

Machine Learning for Optimization of Buy and Sell Signals in Bitcoin"

Machine Learning para Otimização de Sinais de Compra e Venda em Bitcoin

PO-236

Pedro Arthur Ramos Oliveira

October 30nd 2024

Universidade Federal de São Paulo e Instituto Tecnológico de Aeronáutica
Programa de Pós-Graduação em Pesquisa Operacional (PPGPO)



Introdução

Resumo

- Inicialmente idealizado como um sistema de pagamento descentralizado e peer-to-peer, o Bitcoin evoluiu para se tornar uma reserva de valor e um ativo altamente volátil.
- A volatilidade do Bitcoin atraiu investidores e especuladores, mas também trouxe desafios para o desenvolvimento de estratégias de trading eficazes.
- Técnicas de Machine Learning (ML) oferecem uma abordagem inovadora para otimização de sinais de compra e venda de Bitcoin.
- Modelos de aprendizado de máquina podem ser utilizados para identificar padrões e comportamentos históricos, gerando previsões mais precisas e otimizando decisões de trading.
- O uso de ML permite analisar grandes volumes de dados históricos, compreender tendências e ajustar estratégias de trading de forma automatizada e dinâmica.
- O estudo do Bitcoin, aliado à aplicação de técnicas de Machine Learning, revela insights sobre o comportamento deste mercado emergente e explora novas formas de tomada de decisão automatizada.
- Esta combinação proporciona ganhos mais robustos para investidores e traders, trazendo relevância para a adoção dessas técnicas na prática financeira.

Introdução

Introdução

- **Investimento em ativos financeiros:** A previsão de retornos de ativos é fundamental para decisões de investimento, sendo uma atividade de interesse tanto para profissionais quanto para acadêmicos. O uso de machine learning (ML) tem ganhado destaque pela sua capacidade de melhorar progressivamente o desempenho em tarefas analíticas (Ballings et al., 2015; Cavalcante et al., 2016).
- **Machine Learning no trading quantitativo:** Técnicas de ML automatizam a construção de modelos analíticos e permitem identificar padrões complexos em dados financeiros, que seriam difíceis de detectar com métodos tradicionais (Tsantekidis et al., 2017). Essa abordagem possibilita a comparação de diferentes metodologias em períodos variados, utilizando fundamentos e dados de preço (Lo et al., 2000; Jegadeesh & Titman, 1993).
- **Uso de algoritmos avançados como Random Forest:** Algoritmos como o RF têm mostrado eficiência na previsão de preços e tendências de mercado, proporcionando uma vantagem competitiva significativa (Chen et al., 2018).
- **Otimização de trading de Bitcoin:** A aplicação de técnicas de ML está sendo utilizada para otimizar estratégias de compra e venda de Bitcoin, com dados históricos desde 01/01/2012 (Fonte: Kaggle).

Contextualização

Revisão de Literatura

- **Estratégias tradicionais:** A análise técnica estuda padrões de preço e volume (Murphy, 1999), enquanto a análise fundamentalista avalia fatores econômicos para determinar o valor intrínseco (Graham & Dodd, 1934).
- **Machine Learning no trading:** Algoritmos de ML, como redes neurais profundas, SVM e RF, têm mostrado eficácia na previsão de retornos de ações e tendências de mercado (Fischer & Krauss, 2018; Chen et al., 2018).
- **Random Forest:** Algoritmo ensemble que melhora a precisão combinando múltiplas árvores de decisão, sendo eficaz para dados financeiros não lineares (Dietterich, 2000; Breiman, 2001; Henrique et al., 2019).

