Agente de Reinforcement Learning para Decisões Financeiras

Neste trabalho, desenvolvemos um agente capaz de dar suporte em tomadas de decisões financeiras, como compra, venda ou manutenção de ativos do mercado.

Trabalhamos com dados históricos da Vale, Petrobrás e Brasil Foods, utilizando a abordagem Q-learning.

O Algoritmo Q-Learning

Inicialização

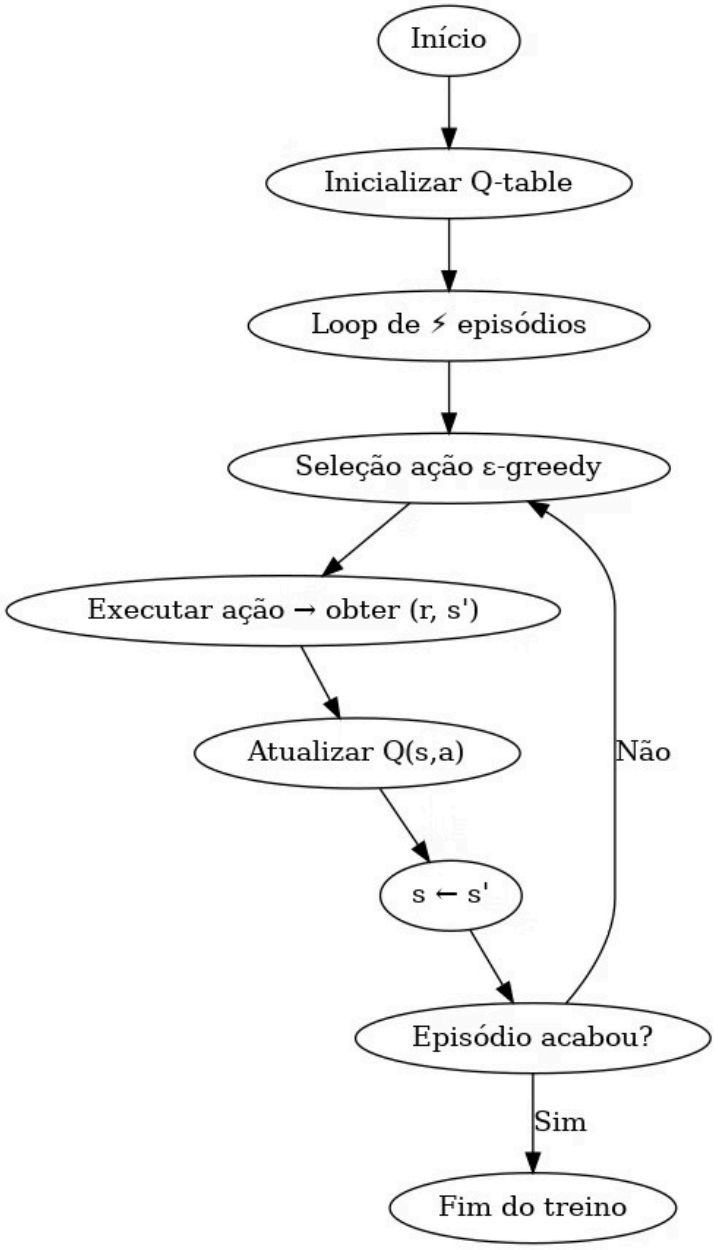
Inicializar Q-table com zeros para todos os pares (estado, ação).

