

**UNIVERSIDADE FEDERAL FLUMINENSE**  
**INSTITUTO DE CIÊNCIA E TECNOLOGIA**  
**BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

**JOÃO MACHADO DA SILVA JUNIOR**

Criação de estimativas para o valor da criptomoeda Bitcoin utilizando  
comentários extraídos de uma rede social

Rio das Ostras – RJ

2020

JOÃO MACHADO DA SILVA JUNIOR

Criação de estimativas para o valor da criptomoeda Bitcoin utilizando comentários  
extraídos de uma rede social

Monografia apresentada ao Curso de  
Bacharelado em Ciência da Computação  
do Instituto de Ciência e Tecnologia,  
como requisito parcial para obtenção do  
Grau de Bacharel.

Orientador:

Prof. Dr. CARLOS BAZÍLIO MARTINS

Rio das Ostras – RJ

2020

JOÃO MACHADO DA SILVA JUNIOR

Criação de estimativas para o valor da criptomoeda Bitcoin utilizando comentários  
extraídos de uma rede social

Monografia apresentada ao Curso de  
Bacharelado em Ciência da Computação  
do Instituto de Ciência e Tecnologia,  
como requisito parcial para obtenção do  
Grau de Bacharel.

Aprovado no mês de \_\_\_\_\_ do ano de \_\_\_\_\_

BANCA EXAMINADORA

---

Prof. Dr. CARLOS BAZÍLIO MARTINS - Orientador

UFF

---

Prof. NOME DO PROFESSOR

INSTITUIÇÃO

---

Prof. NOME DO PROFESSOR

INSTITUIÇÃO

Rio das Ostras – RJ

2020

# Resumo

O Bitcoin é uma criptomoeda descentralizada baseado em uma tecnologia ponto-a-ponto proposto como uma forma de dinheiro eletrônico global.

Como a maioria das moedas contemporâneas o Bitcoin é uma moeda fiduciária. Isso significa que seu valor não está lastreado em nenhum metal ou outro recurso material. Seu valor depende da confiança (fidúcia) das pessoas. Esse fato, adicionado ao fato de não haver nenhuma instituição reguladora que ofereça garantia de qualquer tipo para seus usuários, fizeram com que o Bitcoin se tornasse um ativo financeiro especulativo.

Como todo ativo sujeito a especulação financeira seu valor é influenciado pela opinião que seus investidores tem sobre sua tendência de preço.

Este trabalho apresenta o desenvolvimento de uma aplicação capaz de extrair comentários sobre o Bitcoin de uma rede social, assim como extrair informações sobre o histórico de valor do Bitcoin, e propor múltiplos modelos de análise para estimar o valor futuro do Bitcoin a partir dessas informações extraídas.

A aplicação contém mineração de dados e mineração de texto para abastecer um banco de dados. O desenvolvimento dos modelos propostos inclui processamento de linguagem natural, aprendizado de máquina, regressão logística e análise de sentimentos.

O resultado das análises e uma descrição menos formal do projeto estão disponíveis na rede mundial de computadores, no endereço <http://criptomante.online>.

# Sumário

1. Introdução
2. Referencial Teórico
  - 2.1. Financeiro
    - 2.1.1. Criptomoedas
    - 2.1.2. Bitcoin como Ativo Financeiro
    - 2.1.3. Especulação Financeira
    - 2.1.4. Princípio de Ondas de Elliott
  - 2.2. Web Crawlers
  - 2.3. API REST
  - 2.4. Processamento de Linguagem Natural
    - 2.4.1. Lematização
  - 2.5. Aprendizado de Máquina
    - 2.5.1. Classificação Estatística
    - 2.5.2. Regressão Logística
3. Trabalhos Relacionados
  - 3.1. I Know First
  - 3.2. FinBrain Tecnologias
  - 3.3. Prevendo o mercado de ações com artigos de jornais.
4. Desenvolvimento
  - 4.1. Arquitetura geral
  - 4.2. Tecnologias Utilizadas
    - 4.2.1. Linguagens de Programação
    - 4.2.2. Frameworks
    - 4.2.3. Banco de Dados
    - 4.2.4. Spacy – Processador de Linguagem Natural
    - 4.2.5. Scikit-Learn – Motor de Aprendizado de Máquina
  - 4.3. Mineração de dados
    - 4.3.1. Tecnologias Utilizadas
    - 4.3.2. Mineração de Transações
    - 4.3.3. Mineração de Texto
  - 4.4. Análise por Padrões Numéricos

