


Design Robô	
Nome do Robô	Satoshi
Explicação Nome	Satoshi é uma referência ao criador do Bitcoin, a crypto mais popular entre os investidores
Explicação Lógica Estratégia	A partir de modelos de aprendizado de reforço, é criada uma função que transforma a observação de indicadores utilizados comumente por traders (ex: RSI, MACD) em pesos de um portfólio Long&Short composto por 6 ativos. O modelo utiliza como base de seu aprendizado a maximização do índice de sharpe.
Tipo de Estratégia	Trade
Classe de Ativos	Cripto
Universo	BTCUSDT, ETHUSDT, BNBUSDT, ADAUSDT, XRPUSDT, LTCUSDT
Média Trades por mês	22 trades por mês (mês 22 dias)
Holding Period	1 dia
Qual Plataforma Testou a estratégia	Python
Benchmark Estratégia:	S&P500 e Buy and Hold

1. Definição da Estratégia

Uma das metodologias de construção de estratégias de trade é a análise técnica, que consiste em utilizar apenas dados de tela como preço e volume para se conseguir alguma vantagem sobre o mercado. Desta abordagem, criaram-se indicadores que norteiam os traders a tomarem decisões sobre como reagir a determinado price action. Dentre esses indicadores, temos o RSI (índice de força relativa) e o MACD (Convergência e Divergência de médias móveis). A teoria moderna de portfólio idealizada por Harry Markowitz é um método prático para selecionar investimentos a fim de maximizar seus retornos globais dentro de um nível de risco aceitável. Uma atualização da teoria moderna de portfólio foi introduzida por William F. Sharpe, com o Índice de Sharpe. Esse indicador relaciona os retornos excessivos em relação a um benchmark ponderados pela volatilidade do portfólio.

Modelos de “Policy Gradient” são um tipo de técnica de **Deep Learning** baseados na construção de políticas que levam à maximização de uma função recompensa. Nesses modelos, temos a figura de um **Agente** que irá interagir com um **Estado Observável** através de **Ações**, que o levam a uma **recompensa**. Com isso, a melhor política é aquela que **maximiza a função recompensa**. Nesse sentido, nossa estratégia utiliza os **indicadores clássicos de análise técnica** como o **Estado Observável do Agente** (Satoshi), para que ele decida as **ações** (redistribuir os pesos do portfólio de cripto ativos).

1.1. Estado Observável

Nosso estado observável é composto por 6 indicadores comuns para todos os ativos.

1. **RSI** - O índice de força relativa (RSI) é um indicador de momentum usado na análise técnica. O RSI mede a velocidade e a magnitude das mudanças recentes no preço de um ativo para avaliar as condições sobrevalorizadas ou subvalorizadas no preço. Para o cálculo desse indicador utilizamos uma janela móvel de 14 períodos.
2. **MACD** - A convergência/divergência de médias móveis é um indicador de momentum e tendência que mostra a relação entre duas médias móveis exponenciais (EMAs) do preço de um ativo. Utilizamos a implementação clássica do indicador com a média de longa de 26 períodos e a média curta de 12 períodos.
3. **Diferença de Médias Exponenciais** - Esse indicador é a diferença entre uma média móvel exponencial mais longa (20 períodos) e uma mais curta (5 períodos). Pode ser utilizado para estratégias de trend following ou reversão à média.
4. **Preço Normalizado dos Ativos** - Preço normalizado dos ativos pelo desvio padrão e média, cálculo realizado em uma janela móvel de 20 dias.
5. **DDD** - Daily Drawdown é calculado como a diferença ponderada entre o preço atual (P_{atual}) e máximo (P_{max}) de um ativo para determinado período. Para o cálculo, utilizamos uma janela móvel de 26 períodos.
6. **MDD** - Maximum Drawdown (MDD) é a diferença ponderada entre o preço mínimo (P_{min}) e máximo (P_{max}) de um ativo para determinado período. O Drawdown é um dos principais indicadores de risco de perda (left side volatility). Utilizamos as mesmas condições para cálculo do DDD para o cálculo do MDD.

