UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE ECONOMIA

**LÓGICA FUZZY PARA PREVISÃO DO VOLUME DE NEGOCIAÇÕES EM BOLSA**

Projeto de Monografia de Conclusão de Curso

Aluno: Victor Hugo Alves Duarte

Orientadora: Prof(a). Dr(a). Rosângela Ballini

Campinas

2022

**SUMÁRIO**

**RESUMO4**

1. **REVISÃO BIBLIOGRÁFICA 4**
   1. POPULARIZAÇÃO DA BOLSA DE VALORES 5
   2. EXPECTATIVA, INCERTEZA E CONFIANÇA 6
   3. CORRELAÇÃO E CAUSALIDADE 6
2. **HIPÓTESE 4**
3. **OBJETIVOS 4**
   1. OBJETIVO GERAL5
   2. OBJETIVOS ESPECÍFICOS 6
4. **JUSTIFICATIVA 4**
5. **METODOLOGIA 4**
   1. TIPO DE PESQUISA 5
   2. PERÍODO, ATIVOS E TERMOS DE PESQUISA 6
   3. AQUISIÇÃO E ANÁLISE DOS DADOS 6
   4. CONSTRUÇÃO DA LÓGICA FUZZY 6
   5. MODELO FINAL 6
6. **CRONOGRAMA 4**
7. **REFERÊNCIAS BIBLIOGRÁFICAS 4**

**RESUMO**

Com o aumento da quantidade de dados disponibilizados na internet, o advento e popularização das mídias sociais e o recente aumento da quantidade de investidores pessoa física (PF) na bolsa de valores, surgiu a hipótese de que o volume de negociação de determinados ativos poderia ser influenciado pela popularidade dos mesmos na internet, de modo que os investidores realizem suas compras baseado na influência – dispersa ou centralizada em *infuencers* – que a internet exerce sobre as decisões de investimento.

Dado este contexto, este trabalho de monografia pretende, utilizando métodos estatísticos de análise de correlação e causalidade, explorar a influência da internet – especificamente dos principais mecanismos de busca – nas movimentações de compra na bolsa de valores do Brasil. Além disso, será desenvolvido um sistema lógico fuzzy que buscará indicar a tendência do volume de negociações com base na popularidade do ativo na internet.

**Palavras-Chave: Mercado Financeiro; Lógica Fuzzy; Inteligência Artificial.**

1. **REVISÃO BIBLIOGRÁFICA**
   1. **POPULARIZAÇÃO DA BOLSA DE VALORES**

De acordo com estudo realizado pela B3, o número de investidores pessoa física na bolsa aumentou quase 300% entre 2018, quando o número de contas era de 814 mil, e 2020, quando o número chegou a 3,2 milhões. Ainda, o estudo mostra que a maior parte dos novos investidores recorre às mídias digitais no momento de buscar informações e de aprender sobre os diferentes tipos de investimentos e ativos, estando as palavras-chave “internet”, “influenciadores” e “redes sociais” entre as mais citadas quando é questionado onde se costuma buscar informações sobre a bolsa.

Se, por um lado, o avanço da tecnologia e a popularização das mídias sociais contribuíram para a democratização do acesso aos investimentos, por outro lado, também expuseram os novos investidores aos riscos do mercado de renda variável sem que antes estivessem, necessariamente, plenamente conscientes dos riscos e educados financeiramente para que tomassem decisões compatíveis com seu perfil e momento, os tornando facilmente influenciáveis por personalidades e informações desencontradas nas redes sociais.

* 1. **EXPECTATIVA, INCERTEZA E CONFIANÇA**

[ADAPTADO] Conforme Dequech (1999), as expectativas são previsões sobre como será o futuro. A confiança, por sua vez, se refere ao uso das expectativas como guia de ação prática. Assim, a confiança pode ser definida em dois componentes: a percepção de incerteza, isto é, a possibilidade de que as expectativas estejam erradas, e a aversão à incerteza, isto é, a propensão a agir de acordo com as expectativas mesmo considerando a incerteza que as envolvem. Ainda segundo Dequech, quanto menor a confiança nas expectativas, menor a inclinação de um agente em usar suas expectativas como guia de ação na prática.