- 4.4.1. Estrutura de Dados
    - 4.4.2. Roteiro
  - 4.5. Análise por Repetição de Comentários
    - 4.5.1. Tratamento de Texto
    - 4.5.2. Roteiro
  - 4.6. Análise de Sentimentos
    - 4.6.1. Tratamento de Texto
    - 4.6.2. Roteiro
    - 4.6.3. Precisão Obtida
  - 4.7. Análise Consolidada
  - 4.8. Front-End
  - 4.9. Deploy
- 5. Conclusão
  - 5.1. Desafios
  - 5.2. Resultados Obtidos
  - 5.3. Sugestões de Trabalhos Futuros

# Capítulo 1

## Introdução

Em 2008 uma pessoa chamada Satoshi Nakamoto publicou em um artigo chamado *Bitcoin: Um Sistema de Dinheiro Eletrônico Ponto-a-Ponto* em um grupo de e-mail do site metzdowd.com. Esse artigo propunha uma nova forma de dinheiro. Eletrônico e descentralizado. Independente de qualquer instituição ou governo. Que poderia revolucionar o jeito como pagamos pelo o que compramos. Esse dinheiro eletrônico seria chamado de Bitcoin.

Doze anos já se passaram desde então. O Bitcoin já é razoavelmente bem conhecido em todo o mundo, embora poucas pessoas já tenham adquirido qualquer fração de um bitcoin.

O Bitcoin nunca se tornou uma moeda popular para transações no dia a dia. Se tornou, contudo, um ativo para especulação financeira. É mais comum que seja adquirido como forma de investimento do que como moeda a ser utilizada em uma transação comercial.

Nesse trabalho desenvolvemos uma aplicação capaz de obter de fontes de dados públicas na internet contendo o histórico de preços do Bitcoin. Ao utilizar esse histórico de valores obtidos, em conjunto com comentários de usuários da rede social Reddit, também extraídos pela aplicação desenvolvida foi possível elaborar quatro heurísticas para estimar o valor futuro do Bitcoin. As heurísticas apresentadas utilizam conceitos básicos e simples de especulação financeira, em especial o Princípio de Ondas de Elliot, a partir de seu histórico de valor, análise de sentimentos, aprendizado de máquina e processamento de linguagem natural.

Inicialmente será feita uma apresentação geral sobre o Bitcoin e sobre o Blockchain, que é a tecnologia em que o Bitcoin se apoia. Em seguida, introduzimos conceitos básicos sobre especulação financeira e sobre os conceitos financeiros necessários para entender os fundamentos de cada heurística a ser apresentada.

Em seguida, serão apresentados os conceitos de API e Web Crawlers, necessários para a realização da mineração de dados.

Algumas das heurísticas propostas incluem análise de sentimentos, portanto serão introduzidos conceitos comuns em processamento de linguagem natural e em aprendizado de máquina. Em especial na técnica de regressão logística para classificação estatística.

No capítulo de Trabalhos Relacionados serão apresentadas duas iniciativas comerciais de consultoria de previsão de ações na bolsa de valores utilizando tecnologia. O último trabalho relacionado apresentado será um artigo que propõe análise de sentimentos sobre artigos em jornais americanos como uma heurística para estudar e prever o valor de ações da bolsa americana.

No capítulo de Desenvolvimento serão apresentadas as tecnologias de banco de dados, linguagens de programação, bibliotecas e frameworks utilizados.

A seguir, cada uma das quatro heurísticas propostas será apresentada em detalhe.