1.2. Função Recompensa

A função recompensa é uma **média móvel do Índice de Sharpe** do nosso portfólio, a janela é de **20 períodos (dias)**, ou seja para n menor que 20, a recompensa sempre será igual a 0. Nosso robô receberá um feedback positivo quando essa média móvel aumentar de um dia para o outro, da mesma forma o robô receberá um feedback negativo se essa média diminuir.

1.3. Ações

As ações do robô, como dito anteriormente, são a divisão dos pesos de um portfólio de cripto ativos, baseadas na política ótima calibrada pelo modelo. Estes pesos podem ser **positivos ou negativos**, no entanto o seu somatório em módulo precisa ser igual a um. Nesse modelo de alocação não é possível aproveitar nenhuma porcentagem do valor investido como margem, e portanto, não há a possibilidade de se alavancar. Com isso os pesos estão sujeitos a restrição:

$$\sum_{i=1}^n |w_i| = 1$$

Onde:

- w_i é o peso do ativo i no portfólio
- n é a quantidade de ativos no portfólio

No nosso modelo de investimento, a alocação é feita de maneira **diária**, ou seja, todos os dias há a oportunidade de um rebalanceamento do portfólio. Portanto, cada trade tem duração de um dia.

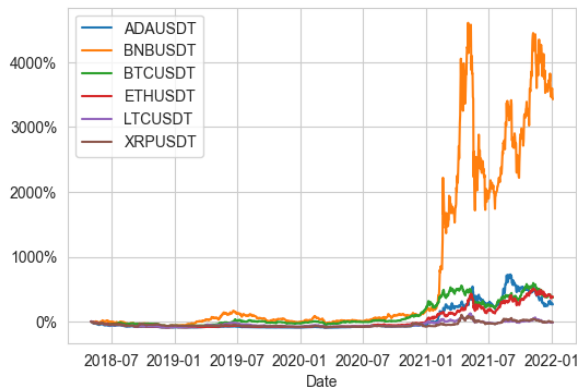
1.4. Modelo de Aprendizado de Reforço

O **Proximal Policy Optimization (PPO)** é um método de gradiente de política popular e eficiente que alcançou resultados impressionantes em vários domínios. É um método on-policy, o que significa que ele usa a mesma política para **exploitation e exploration**. Ele é atualizado iterativamente a política por meio da amostragem de trajetórias do ambiente e do cálculo do gradiente de política, que é a direção que melhora o retorno esperado. No entanto, ao contrário de outros métodos, o PPO não usa uma taxa de aprendizado fixa ou uma região de confiança para controlar o tamanho do treino. Em vez disso, usa uma função de objetivo que penaliza grandes mudanças na política. Dessa forma, o PPO evita o overfit ou o colapso da política para uma solução subótima.

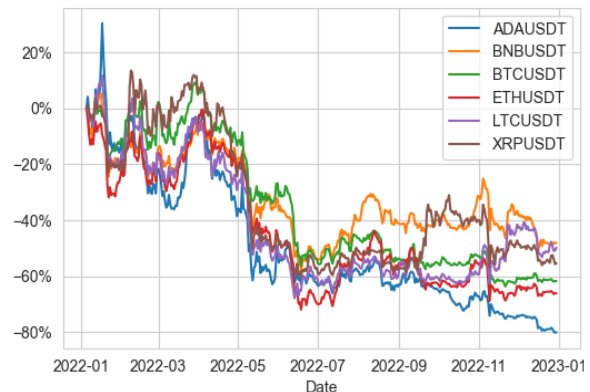
2. Base de Dados

Usamos dados **diários de fechamento do mercado** no período de **2018-05-04 até 2022-12-30**. Esses dados foram obtidos através do **API da corretora Binance** para uma seleção de seis cripto ativos Cardano (**ADA**), Binance Coin (**BNB**), Bitcoin (**BTC**), Ethereum (**ETH**), Litecoin (**LTC**), Ripple (**XRP**). A base monetária negociada é o **USD**. A escolha

destes ativos levou em consideração a quantidade de dados disponíveis e o mkt cap no momento de coleta dos dados. Priorizamos os ativos com mais dados históricos que estivessem no top 100 de mkt cap. Nossos dados foram divididos em um período de **Treino** com 1342 observações e um período de **Teste** com 360 observações totalizando 1702 registros. Abaixo podemos ver a representação gráfica destes retornos, assim como valores anualizados de retorno e risco para os ativos.



