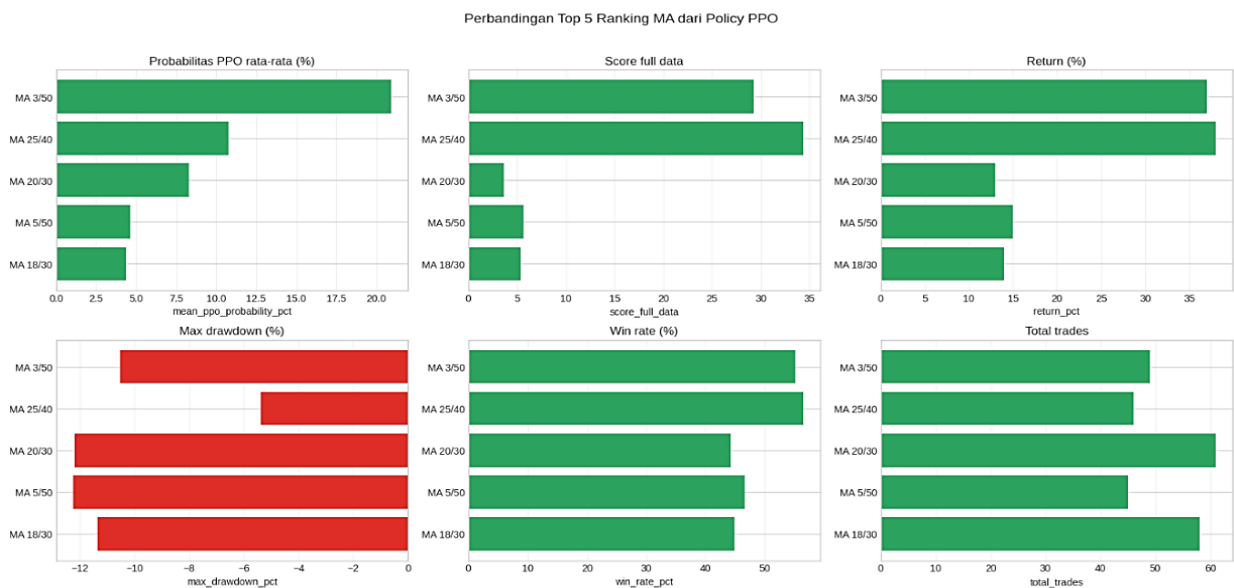
**Tabel 3.** Evaluasi *Crossover MA*

Rank	Pair MA	Win Rate (%)	Max Drawdown (%)	Profit Factor	Total Trades	Net Profit (USD)
1	MA 3/50	55.10	-10.53	1.84	49	137
2	MA 25/40	56.52	-5.38	1.95	46	138
3	MA 20/30	44.26	-12.20	1.19	61	113
4	MA 5/50	46.67	-12.26	1.31	45	115
5	MA 18/30	44.83	-11.38	1.22	58	114

Berdasarkan Tabel 3, MA 25/40 memberikan hasil terbaik pada periode pengujian apabila penilaian difokuskan pada kinerja finansial dan risiko, yaitu net profit 138 USD, profit factor 1,95, win rate 56,52%, serta maximum drawdown -5,38%. Sebaliknya, MA 3/50 menempati peringkat pertama berdasarkan mean PPO probability, tetapi memiliki net profit lebih rendah, profit factor lebih kecil, dan drawdown lebih besar dibandingkan MA 25/40. Perbedaan ini menunjukkan bahwa preferensi policy PPO belum identik dengan kombinasi yang paling menguntungkan dan paling rendah risiko. Oleh sebab itu, ranking PPO perlu dibaca bersama metrik kinerja aktual, bukan digunakan sebagai satu-satunya dasar rekomendasi strategi crossover Moving Average pada pasar cryptocurrency yang volatil.

