



**Politécnico
de Viseu**

Escola Superior
de Tecnologia
e Gestão de Viseu

Análise do sentimento do investidor no mercado brasileiro: evidências empíricas por meio de notícias técnicas

Nataniel Geraldo Mendes Peixoto

Dissertação

Mestrado em Finanças Empresariais

Trabalho efetuado sob a orientação de

Professor Doutor António Pedro Martins Soares Pinto
Professor Doutor Pedro Manuel Nogueira Reis

Janeiro de 2025



**Politécnico
de Viseu**

Escola Superior
de Tecnologia
e Gestão de Viseu

Análise do sentimento do investidor no mercado brasileiro: evidências empíricas por meio de notícias técnicas

Nataniel Geraldo Mendes Peixoto

Dissertação

Mestrado em Finanças Empresariais

Trabalho efetuado sob a orientação de

Professor Doutor António Pedro Martins Soares Pinto
Professor Doutor Pedro Manuel Nogueira Reis

Janeiro de 2025

Agradecimentos

Agradeço primeiramente a Deus, pelo dom da vida, da esperança, da Fé, e pela graça de poder viver os planos que ele tem para nós. A Fé nos sustenta e nos conduz nos caminhos da verdade, nos ajuda a viver em comum unidade e a sermos verdadeiramente filhos e filhas de Deus.

Agradeço aos meus Professores, que me acolheram com muito apreço no decorrer do Mestrado, com prontidão, ouvido atento, disposição para esclarecer dúvidas e partilhar conhecimento. Agradeço em especial aos meus orientadores, Prof. Pedro Pinto e Prof. Pedro Reis, pela cordial atenção e colaboração neste processo da dissertação, desde a escolha do tema até a conclusão final do trabalho.

Agradeço ao Instituto Politécnico de Viseu (IPV), a Escola Superior de Tecnologia e Gestão de Viseu (ESTGV), pela acolhida no intercâmbio estudantil realizado no ano de 2023. Agradeço aos responsáveis pela mobilidade, pelas residências, e todo o corpo docente e administrativo pela educação, cordialidade, olhar sincero, dispostos a ajudar e a tornar esta experiência inesquecível.

Agradeço a Pontifícia Universidade Católica de Minas Gerais (PUC Minas), a Assessoria de Relações Internacionais, por todo auxílio no processo de intercâmbio, desde a seleção inicial até o retorno ao Brasil. Espero ter representado adequadamente a PUC Minas em Viseu, sendo esta uma grande responsabilidade pela notoriedade que a instituição possui no cenário nacional e internacional.

Agradeço aos amigos que fiz neste período, sejam os da PUC Minas que viajaram comigo, os que convivi na residência estudantil ou no arrendamento da casa em Repeses, bem como todas as relações de confiança que construí neste período. Sem dúvidas, são amizades que levarei para a vida.

Agradeço a meus Pais, Edmar Mendes Peixoto e Maria Aparecida Mendes Peixoto, exemplos de garra, Fé, esperança e por todos os ensinamentos que me prepararam para a vida, para uma convivência respeitosa e harmoniosa com todos, respeitando as diferenças e particularidades de cada pessoa.

Resumo

O crescimento das notícias veiculadas por meios eletrônicos tem transformado o modo como os investidores tomam decisões. O conteúdo e o tom dessas notícias impactam diretamente o fluxo de caixa previsional das empresas, refletindo-se tanto no valor das ações quanto no Ibovespa. Em consonância com os princípios das Finanças Comportamentais, tem-se destacado uma área de estudos dedicada a analisar o sentimento do investidor e a compreender o modo como os fatores comportamentais influenciam o mercado.

Nos últimos anos, diversos estudos têm sido realizados com o objetivo de mensurar e compreender esse efeito, utilizando técnicas de *machine learning* ou abordagem baseada em dicionários léxicos. Na literatura brasileira, é pouco analisado o poder preditivo dos notáveis dicionários *Harvard-IV*, *Loughran e McDonald* e *VADER* na análise de sentimento do investidor, aplicado a notícias dos media, principalmente avaliando o impacto de cada dicionário com base em 3 (três) variáveis independentes.

Perante esta lacuna, o objetivo geral desta investigação foi analisar a relação entre o sentimento do investidor, mensurado por meio da aplicação de 3 (três) dicionários léxicos a um conjunto alargado de notícias financeiras, com o comportamento do mercado acionista brasileiro. Foram analisadas 9.515 notícias do período de novembro de 2021 a agosto de 2024, agrupadas em 697 observações diárias. A base das notícias, do tópico “Finanças”, do serviço “Valor +News”, foi adquirida ao jornal *Valor Econômico*, com licenciamento específico para fins acadêmicos, tendo sido processadas na plataforma *Google Colab* por meio de um código *Python* desenvolvido no âmbito do trabalho. Os dados foram modelados por meio da regressão linear múltipla aplicada a séries temporais, e analisados usando o *software Stata*. A validade e robustez do modelo foi confirmada com os testes referentes à estacionariedade, heterocedasticidade, autocorrelação dos resíduos e da multicolinearidade das variáveis.

A extração do sentimento do investidor através dos dicionários *Loughran e McDonald* e *VADER* revelam significância estatística positiva na explicação do comportamento do Ibovespa, reforçando o valor informacional das notícias na formação do sentimento dos investidores e, conseqüentemente, no mercado acionista brasileiro. As variáveis analisadas taxa de câmbio, *Bitcoin* e risco Brasil mostraram-se estatisticamente significativas, enquanto o ouro e a frequência diária de notícias não apresentaram qualquer relevância estatística. Relativamente ao sentimento mensurado pelo dicionário *Harvard-IV* tornou-se significativo apenas após o cálculo da primeira diferença, embora a variável risco Brasil tenha perdido a sua

relevância. Em síntese, os resultados destacam a influência de fatores subjetivos na análise do comportamento dos investidores, em conformidade com os princípios das Finanças Comportamentais.

Palavras-chave: Sentimento do investidor; abordagem léxica; notícias.

Abstract

The growth of electronic news has transformed the way investors make decisions. The content and tone of this news has a direct impact on the forecast cash flow of companies, reflected in both the value of shares and the Ibovespa. In line with the principles of Behavioral Finance, an area of study has emerged dedicated to analyzing investor sentiment and understanding how behavioral factors influence the market.