Rentabilidade no período de Treino



Rentabilidade no período de Teste

Retorno Atualizado

Período	ADAUSDT	BNBUSDT	BTCUSDT	ETHUSDT	LTCUSDT	XRPUSDT
Treino	97,38%	194,36%	61,88%	87,77%	40,96%	57,87%
Teste	-56,39%	-23,97%	-41,27%	-39,05%	-19,47%	-26,71%

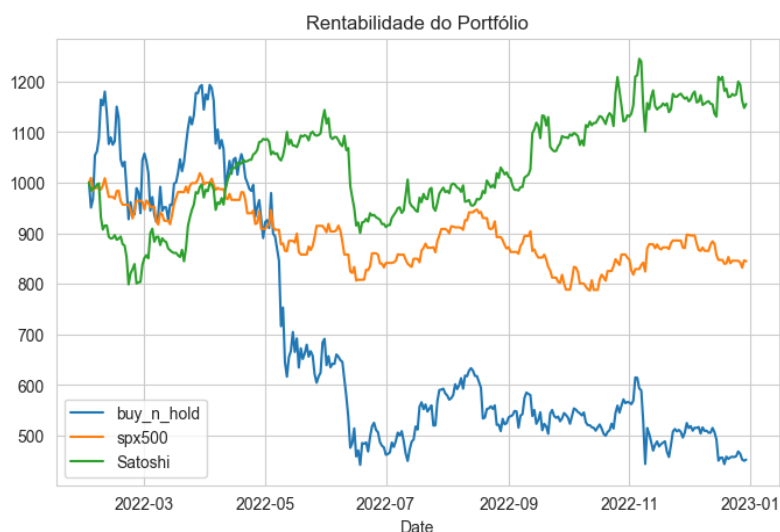
Risco Anualizado

Período	ADAUSDT	BNBUSDT	BTCUSDT	ETHUSDT	LTCUSDT	XRPUSDT
Treino	93,61%	91,69%	60,99%	80,84%	85,49%	98,75%
Teste	77,23%	60,62%	53,13%	72,42%	72,93%	71,70%

3. Resultados Encontrados

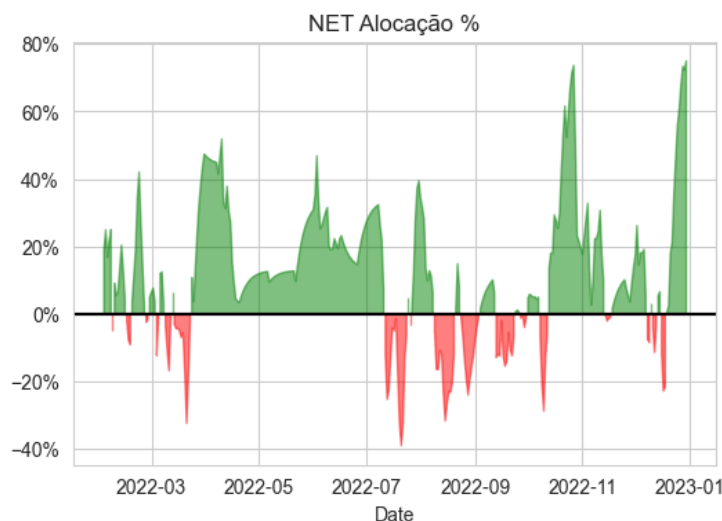
Nessa etapa iremos demonstrar as **medidas de performance da estratégia** junto com todas as métricas relevantes para a avaliação do modelo. Os benchmarks escolhidos são o Índice **S&P500** e um portfólio composto por todos os ativos em proporções iguais