Metodologia i

Criação dos Sinais de Compra e Venda

- Implementação de sinais de compra e venda baseada em médias móveis de curto e longo prazo;
- Comparar duas médias móveis de diferentes períodos para gerar um sinal que indique a tendência de preço e direção esperada do mercado;
- A média móvel de curto prazo usa uma janela de 10 períodos, enquanto a média móvel de longo prazo usa uma janela de 60 períodos.
- Se a média de curto prazo estiver acima da média de longo prazo, isso sugere uma possível tendência de alta, sinalizando compra.
- Caso contrário, a estratégia considera que o ativo está em uma tendência de baixa, recomendando venda ou inatividade.

Metodologia

Com as médias móveis de curto e longo prazo calculadas, definimos uma estratégia para gerar sinais de compra e venda:

Esse código cria uma coluna `signal` onde:

- `signal = 1.0` indica uma tendência de alta (compra) quando a média móvel de curto prazo está acima da média móvel de longo prazo.
- `signal = 0.0` indica uma tendência de baixa ou inatividade (venda) quando a média de longo prazo é maior.

Matematicamente, o sinal $signal_t$ no tempo t é definido como:

$$signal_t = \begin{cases} 1.0 & \text{se } short_t > long_t \\ 0.0 & \text{caso contrário} \end{cases} \quad (1)$$

Feature Engineering

Indicadores Técnicos

- **Média Móvel (Moving Average):** Indica a tendência do preço ao reduzir o "ruído" nos gráficos.
- **Oscilador Estocástico (%K e %D):** Compara o preço de fechamento atual com uma faixa de preços do período, onde %K é o indicador rápido e %D, o lento.
- **Índice de Força Relativa (RSI - Relative Strength Index):** Mede a magnitude das variações de preço para indicar sobrecompra ou sobrevenda.
- **Taxa de Mudança (ROC - Rate of Change):** Mede a variação percentual entre o preço atual e o de n períodos passados.
- **Momentum (MOM):** Indica a velocidade das mudanças no preço ou volume.

Metodologia

Média Móvel Exponencial (EMA)

A Média Móvel Exponencial (EMA) é uma média ponderada que dá maior peso aos preços mais recentes. A EMA é calculada usando a fórmula:

$$EMA_t = \alpha \cdot \text{Close}_t + (1 - \alpha) \cdot EMA_{t-1} \quad (2)$$

onde $\alpha = \frac{2}{n+1}$ e n é o número de períodos. Usamos os seguintes períodos para capturar diferentes tendências:

- **EMA10**: Curto prazo.
- **EMA30**: Médio prazo.
- **EMA200**: Longo prazo (referência para traders).

Metodologia

Taxa de Variação (Rate of Change - ROC)

A Taxa de Variação (ROC) mede a variação percentual do preço em relação a um período anterior:

$$ROC_t = \frac{Close_t - Close_{t-n}}{Close_{t-n}} \times 100 \quad (3)$$

Usamos:

- **ROC10**: Janela de 10 períodos, sensível a mudanças de curto prazo.
- **ROC30**: Janela de 30 períodos, foco em mudanças intermediárias.

Metodologia

Momentum (MOM)

O indicador de Momentum mede a diferença entre o preço atual e o preço de n períodos anteriores, sendo calculado por:

$$\text{MOM}_t = \text{Close}_t - \text{Close}_{t-n} \quad (4)$$

O Momentum (MOM) calcula a diferença do preço atual com o preço de n períodos atrás, medindo a força da tendência de preço.

- **MOM10:** Calcula o momentum para uma janela de 10 períodos, focando em mudanças de curto prazo.
- **MOM30:** Calcula o momentum para uma janela de 30 períodos, considerando tendências intermediárias.

Metodologia

Relative Strength Index (RSI)

O Índice de Força Relativa (RSI) é um oscilador que mede a velocidade e a mudança dos movimentos de preços, indicando condições de sobrecompra ou sobrevenda. Calcula-se o ganho médio (U) e a perda média (D) para um determinado período. Em seguida, calcula-se a razão de força relativa (RS):

$$RS = \frac{\text{Ganho Médio}}{\text{Perda Média}} \quad (5)$$

O RSI é então dado por:

$$RSI = 100 - \frac{100}{1 + RS} \quad (6)$$

Metodologia

Oscilador Estocástico (%K e %D)

O Oscilador Estocástico avalia a posição do preço de fechamento em relação aos máximos e mínimos de um período n , ajudando a identificar sobrecompra ou sobrevenda. É composto por:

- **%K**: linha principal.
- **%D**: média móvel de %K para suavizar o sinal.