Ao final do capítulo, haverá uma breve descrição sobre a criação de um site para demonstrar os resultados obtidos com as análises.

A conclusão conterá alguns dos desafios encontrados ao longo do desenvolvimento do trabalho. Assim como os resultados obtidos. Tanto no viés tecnológico quanto na precisão apurada das heurísticas propostas.

O site com os resultados obtido pode ser acessado no endereço <http://criptomante.online/>

O código-fonte da aplicação está disponível no endereço [https://github.com/granpk/Criptomante\\_python](https://github.com/granpk/Criptomante_python)

# Capítulo 2

## Referencial Teórico

### 2.1 Financeiro

Este projeto trata-se de uma proposta de método de estimativa para o valor de mercado do Bitcoin. Portanto, será necessário introduzir conceitos básicos sobre criptomoedas, e sobre o Bitcoin em especial.

Como este projeto irá explorar o Bitcoin como um ativo especulativo, também será necessário introduzir conceitos básicos sobre o mercado financeiro e sobre especulação financeira.

#### 2.1.1 Criptomoedas

Em termos econômicos, moeda é tudo aquilo que é geralmente aceito para liquidar débitos e transações, isto é, para pagar pelos bens e serviços e para quitar obrigações. Ela é considerada o instrumento básico para que se possa operar no mercado. Pois a moeda atua como meio de troca.

As moedas contemporâneas são chamadas de moedas fiduciárias porque seu funcionamento baseia-se na confiança (fidúcia) das pessoas nas instituições que emitem e regulam a moeda. Ao contrário de outros tipos de moedas mais antigos, moedas fiduciárias não tem seu valor lastreado em nenhum metal e não possui nenhum valor intrínseco. [2]

Em 2008 um artigo chamado *Bitcoin: Um Sistema de Dinheiro Eletrônico Ponto-a-Ponto* foi publicado em uma lista de e-mails aberta pertencente ao site metzdowd.com. O artigo era assinado por alguém chamado Satoshi Nakamoto (presumidamente um pseudônimo) e descrevia um sistema de dinheiro eletrônico baseado em transações ponto-a-ponto. [3]