### 3.5 Analisis Visual Ranking dan Performa MA PPO

Hasil evaluasi *ranking* PPO menunjukkan perbedaan signifikan antara kombinasi *Moving Average* yang paling sering dipilih oleh agen dan performa finansial aktual setiap kombinasi. Dari sisi *mean PPO probability*, MA 3/50 menjadi kombinasi yang paling sering dipilih oleh PPO, menunjukkan bahwa agen memberikan preferensi adaptif terhadap kombinasi ini dalam berbagai segmen pasar. Namun, jika dilihat dari metrik *return*, *profit factor*, dan *max drawdown*, MA 25/40 menunjukkan performa finansial yang lebih optimal, sedangkan MA 3/50 memiliki *max drawdown* lebih besar, menandakan risiko yang lebih tinggi.



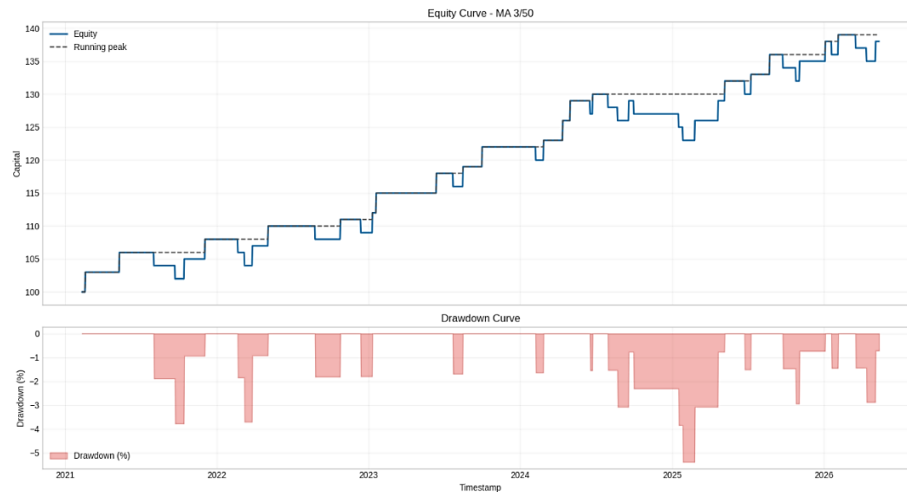
**Gambar 3.** Perbandingan top 5 ranking

Analisis lanjutan dari grafik memperlihatkan bahwa setiap kombinasi Moving Average memiliki karakteristik risiko, frekuensi transaksi, dan performa yang berbeda. MA 25/40 menghasilkan win rate dan profit factor tertinggi sekaligus drawdown terendah pada lima kandidat yang ditampilkan, sehingga lebih kuat sebagai kandidat strategi pada data pengujian ini. Sementara itu, tingginya mean PPO probability pada MA 3/50 menunjukkan preferensi agen, tetapi belum membuktikan keunggulan finansial. Visualisasi tersebut mempertegas perlunya menggabungkan skor kebijakan

PPO dengan evaluasi net profit, profit factor, dan maximum drawdown sebelum suatu kombinasi MA direkomendasikan.

### 3.6 Evaluasi Pertumbuhan Modal dan Implikasi Strategi

Hasil evaluasi kinerja MA 3/50 ditunjukkan melalui *equity curve* dan *drawdown curve* yang merepresentasikan perkembangan modal selama periode pengujian. Grafik ini menampilkan bagaimana modal awal 100 USD berkembang seiring waktu berdasarkan sinyal yang dihasilkan oleh agen PPO, serta menunjukkan fluktuasi drawdown sebagai indikator risiko terbesar yang dialami selama periode tersebut. *Equity curve* menunjukkan tren naik secara konsisten dengan beberapa penurunan sementara, sedangkan *drawdown curve* menyoroti periode penurunan modal maksimum, yang penting untuk menilai stabilitas strategi.



**Gambar 4.** *Equity Curve* dan *Drawdown*

Dari Gambar 4, *equity curve* MA 3/50 memperlihatkan pertumbuhan modal selama periode pengujian, tetapi interpretasinya perlu dikaitkan dengan Tabel 3. Walaupun MA 3/50 memiliki probabilitas pemilihan PPO tertinggi, maximum drawdown-nya (-10,53%) masih lebih besar daripada MA 25/40 (-5,38%). Dengan demikian, grafik ini menunjukkan bahwa MA 3/50 dapat menghasilkan pertumbuhan modal dalam simulasi, namun belum menjadi pilihan paling efisien dari sisi risiko dan profitabilitas.

