



MINERAÇÃO DE DADOS

PROJETO PRÁTICO

Prever o retorno das criptomoedas com o sentimento do Twitter

Grupo 1:

Rui Monteiro, Rodrigo Rodrigues, José Gonçalves
PG50739 PG50726 PG50519

ÍNDICE



01

Motivação e Objetivos

Prever o retorno das criptomoedas com o sentimento do Twitter

Fontes de dados

Quais dados, de quais fontes serão importantes para utilizar de forma a conseguir atingir os objetivos pretendidos

02

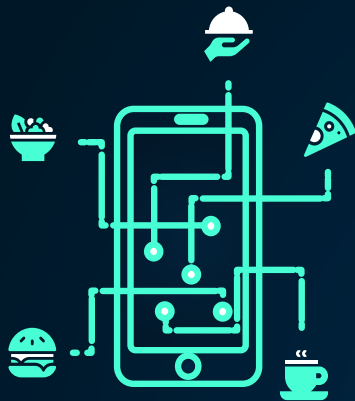


Metodologia e Plano de Desenvolvimento

Metodologia utilizada e plano de desenvolvimento para o obter



03



1. MOTIVAÇÃO E OBJETIVOS

- As criptomoedas são ativos altamente voláteis e imprevisíveis, tornando a previsão de preços desafiadora.
- O Twitter é uma fonte rica de dados não estruturados e em tempo real que pode ser usada para capturar a opinião pública sobre criptomoedas.
- Usando técnicas de análise de sentimento em dados do Twitter, é possível obter *insights* valiosos sobre as expectativas e emoções em relação às criptomoedas.
- O objetivo deste trabalho é desenvolver um modelo de previsão que utilize dados do Twitter para prever os retornos das criptomoedas, permitindo assim uma medida de corroboração face às opiniões dos utilizadores.



Cryptocurrency Price Predictions @cryptojohnsmith · 50 min
ETHEREUM PRICE PREDICTION

PRICE : FALL 📉 in 60 min.
SELL 2.77 ETH for \$1.804.0 apiece on Binance
Target price : \$1,767.92
Amount: \$4,997



That Martini Guy B @MartiniGuyYT · 24 de mar
#Bitcoin 🍷 looks spicy

Its warming up ☹️

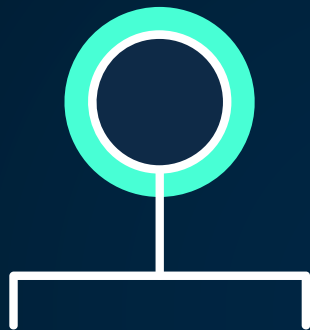


Crypto Rover @rovercrc · 27 de mar
#Bitcoin 🍷 Volatility is about to pick up! ☹️



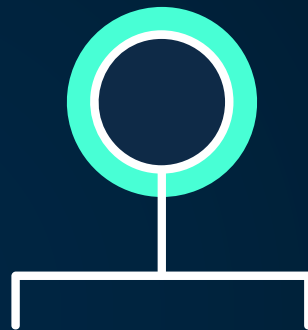
BitQuant @paraboloic · 29 de mar
I don't know whether this is good news or bad news for you, but according to Fibonacci levels, #Ethereum is expected to reach \$16,700. The timeframe is likely the same as for #Bitcoin 🍷, I suppose. Meaning this year 🤔

2. FONTES DE DADOS



Twitter Scraper

Recorrer a uma biblioteca python para obter Tweets relativamente às diferentes keywords (twitter scraping).