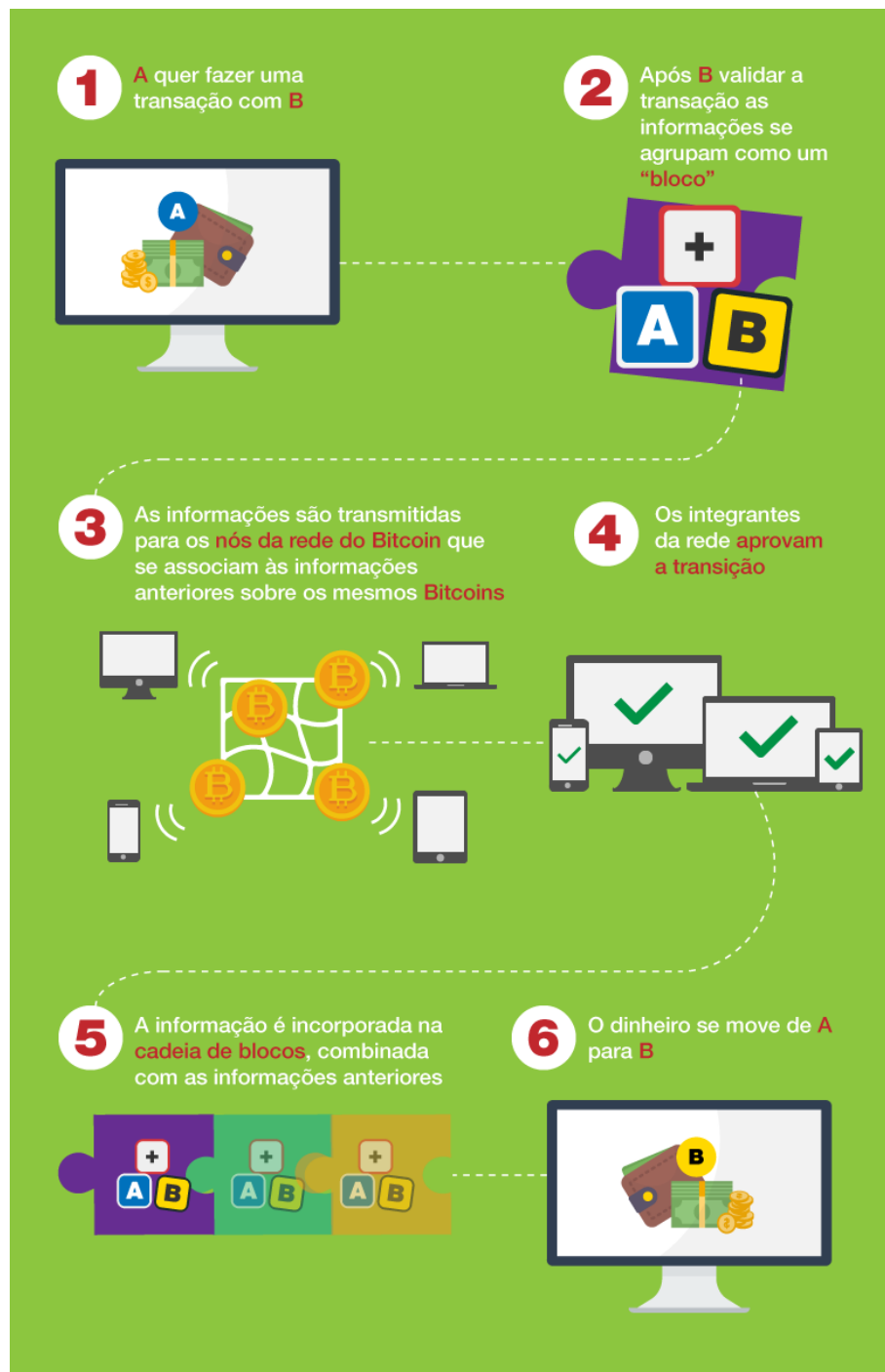
No mesmo artigo, Nakamoto propõe uma tecnologia contendo um modelo banco de dados distribuído e seguro que posteriormente veio a ser chamado de Blockchain. Segundo o artigo, o Blockchain funcionaria como um livro-razão público que contém todas as transações com a criptomoeda. Sua segurança está baseada no fato de que as transações armazenadas na rede distribuída podem ser validadas por qualquer nó da rede utilizando certos algoritmos. E, portanto, para que houvesse qualquer fraude seria necessário obter o consenso de outros nós da rede.

O escopo do Blockchain proposto por Nakamoto pode ser resumido da seguinte forma

- Transações podem ser agrupadas. Esse agrupamento se chama Bloco
- Cada Bloco possui referência para um bloco mais antigo. Dessa forma, é possível percorrer uma cadeia de blocos desde o bloco atual até o primeiro bloco da rede (chamado de bloco genesis)
- Todos os nós da rede são capazes de validar um bloco a partir de sua cadeia.
- Após um nó validar um novo bloco, ele propaga essa informação pela rede. Isso é conhecido como adicionar o bloco à Blockchain
- A propagação do bloco evita o problema econômico chamado de gasto-duplo. Onde o mesmo ativo financeiro é usado como troca em duas transações distintas.
- Existem sites que fornecem o serviço de consulta de quantas ‘validações’ um bloco recebeu. Não há regra sobre a quantidade necessária de validações para que um bloco seja considerado válido. Mas a maioria das plataformas consideram que um bloco com seis validações é um bloco válido.
- Os usuários recebem Bitcoins gratuitos como incentivo por fornecerem seus computadores para serem nós da rede, validando novos blocos e realizando outras ações de rotina. Essa atividade é chamada de mineração de Bitcoin, e não terá seu processo aprofundado nesse projeto.