3.1. Performance



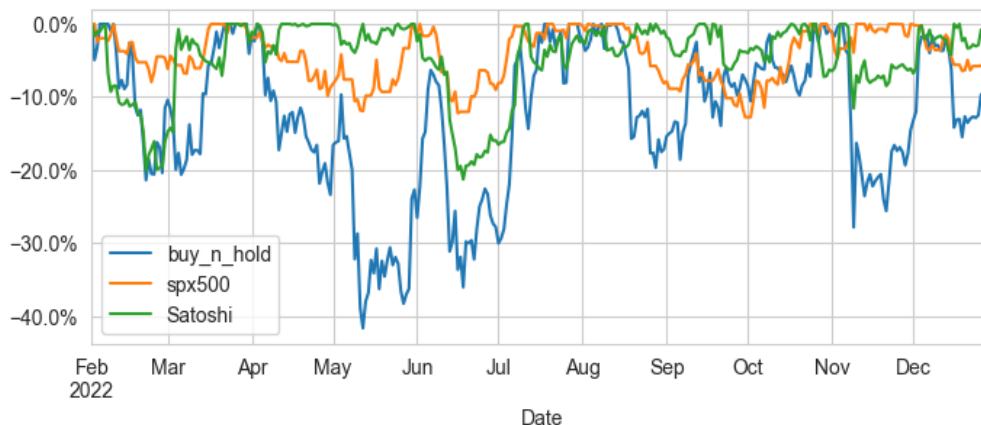
Rentabilidade da Estratégia: Comparação com os benchmarks

A partir do gráfico de rentabilidade constatamos que nosso robô **manteve uma performance acima dos seus pares** por quase todo o intervalo de tempo. Perdendo apenas no início dos nossos testes. É interessante ressaltar que o robô conseguiu antever possíveis correções no mercado cripto evitando o drawdown sofrido após maio de 2022. Porém, quando a correção se estabeleceu nos dois benchmarks, o robô não conseguiu antever esse movimento como o que ocorreu em meados de junho de 2022 em meio ao aumento da taxa de juros americano.



Resultante de Alocação do Portfólio

No gráfico da exposição Long & Short do portfólio, evidencia-se que a alocação Long na maior parte dos dias é preferida pelo robô, apesar de eventuais exposições Net short. Como consequência, nosso robô conseguiu controlar a volatilidade da carteira.



Evolução do Drawdown: Comparação com Benchmarks

Sendo assim, a alocação direcional não ultrapassou em quase todo o backtest o nível de 25%. Como resultado, nosso robô conseguiu reconhecer um dos fatores que maximizam a sua função recompensa obtendo grandes ganhos e reduções de drawdowns implementando uma espécie de hedge das suas posições. Se analisarmos em conjunto os gráficos de drawdowns e da resultante de alocação do portfólio, essa hipótese fica evidente.

Em síntese, conseguimos caracterizar a performance da nossa estratégia com a tabela abaixo.

obs: Para o cálculo do índice de sharpe utilizamos uma taxa livre de risco de 4,07% a.a baseada no título de dívida americano US10Y.

Indicadores de Performance Anualizados			
Vars	Satoshi	Buy&Hold	S&P500
Retorno	16,23%	-33,54%	-10,17%
Risco	28,51%	61,72%	20,50%
Sharpe Ratio	0,3878	-0,7261	-0,71788
MDD	-21,28%	-41,65%	-12,76%

3.2. Alocação

Abaixo a tabela com a mediana de alocação para cada um dos meses dos dados de teste. Na tabela, os ativos com os pesos mais positivos são BNB e BTC. Em relação aos mais negativos ficamos com ADA e XRP. Constatamos que o robô segue uma política previsível de alocação na maioria dos casos, no entanto em algumas oportunidades o algoritmo muda repentinamente a alocação, como foi o caso dos meses, julho-22, ago-22 e dez-22.