É comum, no mercado financeiro, a noção de que existem heurísticas que influenciam na tomada de decisão dos investidores. Dentre estas, existe a heurística da disponibilidade, que descreve como a disponibilidade dos eventos, isto é, a facilidade de encontrá-los na memória recente, induzem o investidor a julgar situações e tomar sua decisão. Assim, a frequência com que um ativo é comentado positivamente ou indicado na mídia pode contribuir para que esta heurística seja utilizada, de modo que a recorrência da citação do ativo pode indicar, para o investidor, que o mesmo deve ser adquirido tão somente por influências de terceiros.

[ADAPTADO] De acordo com Keynes (1936), o investidor financeiro não compra uma ação porque acredita que ela irá valorizar por conta dos fundamentos da empresa, mas porque acredita que os outros investidores também acreditam que a ação irá valorizar (e assim por diante), fazendo com que o preço suba, neste caso, ou caia, em um cenário contrário. Devido à incerteza, isto é, a falta de confiança nas expectativas, uma grande parte das decisões tomadas pelos homens – como as de investimento – dependem em certa medida da crença de que outros investidores farão o mesmo e de que o sentimento do mercado é compartilhado pela maioria dos investidores.

Deste modo, infere-se que a tomada de decisão de investidores é tomada por fatores não plenamente racionais e é fortemente influenciada pela percepção geral do mercado, uma vez que o ambiente é de incerteza e as expectativas são, muitas vezes, ancoradas na disponibilidade dos eventos ou citações.

* 1. **CORRELAÇÃO E CAUSALIDADE**

Causalidade de Granger e teste de cointegração (ler artigo do Carneiro).

No entanto, cabe ressaltar que um modelo probabilístico como se pretende construir não tem como pressuposto a causalidade entre as variáveis, tendo como objetivo final somente o alto desempenho nas métricas de avaliação, isto é, grau satisfatório de capacidade preditiva. Portanto, ainda que não seja comprovada relação de causalidade entre os eventos, será possível construir o modelo, desde que as séries temporais estejam correlacionadas.

1. **HIPÓTESE**

Com base nas referências a respeito da incerteza, expectativas e finanças comportamentais, surgiu a hipótese de que as decisões de investimento na bolsa de valores sofrem forte influência das mídias digitais, de modo que exista correlação positiva entre a quantidade de menções a determinado ativo na internet e seu volume de negociação em bolsa. Confirmando a hipótese inicial, seria possível construir um modelo baseado em lógica fuzzy para prever o volume de negociações de um ativo específico com base em sua popularidade na internet.

1. **OBJETIVOS**
   1. **OBJETIVO GERAL**

Analisar a correlação e causalidade entre a quantidade de menções de um ativo na internet e seu volume de negociação no mercado fracionário na bolsa e construir um sistema lógico fuzzy que tem como saída decisões de tendência (alta, estabilidade e queda) do volume de negociações do ativo com base em sua popularidade na internet.

* 1. **OBJETIVOS ESPECÍFICOS**
* Realizar uma revisão de literatura acerca dos conceitos de expectativas, incerteza e confiança aplicados ao mercado de investimentos mobiliários;
* Analisar a correlação entre a popularidade de um ativo na internet e seu volume de negociação;
* Aplicar testes de causalidade sobre a relação descrita acima;
* Construir o modelo final baseado em lógica fuzzy.

1. **JUSTIFICATIVA**

O trabalho justifica-se a partir do entendimento de que a integração entre as áreas de finanças comportamentais e ciência de dados é um campo de estudo novo e que, devido às potencialidades, apresenta diversas frentes de exploração e carece de contribuições.

Além disso, a originalidade do tema proposto pode dar início a uma nova linha de pesquisa nas frentes de sistemas nebulosos, modelos preditivos, análise de causalidade, economia comportamental, mercados eficientes e até mesmo de estudos regulatórios em mercado de capitais.

1. **METODOLOGIA**
   1. **TIPO DE PESQUISA**

Pesquisa em métodos quantitativos. Mais especificamente, em análise de causalidade e sistemas fuzzy.

* 1. **PERÍODO, ATIVOS E TERMOS DE PESQUSIA**

O período selecionado para a análise e construção do modelo foi de 01 de janeiro de 2021 a 31 de dezembro de 2021, agrupando os dados semanalmente de forma que seja possível utilizar os dados do ano completo mais recente.

O ativo selecionado foi o BOVA11 – iShares Ibovespa Fundo de Índice por sua característica de utilizar o Índice Ibovespa como referência.