Segue abaixo um infográfico



Infográfico com transação com Bitcoin. Fonte: [Wirecard](#)

O código-fonte relativo a primeira versão do banco de dados Blockchain foi tornado aberto por Satoshi Nakamoto. Desde então, o Blockchain ganhou popularidade e contribuidores ao redor do mundo. Eventualmente, bancos de dados derivados do Blockchain de Nakamoto foram usados como base para a criação de outras criptomoedas (Popularmente chamadas de altcoin).

As motivações de novas criptomoedas são as mais diversas. Por exemplo, a criptomoeda *Lightcoin* utiliza uma versão mais leve do Blockchain e isso permite maior velocidade para as transações. Já a *Gridcoin* utiliza uma versão da Blockchain onde parte de sua rede de nós executam tarefas para projetos científicos de pesquisadores que precisam de poder computacional, mas não possuem os recursos financeiros necessários.

Segue abaixo uma tabela com as principais criptomoedas contemporâneas. Ordenadas pelo volume financeiro que representam.

Criptomoeda	Volume de Mercado (Em Dólar)
Bitcoin	168.883.239.154
Ethereum	26.338.014.276
Tether	9.169.814.972
XRP	8.910.693.541
Bitcoin Cash	4.147.693.340
Cardano	3.196.402.751
Bitcoin SV	3.190.222.787
Chainlink	2.806.286.910
Litecoin	2.763.661.163
Binance Coin	2.664.209.051

Tabela com as principais criptomoedas e seu Volume de Mercado

### 2.1.2 Bitcoin como Ativo Financeiro

A primeira transação comercial conhecida utilizando Bitcoin ocorreu em 2010 quando o programador Laszlo Hanyecz comprou duas pizzas por B10.000.

Por sua natureza descentralizada e de difícil rastreabilidade, o Bitcoin ganhou popularidade como moeda para tráfico de produtos ilegais e outras atividades criminosas. A plataforma Silk Road utilizou exclusivamente Bitcoins entre fevereiro de 2011 até outubro de 2013, quando seus responsáveis foram identificados pelo FBI. Durante esse período, estima-se um movimento de 9,9 milhões de bitcoins. Essas atividades ilegais contribuíram para a valorização da moeda. Que avançou de \$0,30 em 2011 para \$ 770 em 2013.

O valor do Bitcoin teve ciclos de quedas e aumentos acentuados desde então. Esses movimentos foram causados principalmente por fenômenos que afetavam a confiabilidade que o público tinha sobre a moeda. Um bug no Blockchain causou uma queda no valor do Bitcoin de 23% em um único dia em 2013. Outras quedas foram causadas por consequência da proibição do governo chinês de transações com Bitcoin em seu território.

Houve pelo menos uma atualização do software do Blockchain que causou um aumento acentuado. Essa foi a atualização chamada de SetWit, que melhorava a escalabilidade da rede. Na semana em que essa atualização foi integrada ao Blockchain houve um aumento de 50% no valor do Bitcoin.

De acordo com o economista Mark T. Williams, a volatilidade do Bitcoin é sete vezes maior do que a do ouro e dezoito vezes maior do que a do Dólar Americano [4]. A Fundação Bitcoin, formada por defensores e entusiastas da criptomoeda, afirma que sua alta volatilidade se deve a sua baixa liquidez.

Atualmente há inúmeros sites que oferecem o serviço de corretagem de bitcoin. Esses sites oferecem uma interface gráfica para a venda e compra da criptomoeda. No Brasil, alguns sites oferecem inclusive informes de rendimento para auxiliar na declaração de imposto de renda. Nesse projeto, estudaremos principalmente transações no site <https://www.bitstamp.net/>, em operação desde 2011.

### 2.1.3 Especulação Financeira

Especulação Financeira é a compra de ativos com objetivos que não sejam o uso do ativo, mas sim a expectativa de obtenção de lucros em uma venda futura de maior valor.

Muitos especuladores prestam pouca atenção ao valor intrínseco do ativo que estão comprando. Esses se interessam principalmente pelo movimento nos preços dos ativos para obter lucros.