Date	ADAUSDT weights	BNBUSDT weights	BTCUSDT weights	ETHUSDT weights	LTCUSDT weights	XRPUSDT weights
fev-22	4,2%	17,3%	8,7%	5,3%	0,5%	-10,5%
mar-22	-15,0%	16,0%	9,1%	9,6%	4,3%	-14,1%
abr-22	-23,4%	21,7%	23,4%	9,6%	-0,6%	-16,5%
mai-22	-23,4%	23,4%	23,4%	9,6%	-3,2%	-16,9%
jun-22	-21,7%	23,4%	23,4%	9,6%	-3,4%	-1,1%
jul-22	10,2%	23,4%	16,0%	-16,0%	-16,9%	-5,6%
ago-22	-15,7%	13,8%	-21,1%	-10,7%	7,6%	-11,3%
set-22	-23,4%	23,4%	23,4%	6,8%	-3,2%	-16,8%
out-22	-12,6%	17,5%	17,6%	16,9%	3,0%	1,2%
nov-22	-23,4%	23,4%	23,4%	-8,5%	-3,2%	-10,5%
dez-22	17,3%	12,9%	5,0%	1,1%	4,4%	-0,5%

Evolução Mensal das Alocações

Para examinarmos o aprendizado do robô e definir a importância de cada variável que foi utilizada na construção da política ótima utilizamos um **modelo de regressão linear**. Acreditamos que apesar de modelos de redes neurais possuírem componentes não lineares, uma avaliação preliminar através de um modelo mais simples nos levará a insights importantes sobre a alocação do modelo proposto.

Para verificar a importância de cada uma das variáveis optamos por aplicar uma normalização. Essa normalização leva o valor dos inputs para uma escala de 0 a 100, mantendo a proporção que cada registro representa na sua respectiva coluna. Além disso, tomamos o módulo nos indicadores que medem drawdown para facilitar a interpretação.

Indicadores	ADAUSDT	BNBUSDT	BTCUSDT	ETHUSDT	LTCUSDT	XRPUSDT
normalized_fech	0,4099	0,0696	-0,1674	0,2307	0,2145	-0,0158
	0,0015	0,4830	0,2392	0,1076	0,0831	0,9014
macd	0,2401	-0,0672	-0,2437	0,3488	0,1035	0,0464
	0,0006	0,2280	0,0014	0,0001	0,0992	0,5277
rsi	-0,3771	0,2047	0,2493	-0,5552	-0,2298	-0,1505
	0,0061	0,1501	0,1493	0,0026	0,2049	0,3810
ewma_diff	-0,5161	0,0467	0,2290	0,1791	0,0031	-0,0443
	0,0000	0,4794	0,0022	0,0127	0,9674	0,5260
ddd	0,1101	0,2779	0,2587	-0,1545	0,0570	0,0012
	0,2023	0,0000	0,0005	0,0340	0,5088	0,9889
mdd	0,0741	-0,0543	-0,0107	-0,0274	-0,2100	-0,0027
	0,1737	0,1505	0,8149	0,5807	0,0000	0,9573
F-Statistic	18,71	95,87	54,83	3,64	9,33	15,54
R-Squared	0,26	0,64	0,50	0,06	0,15	0,22

Resultado Modelo de Regressão

Na tabela acima, mostramos que apenas os quadrantes pintados em cinza são estatisticamente significativos com nível de 5%. Além disso, o módulo de cada parâmetro na regressão, em virtude da normalização, indica a intensidade que cada variável irá intervir no peso do ativo. Ou seja, para o ativo ADA, por exemplo, a diferença entre médias

exponenciais interfere mais do que o RSI nas decisões de alocação desse ativo. Nesse sentido temos as seguintes conclusões quanto a alocação do robô:

ADA: À medida que o fechamento normalizado e o MACD aumentam, a alocação na compra é ampliada. Por outro lado, o aumento do RSI e da diferença de médias móveis exponenciais resulta em uma maior posição vendida no ativo.

BNB: Um maior drawdown diário leva a uma posição comprada mais forte. Nesse caso, o robô opta por uma estratégia semelhante ao "dollar cost average" com base no DDD.

BTC: O aumento na diferença entre as médias móveis exponenciais e no daily drawdown resulta em uma posição comprada maior. Por outro lado, um valor MACD mais elevado leva a uma posição vendida maior no ativo.

ETH: À medida que o MACD e a diferença entre as médias móveis exponenciais aumentam, a posição comprada se fortalece. Em contrapartida, um aumento no RSI e no DDD leva a uma posição vendida maior no ativo.

LTC: Um maior drawdown máximo observado na janela móvel selecionada resulta em uma posição vendida maior.