In recent years, several studies have been carried out with the aim of measuring and understanding this effect, using machine learning techniques or approaches based on lexical dictionaries. In the Brazilian literature, there is little analysis of the predictive power of the notable Harvard-IV, Loughran and McDonald and VADER dictionaries in investor sentiment analysis, applied to news from the media, mainly evaluating the impact of each dictionary based on 3 (three) independent variables.

Given this gap, the general objective of this research was to analyze the relationship between investor sentiment, measured by applying three (3) lexical dictionaries to a wide range of financial news, and the behavior of the Brazilian stock market. We analyzed 9,515 news items from November 2021 to August 2024, grouped into 697 daily observations. The news base, from the “Finance” topic of the “Valor +News” service, was acquired from the Valor Econômico newspaper, with specific licensing for academic purposes, and was processed on the Google Colab platform using Python code developed within the scope of the work. The data was modeled using multiple linear regression applied to time series, and analyzed using Stata software. The validity and robustness of the model was confirmed with tests for stationarity, heteroscedasticity, autocorrelation of the residuals and multicollinearity of the variables.

The extraction of investor sentiment using the Loughran and McDonald and VADER dictionaries revealed positive statistical significance in explaining the behavior of the Ibovespa, reinforcing the informational value of news in shaping investor sentiment and, consequently, the Brazilian stock market. The analyzed variables exchange rate, Bitcoin and Brazil risk were statistically significant, while gold and the daily frequency of news were not statistically significant. The sentiment measured by the Harvard-IV dictionary became significant only after calculating the first difference, although the Brazil risk variable lost its relevance. In summary, the results highlight the influence of subjective factors in the analysis of investor behavior, in line with the principles of Behavioral Finance.

Keywords: Investor sentiment; lexical approach; news.

Índice geral

Agradecimentos	ii
Resumo	iii
Abstract	v
Índice geral.....	vi
Índice de tabelas	vii
Índice de gráficos.....	viii
Índice de quadros	ix
Lista de siglas/abreviaturas	x
1. Introdução.....	1
1.1. Contextualização	1
1.2. Problema de investigação	2
1.3. Objetivos da investigação	3
1.4. Relevância e justificação do tema	4
1.5. Estrutura da dissertação.....	5
2. Revisão de literatura	6
2.1. Evolução das finanças	6
2.2. Tecnologia e finanças	9
2.3. Sentimento do investidor	11
3. Dados e metodologia.....	23
3.1. Variáveis do modelo	23
3.2. Validação do modelo -Testes estatísticos.....	24
3.3. Processamento das notícias e o sentimento do investidor.....	25
4. Análise dos resultados	29
4.1. Estatística descritiva	29
4.2. Pressupostos do modelo de regressão	33
4.3. Matriz de correlação	34
4.4. Regressão linear múltipla	36
Conclusão	42
Referências bibliográficas.....	44
Apêndices	52
Apêndice A – Código <i>Python</i>	53

Índice de tabelas

Tabela 1 - Estatística descritiva das variáveis do modelo.....	29
Tabela 2 - Resultados do teste <i>Dickey–Fuller</i>	33
Tabela 3 - Resultados teste <i>Breusch–Pagan</i>	33
Tabela 4 - Resultados teste <i>Breusch–Godfrey LM</i>	34
Tabela 5 - Resultados teste VIF	34
Tabela 6 - Matriz de correlações entre as variáveis do modelo	35
Tabela 7 - Regressão Linear Múltipla.....	37
Tabela 8 - Regressão Linear Múltipla – 1º Diferença para o léxico <i>Harvard-IV</i> (sent_HIV) .	41

Índice de gráficos

Gráfico 1 - Comportamento sentimento do investidor de acordo com diferentes indicadores	30
Gráfico 2 - Gráfico de dispersão da variável “risco brasil”	32

Índice de quadros

Quadro 1 - Resumo trabalhos relacionados.....	18
---	----

Lista de siglas/abreviaturas

ARDL- *Autoregressive Distributed Lag*

CDS- *Credit Default Swap*

EMBI- *Emerging Markets Bond Index*

ESTGV- Escola Superior de Tecnologia e Gestão de Viseu

ETFs- *Exchange-Traded Funds*

FMI- Fundo Monetário Internacional

GI- *General Inquirer*

HME- Hipótese de Mercados Eficientes

ICC- Índice de Confiança do Consumidor

ICE- Índice de Confiança Empresarial

IPCA- Índice Nacional de Preços ao Consumidor Amplo

IPV- Instituto Politécnico de Viseu

JSON- *JavaScript Object Notation*

OLS- *Ordinary Least Squares*

PUC Minas- Pontifícia Universidade Católica de Minas Gerais

VADER- *Valence Aware Dictionary and Sentiment Reasoner*

VIF- *Variance Inflation Factor*

1. Introdução

Esta seção apresenta a contextualização do tema do trabalho, a formulação das questões de investigação, os objetivos gerais e específicos, a relevância do estudo, além de descrever a estrutura de organização da dissertação.

1.1. Contextualização

Os media especializados têm-se tornado cada vez mais a fonte de informações para as empresas, nomeadamente para aquelas que são negociadas no mercado de capitais, no qual, prevalece o pressuposto da equidade das informações divulgadas (Galdi e Gonçalves, 2018). Os agentes do mercado utilizam um elevado volume de informações que é diariamente difundida por diferentes fontes e emergem de modo imprevisível. Deste modo, “não parece equivocado pensar que os investidores vejam valor no conteúdo das notícias, mas quantificar esse valor torna-se mais complicado“. (Costa, 2021, p. 9)

De acordo com Silva (2017), a utilização de informações textuais tem sido o foco de investigações empíricas nos últimos anos. A autora cita trabalhos que utilizaram diferentes fonte de informação, a saber: jornais, relatórios da administração, mensagens de chat, *Twitter*, *Facebook* e até mesmo mensagens de *Spam*.

A hipótese de eficiência de mercado, formulada por Fama (1970), preconiza que o preço de um ativo tende a refletir as informações disponíveis da instituição emissora, não sendo possível aos investidores obterem ganhos anormais, isto é, retornos superiores ao retorno ajustado ao risco de determinado ativo. Entretanto, estudos posteriores, ao detetar anomalias nessa hipótese, contribuíram para o desenvolvimento de uma nova área de investigação associada às Finanças Comportamentais (Mussa et al., 2008).

A associação entre a divulgação de notícias e o comportamento do mercado tem sido objeto de diversos estudos, surgindo assim uma linha de investigação relacionada com o sentimento do investidor (Galdi e Gonçalves, 2018). A análise de sentimentos pode ser entendida como um campo de estudo que analisa opiniões, sentimentos, avaliações e emoções das pessoas em relação a determinado produto, serviço, organização, eventos, entre outras questões (Liu, 2012).