### 3.7 Visualisasi Entry dan Exit pada Strategi MA

Strategi yang digunakan menggabungkan pergerakan harga BTC/USDT harian dengan dua *Moving Average* (MA 25 dan MA 40) serta titik *entry* dan *exit* yang dieksekusi oleh strategi *crossover* yang dipelajari agen PPO. Titik *entry long* ditandai dengan segitiga hijau ke atas, *entry short* dengan segitiga ungu ke bawah, dan *exit* dengan tanda silang merah. Visualisasi ini membantu melihat bagaimana kombinasi MA memicu sinyal beli atau jual secara dinamis, serta bagaimana strategi merespons perubahan tren di pasar *cryptocurrency* yang volatil.

Dari Gambar 5, titik *entry* dan *exit* menunjukkan bagaimana aturan *crossover* MA 25/40 diterjemahkan menjadi keputusan transaksi sepanjang periode pengujian. Visualisasi ini membantu menelusuri hubungan antara perpotongan MA dan eksekusi strategi, tetapi tidak dengan sendirinya membuktikan bahwa agen mampu beradaptasi secara real-time atau mengurangi kerugian pada implementasi nyata. Penilaian terhadap efektivitas strategi tetap harus didasarkan pada metrik pengujian, validasi out-of-sample, serta simulasi biaya transaksi dan slippage.



**Gambar 5.** *Entry/Exit* MA 25/40



### 3.8 Pembahasan

Hasil penelitian menunjukkan bahwa PPO mampu membedakan preferensi kombinasi Moving Average berdasarkan kondisi pasar, tetapi preferensi probabilitas agen tidak selalu identik dengan performa finansial tertinggi. MA 3/50 memperoleh *mean PPO probability* tertinggi, sedangkan MA 25/40 memberikan hasil finansial lebih baik berdasarkan *profit factor*, *net profit*, dan *maximum drawdown*. Temuan ini menunjukkan bahwa keluaran PPO perlu dibaca sebagai indikator preferensi kebijakan yang adaptif, bukan sebagai satu-satunya ukuran keuntungan. Dengan demikian, evaluasi strategi tetap perlu menggabungkan metrik probabilitas, profitabilitas, risiko, dan stabilitas transaksi.

Temuan tersebut sejalan dengan (Chiu et al., 2023), yang menunjukkan bahwa variasi kombinasi moving average dapat menghasilkan performa berbeda dan tidak ada satu parameter yang selalu unggul untuk seluruh kondisi pasar. Perbedaannya, penelitian Chiu et al. menekankan eksplorasi banyak kombinasi melalui *heatmap*, sedangkan penelitian ini menggunakan PPO untuk mempelajari preferensi kombinasi MA secara otomatis berdasarkan *state* pasar dan *reward* historis. Hasil ini juga mendukung (Romo et al., 2025), yang menekankan pentingnya optimasi adaptif pada strategi dual moving average, tetapi penelitian ini memperluas pendekatan tersebut dengan mekanisme *reinforcement learning* yang memungkinkan proses pemilihan parameter menjadi bagian dari pembelajaran agen.

Dibandingkan dengan penelitian *deep reinforcement learning* sebelumnya, hasil ini memperlihatkan bahwa integrasi indikator teknikal masih relevan ketika indikator tersebut tidak digunakan sebagai aturan statis, tetapi sebagai representasi informasi dalam proses pembelajaran. (Kochliaridis et al., 2023) menunjukkan bahwa kombinasi *deep reinforcement learning*, analisis teknikal, dan mekanisme pengendalian risiko dapat meningkatkan kualitas keputusan trading pada pasar cryptocurrency. (Ospina-Holguín & Padilla-Ospina, 2025) juga menegaskan bahwa *reinforcement learning* dapat dikombinasikan dengan aturan moving average untuk menghasilkan *alpha* yang lebih optimal. Penelitian ini memperkuat arah tersebut karena PPO tidak hanya memakai moving average sebagai sinyal tambahan, tetapi mengevaluasi langsung beberapa kombinasi MA sebagai kandidat strategi.

