Olá!

Eu me chamo **João Machado**, sou aluno do curso de Ciência da Computação da Universidade federal Fluminense, campus de Rio das Ostras.

O site **Criptomante** faz parte do meu Projeto Final.

Seu objetivo é realizar previsões para o valor da bitcoin. Essas previsões são baseadas em seu histórico de valores e no que as pessoas conversam na internet sobre a criptomoeda.

Todas as previsões e análises são baseadas na cotação do Bitcoin em dólar americano.

Todas as análises se baseiam no princípio de que todo comportamento humano, em perspectiva coletiva, possui tendências cíclicas.

Esse princípio é extensivamente explorado pelo Princípio de Ondas de Elliott, descrito por Ralph Nelson Elliott.

A parte técnica do projeto inclui:

* Programação orientada a objetos, utilizando Python
* Banco de dados, utilizando Postgresql
* Queries avançadas em SQL, específicas do Postgresql
* Automatização de leitura de páginas WEB e extração de informação (Web Crawling). Utilizando Selenium
* Extração de informação de arquivos CSV
* Consumo de APIs REST
* Multithreading
* Interpretação de linguagem natural (inglês), utilizando Spacy
* Aprendizado de Máquina, utilizando Scikit-Learn
* Produção de um website, utilizando Django, Html5, Css3 e javaScript

**Extração do histórico de cotações do Bitcoin**

Para que possamos realizar previsões sobre o valor do Bitcoin, é importante que saibamos seu valor atual e passado.

A plataforma Bitcoincharts é a maior fornecedora de informações técnicas e financeiras sobre o Bitcoin do mundo. A plataforma disponibilizada uma API-Json, com informações de todas as transações realizadas nos últimos cinco dias nos principais fazedores de mercado (sites onde é possível comprar ou vender bitcoin) do mundo.

Para informações anteriores aos últimos cinco dias, a plataforma Bitcoincharts disponibiliza os dados em formato CSV.

Nesse projeto, nos concentramos nas transações feitas no fazedor de mercado Bitstamp. O Bitstamp é uma das maiores plataformas de troca entre Bitcoin e Dólar Americano ainda ativa. Seu histórico possui mais de 21 milhões de transações desde setembro de 2011.

Para realizar o consumo da API, utilizamos a biblioteca Requests. E então interpretamos o arquivo CSV de resposta, e alimentamos nosso banco de dados com as informações das transações.

Agrupamos as transações em intervalos de 60 minutos. O valor médio considerado é o valor médio das transações, independente do volume de cada transação.

**Extração de conversas sobre o Bitcoin na internet**

Além de analisar o histórico de cotações do Bitcoin, também analisados o conteúdo publicado em plataformas de conversação sobre a criptomoeda.

A plataforma que observados é a categoria */r/BitcoinMarkets* da rede social Reddit.

A extração de dados funciona em duas partes. Primeiro, utilizamos a API da plataforma Pushshift, para obter os tópicos publicados na plataforma. Em seguida, utilizando Selenium, executamos um crawler para ler os conteúdos dos tópicos e as mensagens dos usuários.

**Análise Numérica**

Após ter todas as informações relevantes em nosso banco de dados, realizamos o que chamamos de análise numérica.

Seguindo o Princípio de Ondas de Elliott, criamos uma estrutura de dados para o atual contexto do Bitcoin, e comparamos com os contextos históricos. Para assim, prevermos seu valor futuro.

Essa análise é composta pelos seguintes passos:

1. Definimos uma estrutura de dados, contendo a variação do bitcoin entre diferentes intervalos de tempo vizinhos a um momento específico, chamado de momento-base. Por exemplo, sendo o momento-base 25/04/2020 11:00, observamos a variação do bitcoin entre esse momento e 24/04/2020 11:00. E entre 25/03/2020 11:00 e 25/04/2020 11:00. Chamamos essa estrutura de dados de Snapshot
2. Calculamos todos os Snapshots para todos os valores do Bitcoin, desde 2011
3. Para todos os snapshots históricos, armazenamos como o bitcoin variou nos dias seguintes ao momento-base. (Por exemplo, se o momento-base for 25/04/2020 11:00, sabemos que 24 horas depois sua variação foi de +0.5%). Chamamos isso de variação futura do snapshot
4. Usamos uma métrica para definir a semelhança entre dois Snapshots
5. Finalmente, elaboramos um snapshot com o momento-base sendo o momento atual. Usamos a métrica do passo 4 para definirmos peso para cada Snapshot histórico. E então, prevemos que a variação futura do Snapshot atual (Passo 3) será igual a média aritmética ponderada dos Snapshots históricos (Passo 2).