Para Baker e Wurgler (2007) o sentimento do investidor é uma crença sobre fluxos de caixa futuros e riscos de investimento que não é justificado pelos fatores de mercado. Este efeito

tem vindo a ser estudado desde a década de 90, destacando-se os trabalhos conduzidos por Shleifer e Summer (1990) e Barberis et al. (1998). Contudo, a terminologia “análise de sentimento” (*sentiment analysis*) foi utilizada, possivelmente, pela primeira vez por Nasukawa e Yi (2003). Estes autores argumentam que a grande questão associada a esta temática, decorre da necessidade de identificar o modo como os sentimentos são expressos através das notícias divulgadas e, se estas exprimem uma opinião negativa ou positiva.

Indurkha e Damerou (2010) destacam a dificuldade de o leitor humano encontrar fontes relevantes de informações, extrair frases relacionadas com opiniões, lê-las, resumi-las e organizá-las de modo estruturado. Perante este facto, são necessários sistemas automatizados para sumarizar o grande volume de opiniões. A análise de sentimento emerge como uma resposta a essa necessidade, impulsionando, inclusive, o surgimento de empresas especializadas na prestação deste serviço.

1.2. Problema de investigação

Nos últimos anos, um conjunto de trabalhos relacionados com a análise de sentimento no Brasil tem utilizado *proxies* quantitativas para avaliar o sentimento do investidor, tais como o Índice de Confiança do Consumidor (ICC) (Piccoli et al., 2018) e a metodologia do notável trabalho de Baker e Wurgler (Mello, 2019; Duarte, 2023), em detrimento das notícias divulgadas pelos media. Outras fontes de notícias também foram consideradas, tais como *Twitter* (Carosia et al., 2019; Souza e Martins, 2020; Santos, 2022) e relatórios de reporte financeiro anuais das empresas (Pagliarussi et al., 2016). No que concerne às técnicas de análise de sentimento, destacam-se trabalhos que utilizaram *machine learning* (Souza e Martins, 2020) e a abordagem baseada em dicionário léxico (Santos, 2022).

Na abordagem baseada em léxico, a literatura adota diferentes estratégias para a utilização de dicionários. Alguns estudos empregam dicionários próprios, como os de Silva (2017), Pagliarussi et al. (2016) e Silva e Machado (2019). Outros recorrem a dicionários de pesquisas anteriores, como Costa (2021), que utilizou o modelo proposto por Silva (2017). Também se destacam dicionários abrangentes que consolidam 3 (três) dos mais utilizados na área, como no trabalho de Valcareggi (2022), que utilizou o dicionário de Shapiro et al. (2020). Já Castro (2023) optou por um dicionário amplamente conhecido, o VADER, para analisar notícias do *Twitter*, destacando sua aplicabilidade em contextos de media social.

Porém, é pouco discutido o poder preditivo dos notáveis dicionários *Harvard-IV*, *Loughran e McDonald* e *VADER* na análise de sentimento do investidor, aplicado a notícias financeiras, avaliando o impacto de cada dicionário através de 3 (três) variáveis independentes na mesma regressão, com uma variável para cada dicionário. Embora sejam todos classificados como dicionários de léxicos, foram desenvolvidos em diferentes momentos de tempo. De acordo com Tahmasebi et al. (2021), os léxicos alteram-se no decorrer dos anos, impulsionado, em parte, pelas necessidades de comunicação e do desenvolvimento do mundo ao nosso redor.

Neste sentido, pretende-se responder à seguinte questão: “O sentimento do investidor no mercado brasileiro, mensurado a partir de notícias e pela aplicação da abordagem léxica, é estatisticamente significativo na explicação do comportamento do índice Ibovespa?”. A hipótese a ser testada é se há evidências estatisticamente significativas da influência do sentimento no comportamento do mercado acionista brasileiro e, se esta relação é positiva, em linha com o trabalho de Silva (2017). Galdi e Gonçalves (2018) destacam que, ao considerar o mercado de modo agregado, é espetável que o tom (conteúdo) das notícias num dia específico esteja relacionado com o desempenho e com a volatilidade dos índices de mercado.

1.3. Objetivos da investigação

O presente trabalho tem como principal objetivo investigar se o sentimento do investidor no mercado brasileiro, mensurado a partir de notícias divulgadas nos media e pela aplicação da abordagem de léxicos, tem poder explicativo no comportamento do índice de mercado Ibovespa. A partir deste objetivo geral foram definidos os seguintes objetivos específicos: i) analisar a relação entre o sentimento do investidor e o comportamento do mercado financeiro brasileiro; ii) avaliar quais dos dicionários de léxicos, de entre os 3 (três) selecionados, tem maior relevância na caracterização do sentimento do investidor e; iii) analisar a relação entre as demais variáveis independentes do modelo com o mercado financeiro brasileiro.

Para desenvolvimento da investigação empírica, serão utilizadas notícias do tópico “Finanças”, do serviço “Valor +News”, de um site especializado no Brasil, o *Valor Econômico*. Os dados serão adquiridos diretamente à empresa, mediante licenciamento prévio, com destino exclusivo à dissertação ou qualquer artigo derivado da pesquisa. O “Valor +News” é um serviço que oferece uma seleção diária de cerca de 40 notícias, abrangendo temas como Política, Finanças, Brasil, Empresas e Internacional. O conteúdo é elaborado por mais de 40 profissionais especializados, com correspondentes no Brasil e no mundo (Valor Econômico, n.d.).

O sentimento do investidor será mensurado através da abordagem de léxicos, sendo utilizados 3 (três) dicionários amplamente difundidos na literatura (*Harvard-IV*, *Loughran e McDonald* e *VADER*). O tratamento da base de notícias e a aplicação dos dicionários de léxicos serão realizados com o recurso a um código *Python* desenvolvido no âmbito deste trabalho, bem como pela integração das demais bibliotecas necessárias.

A variável dependente, retorno diário do Ibovespa, e as variáveis independentes: sentimento do investidor, taxa de câmbio, cotação do ouro, bitcoin, risco país e frequência diária de notícias publicadas, serão analisadas com recurso à regressão linear múltipla em séries temporais, e ao software *Stata BE 18.5*. Os dados serão coletados no período de novembro de 2021 a agosto de 2024, considerando que esse é o período disponível para as notícias na base do serviço “Valor +News”.

