

Bitcoin: Predição da valorização da criptomoeda

1st José Carlos da Silva Cruz

Centro de informática da Universidade Federal de Pernambuco (CIn-UFPE)

Recife, Brasil

jcsc@cin.ufpe.br

Abstract—Este documento descreve a proposta de projeto para a predição da taxa de valorização do bitcoin após uma semana, analisando o valor da criptomoeda em um intervalo de tempo antes da data que deseja ser analisada. O método utilizado será machine learning com o algoritmo ingênuo de Bayes.

Index Terms—Bayes, Machine, learning, bitcoin, criptomoedas, estatística, big, data

I. INTRODUÇÃO

Este documento descreve toda a proposta planejada para a realização do aprendizado de máquina sobre a predição da valorização do bitcoin. Logo, será descrita a forma utilizada para a construção do banco de dados, métodos de aprendizagem e análise dos dados e, conseqüentemente, o resultado final e modos de uso da aplicação em estudo.

II. BANCO DE DADOS

A. Buscando informação

A busca por dados para serem analisados é a principal preocupação de todo analista de dados, e para a análise do nosso objetivo não será diferente. Há várias plataformas que armazenam dados para treinamentos, tais como a "UCI Machine Learning" [1], contudo, não há nesta plataforma dados sobre o bitcoin ou qualquer outra criptomoeda. Contudo, a fonte de dados sobre o valor da moeda em todo seu histórico pode ser facilmente encontrada na invest.com [2]. A informação é armazenada e pode ser filtrada por períodos, com preços de máxima e mínima para cada dia desde o nascimento da criptomoeda.

B. Filtrando os dados

A forma de retirar informações de uma plataforma online pode ser facilmente realizada com técnicas famosas chamadas "webscrapping", técnica utilizada para filtrar informações da internet por meio de softwares automatizados. Logo, para realizar esta tarefa, será utilizado a linguagem python [3] e a ferramenta selenium [4]. Portanto, apenas com essas duas ferramentas seremos capazes de montar nosso próprio banco de dados e utilizá-lo na aprendizagem de máquina.

III. ANÁLISE E TRATAMENTO DOS DADOS

Após armazenar todos os dados que serão utilizados no treinamento, deveremos levar em consideração como eles se comportam, removendo qualquer viés que possa atrapalhar a análise. Um exemplo claro é desconsiderar alguns dados que constam no início do surgimento da criptomoeda, pois é perceptível que o cenário de compra foi completamente

modificado com o avanço da tecnologia e inclusão digital no mundo todo. Algumas das técnicas utilizadas para análise dos dados serão: histograma, gaussiana e/ou redes neurais.

A. histogramas

O histograma será utilizado para contabilizar a frequência de um determinado intervalo de valorização da moeda, sabendo qual a valorização em um intervalo anterior ao dia em estudo. Este histograma será utilizado para suavizar resultados que não estão presentes no banco de dados, tendo em vista que nosso campo de estudo é classificar resultados no conjunto dos números reais.

B. gaussiana

A gaussiana será uma provável resposta na suavização dos dados. Assumindo que os dados obedecem a curva normal, poderemos prever o resultado esperado para um determinado valor real que não estava presente no conjunto de dados. Aumentando assim o poder de informação sobre o conjunto de dados analisados.

C. redes neurais

Portanto, caso os dados não obedeçam a curva normal, precisamos encontrar um função que classifique aproximadamente cada valor real com a sua projeção de valorização da criptomoeda. Portanto, para tal atividade, faz-se necessário o uso de técnicas de redes neurais. A rede neural que poderá ser utilizada é apenas um simples conjunto baseado na "back-propagation", onde a rede aprende sobre os dados analisados e consegue prever um resultado que pode ou não condizer com o seu rótulo.

IV. ALGORITMO INGÊNUO DE BAYES

O algoritmo ingênuo de Bayes, baseado no teorema de Bayes, é uma análise probabilística sobre os dados apresentados. O algoritmo analisa a amostra de dados e calcula qual a probabilidade de um evento acontecer considerando a existência de um outro evento qualquer. Em termos matemáticos, seja **A** um evento desejável, qual a probabilidade de **A** acontecer, considerando que o evento **B** aconteceu?

O teorema de Bayes responde essa pergunta da seguinte forma:

$$P(A|B) = \frac{P(B|A) * P(B)}{P(A)} \quad (1)$$

Em que $P(A|B)$ é a probabilidade do evento **A** acontecer, considerando que o evento **B** já aconteceu, $P(B|A)$ é a

- [1] sem autor. UCI Machine Learning. Disponível em:
<https://archive.ics.uci.edu/ml/datasets.php?format=&task=cla&att=&area=&numAtt=&>
Acesso em: 11, novembro de 2021.
- [2] sem autor. Dados históricos do bitcoin. invest. Disponível em:
<https://br.investing.com/crypto/bitcoin/historical-data>. Acesso em: 11,
novembro de 2021.
- [3] sem autor. Python documentation. Disponível em:
<https://www.google.com/search?q=python&oq=python&aqs=chrome..69j57j0i512j0i43j0l7j8>. Acesso em: 11, novembro de 2021.
- [4] sem autor. Selenium documentation. Disponível em:
<https://www.selenium.dev/documentation/>. Acesso em: 11, novembro
de 2021.