XRP: Não foi possível estabelecer uma relação linear significativa entre a alocação e os indicadores.

4. Back Test

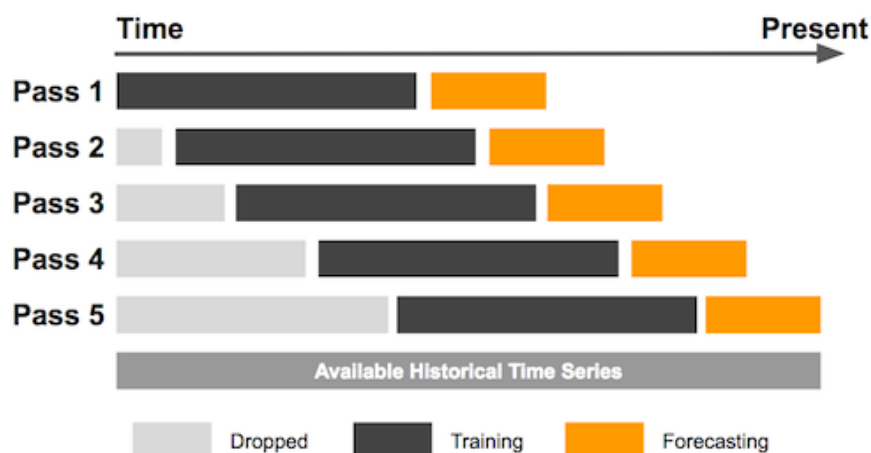
Neste trabalho aplicamos a metodologia de backtest **Walk Forward** de duas formas, primeiro utilizamos um horizonte temporal mais longo, cujos resultados foram mostrados no módulo dos resultados, e um na forma de **Cross Validation** adaptado para séries temporais que será introduzido nesta seção. No Walk Forward conseguimos extrair insights sobre como a alocação do robô é feita, além de dar uma interpretação dos resultados. O outro é importante para verificar a robustez do modelo em diferentes conjuntos de dados de treino e teste.

Nos nossos backtests, estamos interessados em verificar a robustez da estratégia e a probabilidade de ganharmos dinheiro no futuro, para isso é importante atentar-se aos principais vieses observados em backtests. Mostraremos resultados relacionados aos vieses de overfit e lookahead bias.

4.1. Cross Validation

Para a construção das janelas de treino e teste, optamos por utilizar a técnica de janelas móveis, visto que a quantidade de dados de treino está intimamente ligada com a quantidade de tempo necessário para o algoritmo PPO aprender, como não existe uma quantidade ótima que relaciona essas duas variáveis, optamos por fixar a quantidade de

dados de treino e teste. Nesse modelo dividimos nosso dataset em 8 compartimentos com 120 dias de treino e 24 dias de avaliação do modelo.



4.2. Probabilistic Sharpe Ratio (PSR): Como não ser enganado pelo acaso.

Para concluirmos que nossa estratégia possui algum tipo de vantagem sobre o mercado, costumeiramente fazemos referência ao índice de sharpe. No entanto, apenas este valor não entrega toda a informação necessária para esclarecer sobre a superioridade do modelo, visto que a performance superior pode ser apenas uma coincidência (overfit). Por esse motivo introduzimos o **PSR**[5]. Essa metodologia utiliza de propriedades estatísticas dos retornos do portfólio para medir a probabilidade de nosso modelo superar determinado benchmark em relação ao SR no longo prazo. Nesse experimento consideramos como benchmark um sharpe ratio de 0.

Para o caso do backtest Walk Forward, obtemos um valor para o **PSR de 72,67%**. Isso significa que podemos afirmar que nossa estratégia possui uma vantagem sobre o mercado com 72,67% de probabilidade de estarmos corretos. Para testarmos a robustez deste resultado apresentaremos o resumo dos indicadores calculados no backtest usando cross validation.

k-fold	Retorno	Risco	SR	PSR
1	-54,24%	43,03%	-1,81	27,32%
2	334,12%	40,23%	3,66	90,16%
3	342,18%	59,36%	2,52	78,18%
4	161,32%	33,15%	2,90	83,30%
5	81,84%	31,35%	1,91	73,81%
6	-25,60%	20,67%	-1,43	32,62%
7	153,84%	34,39%	2,71	83,13%
8	19,38%	17,01%	1,04	63,48%

A partir dos resultados obtidos, nota-se que existe alta variância nos indicadores, o que é **um risco para a estratégia**, visto que isso evidencia que existe correlação entre a performance e a janela que treinamos o modelo. Dito isso, a média do PSR para os elementos explicitados foi de **66,50%**.