1.4. Relevância e justificação do tema

A ampla divulgação de notícias por meio eletrônico influencia o modo como o investidor toma decisões. O teor e o tom (conteúdo) das notícias, sendo positivas ou negativas relativamente a determinado setor ou empresa, tem impacto direto no fluxo de caixa previsionais (Galdi e Gonçalves, 2018). Como resultado, observa-se uma valorização ou desvalorização das ações da empresa, dando origem à compra ou venda de posições acionistas que passam a apresentar algum risco para o investidor.

Nestas circunstâncias, as notícias divulgadas nos diferentes meios de comunicação podem ter impacto direto na volatilidade das ações da empresa e, conseqüentemente no índice de mercado (Mitra e Mitra, 2011). Momentos de otimismo ou pessimismo do investidor podem resultar em enviesamentos de irracionalidade, podendo induzir o risco sistemático e afetar os preços em equilíbrio (Naik e Padhi, 2016). Assim, torna-se relevante compreender os fatores comportamentais que estão latentes no mercado, permitindo uma maior compreensão dos desdobramentos económicos e de formação de preços (Souza et al., 2023).

O presente estudo procura contribuir para a academia ao propor-se avaliar o sentimento do investidor a partir de uma variável quantitativa, extraída do conteúdo informativo das notícias. Para além disso, o estudo inova ao considerar notícias exclusivas para assinantes do serviço “Valor +News” do jornal *Valor Econômico*, utilizadas com a permissão da empresa para fins meramente académicos. Assim, espera-se avaliar, com o recurso ao conteúdo das notícias, se as informações técnicas geradas pelos especialistas e consumidas por um público restrito

ajudam a explicar o comportamento do mercado acionista brasileiro, bem como a sua relação com as demais *proxies* a serem utilizadas.

A presente pesquisa também inova ao incorporar novas *proxies* passíveis de explicar os movimentos do Ibovespa, tal como a cotação do ouro, da bitcoin e do risco do país Brasil, que são pouco utilizadas na literatura que tem abordado esta temática, na expectativa de melhorar o poder explicativo do modelo.

Trabalhos como de Tetlock (2014), ao questionarem o real impacto dos media sobre a comportamento dos preços, reforçam a necessidade de aprofundar os estudos nesta área de investigação. Tal cenário é ainda corroborado pela constante evolução dos media, com o surgimento de novas tecnologias, novas redes sociais e, novas formas de fazer chegar ao utilizador as informações que impactam o seu poder de decisão (Colombo et al., 2023). Investigar o efeito da psicologia do investidor no retorno do mercado de ações contribui para o entendimento do próprio mercado e do poder preditivo do sentimento (Reis e Pinho, 2020c).

1.5. Estrutura da dissertação

A presente dissertação está dividida em 4 (quatro) capítulos, para além desta parte introdutória. O estudo contempla a revisão da literatura; a recolha de dados e metodologia; a análise de resultados; e por fim a conclusão. De seguida, são apresentadas as referências bibliográficas e apêndices do estudo.

A introdução contextualiza o tema, os problemas de pesquisa, os objetivos do estudo, a relevância do tema para a literatura e explicada de modo breve a metodologia e a estrutura da dissertação. A revisão da literatura, apresenta uma breve evolução das finanças comportamentais, a relação entre tecnologia e finanças e os principais trabalhos realizados nos últimos anos relacionadas com a temática da análise de sentimento. O capítulo seguinte, apresenta de forma detalhada a metodologia do trabalho, a definição das variáveis em estudo do modelo, os testes estatísticos realizados, bem como das técnicas de processamento da base de notícias e do sentimento do investidor. Posteriormente, são apresentados os resultados do estudo, discutidos os pressupostos do modelo de regressão e realizados os testes estatísticos, na procura da resposta às perguntas de investigação. Por fim dá-se conta das conclusões do estudo, bem com das principais contribuições, limitações, e pistas para trabalhos futuros.

2. Revisão de literatura

Neste capítulo é apresentada uma revisão da literatura das principais temáticas que fundamentam o estudo, abordando conceitos, evolução histórica e trabalhos relacionados. Por meio desta seção é possível explorar lacunas e possibilidades de contribuições na literatura.

2.1. Evolução das finanças

De acordo com Haugen (2000), citado por Famá et al. (2008, p. 69), a evolução das finanças pode ser dividida em 3 (três) fases: Finanças antigas, compreendendo o período de 1930 a 1950; Finanças Modernas, do período de 1950 a 1990 e Finanças Comportamentais, de 1990 em diante.

Um das principais premissas do modelo das Finanças Modernas é a racionalidade do investidor. Este paradigma foi fundamentado em relevantes trabalhos nomeadamente: de irrelevância da estrutura de capitais de Modigliani e Miller (1958, 1961), da diversificação do portfólio de Markowitz (1952), da precificação de ativos de William Sharpe (1964), da precificação de opções de Myron Scholes e Fisher Black (1972) e na Hipótese de Mercados Eficientes (HME) de Eugene Fama (1970) (Famá et al., 2008).

A intensificação dos estudos relacionados com a Hipótese de Eficiência de Mercado levantou questões relativamente à racionalidade do investidor, nomeadamente associadas à prevalência de evidências de comportamentos anormais em relação ao retorno dos ativos. Surge então, a corrente denominada Finanças Comportamentais, com o intuito de aperfeiçoar e ultrapassar os constrangimentos do modelo das Finanças Modernas (Mussa et al., 2008)

A problemática das Finanças Comportamentais ganhou relevância no meio académico no final de década de 70, após a publicação do trabalho dos psicólogos Kahneman e Tversky (1979), vencedor do Prémio Nobel. A partir de então, o processo de tomada de decisão do ser humano, assume maior relevância, quando sujeito a situações de risco (Famá et al., 2008).

De acordo com Shleifer e Summers (1990), as Finanças Comportamentais são sustentadas por basicamente dois pilares. O primeiro refere-se aos limites à arbitragem, sendo esta uma operação arriscada e, portanto, limitada e, o segundo incorpora os aspetos psicológicos dos investidores, relacionado com o processo de tomada de decisão e os vieses aos quais os agentes estão sujeitos.

Mussa et al. (2008) destacam que: a crença na racionalidade do investidor e, as finanças constituírem-se como uma ciência exata, pode ser considerada uma falácia. De acordo com Milanez (2003), esta nova abordagem incorpora conceitos da Psicologia, da Sociologia e de outras ciências. Kimura (2003) destaca que as Finanças Comportamentais estão pautadas pela possibilidade de os investidores terem atitudes irracionais, impactando consistentemente o comportamento do mercado.