Keterkaitan antara volume, volatilitas, dan return yang dijelaskan oleh (Albana et al., 2026) juga terlihat dalam penelitian ini melalui pentingnya volatilitas dan perubahan harga sebagai dasar pembentukan *state*. Pada pasar yang sangat volatil, kombinasi MA yang terlalu responsif dapat menangkap peluang lebih cepat, tetapi berisiko meningkatkan *drawdown*, sedangkan kombinasi yang lebih moderat dapat menghasilkan performa lebih stabil. Karena itu, hasil MA 25/40 yang unggul dari sisi *profit factor* dan *drawdown* menunjukkan bahwa strategi adaptif tidak cukup dinilai dari frekuensi sinyal, tetapi juga dari kualitas risiko yang dihasilkan. Namun, sebagaimana diingatkan oleh (Gort et al., 2023), hasil *backtest* pada cryptocurrency tetap perlu diuji dengan biaya transaksi, *slippage*, interval data berbeda, dan pasangan aset lain agar generalisasi model lebih kuat.

## 4. KESIMPULAN

Kesimpulan penelitian ini menunjukkan bahwa PPO dapat digunakan untuk memeringkat kandidat parameter crossover Moving Average pada BTC/USDT, tetapi preferensi policy tidak otomatis identik dengan hasil finansial terbaik. MA 3/50 memperoleh *mean PPO probability* tertinggi, sedangkan MA 25/40 memberikan performa simulasi yang lebih baik berdasarkan *net profit*, *profit factor*, *win rate*, dan *maximum drawdown*. Temuan tersebut menunjukkan perlunya menyelaraskan fungsi reward dengan tujuan keuntungan bersih dan pengendalian risiko. Karena pengujian masih menggunakan satu pasangan aset, data harian historis, dan belum memasukkan biaya transaksi maupun *slippage*, hasil profitabilitas belum dapat dianggap merepresentasikan kinerja live trading. Penelitian selanjutnya perlu menggunakan validasi out-of-sample dan walk-forward, beberapa aset dan interval waktu, pengujian beberapa random seed, serta reward yang memasukkan friksi transaksi dan risiko agar rekomendasi strategi lebih robust.

## REFERENCES

- Abbasi, M., Al-Sadat Mousavi, S., Author, C., & Jafari-Nodoushan Assistant, A. (2024). Identification and Evaluation of Profitable Technical Trading Rules in the Cryptocurrency Market: A Mixed Method Approach. *Persian Financial Research Journal*, 26(3), 525–546. <https://doi.org/10.22059/FRJ.2024.358549.1007462>
- Albana, M. F. A., Safitri, M., Waluyo, D. E., & Chasanah, A. N. (2026). Pengaruh Volume, Volatilitas, dan Market Cap terhadap Return Bitcoin, Ethereum, BNB 2022-2024. *Jurnal Manajemen Bisnis Dan Keuangan*, 7(1). <https://doi.org/10.51805/jmbk.v7i1.500>
- Bandarupalli, E. (2025). *Risk-Aware Deep Reinforcement Learning for Crypto and Equity Trading Under Transaction Costs*. <https://doi.org/10.2139/ssrn.5662930>
- Chiu, C. L., Ni, Y., Hu, H. C., Day, M. Y., & Chen, Y. (2023). Enhancing Crypto Success via Heatmap Visualization of Big Data Analytics for Numerous Variable Moving Average Strategies. *Applied Sciences (Switzerland)*, 13(23). <https://doi.org/10.3390/app132312805>
- Edy Prasetyo, R., Chaidir, I., & Supriyatna, A. (2025). Reinforcement learning for bitcoin trading: A comparative study of PPO and DQN. *Jurnal Mandiri IT*, 14(2), 159–169. <https://doi.org/10.35335/mandiri.v14i2.455>
- Firsov, D. V., Silvestrov, S. N., Kuznetsov, N. V., Zolotarev, E. V., & Pobyvaev, S. A. (2023). Using PPO Models to Predict the Value of the BNB Cryptocurrency. *Emerging Science Journal*, 7(4), 1206–1214. <https://doi.org/10.28991/ESJ-2023-07-04-012>