As variações incluem:

- **%K10** e **%D10**: 10 períodos, captam tendências de curto prazo.
- **%K30** e **%D30**: 30 períodos, sinais de médio prazo.
- **%K200** e **%D200**: 200 períodos, análise de longo prazo.

Metodologia

Média Móvel Simples (MA)

A Média Móvel Simples calcula a média dos preços de fechamento sobre um intervalo específico. A fórmula é:

$$MA_t = \frac{1}{n} \sum_{i=0}^{n-1} \text{Close}_{t-i} \quad (7)$$

Metodologia

Modelos de Machine Learning

Testamos diferentes algoritmos de Machine Learning para a tarefa de classificação, medindo o desempenho em termos de acurácia com validação cruzada de 10 folds.

- **LR**: 89.97% (0.011)
- **LDA**: 90.77% (0.002)
- **KNN**: 93.40% (0.002)
- **CART**: 90.28% (0.002)
- **NB**: 79.77% (0.005)
- **NN**: 81.71% (0.083)
- **AB**: 91.36% (0.003)
- **GBM**: 92.39% (0.002)
- **RF**: 93.47% (0.002)

Metodologia

Random Forest

Estratégia de trading utilizando algoritmos de Machine Learning para analisar movimentos de mercado com base em indicadores técnicos. O **Random Forest** é um modelo de *ensemble* que combina múltiplas árvores de decisão para aumentar a precisão e evitar *overfitting*. Cada árvore utiliza uma amostra aleatória dos dados e um subconjunto de variáveis, reduzindo a variância do modelo.

Os passos incluem:

- **Otimização de Hiperparâmetros:** Utilizando *Grid Search* para ajuste dos modelos.
- **Validação Cruzada:** Avaliação do desempenho dos modelos com 10 folds.

Metodologia

Otimização de Hiperparâmetros com Grid Search

Padronização: Os dados de treinamento foram padronizados usando *StandardScaler* para garantir média zero e desvio padrão unitário em cada variável X .

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}$$

- $\mathbf{X} \in \mathbb{R}^{n \times p}$, com n amostras e p variáveis
- \bar{x}_j, σ_j : média e desvio padrão da variável x_j

Random Forest: Hiperparâmetros testados

- **n_estimators**: número de árvores na floresta. Testamos os valores 20 e 80.
- **max_depth**: profundidade max de cada árvore, para evitar overfitting. Testamos a 5 e 10.
- **criterion**: critério de divisão, podendo ser gini (impureza de Gini) ou entropy (ganho de informação).

Validação Cruzada: Utilizando 10 *folds*, a média de acurácia é obtida em cada execução para determinar o melhor conjunto de hiperparâmetros.