O uso do Bitcoin como moeda de troca, com o objetivo de realizar transações comerciais não se enquadra como especulação financeira. Porém, a aquisição de bitcoins com o objetivo de obter lucros em uma venda futura de maior valor se enquadra como uma operação de especulação financeira.

Como o Bitcoin é uma criptomoeda fiduciária sem nenhum governo que possa intervir em sua circulação, o Bitcoin passou a ser considerado um ativo sem nenhum valor intrínseco, e, portanto, ideal para a especulação financeiro. O movimento em seus valores se deve em grande parte as expectativas de seus especuladores. Como todo ativo especulativo, o preço do Bitcoin é alterado conforme há revoluções na relação de oferta e demanda.

Existem diversas técnicas de análise para realizar o estudo de viabilidade de lucro com a compra e venda de determinado ativo. Em geral essas técnicas podem ser classificadas em três tipos: Análise Técnica, Análise Fundamentalista e Análise de Sentimentos.[\[5\]](#)

A Análise Fundamentalista é o estudo da situação financeira, econômica e mercadológica de uma empresa, setor econômico, moeda ou qualquer outro ativo. Para o estudo de uma empresa, por exemplo, seria necessário estudar a área de atuação da empresa, seus concorrentes, suas demonstrações contábeis e as possíveis novidades tecnológicas que poderiam afetar sua área de negócio[\[6\]](#).

Quando aplicada ao Bitcoin, a Análise Fundamentalista estudaria o caráter tecnológico da criptomoeda. Como suas características de escalabilidade e de segurança. Outro fator que pode ser estudado é a evolução do Bitcoin como moeda para transações comerciais, que aumentaria a sua demanda e poderia diminuir a volatilidade de seu preço.

A Análise Técnica é o estudo dos movimentos de preço de determinado ativo, baseado nos ciclos de oferta e procura. Pode ser vista tanto como uma forma de psicologia social aplicada como pesquisa de opinião, onde padrões gráficos visuais ou estatísticos percebidos, seriam como fotografias do comportamento dos participantes do mercado em determinado momento.

Como exemplo de técnica de análise técnica, segue abaixo a aplicação da técnica Bandas de Bollinger em um gráfico. Nesse gráfico, o ativo em estudo são ações da empresa Vale do Rio Doce. A variação das ações ao longo de um dia é exibida no formato de um retângulo verde ou vermelho. As bandas são dadas pelas seguintes fórmulas:

Banda superior: Média Móvel Simples (20 dias) + (2xDesvio Padrão (20 dias))

Banda inferior: Média Móvel Simples (20 dias) - (2xDesvio Padrão (20 dias))

Linha central: Média Móvel Simples (20 dias)