**Análise por Repetição de Comentários**

Como todo ativo financeiro sujeito a especulação, o valor do Bitcoin é afetado pela opinião de seus negociadores.

Segundo o Princípio de Ondas de Elliott, essas opiniões são cíclicas, e consequentemente, afetam o valor do Bitcoin de forma cíclica.

Nesse projeto desenvolvemos duas abordagens para estudar os ciclos de opinião, e tentar estimar seu impacto no valor do Bitcoin.

A primeira abordagem, que é bem simples, é chamada de Análise por Repetição de Comentários.

Essa análise é composta pelos seguintes passos:

1. Listamos todas as mensagens enviadas nas últimas 24 horas.
2. Utilizamos o processador de linguagem natural Spacy para dividir as mensagens em frases isoladas, utilizando as regras gramaticais do idioma inglês.
3. Buscamos, em nosso banco de dados, outras ocorrências dessas frases. Para cada ocorrência anterior, verificamos se o momento da ocorrência precedeu um aumento ou queda no valor do Bitcoin de pelo menos 5%. Dessa forma, para cada frase, obtemos a razão:  
   Positividade = <Ocorrências que precederam aumentos>/<Quantidade de ocorrências>
4. Calculamos a Positividade atual como a média da Positividade das sentenças postadas nas últimas 24h
5. Por fim, ajustamos os valores estimados na Análise Numérica em 5% para mais ou para menos, dependendo se a Positividade Atual é maior ou não é maior do que 0.5, respectivamente

**Análise Semântica de Comentários**

Além da Análise por Repetição de Comentários, nesse projeto será apresentada uma segunda abordagem para identificar ciclos de opiniões sobre o Bitcoin.

Essa abordagem, chamada de Análise Semântica de Comentários, utiliza lematização de comentários antigos e o histórico do valor da bitcoin para obter o valor futuro do Bitcoin utilizando a técnica de Regressão Logística com a biblioteca Scikit-Learn.

Essa análise é composta pelos seguintes passos:

1. Para cada mensagem em nosso banco de dados, utilizamos o processador de linguagem natural Spacy para dividir as mensagens em frases isoladas, utilizando as regras gramaticais do idioma inglês
2. Para cada frase obtida, realizamos o processo de lematização utilizando o processador Spacy. E armazenamos a lematização da frase em nosso banco de dados.
3. Para cada mensagem em nosso banco de dados, atribuímos uma classificação binária indicando se a data de publicação da mensagem precedeu um aumento (1) ou uma queda (0) do valor da bitcoin de pelo menos 5%. Mensagens que não tenham precedido aumentos ou quedas de pelo menos 5% são ignoradas.
4. Utilizando a lematização obtida no passo 2, e a classificação obtida no passo 3, obtemos um dataset (Conjunto de dados) classificados. Dividimos aleatóriamente nosso dataset em dois. Sendo um dataset contendo 90% das frases lematizadas, e sua classificação. E sendo outro dataset contendo 10% das frases lematizadas.
5. Utilizamos o primeiro dataset para criar um modelo classificador binário por Regressão Logística utilizando a biblioteca Scikit-Learn. E então usamos esse modelo para classificar o segundo dataset.
6. Como o segundo dataset foi feito a partir de frases antigas, cuja classificação é conhecida e foi obtida no passo 3, comparamos a classificação prevista no passo 5 com a classificação real. Assim obtemos os indicadores de Verdadeiros Positivos, Falsos Positivos, Verdadeiros Negativos e Falsos Negativos. Onde a classificação positiva significa: “O valor do Bitcoin, caso tenha variação de 5% dentro de 24 horas após a publicação da mensagem, essa variação será um aumento”. Esses indicadores estão presentes no Dashboard.
7. Por fim, listamos e lematizamos todas as mensagens recentes publicadas sobre o Bitcoin, e então utilizamos nosso modelo para estimar se a tendência de variação do Bitcoin é positiva ou negativa. Obtendo essa tendência, manipulamos o gráfico principal do Dashboard em 5%, para mais ou para menos.