

Detecção de Anomalias em Criptomoedas

Uma análise aprofundada sobre a aplicação de Inteligência Artificial e Machine Learning na identificação de eventos atípicos em mercados financeiros digitais.



Discente

Martiniano Gomes Barros Cirqueira Neto

Este projeto de avaliação final explora a detecção de anomalias em dados financeiros de criptomoedas, focando em eventos críticos de mercado.

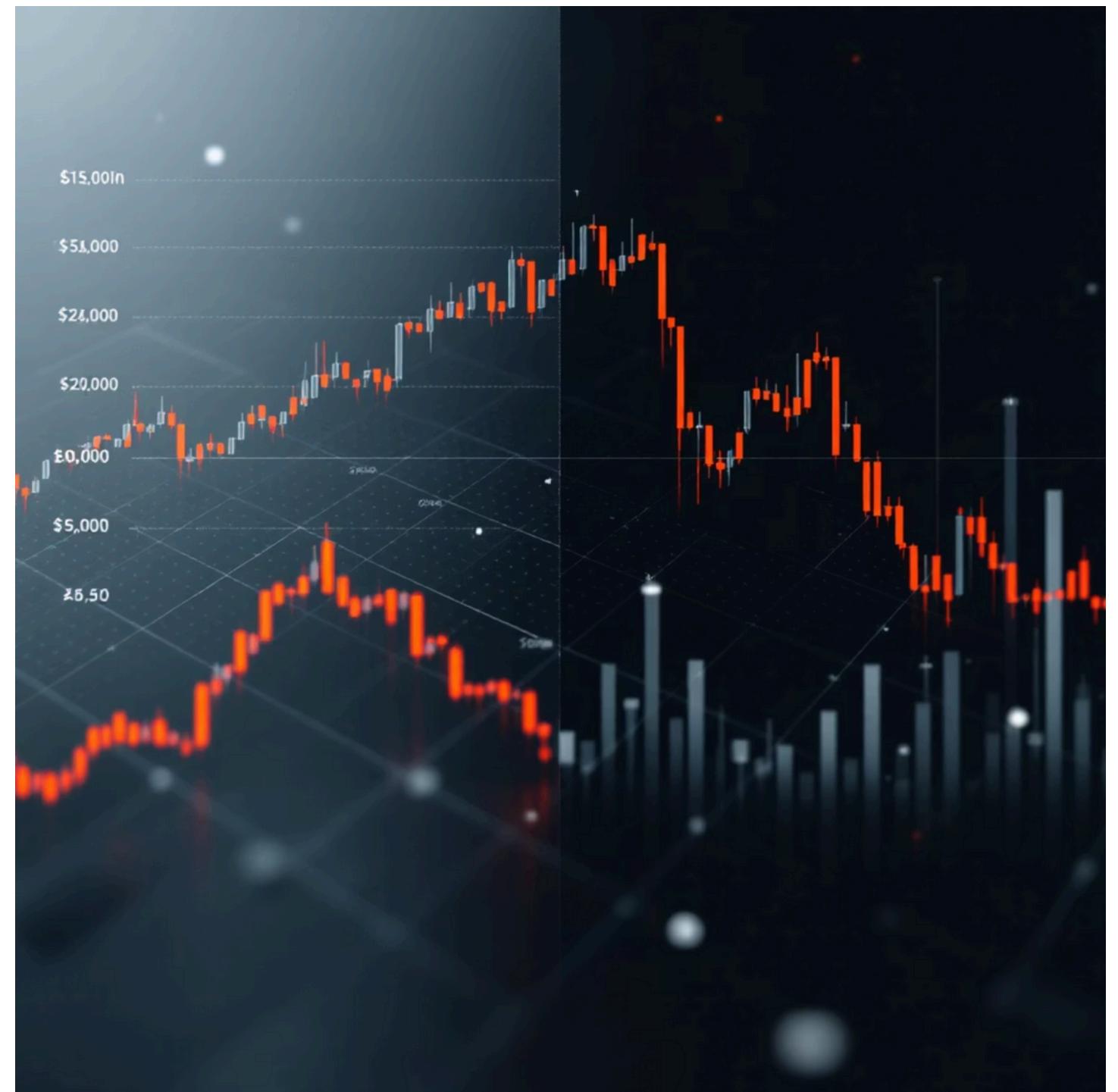


Contexto e Objetivo do Projeto

Este trabalho aplica técnicas de Inteligência Artificial e Machine Learning para detectar anomalias em séries temporais financeiras, com foco em eventos como **pump-and-dump**, **flash crashes** e **manipulações de mercado**.