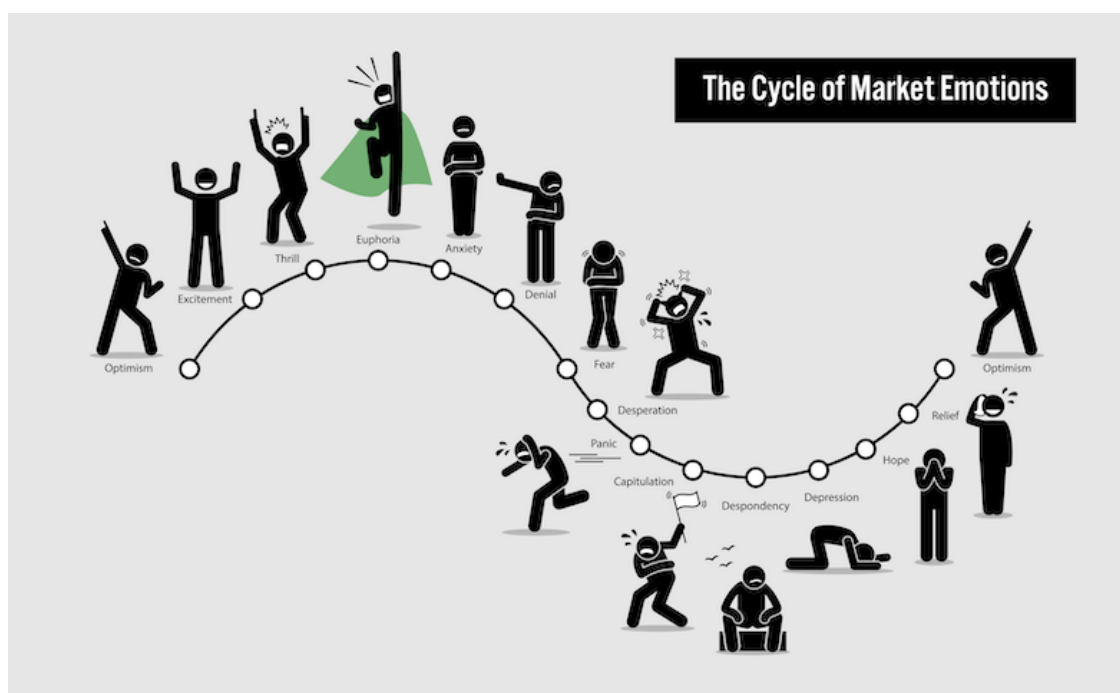


Exemplo de aplicação de Bandas de Bollinger. [7]

Uma interpretação comum para o gráfico para cima é comprar o ativo quando o valor do ativo tocar a banda inferior, e vender o ativo quando o valor do ativo tocar a banda superior. Note que não há interesse entender o motivo de variação no preço do ativo.

A análise proposta nesse trabalho nomeada de *Análise por Padrões Numéricos* se classifica como uma técnica de Análise Técnica. Existem diversas técnicas de análise técnica disponíveis na literatura financeira.

A Análise de Sentimentos se aproveita do fato de que o mercado não é movido apenas por mudanças fundamentais dos ativos, mas também pelos sentimentos e opiniões das pessoas, frequentemente de curto-prazo. O uso de inteligência artificial, aprendizado de máquina e processamento de linguagem natural pode fornecer técnicas sistematizadas de análises de sentimentos. Com exceção da Análise por Padrões Numéricos, todas as outras três análises propostas nesse trabalho aplicam, em algum grau, a Análise de Sentimentos.



Ciclo de emoções de um investidor. [8]

#### 2.1.4 Princípio de Ondas de Elliot

O Princípio de Ondas de Elliot é um conceito de Análise Técnica proposto por Ralph Nelson Elliott na década de 1930 que propõe que há ciclicidade não apenas no valor de ativos do mercado financeiro, mas também na psicologia dos investidores, nas opiniões populares e em outros fatores coletivos e sociais. Elliot. Atualmente, esse princípio é bem aceito na literatura financeira e é um conceito importante para a análise técnica.

Elliott propôs que a psicologia coletiva se move entre otimismo e pessimismo em ciclos perceptíveis. E que esse movimento de emoções gera padrões cíclicos nos valores de ações na bolsa de valores. No mesmo livro em que Elliott propôs esse princípio, ele também propõe um modelo de análise técnica para identificar esses ciclos e como usá-los para prever o valor futuro de um ativo qualquer. [9]

Sua base de raciocínio é a de que a emoção surge primeiro que a ação. Por isso, a representação gráfica de uma série histórica de cotações de um ativo nada mais é do que a oscilação de humor do coletivo numa tentativa desesperada de encontrar sua precificação.

Segundo Elliott, o público age de forma emocional, subjetiva e impulsiva, tomando decisões em condições de ignorância e incerteza, e, na maioria das vezes, assumindo a chamada *atitude de manada*.

Existem diversas técnicas de análise fundamentadas nesse princípio em maior ou em menor grau, como por exemplo o Modelo de Sentimentos de Thovallo [10]. De forma semelhante, nesse trabalho o Princípio de Ondas de Elliot será um conceito importante na construção das análises que o trabalho propõe. Contudo o seu modelo matemático não será usado.