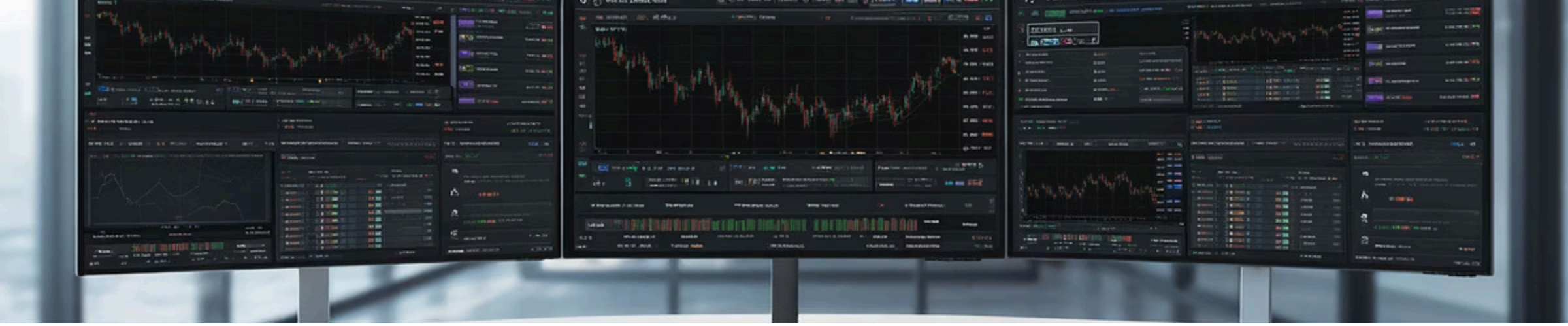
Seleção de Ação

Selecionar ação usando política ε-greedy baseada em Q(s,·).

Atualização

$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max Q(s',a') - Q(s,a)]$





Ambiente de Negociação

Saldo Inicial

O agente começa com R$10.000 para investir.

Ações Possíveis

Manter posição (0), Comprar (1) ou Vender (2) para cada ativo.

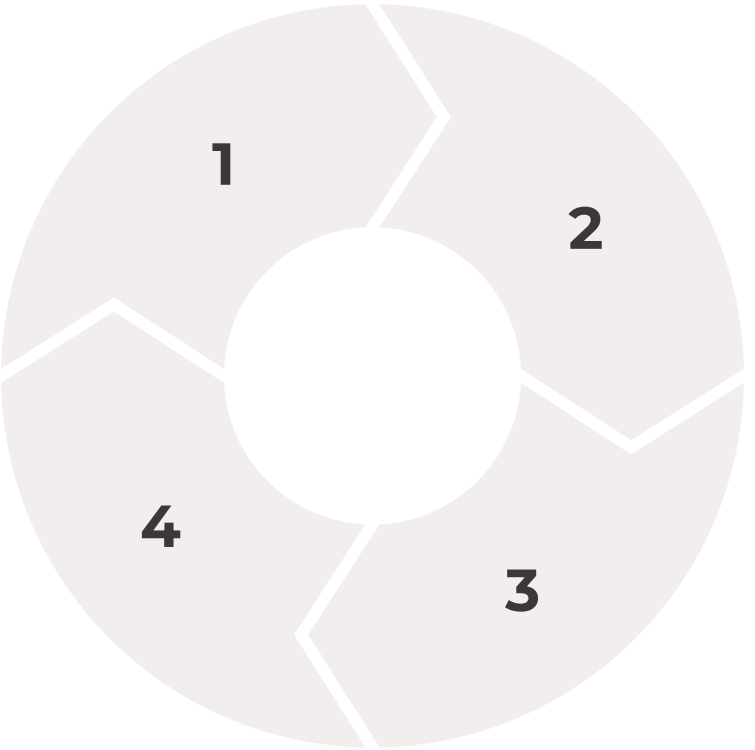
Recompensa

Variação do valor da carteira a cada passo (dia).

Agente Q-Learning

Observação

Recebe estado atual (saldo + preços)