Os termos de pesquisa inicialmente selecionados foram “IBOVESPA” o nome do índice; “IBOV”, abreviação popular do índice; e “BOVA11”, nome do ETF que será utilizado como ativo.

* 1. **AQUISIÇÃO E ANÁLISE DOS DADOS**

A princípio, o projeto será desenvolvido utilizando as linguagens Python e R para a coleta dos dados. O tratamento dos dados, análises exploratórias, descritivas e gerais e modelo final serão desenvolvidos em Python.

Os dados serão coletados de duas fontes distintas (Twitter e Google Trends) por meio de API’s (“Application Programming Interface”). Para o Google Trends, a coleta dos dados será feita através da API *pytrends*, enquanto para o Twitter a coleta será feita através da API *tweepy*. Serão coletadas informações sobre o volume de pesquisas relacionadas ao ativo escolhido e seu comportamento no período de tempo previamente definido. Como contingência, caso uma das API’s apresente instabilidade, a análise será feita somente com uma das fontes de dados. Em último caso, os dados serão coletados manualmente por meio da interface do Google Trends.

Os dados de volume de transações e volume de negociações serão coletados diretamente nas fontes disponibilizadas pela B3.

Após esta etapa inicial, será feita a transformação dos dados para séries temporais e serão feitas análises a fim de se compreender o grau de correlação e de causalidade entre a quantidade de menções sobre um ativo na internet e a quantidade de transações em bolsa. A princípio, o teste de causalidade de Granger é suficiente para os objetivos propostos.

* 1. **CONSTRUÇÃO DA LÓGICA FUZZY**

Nesta etapa, será desenvolvido um sistema lógico fuzzy que terá como saída a decisão de aumento, estabilidade ou queda do volume de negociações de acordo com a popularidade do ativo na internet.

* 1. **MODELO FINAL**

O modelo final será desenvolvido em Python com o auxílio da biblioteca, bem como de bibliotecas auxiliares para extração, tratamento, modelagem e exploração de dados. Algumas etapas referentes às análises de correlação e causalidade serão desenvolvidas com o auxílio da linguagem R.

Ao finalizar o modelo, o mesmo será exposto a testes e serão apresentadas métricas de avaliação a fim de se constatar suas capacidades preditivas.

1. **CRONOGRAMA**

Tabela

1. **REVISÃO BIBLIOGRÁFICA**

B3. **A DESCOBERTA DA BOLSA PELO INVESTIDOR BRASILEIRO**. 2020. Disponível em: https://www.b3.com.br/pt\_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/perfil-pessoas-fisicas/perfil-pessoa-fisica/. Acesso em: 18 set. 2022.

DEQUECH, David. Expectations and Confidence under Uncertainty. **Journal Of Post Keynesian Economics**, [S.L.], v. 21, n. 3, p. 415-430, mar. 1999. Informa UK Limited. http://dx.doi.org/10.1080/01603477.1999.11490205.

KEYNES, J. M. (1936). **A Teoria Geral dos Juros, do Emprego e da Moeda**. São Paulo: abril, 1983.

**AUXILIAR**

B3. **Histórico de pessoas físicas**. Disponível em: <http://www.b3.com.br/pt\_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/historico-pessoas-fisicas/>. Acesso em: 17 set. 2020.

HENTSCHEL, Martin; ALONSO, Omar. **Follow the money:** A study of cashtags on twitter. First Monday. Chicago, maio 2014. Disponível em: <https://firstmonday.org/ojs/index.php/fm/article/view/5385/4109>. Acesso em: 17 set. 2020.

Valor Investe. **O que os dados da B3 sobre pessoas físicas na bolsa revelam**. Disponível em: <https://valorinveste.globo.com/blogs/fernando-torres/post/2020/05/o-que-os-dados-da-b3-sobre-pessoas-fisicas-na-bolsa-revelam.ghtml>. Acesso em17 set. 2020.Seu Dinheiro. **20% dos usuários do Twitter são investidores**. Disponível em: <https://www.seudinheiro.com/2019/investimentos/20-dos-usuarios-do-twitter-sao-investidores-sera-este-o-perfil-do-fintwit/>. Acesso em: 17 set. 2020.

Valor. **Twitter contribui para inflar giro das ações**. Disponível em: <https://valor.globo.com/financas/noticia/2020/08/11/twitter-contribui-para-inflar-giro-das-acoes.ghtml>. Acesso em: 17 set. 2020.