

Exploring Machine Learning Techniques to Identify Fraudulent Cryptocurrency Transactions: A Case Study at Mercado Bitcoin

Kevin de Santana Araujo

`kevin.araujo@mb.com.br`

`https://www.mercadobitcoin.com.br`

Mercado Bitcoin (MB)

Anomalias em Transações de Criptomoedas
Brasília, Dezembro 2023

Sumário I

Introdução

Contexto

Problema

Pesquisa

Pergunta de Pesquisa

Objetivos

Objetivos Específicos

Métodos da Pesquisa

Trabalhos Correlatos 1/2

Trabalhos Correlatos 2/2



Sumário II

Resultados

Exploração Inicial

Discussão

Mercado Bitcoin

Relevância na Instituição

Referências

Mercado Bitcoin

- Principal corretora de criptoativos da américa latina;
- R\$ 50 bilhões em volume negociado (trading in) na plataforma desde 2013;
- Mecanismos existentes para detectar e combater fraudes;
- Detecção de transações fraudulentas.

Fraudes e Manipulações

- US\$ 2.57 bilhões fraudados em 2022 com criptomoedas [1];
- MB: Comportamento de robôs na plataforma representam 90% de todas as requisições da plataforma;

Fraude

- O que é
- Como caracterizar?
- Quando acontece?



Pergunta de Pesquisa

- Considerando um conjunto de dados composto de dados de negociação em uma corretora de criptomoedas, podemos determinar se há anomalias nos dados e classificá-las?

Objetivos

- É possível modelar um mecanismo de detecção para detectar um comportamento de manipulação de mercado?
- Quais técnicas de aprendizado de máquina não supervisionadas são comumente empregadas para detectar anomalias em séries temporais?
- Quais técnicas de aprendizado de máquina não supervisionadas para detecção de anomalias são adequadas para o contexto específico que estamos estudando?



Objetivos Específicos

- Coletar dados de negociação da API pública do Mercado Bitcoin;
- Preparar e transformar os dados coletados;
- Aplicar técnicas de *fine-tuning* em modelos já treinados;
- Explorar algoritmos de *machine learning*, *deep learning* e *large language models*;
- Avaliar estes modelos inserindo deliberadamente dados fraudulentos e verificar se o modelo detectará ou não essas manipulações;
- Comparar os resultados com modelos bem conhecidos e estabelecidos;
- Coletar e debater os resultados.

Métodos da Pesquisa

- Coleta de dados de negociação a partir da API pública do Mercado Bitcoin e pré-processamento dos dados;
- Feature Engineering
- Arquitetura do ambiente a ser utilizado ao longo da pesquisa;
- Arquitetura do(s) modelo(s);
- Treinamento e fine-tuning;
- Demonstração e avaliação do(s) modelo(s);
- Design Science Research Process (DSR) aplicado ao longo de toda a pesquisa;

Trabalhos Correlatos 1/2

Detecting Financial Market Manipulation with Statistical Physics Tools

- Dados do *flash crash* do par **LUNA/USD** de Maio de 2022;
- Ordens de um livro de ofertas são modelados como partículas físicas, criação de ordens são partículas que adentram o sistema, cancelamento de ordens são partículas que deixam o sistema e uma transação entre ordens é a destruição de uma partícula;
- A mudança do *momentum* no sistema é utilizado para calcular o impacto de cada ordem;
- Foi capaz de detectar *spoofing* e *layering*, duas das mais comuns manipulações de mercado.

Trabalhos Correlatos 2/2

Adaptive Hidden Markov Model With Anomaly States for Price Manipulation Detection

- Detecção de manipulação de preços em ativos do mercado de ações;
- Uma janela de tempo é selecionada, e então um dataset com os preços de compra e venda do ativo são selecionados. O vetor de features é construído utilizando-se wavelet e outras técnicas de análise estatística;
- Uma *Adaptive Hidden Markov* é construída extraíndo-se as *probability density function* das features;
- Os estados ocultos são as anomalias.

Dataset 1 - Negociações Realizadas

- Trades (execuções), histórico de negociações **realizadas**;
- Apresenta a intenção concretizada das negociações, diferentemente do *orderbook* que apresenta as intenções de compra e venda.
- Dados de transações do ativo BTC até o começo de outubro;
- Jupyter notebook

Dataset 2 - Livro de Ofertas

- Trades (execuções), histórico de negociações **realizadas**;
- Livro de ofertas é composto por duas listas: (1) uma lista com as ofertas de compras ordenadas pelo maior valor; (2) uma lista com as ofertas de venda ordenadas pelo menor valor;
- Uma oferta é constituída por uma ou mais ordens, sendo assim, a quantidade da oferta é o resultado da soma das quantidades das ordens de mesmo preço unitário. Caso uma oferta represente mais de uma ordem, a prioridade de execução se dá com base na data de criação da ordem, da mais antiga para a mais nova;
- 1000 transações (compra e venda) mais recentes do ativo BTC;
- Jupyter notebook

Discussão

- Manipulações podem ser confundidas com comportamentos normais do mercado;
- Dados auxiliares;
- Outras corretoras;
- Dataset(s).

Pontos em Aberto

- Comportamento de robôs na plataforma representam 90% de todas as requisições da plataforma;
- Mercado de ações convencionais;
- Como caracterizar uma manipulação?

Manipulação de Mercado e Mercado Bitcoin

- Mecanismo de detecção preliminar de uma possível manipulação de mercado;
- Transações suspeitas precisam passar por um back-office;
- Acontece após a execução de uma ordem (pós-trading).

- [1] Federal Bureau of Investigation (FBI). *Internet Crime Report for 2022*. URL: https://www.ic3.gov/Media/PDF/AnnualReport/2022_IC3Report.pdf.



Agradecimentos

O projeto conta com o apoio do Mercado Bitcoin.



Obrigado! Perguntas?