

Previsão da Direção de Criptomoedas Utilizando Dados de Mercado e On-Chain com XGBoost

Autores:

- Bruno Augusto Lopes Fevereiro
- Carlos Alberto Matias da Costa
- Gabriel Bianconi

Resumo

Este trabalho apresenta um pipeline automatizado para prever a direção de preços de criptomoedas, combinando dados históricos de mercado e informações on-chain do Bitcoin. Utilizando técnicas de ciência de dados e machine learning, os dados são coletados, processados e transformados em features preditivas, com modelos XGBoost treinados para prever movimentos de alta ou baixa no próximo dia. A abordagem integra indicadores técnicos, lags temporais e retornos percentuais, permitindo análises tanto para criptomoedas com dados on-chain (BTC) quanto para aquelas sem informações de blockchain (ETH, SOL).

Introdução

O mercado de criptomoedas é altamente volátil e sensível a fatores técnicos, econômicos e de rede. Prever a direção de preços é desafiador, dado que fatores externos influenciam significativamente os valores das moedas. Este estudo propõe um pipeline automatizado que integra:

1. Dados de mercado: preços históricos e volume de negociação via Yahoo Finance.
2. Dados on-chain: métricas da rede Bitcoin, como número de transações e hash rate, fornecendo insights sobre a atividade da rede.
3. Indicadores técnicos: RSI, MACD, Bollinger Bands, ADX e médias móveis exponenciais (EMAs).

O objetivo é treinar modelos XGBoost para classificação da direção do preço no próximo dia, utilizando validação temporal e treinamento final em todos os dados disponíveis.

Metodologia

Coleta de Dados

Mercado: Dados históricos de preço e volume (Open, High, Low, Close, Adj Close, Volume) são obtidos via yfinance.

On-Chain (Bitcoin): Dados da blockchain são obtidos via API do Blockchain.com, incluindo:

- n-transactions: número diário de transações.

- hash-rate: poder computacional total da rede.
Os dados são coletados ano a ano, concatenados e convertidos de timestamps UNIX para datas normalizadas.

Integração: Os datasets de mercado e on-chain são combinados via left join, garantindo que todas as datas do mercado estejam presentes. Lacunas são preenchidas por interpolação (bfill e ffill).

Engenharia de Features

A função `create_features_pipeline` transforma os dados brutos em variáveis preditivas:

- Indicadores técnicos: RSI, MACD, Bollinger Bands, ADX e EMAs (50 e 200 dias).
- Lags temporais: valores passados de 1, 2, 3, 7 e 14 dias para Close, Volume, RSI e, quando disponíveis, dados on-chain.
- Retornos percentuais: variações de preço em 1 e 7 dias.
- Limpeza: remoção de linhas com valores ausentes em colunas essenciais.

Modelagem

O modelo é XGBoost Classifier, treinado para prever subida ou queda do preço no dia seguinte:

1. Validação temporal 80/20: primeiros 80% dos dados para treino, últimos 20% para teste.
2. Treinamento final: todo o conjunto de dados para produção.

As features são padronizadas via StandardScaler, e a importância das variáveis é analisada com gráficos de barras.

Aplicação Web

Uma interface com Streamlit permite:

- Seleção da criptomoeda (BTC, ETH, SOL).
- Exibição de previsões de direção com probabilidade.
- Visualização do histórico de preços e métricas on-chain (quando disponíveis).

A aplicação é hospedada via ngrok, permitindo acesso público sem servidor dedicado.

Resultados e Discussão

O pipeline demonstra que é possível combinar dados de mercado e on-chain para gerar previsões diárias da direção de preços das criptomoedas. As métricas de importância de features indicam que indicadores técnicos como

RSI, EMA e MACD são fortemente relevantes, enquanto dados on-chain, quando disponíveis, contribuem para melhorar a acurácia do modelo no Bitcoin.

Observou-se que, embora o XGBoost consiga capturar tendências e padrões de curto prazo, a previsão de preços em criptomoedas continua limitada devido à alta volatilidade e a fatores externos não mensurados.

Conclusão

Este estudo apresenta um pipeline completo para previsão de direção de criptomoedas, integrando dados de mercado, métricas on-chain e indicadores técnicos em modelos XGBoost. Apesar de demonstrar capacidade de previsão, os resultados devem ser interpretados com cautela, dado o caráter altamente volátil e imprevisível do mercado de criptomoedas.

Referências

1. Yahoo Finance API: <https://pypi.org/project/yfinance/>
2. Blockchain.com Charts API: <https://www.blockchain.com/charts>
3. Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
4. Pandas TA Documentation: <https://github.com/twopirllc/pandas-ta>
5. Streamlit Documentation: <https://docs.streamlit.io>