Ativos Analisados

- **Bitcoin (BTC-USD)**: Mercado maduro, líquido e mais estável.
- **TrumpCoin (TRUMP-USD)**: Memecoin altamente volátil e especulativa.



Dados Utilizados

A base de dados para esta análise foi obtida do Yahoo Finance (yfinance), com granularidade horária para capturar as flutuações intradiárias.

Detalhes dos Ativos

- **Bitcoin:** Aproximadamente 8.700 registros, cobrindo um período de 1 ano.
- **TrumpCoin:** Cerca de 17.000 registros, abrangendo todo o histórico máximo disponível.

Variáveis Principais

- OHLC (Abertura, Máxima, Mínima, Fechamento) e Volume.
- Indicadores técnicos: RSI, MACD, Bollinger Bands.
- Métricas derivadas: variação de preço, volume e volatilidade.
- Features temporais: hora do dia e dia da semana para capturar padrões.



Pré-processamento e Engenharia de Features

Um rigoroso processo de pré-processamento foi fundamental para garantir a qualidade dos dados e a robustez dos modelos.

Remoção de Valores Inválidos

Eliminação de NaN e valores infinitos para evitar distorções na análise.

Clipping de Extremos

Limitação de outliers nos percentis de 1% a 99% para mitigar o impacto de dados anômalos extremos.

Normalização

Utilização do RobustScaler, um método resistente a outliers, para padronizar as features.

Labels Sintéticos

Criados apenas para avaliação, baseados em quantis estatísticos, pois o problema é tratado como **detecção de anomalias não supervisionada**, sem *ground truth* real.

Algoritmos Avaliados

Quatro modelos clássicos de detecção de anomalias foram selecionados e configurados para esta análise, todos com uma taxa de contaminação (contamination / nu) de aproximadamente 1,5%, consistente com a raridade esperada de eventos anômalos.



Isolation Forest

Algoritmo eficiente para isoler anomalias devido à sua natureza de divisão aleatória.



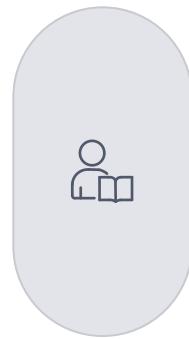
Local Outlier Factor (LOF)

Identifica anomalias com base na densidade local de cada ponto de dados.



One-Class SVM

Construção de um limite para dados "normais", classificando o que está fora como anomalia.



Isolation Forest (Deep)

Variante aprimorada do Isolation Forest, que incorpora maior diversidade de árvores para detecção mais robusta.



Resultados – Bitcoin

O Bitcoin, um mercado mais estável e líquido, apresentou desafios e padrões distintos na detecção de anomalias.

1

Desempenho do Isolation Forest

Apresentou o melhor equilíbrio entre **precisão** e **recall**, indicando sua robustez em cenários de mercado maduro.

2

Desempenho do LOF

Teve performance inferior, sugerindo menor eficiência em ambientes com densidade de dados mais homogênea.

3

Baixos Valores de F1-score

- Refletem a **raridade real** das anomalias.
- Indicam a **natureza aproximada** dos labels sintéticos.