Decisão

Escolhe ação via estratégia ε-greedy

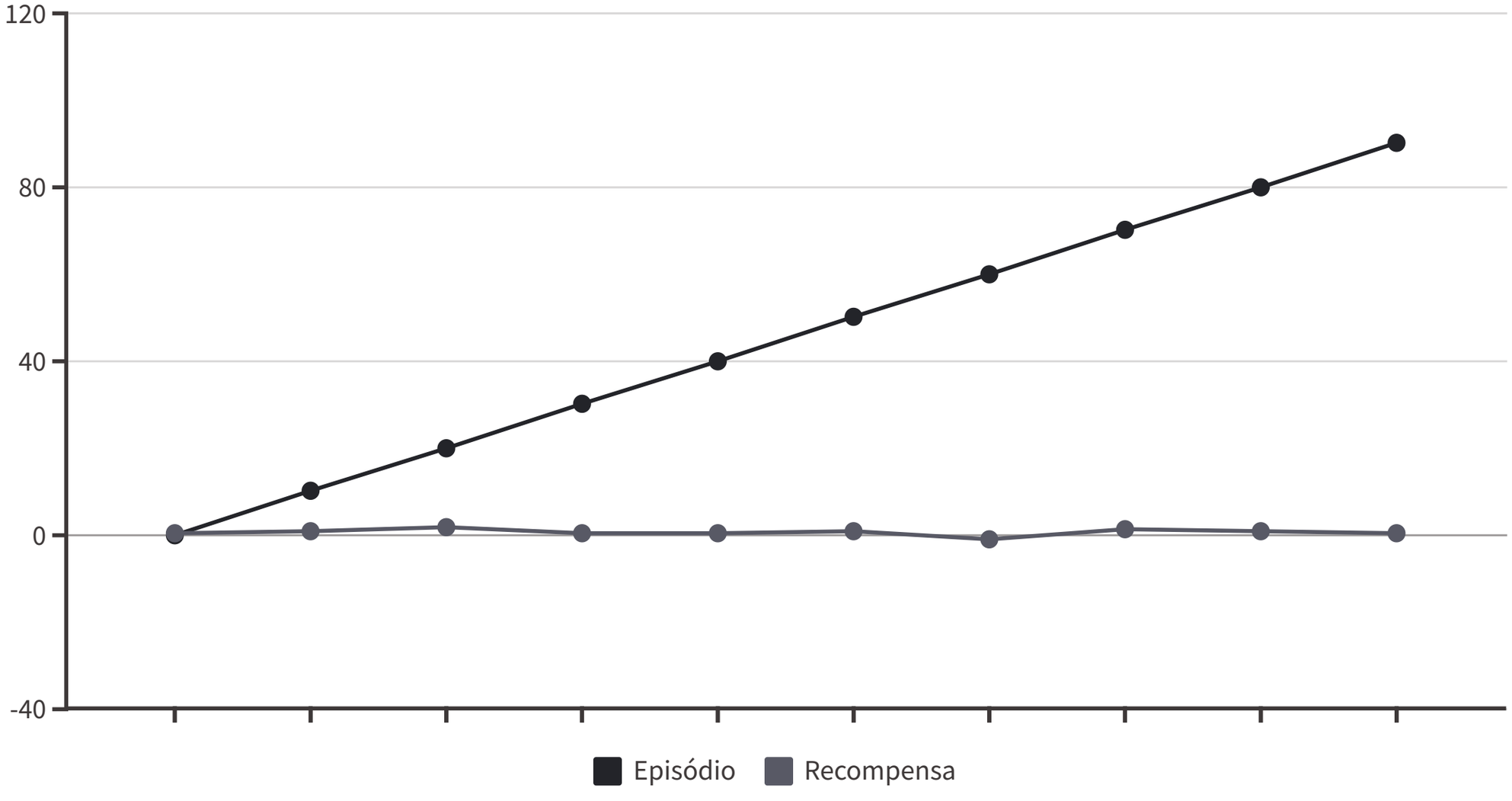
Execução

Realiza compra, venda ou mantém posição

Aprendizado

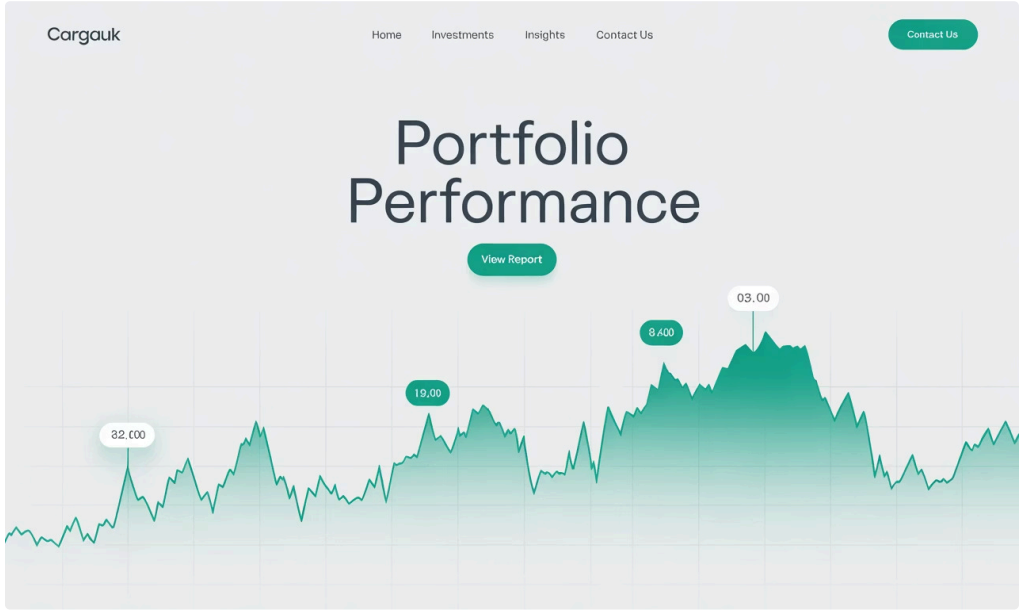
Atualiza Q-table com nova experiência

Treinamento Inicial



Após 100 episódios, observamos recompensas variáveis, com média de 0.69. O agente ainda não apresenta desempenho consistente.