4.3. Lookahead Bias

O **Look-ahead bias** ocorre quando utilizamos informações do futuro para tomar decisões do presente. Ou seja, esse tipo de viés ocorre quando há o vazamento de informações futuras para dentro do modelo. Nesse caso existe um incremento de performance levando a julgamentos imprecisos sobre a validade de estratégias quantitativas. No nosso caso, o vazamento de informação futura poderia ocorrer no cálculo dos indicadores que utilizam o preço de fechamento de mercado como base. Como exemplo, pegamos o indicador simples de uma média móvel exponencial. O valor do EWMA para um determinado dia inclui o preço de fechamento daquele dia, porém, na prática não sabemos o fechamento de mercado de hoje, apenas o de ontem. Nesse sentido, para utilizarmos qualquer indicador que utilize o fechamento de mercado como input, há a necessidade de adicionarmos um atraso entre o preço que será tomado a decisão do trade e o indicador. Portanto, para evitar essa questão, adicionamos um atraso em todos os indicadores em relação ao preço de fechamento, dessa forma, não há vazamento de informação futura e eliminamos totalmente esse viés.

5. Conclusão

Nesse relatório exploramos uma estratégia derivada de modelos de **deep learning**, mais especificamente o algoritmo de gradiente de política **PPO**. Durante a apresentação dos resultados constatamos que nosso modelo, no período analisado, garantiu resultados muito superiores aos benchmarks escolhidos. Também mostramos que o robô aprendeu a controlar a volatilidade e realizar alocações pontuais para maximizar o retorno, **baseado apenas em análise técnica**. Ao avançar para o módulo de Backtest, apresentamos o indicador PSR em conjunto com a metodologia de amostragem cross validation. Com esse esforço, testamos a robustez do nosso modelo. Baseado nas evidências desta sessão, constatamos que nossa estratégia não conseguiu superar o nível conservador de **95% de confiança**, atualmente permanecemos entre o intervalo **66,5% e 72,67%**. Nesse caso o indicado seria não implementar a estratégia. Apesar de não ter passado no teste de robustez, ficamos animados com a capacidade de aprendizado do robô, e imaginamos o potencial que ele poderia chegar usando um novo conjunto de espaço observável. Para melhorar a performance do modelo, também é possível ajustar os hiperparâmetros do modelo, incluindo a janela móvel de cálculo dos indicadores, e da recompensa.

6. Referências bibliográficas

- [1] Y. Liu, Q. Liu, H. Zhao, Z. Pan and C. Liu, "Adaptive quantitative trading: An imitative deep reinforcement learning approach," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 2, pp. 2128-2135, 2020. URL: <https://ojs.aaai.org/index.php/AAAI/article/view/5587>
- [2] Z. Jiang, D. Xu and J. Liang, "A deep reinforcement learning framework for the financial portfolio management problem," arXiv, vol. 1706.10059, pp. 1-31, 2017. URL: <https://arxiv.org/pdf/1706.10059.pdf>
- [3] L. T. Hieu, "Deep reinforcement learning for stock portfolio optimization," International Journal of Modeling and Optimization, vol. 10, no. 5, pp. 139-144, 2020. URL : <https://arxiv.org/ftp/arxiv/papers/2012/2012.06325.pdf>
- [4] H. Zhang, Z. Jiang and J. Su, "A deep deterministic policy gradient based strategy for stocks portfolio management," arXiv, vol. 2103.11455v1, pp. 1-8, 2021. URL: <https://arxiv.org/pdf/2103.11455.pdf>
- [5] Marcos López de Prado and David Bailey (2012). The Sharpe ratio efficient frontier. URL: <https://www.davidhbailey.com/dhbpapers/sharpe-frontier.pdf>