Resultados – TrumpCoin

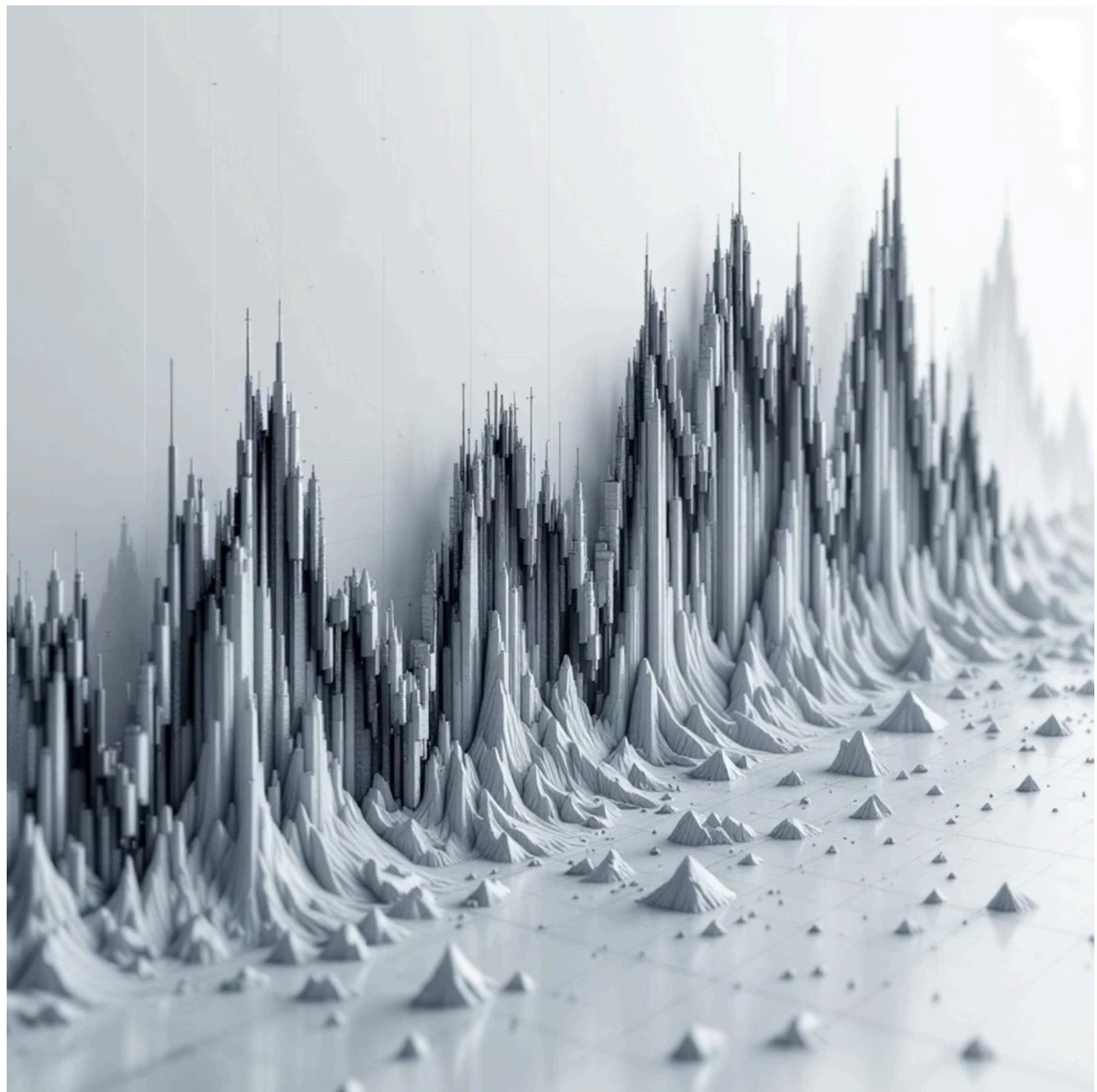
A TrumpCoin, uma memecoin de alta volatilidade, expôs as limitações das métricas tradicionais em mercados extremamente ruidosos.

Características do Ativo

- Alta volatilidade e comportamento especulativo.
- Grande quantidade de eventos extremos, com labels sintéticos indicando anomalias em aproximadamente 40% dos casos.

Desempenho dos Algoritmos

- **LOF e Isolation Forest (Deep)** mostraram maior precisão relativa.
- Métricas tradicionais de avaliação tornam-se menos confiáveis devido ao **ruído extremo**.



Conclusão

Este trabalho demonstra que a eficácia da detecção de anomalias depende intrinsecamente do perfil do ativo, do regime de mercado e do objetivo da análise.



Diversidade de Modelos

Não existe um modelo único ideal para todos os mercados financeiros.



Robustez do Isolation Forest

O Isolation Forest provou ser uma escolha robusta como **baseline** para a maioria dos cenários.



Memecoins e Contexto

Memecoins exigem abordagens mais cuidadosas e **interpretação contextual** devido à sua natureza volátil.