Avaliação de Performance

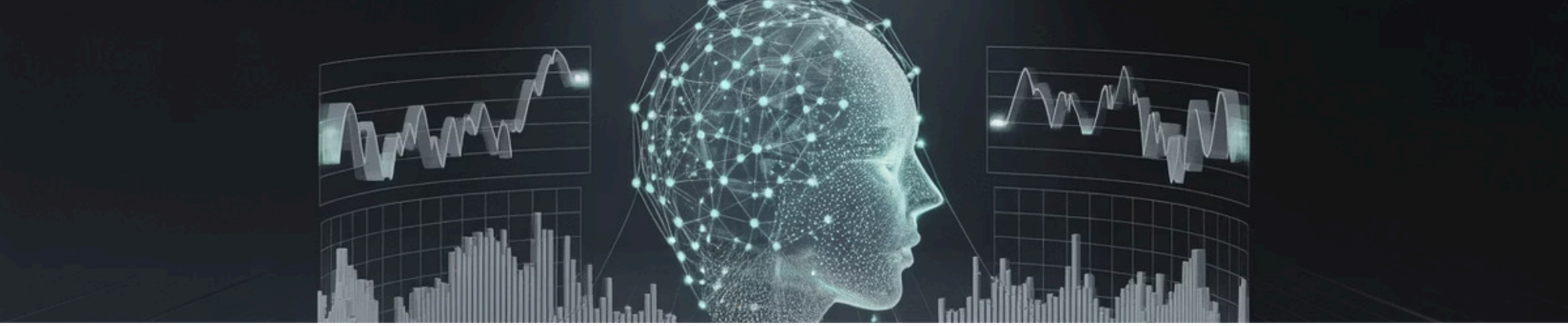


Resultados Iniciais

O agente aprendeu a variar decisões, mas não a gerar lucro consistente.

Sharpe Ratio próximo de zero indica baixa efetividade da estratégia aprendida.

Comportamento ainda próximo do aleatório, sugerindo necessidade de ajustes.



Evolução do Modelo

Indicadores Técnicos

Adição de médias móveis (curta e longa) e retorno percentual diário.

Normalização

Transformação dos valores para média 0 e desvio padrão 1.

Reconfiguração

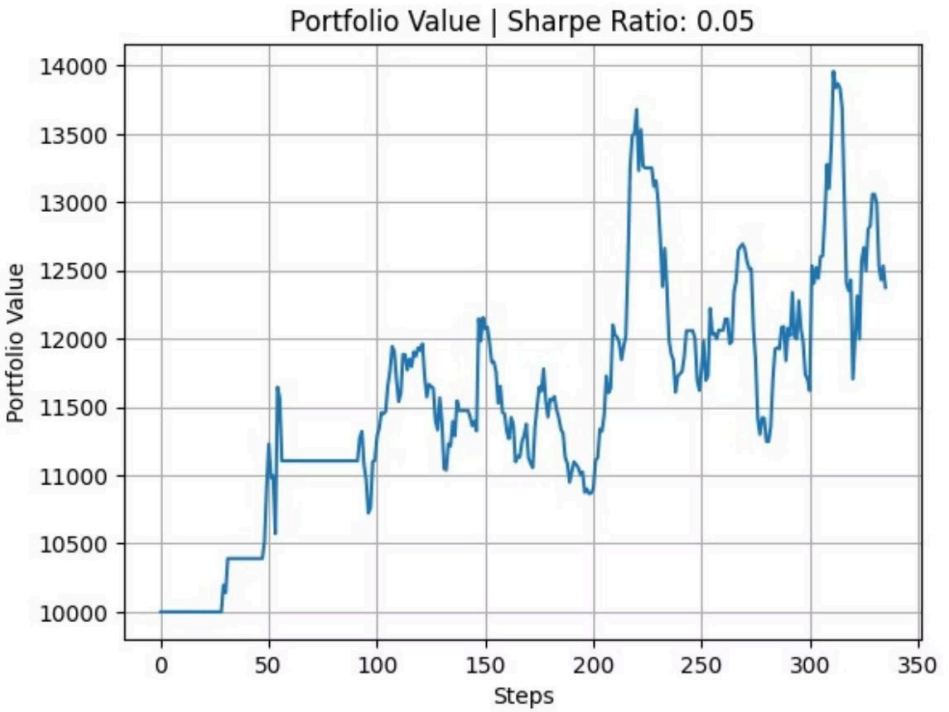
Ajuste do ambiente para usar novas informações como estado.