- Gort, B. J. D., Liu, X.-Y., Sun, X., Gao, J., Chen, S., & Wang, C. D. (2023). Deep Reinforcement Learning for Cryptocurrency Trading: Practical Approach to Address Backtest Overfitting. *ArXiv*. <https://doi.org/10.48550/arXiv.2209.05559>
- Khujamatov, E. H., Ismanov, K., Mallaev, O. U., & Sattarov, O. (2026). Optimizing Crypto-Trading Performance: A Comparative Analysis of Innovative Reward Functions in Reinforcement Learning Models. *Mathematics*, *14*(5). <https://doi.org/10.3390/math14050794>
- Kochliaridis, V., Kouloumpiris, E., & Vlahavas, I. (2023). Combining deep reinforcement learning with technical analysis and trend monitoring on cryptocurrency markets. *Neural Computing and Applications*, *35*(29), 21445–21462. <https://doi.org/10.1007/s00521-023-08516-x>
- Kuncara, T., & Anugrah, K. P. (2023). Analisis Volatilitas Cryptocurrency Sebelum Pandemi dan Saat Pandemi Covid-19 dengan Metode Return pada Bitcoin dan Ethereum. *KEUNIS*, *11*(1), 86. <https://doi.org/10.32497/keunis.v11i1.3981>
- Li, X. (2025). Forecasting and Anomaly Detection in Bitcoin Historical Data. *Applied and Computational Engineering*, *139*, 1–9. <https://doi.org/10.54254/2755-2721/139/2025.21901>
- Liu, F., Li, Y., Li, B., Li, J., & Xie, H. (2021). *Bitcoin Transaction Strategy Construction Based on Deep Reinforcement Learning*. *Applied Soft Computing*, *113*(December), doi: <https://doi.org/10.1016/j.asoc.2021.107952>
- Ospina-Holguín, J. H., & Padilla-Ospina, A. M. (2025). Reinforcement learning meets technical analysis: combining moving average rules for optimal alpha. *Cogent Economics and Finance*, *13*(1). <https://doi.org/10.1080/23322039.2025.2490818>
- Purnawa, I. B. D. P., Dharna, E. M., & Sudiatmika, I. B. K. (2022). Trading Cryptocurrency Agent Menggunakan Deep Reinforcement Learning dengan Model yang Berbasis Q-Learning Menggunakan DQN, Double DQN, dan Dueling DQN pada Studi Kasus Bitcoin. *Jurnal Teknologi Informasi Dan Komputer*, *8*(3). <https://doi.org/10.36002/jutik.v8i3.2075>
- Rahmat, M. R. A., & Iswardhani, I. (2025). Volatilitas Cryptocurrency di Indonesia: Investasi Keuangan Pada Bitcoin Sebelum dan Sesudah Peresmian Bursa Kripto. *SYNERGY: Jurnal Bisnis Dan Manajemen*, *5*(1), 65–69. <https://doi.org/10.52364/synergy.v5i1.65>
- Romo, A., Soto, R., Vega, E., Crawford, B., Salinas, A., & Becerra-Rozas, M. (2025). Adaptive Optimization of a Dual Moving Average Strategy for Automated Cryptocurrency Trading. *Mathematics*, *13*(16). <https://doi.org/10.3390/math13162629>
- Tsoku, J. T., & Makatjane, K. (2026). Deep learning-based pairs trading: real-time forecasting of co-integrated cryptocurrency pairs. *Frontiers in Applied Mathematics and Statistics*, *12*. <https://doi.org/10.3389/fams.2026.1749337>
- Xia, Z., Shi, M., & Lin, C. (2023). Stock Trading Strategy Developing Based on Reinforcement Learning. *Proceedings of the 2nd International Academic Conference on Blockchain, Information Technology and Smart Finance (ICBIS 2023)*, 156–164. [https://doi.org/10.2991/978-94-6463-198-2\\_18](https://doi.org/10.2991/978-94-6463-198-2_18)
- Yang, H., Liu, X.-Y., Zhong, S., & Walid, A. (2025). Deep Reinforcement Learning for Automated Stock Trading: An Ensemble Strategy. *ACM Transactions on Intelligent Systems and Technology*. <http://arxiv.org/abs/2511.12120>