Silva (2017) destaca que as explicações para a ineficiência do mercado estão relacionadas com a noção de que o investidor não possui a racionalidade preconizada pela Hipótese de Mercados Eficientes. Pelo contrário, possui uma racionalidade limitada, influenciada por fatores psicológicos, conhecidos como vieses cognitivos, como o viés da negatividade.

As Finanças Comportamentais não têm o intuito de rejeitar totalmente a teoria moderna de finanças, mas sim, mostrar que, as explicações tradicionais não são suficientes para ajudar a compreender muitos dos acontecimentos que ocorrem no mercado financeiro (Milanez, 2003). Diversos trabalhos nas últimas décadas estudaram empiricamente diferentes anomalias presentes no mercado (Banz, 1981; Jaffe et al., 1989; Haugen e Jorion, 1996). Mussa et al. (2008) elenca as seguintes anomalias: efeito dia da semana, efeito mês do ano ou efeito janeiro, efeito momento, otimismo e pessimismo exagerado e aversão à perda. Duarte (2023) cita ainda, enquanto enviesamentos de comportamento, o efeito âncora e excesso de confiança.

Famá et al. (2008) destaca que há três correntes de pensamento relativamente às Finanças Comportamentais: conservadora, moderada e radical. A corrente conservadora defende que as Finanças Modernas continuarão a ser o modelo predominante. A ala moderada defende que as Finanças Comportamentais irão complementar as Finanças Modernas. E por fim, a corrente radical aposta nas Finanças Comportamentais como novo paradigma.

Para Shefrin (2015), citado por Pedro (2018, p. 22) existe uma clara tendência para que os aspetos comportamentais passem a ter maior preponderância nos estudos relacionados com as finanças e na influência que têm na tomada de decisões de índole financeira. Sendo assim, é possível considerar diversas variáveis intrínsecas na tomada de decisões, a saber: “estrutura do cérebro, estrutura hormonal, padrões genéticos e exposição a diversas relações sociais”. O autor destaca ainda que, o “emprego dos aspetos psicológicos é a nova fronteira do conhecimento em finanças.”.

Estudos como o de García (2013) evidenciam a relação latente entre psicologia e economia comportamental, deixando claro a importância das notícias como fonte de informações para a tomada de decisões. O autor encontrou evidências de uma relação entre o sentimento de mercado e o conteúdo das notícias e os media em geral, sendo mais acentuada em períodos de incertezas, ou seja, identificou uma maior sensibilidade dos investidores em tempos difíceis. Os períodos de retração têm o poder de alterar o modo como os agentes de mercado tomam decisões, relativamente aos períodos de expansão económica.

De acordo com Pedro (2018) existe um contra-argumento das Finanças Comportamentais para cada argumento das Finanças Clássicas. Estudos recentes evidenciam a existência de fenómenos de mercado pautados pela irracionalidade dos investidores, tais como: os efeitos manada e as bolhas imobiliárias, assim como fatores relacionados como humor, feriados, clima, entre outros. Tal constatação evidencia que o investidor moderno não se encaixa nas características preconizadas pelas Finanças Modernas e que, os fatores comportamentais passam a ter um destaque considerável no processo decisório.

Shiller (2000), citado por Famá et al. (2008, p. 66), fala das bolhas especulativas. Para o autor, este fenómeno pode ser identificado como uma “situação em que preços temporariamente altos são sustentados em grande parte pelo entusiasmo dos investidores e não por uma estimativa consistente do valor real”. Este raciocínio questiona a racionalidade do investidor. Um exemplo notável de “bolha especulativa” foi a crise do *subprime* em 2008 (Souza et al., 2023).

Souza et al. (2023) desenvolveram um trabalho com o propósito de detetar bolhas especulativas no mercado brasileiro, no período de 1993 a 2020, a partir de notícias informativas sobre o mercado. Os resultados indicaram a existência de cinco janelas anómalas no período estudado, corroborando os estudos que conciliam a perspectiva comportamental e a prevalência de bolhas especulativas no mercado.

A heterogeneidade dos investidores assume também um aspeto relevante (Silva, 2017), na medida que, se verificam diferentes níveis de assimilação do conteúdo informativo das notícias. Isso ocorre devido ao fato, de os agentes serem vulneráveis a desvios cognitivos, levando à equalização das crenças de modo diferente do que preconiza a racionalidade esperada. Assim, a racionalidade limitada, conjugada com fatores psicológicos e o viés da negatividade, afetam diretamente a homogeneidade das expectativas dos agentes, provocando ruídos no processo de ajustamento dos preços às informações.

2.2. Tecnologia e finanças

O desenvolvimento da internet permite gerar diariamente um grande volume de informações, contribuindo de forma significativa para o aumento de estudos que as utilizam. A procura de informações no passado era realizada entre amigos e familiares e, no caso das empresas, por estudos de opinião, elaborados a partir de inquéritos. Com o crescimento das mídias sociais (fórum de discussão, blogs, redes sociais), os indivíduos e as organizações estão a recorrer cada vez mais ao conteúdo da informação divulgada pelos meios de comunicação para tomarem decisões (Liu, 2012).

Santos (2022) destaca o surgimento constante de alternativas viáveis que procuram entender os fenómenos económicos, seja através da previsão de comportamentos futuros, na antecipação de acontecimentos ou, na previsão das movimentações de mercado. De igual modo, a evolução tecnológica tem contribuído para o desenvolvimento de novos métodos computacionais e quantitativos que contribuem significativamente para estas análises, com destaque para o estudo do sentimento e mineração de dados, detetando emoções implícitas em textos, livros, artigos, notícias e restantes fontes de informação.

Souza e Martins (2020) destacam a importância da tecnologia na mensuração do sentimento do investidor. De acordo com os autores, essa nova tendência possibilita a captação do sentimento de forma quase instantânea, permitindo resolver problemas como o identificado por Klibanof et al. (1998), quando referem que, um número muito significativo de trabalhos na literatura avalia o comportamento do investidor de forma retrospectiva, isto é, *ex-post* a ocorrência do fato, avaliando muitas vezes os seus determinantes.

De acordo com Medhat et al. (2014) as técnicas de classificação de sentimento podem ser divididas em: abordagem de aprendizagem de máquina (*Machine Learning Approach*), abordagem baseada no léxico (*Lexicon Based Approach*) e abordagem híbrida (*Hybrid Approach*). A abordagem baseada na aprendizagem de máquina aplica algoritmos e utiliza recursos linguísticos. A abordagem baseada no léxico utiliza uma coleção de termos de sentimento conhecidos e pré-compilados, procurando identificar uma polaridade negativa ou positiva. A abordagem híbrida combina as duas abordagens anteriores.