Resultados do Grid Search e Avaliação do Modelo

O modelo foi então treinado com esses parâmetros no conjunto de dados completo e testado no conjunto de validação. A avaliação de desempenho foi realizada através de métricas de classificação:

```
0.92205
[[9086 807]
 [ 752 9355]]

              precision    recall  f1-score   support

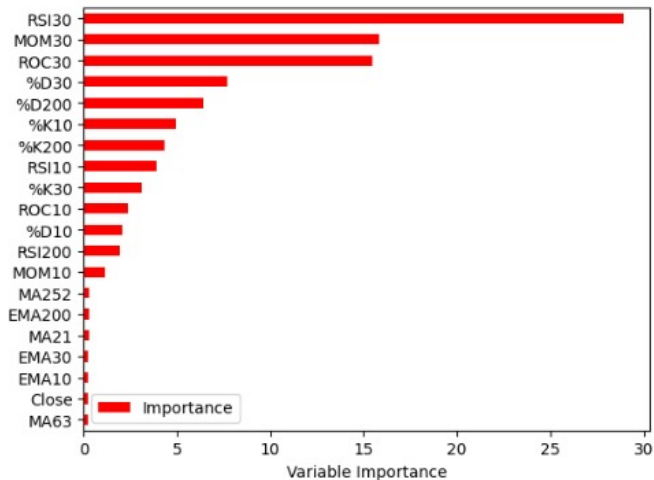
         0.0         0.92         0.92         0.92         9893
         1.0         0.92         0.93         0.92        10107

 accuracy                   0.92         20000
 macro avg         0.92         0.92         0.92         20000
 weighted avg      0.92         0.92         0.92         20000
```

Figure 1: Results

Resultados do Grid Search e Avaliação do Modelo

Variable Importance



Backtesting

Backtest comparando o retorno cumulativo da estratégia baseada nas previsões do modelo com o retorno do mercado.

- **Market Returns:** Retorno diário do mercado, calculado como:

$$R_t^{\text{market}} = \frac{P_t - P_{t-1}}{P_{t-1}}$$

onde P_t é o preço de fechamento no dia t .

- **Actual Returns:** Retorno real do mercado ajustado pelo sinal verdadeiro de posição do dia anterior:

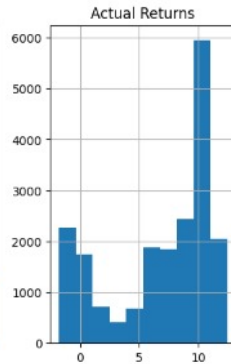
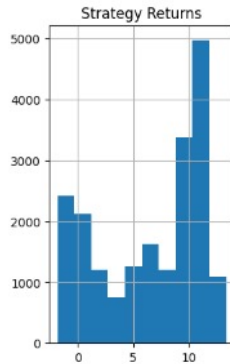
$$R_t^{\text{actual}} = R_t^{\text{market}} \times \text{signal}_{t-1}^{\text{actual}}$$

- **Strategy Returns:** Retorno da estratégia com base nas previsões do modelo:

$$R_t^{\text{strategy}} = R_t^{\text{market}} \times \text{signal}_{t-1}^{\text{pred}}$$

Resultados do Grid Search e Avaliação do Modelo

Results



Metodologia

Conclusão

- Enquadramos o problema por meio de engenharia de características e transformação de rótulos, alinhando-os ao objetivo de investimento.
- A criação de características intuitivas, como tendências e momentum, aumentou o poder preditivo do modelo.
- Para avaliação de uma estratégia de trading, acurácia ou AUC são adequadas. Contudo, se a prioridade é minimizar falsos positivos em posições compradas, o *recall* é preferível.
- Implementamos o *backtesting*, permitindo simular estratégias com dados históricos para analisar risco e rentabilidade sem comprometer capital real.

Referências Bibliográficas

Referências

- **Chen, Junwei.** *Analysis of Bitcoin Price Prediction Using Machine Learning.* Journal of Risk and Financial Management, MDPI. Disponível em: <<https://www.mdpi.com>>
- **Sossi-Rojas, Stefano; Velarde, Gissel; Zieba, Damian.** *A Machine Learning Approach for Bitcoin Forecasting.* Engineering Proceedings, MDPI. Disponível em: <<https://www.mdpi.com>>
- **Jabbar, Abdul; Jalil, Syed Qaisar.** *A Comprehensive Analysis of Machine Learning Models for Algorithmic Trading of Bitcoin.*
- **Silva, Maciel Vicente da.** *Algoritmos de machine learning em estratégias de trading: uma análise da eficiência e aplicabilidade no mercado financeiro / Maciel Vicente da Silva.* - 2024. 159 f.