Evolução do Agente de Trading

Tentamos evoluir o agente aumentando a granularidade e complexidade. Usamos indicadores técnicos: média móvel curta e longa, e retorno percentual diário.

Inicialmente tentamos normalizar os dados, mas após testes percebemos que isso prejudicava o desempenho.

Reconfiguramos o ambiente para usar as novas informações como estado.



O Sharpe ratio melhorou, mas continua próximo de zero. Isso ainda não justifica o uso do modelo devido ao baixo retorno e alto risco.

Validação Cruzada Temporal

0.0567

Split 1

Sharpe do Agente (2020-2021)

Superou o mercado (-0.0962)

0.0550

Split 2

Sharpe do Agente (2021-2022)

Ligeiramente superior ao mercado (0.0518)

-0.0557

Split 3

Sharpe do Agente (2022-2024)

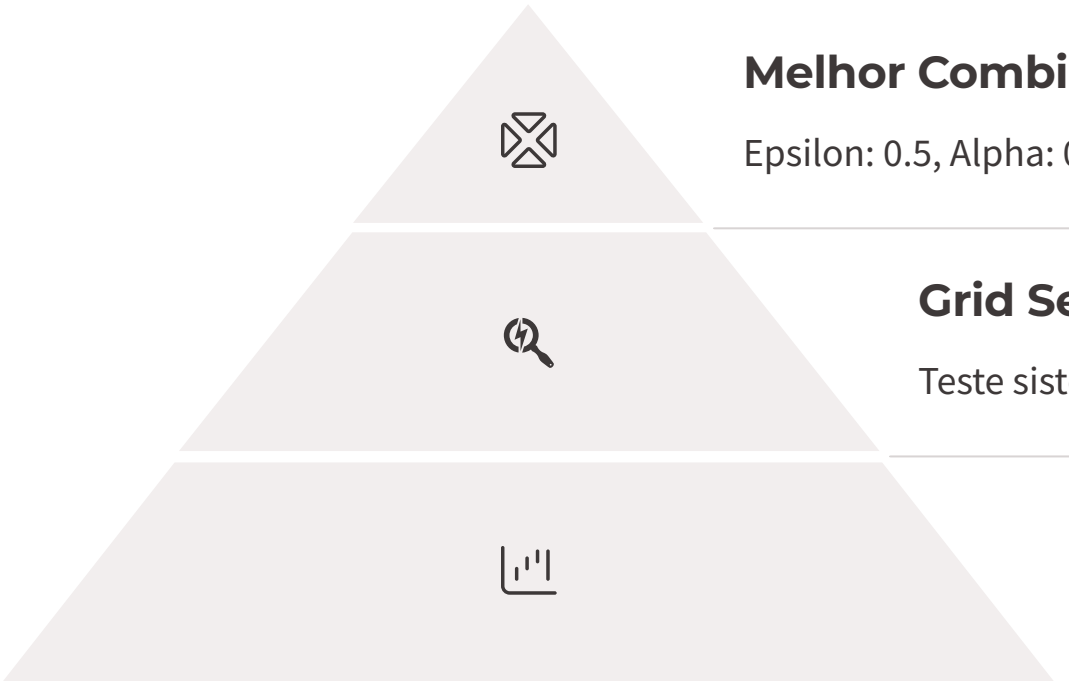
Inferior ao mercado (0.0521)

0.0187

Média

Sharpe Ratio médio

Otimização de Hiperparâmetros



Melhor Combinação

Epsilon: 0.5, Alpha: 0.01, Gamma: 0.9

Grid Search

Teste sistemático de combinações de parâmetros

Validação Temporal

Avaliação em diferentes períodos de mercado

Conclusões

Desafios Encontrados

Falta de generalização e possível overfitting. Limitações do Q-Learning tabular em ambientes complexos.

Validação Temporal

Essencial para revelar fragilidades da estratégia. Mostrou que o agente não transfere aprendizado entre períodos.

Indicadores

Apesar da inclusão de médias móveis e momentum, o agente não demonstrou uso efetivo desses sinais.

Próximos Passos

Deep Q-Networks

Implementar DQN para lidar com alta dimensionalidade

Custos de Transação

Incluir custos reais e penalidades por excesso de operações

Ampliar Base de Dados

Testar diferentes ativos e períodos de mercado