Na literatura brasileira podem ser identificados trabalhos que utilizam as diversas técnicas de classificação de sentimento. A abordagem de aprendizagem de máquina foi utilizada por Igarashi et al. (2021), com auxílio do *Amazon Comprehend* e por Souza e Martins (2020) quando recorreram ao *Google Cloud Natural Language API*.

Os léxicos de palavras são muito úteis para análise de sentimentos (Huto e Gilbert, 2014) e, diferentes léxicos têm sido desenvolvidos e aperfeiçoados nos últimos anos. Liu (2012) destaca que o indicador mais importante de sentimento pode ser avaliado por palavras, que expressam um sentimento positivo ou negativo. Os trabalhos nesta área têm utilizado basicamente 3 (três) dicionários de palavras: *Harvard-IV*, *Loughran e McDonald* e *VADER*.

Silva e Machado (2019) destacam que o processo de mensuração do sentimento textual está relacionado com a análise do conteúdo linguístico dos textos, sendo analisadas as palavras utilizadas no processo de comunicação, com o objetivo de extrair o posicionamento de otimismo ou pessimismo presente no texto.

Stone et al. (1966) desenvolveram o *General Inquirer* (GI), uma aplicação de análise de texto com um dos mais antigos léxicos construídos manualmente e que, ainda continua a ser utilizado, num processo de desenvolvimento e aperfeiçoamento contínuo desde 1966. O léxico *Harvard-IV* tem sido amplamente utilizado em diferentes trabalhos (Esuli e Sebastiani, 2005; Kamps, Mokken, Marx, e de Rijke, 2004; Turney e Littman, 2003) para determinar as propriedades de sentimento do texto, como referem Huto e Gilbert (2014). Jurafsky e Martin (2024) destacam que o GI é o mais antigo léxico que representa as palavras de forma binária, sendo uma lista de palavras positivas e uma lista de palavra negativas.

Loughran e McDonald (2011) analisaram relatórios 10-k (reportes financeiros das empresas) no período de 1994 a 2008 e, concluíram que aproximadamente três quartos das palavras identificadas como negativas no *Harvard-IV* (dicionário produzido por outras áreas científicas), não são consideradas como negativas na área financeira. Nestas circunstâncias, os autores desenvolveram uma lista de palavras que refletem melhor o tom (sentimento) das notícias financeiras, propondo um novo dicionário que passou a predominar entre os estudos empíricos sobre mensuração de sentimento textual (Silva, 2017).

Huto e Gilbert (2014) construíram o *VADER* (*Valence Aware Dictionary and Sentiment Reasoner*) que atribui valores para sentimentos positivos, negativos e neutros. Este trabalho adiciona informações aos dicionários mais utilizados. O *VADER* tornou-se no dicionário base para diferentes estudos de processamento de linguagem natural, devido ao seu elevado rigor em textos de diferentes conteúdos. Tal dicionário foi utilizado, por exemplo, no trabalho de Santos (2022).

Um dos maiores benefícios na utilização de uma abordagem computadorizada na análise de sentimento, reside na possibilidade de se analisar milhares de documentos de forma muito

mais otimizada do que uma abordagem manual. Além disso, uma questão desafiadora em finanças reside no fato do desenvolvimento de um modelo capaz de explicar as flutuações do mercado de capitais. A conexão entre análise de sentimento e mercado de ações tem sido um tema recorrentemente explorado na literatura, e a tecnologia tem contribuído para esta interlocução entre áreas do conhecimento (Pagliarussi et al., 2016).

2.3. Sentimento do investidor

De um modo geral, os estudos das análises de sentimento procuram relacionar as opiniões expressas através de textos com algumas métricas de mercado (Igarashi et al., 2021). Estes estudos têm um profundo impacto na política, na economia e nas ciências sociais, na medida que são influenciadas pelas opiniões dos agentes económicos (Liu, 2012).

A partir da década de 1980, a literatura passou a reconhecer que determinados fenómenos e padrões de mercado, conhecidos como anomalias, poderiam ser atribuídos a momentos de irracionalidade dos investidores, influenciados por seu sentimento (Ferreira et al., 2021). Silva (2017) destaca que a ausência de alinhamento das crenças implica ruídos no processo de ajuste dos preços às informações que vão aparecendo. Logo, podem surgir ineficiências temporais no mercado, com reações *underreaction* e *overreaction*, devidas ao sentimento do investidor.

Faustino (2013) destaca que, num segundo momento, início dos anos 2000, surgiram trabalhos que procuravam absorver o efeito semântico das notícias, isto é, se são positivas, negativas ou neutras, e se transmitiam aos investidores tendências ou indícios de compra ou venda de ações.

Baker e Wurgler (2006) ao contrário do que preconizava a teoria clássica, consideraram o sentimento do investidor como um fator que impacta o mercado americano. Os autores desenvolveram um indicador de sentimento baseado em seis *proxies* utilizadas frequentemente para avaliar o sentimento, medido anualmente entre 1962 e 2001. Os resultados indicaram que o sentimento tem influência sobre o retorno das ações, na medida em que, um aumento de sentimento tende a elevar o valor das ações mais difíceis de avaliar e arbitrar. Foi observado também, uma relação entre menor sentimento e maior retorno subsequente.

Tetlock (2007) utilizou a análise de componentes principais para construir uma medida simples de pessimismo extraída de um conjunto de notícias. Foram utilizadas notícias diárias

da coluna *Abreast of the Market* do *Wall Street Journal*, analisadas pelo dicionário *Harvard-IV*. Os resultados indicaram que elevados níveis de pessimismo foram capazes de prever, de forma consistente, uma pressão de baixa sobre os preços de mercado, seguida por uma reversão aos fundamentos, e que valores excepcionalmente elevados ou baixos de pessimismo anteciparam um alto volume de negociações no mercado. Em síntese, os autores identificaram que as medidas de conteúdo extraídas das notícias funcionam como *proxy* para o sentimento dos investidores.

Pagliarussi et al. (2016) utilizaram a técnica de análise de sentimento para extrair o tom (conteúdo) de 829 relatórios de gestão de empresas brasileiras entre 1997 e 2009, relacionando com variáveis de mercado (retorno anormal, volume de negócios e volatilidade no preço das ações). Assim como Loughran e McDonald (2011), Pagliarussi et al. (2016) construíram um dicionário de palavras para análise dos textos. Contraponto aos estudos realizados no mercado norte americano, não identificaram evidências significativas de que, os relatórios de gestão, tenham influência nas transações do mercado acionista no Brasil.