API de criptomoedas

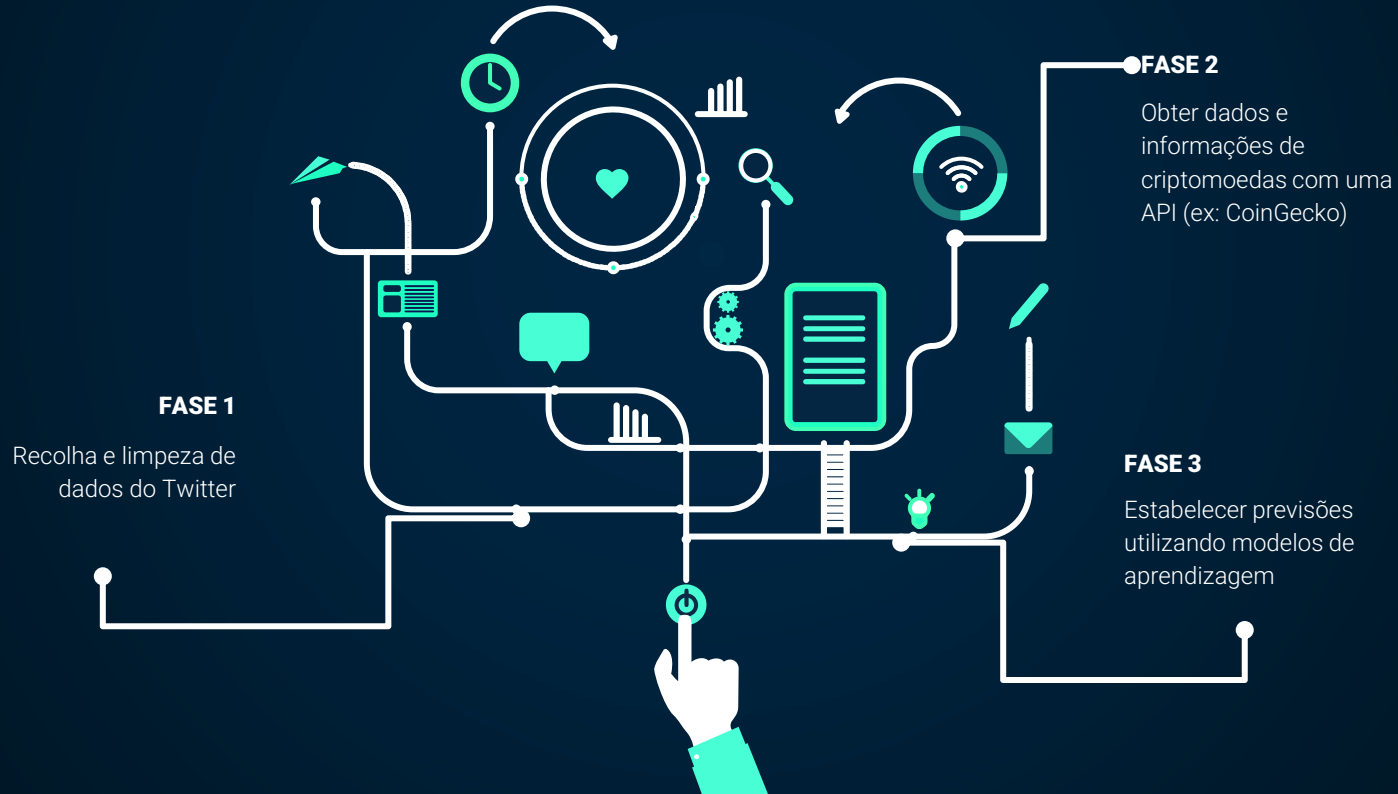
Obter dados através de uma API que permita filtrar por criptomoeda, intervalo de tempo, entre outros.

3.1 METODOLOGIA E PLANO DE DESENVOLVIMENTO

CRISP-DM

1. **Business Understanding**: Definir claramente os objetivos e os requisitos do projeto.
2. **Data Understanding**: recolher e explorar os dados do Twitter sobre criptomoedas.
3. **Data Preparation**: limpar os dados do Twitter e transformá-los num formato adequado para análise de sentimento.
4. **Modelling**: utilizar modelos de previsão que utilizem os dados do Twitter para prever os retornos das criptomoedas, permitindo comparar com os resultados reais obtidos através de uma API.
5. **Avaliação**: avaliar o desempenho do modelo de previsão (pode ser feito usando técnicas de validação cruzada e análise de sensibilidade), extraíndo conhecimento através de um conjunto de métricas.

3.2 FASES DO PROJETO



FASE 1



01

Recolher dados

Obter tweets relativos à criptomoeda e armazená-los

Limpar dados

Aplicar pré-processamento ao texto para obter dados limpos e standardizados (com efeitos na polaridade)

02



03

Derivar sentimento

Usar ferramentas como *TextBlob* ou *VADER algorithm* para obter o sentimento dos Tweets

Polaridade dos tweets (emoções expressas — positivo, negativo, ou neutro)

Rácio *bullish* do tweet (nº de tweets positivos sobre nº tweets negativos)

FASE 2



01

Obter dados históricos relativos ao preço e tendências de criptomoedas

Recorrer a uma API que permita obter preços de uma determinada criptomoeda, assim como estabelecer filtros temporais

Limpar dados

Aplicar técnicas de pré-processamento aos dados obtidos através da API de forma a poder utilizá-los posteriormente

02



FASE 3

01 **Aplicar modelos que permitam tirar conclusões e corroborar expectativas**

Aplicar modelos de aprendizagem de forma a prever de forma satisfatória os retornos das criptomoedas em função dos dados obtidos. Através dos resultados obtidos, pretende-se efetuar uma análise estatística que permita obter dados como:

- Correlação entre tweets e interações em tweets com os resultados dos mercados (*Cross-correlation Analysis*), de modo a verificar métricas como a credibilidade dos tweets.
- Aplicar testes de causalidade (p.e. Granger Causality Test), para obter métricas como a quantidade de interações com o tweet significar maior fidedigno

Com isto, será possível utilizar uma RNN (p.e. LSTM ou GRU) ou um *Vector Autorregressive Model* (VAR) para prever o valor de criptomoedas.



MINERAÇÃO DE DADOS

PROJETO PRÁTICO

Prever o retorno das criptomoedas com o sentimento do Twitter

Grupo 1:

Rui Monteiro, Rodrigo Rodrigues, José Gonçalves
PG50739 PG50726 PG50519