Silva (2017) elaborou um índice de sentimento utilizando um dicionário com léxico próprio, composto por palavras positivas e negativas, com o objetivo de analisar o impacto das notícias sobre a volatilidade dos preços no mercado acionista brasileiro, considerando um cenário de incertezas económicas. Foram utilizados valores diários dos índices Ibovespa, DJIA, VIX e S&P500, e o sentimento textual foi construído com base em notícias do jornal *Valor Econômico*, com 45.304 acontecimentos, do período de 25/07/2011 a 30/06/2017. Os resultados indicaram que, em períodos de incerteza económica, se verifica uma relação direta entre pessimismo e reduções de retorno. Foram encontradas evidências de que o sentimento textual é um preditor da volatilidade, e que a informação proveniente de jornais no Brasil tem influência sobre a percepção dos investidores face às incertezas no mercado e na economia.

Galdi e Gonçalves (2018) investigaram a relação entre o teor das notícias diárias divulgadas no jornal *Valor Econômico* com a rentabilidade e volatilidade dos índices de mercado. A base de dados composta por 1.237 notícias diárias entre 02/01/2012 e 30/12/2016, revela que o mercado avalia com maior peso as palavras de incerteza e negativas divulgadas nas notícias. Termos com conotação de incerteza tiveram associação negativa com a rentabilidade e, foram identificados fracos indícios de relação positiva entre o tom negativo das notícias e volatilidade do Ibovespa. Nestas circunstâncias, os autores encontram evidências de conteúdo informativo nas notícias veiculadas pelos media especializados no Brasil.

Picolli et al. (2018) avaliaram a influência do sentimento do investidor na relação risco-retorno no mercado de ações brasileiro, no período de 2002 a 2015. O Índice de Confiança do Consumidor (ICC) foi utilizado como substituto do nível de sentimento do investidor. Os resultados indicaram que a relação entre variância condicional e retorno das ações é positivo (negativo) em períodos de baixo (alto) sentimento. Além disso, concluíram que a deterioração da relação positiva entre risco e retorno quando o sentimento é elevado resulta de um rápido crescimento no número de investidores menos sofisticados.

Mello (2019) investigou se o sentimento do investidor (*Investor Sentiment*) afeta os retornos da bolsa de valores brasileira. O índice de sentimento do investidor foi construído com base na metodologia desenvolvida por Baker e Wurgler (2006), com adaptações e, calculado numa base mensal com recurso ao método de componentes principais, para o período de 2003 a 2018. Os resultados indicam que as ações de elevada volatilidade são mais afetadas pelo sentimento do investidor e, que o sentimento do investidor apresenta boa correlação com o Índice de Confiança do Consumidor.

Silva e Machado (2019, p. 27) analisam se o risco sistemático das ações preferenciais da Vale (VALE5) sofreu influência do tom e volume das notícias associadas ao desastre ambiental no Brasil ocasionada pela Samarco, sendo esta “uma empresa de capital fechado gerida mediante uma joint-venture entre as empresas Vale S.A. e a *BHP Billiton*”. As matérias financeiras foram extraídas do site *Valor Econômico* e, o tom das notícias analisado através da aprendizagem baseada em dicionários de palavras, tendo sido construído uma lista própria de palavras. O risco sistemático, medido pelo beta dinâmico, foi estimado com modelos autorregressivos de heterocedasticidade condicional (ARCH-GARCH). Evidências empíricas entre o 5º e 9º decil indicaram que o volume e o tom das notícias divulgadas têm influência sobre o beta da ação, indicando que o risco sistemático tem conexão com as notícias divulgadas pelos media.

Reis e Pinho (2020a) investigam o impacto da racionalidade ou irracionalidade nos retornos das ações de diferentes países e setores, utilizando como proxy para racionalidade a pandemia de Covid-19. Foi utilizado o dicionário *Merriam-Webster* para identificar as palavras negativas, do período de 2004 até abril de 2020 e, em seguida, combinado os termos para pesquisa do *Google Trends* no período de 2015 até abril de 2020. Os resultados indicaram que os retornos dos EUA são mais sensíveis ao sentimento, e por consequência, mais propensos a fatores irracionais, quando comparados com os retornos na Europa. Além disso, referem que o fator “país” influencia os retornos de forma diferente. Na Itália, Espanha, e Reino Unido, países

muito afetados pela pandemia, os índices setoriais foram mais reativos a casos de Covid-19, encontrando-se evidências de racionalidade.

Igarashi et al. (2021, p. 470) avaliam a influência de notícias e opiniões na movimentação do mercado, “confrontando a polaridade destes sentimentos com a tendência preconizada a partir da aplicação de técnicas de análise como as médias móveis”. Utilizaram técnicas de inteligência artificial para processar linguagem natural, nomeadamente o *Amazon Comprehend*. As notícias foram extraídas de portais especializados no mercado financeiro, procurando diminuir possíveis ambiguidades presentes em situações de informalidade. Os resultados obtidos evidenciam uma correlação moderada entre os sentimentos das notícias e a movimentação do preço da ação.

Reis e Pinho (2020b) apresentam um novo índice de sentimento do investidor europeu, com base em novas *proxies* de sentimento individuais, bem como a existência de repercussão entre Estados Unidos e Europa. Além disso, analisaram também a relação deste novo índice de sentimento na volatilidade e no retorno das ações. Foram recolhidos dados do período de janeiro de 1973 a julho 2019. Os resultados evidenciam que o índice de sentimento proposto possui relação muito próxima com o índice universalmente conhecido em estudos anteriores, revelando forte co-movimento entre os mercados de ações dos Estados Unidos e Europa. Além disso, identificam também que, o otimismo e boas notícias tem um impacto mais forte na volatilidade do que o pessimismo e más notícias.

Souza e Martins (2020) estudam a relação entre as informações divulgadas no *Twitter* e o retorno e volume de negócios no mercado brasileiro. Foram recolhidos 314.864 *tweets* relacionados com as ações negociadas no Brasil entre 01 de janeiro de 2017 e 31 de dezembro de 2018. O índice de sentimento foi desenvolvido com auxílio de técnicas de *machine learning* por meio do *Google Cloud Natural Language API*. Os resultados indicaram uma relação positiva entre otimismo e retornos positivos em tempos contemporâneos, com tendência de inversão com o passar dos dias, em que um sentimento otimista atual está associado a retornos inferiores no futuro. Além disso, há associação significativa entre o volume diário de *tweets* e o volume de negócios realizados. Quanto maior o número de mensagens com conotação negativa, maior também tende a ser o volume. Neste sentido, foi possível concluir que existe associação entre as informações divulgadas em redes sociais, nomeadamente o *Twitter*, e os movimentos de mercado na realidade brasileira, principalmente em períodos de pessimismo.

Reis e Pinho (2020c) estudam a adequação de 13 *proxies* de sentimento do investidor, normalmente utilizadas como indutoras dos retornos mensais de ações do índice europeu S&P 350 ao longo de 45 anos. A amostra foi composta por 362 empresas de 16 países europeus. Os resultados identificam que diferentes medidas de sentimento do investidor explicam o retorno das ações. Medidas de sentimento de volatilidade juntamente com indicadores de sentimento económico e de confiança, apresentaram efeito significativo na previsão de comportamentos irracionais de investidores para determinar o retorno das ações.

Ferreira et al. (2021) analisam o efeito do sentimento do investidor sobre a volatilidade do mercado de ações brasileiro, considerando empresas que possuem características de difícil precificação. A volatilidade foi avaliada pelo índice IBRX100 e, o índice de Miranda (2018) foi utilizado como *proxy* para sentimento do investidor. Os resultados indicaram uma relação negativa e significativa entre o sentimento e a volatilidade do mercado acionista.

Por sua vez, Costa (2021) procura compreender a relação entre as notícias, consideradas mais complexas e, a reação do mercado a um conjunto de informação. Foram utilizados artigos de notícias financeiras recolhidas dos media especializado, com recurso a *web scraping* próprios, procurando obter de três tipos de variáveis: variáveis de complexidade textual, de sentimento textual e de atenção dos media especializada, com recurso ao dicionário desenvolvido por Silva (2017). Os resultados indicaram que as notícias menos complexas estão relacionadas com maiores volumes de negociação e com maiores níveis de investimentos corporativos e, que o sentimento textual tem prevalência de curtíssimo prazo e, que a complexidade afeta o mercado por períodos mais longos.

Nogueira et al. (2021) constroem um índice de sentimento do investidor para o mercado brasileiro no período de 2008 a 2017. A amostra foi composta por 150 empresas não-financeiras. A técnica estatística escolhida foi de Análise de Componentes Principais e, selecionadas como *proxies* de sentimento do investidor variáveis de mercado, tais como: liquidez, prémio de dividendos, proporção de *put-call*, entre outros, à semelhança de Baker e Wurgler (2006). Os resultados apontaram para uma correlação positiva e estatisticamente significativa entre o índice de sentimento e o Ibovespa, indicando que níveis de sentimento mais altos são acompanhados por tendência de crescimento no Ibovespa. Além disso, a evolução do índice de sentimento revelou-se condizente com os momentos económicos do País.

Valcareggi (2022) investiga o impacto das notícias económicas no Índice Bovespa, considerando diversos cenários económicos. O índice de sentimento foi construído com base

em 104.290 notícias sobre a economia brasileira em jornais de grande circulação. Foi utilizado o dicionário léxico proposto por Shapiro et al. (2020). Os resultados indicaram que, em períodos de estabilidade, os índices de sentimento e Ibovespa reagem na mesma proporção e direção. Face a um cenário económico favorável, as notícias pessimistas de 2004 a 2009 não foram capazes de influenciar a evolução crescente do Ibovespa. Por outro lado, diante de um cenário económico desfavorável, as notícias boas entre 2009 e 2017 não foram capazes de impulsionar o Ibovespa.

Santos (2022) utiliza a análise de sentimento das notícias económicas e financeiras, procurando compreender o seu impacto sobre a economia e os agentes económicos. Foi utilizado a técnica de *web scraping* para recolha dos dados, através da linguagem *Python*, e de vetores autorregressivos para análise. Os dicionários de léxicos escolhidos foram o LM-SA-2020, desenvolvido no trabalho de Terblanche e Marivate (2021), como cita Santos (2022) e, o dicionário VADER. A fonte de notícias foram as postagens no *Twitter* da Folha de São Paulo, no período de 2011 a 2022. Os resultados indicaram que o dicionário LM-SA-2020 apresentou um desempenho superior relativamente ao VADER na explicação da evolução das variáveis macroeconómicas, comportamento já esperado atendendo à sua fundamentação teórica e económica. Períodos de expansão da atividade económica coincidem com alta dos preços e juros no longo prazo. O índice de sentimento mensurado com base em notícias mostrou-se relevante para compreensão dos fenómenos económicos, bem como a importância da escolha de dicionários léxicos, na medida em que, revelam uma elevada capacidade de adequação ao cenário em estudo.

Duarte (2023) elaborou um índice de sentimento do investidor para o mercado brasileiro, tendo por base o trabalho de Baker e Wurgler (2007) e, propôs um índice alternativo com recurso a outras *proxies* que procuram capturar o sentimento de forma mais tempestiva. Os dados recolhidos referem-se a empresas cotadas em bolsa para o período de 2010 a junho 2022. Os resultados indicaram que em períodos de pré-crise o sentimento é mais acentuado e, que a alteração de sentimento é capturada de forma mais correta a partir da *proxy* captação líquida de fundos de ações.

Castro (2023) investigou o modo como, os sentimentos dos investidores influenciam o preço de encerramento do *Bitcoin*, face às publicações na rede social *Twitter* no período de 10/02/2021 a 06/01/2023. O VADER foi utilizado como dicionário léxico para mensurar o sentimento. Os resultados indicaram que os *tweets* influenciam o preço do *Bitcoin*, porém divergem relativamente às abordagens propostas: o modelo *Autoregressive Distributed Lag*

(ARDL) identifica a prevalência de relação entre as variáveis, enquanto o modelo de regressão linear não capta qualquer relação. O autor, considera o modelo ARDL mais adequado para o estudo, atendendo à relação estabelecida entre as variáveis.

Carvalho et al. (2023) analisaram a evolução dos estudos com abordagem textual para prever o retorno das ações no mercado acionista. A amostra final contou com 78 artigos, do período de 2006 a 2021. De um modo geral, os resultados evidenciaram a capacidade do sentimento textual explicar o comportamento no mercado. Além disso, entre os artigos analisados, apenas 2,6% tinham o Brasil como foco de investigação. A predominância dos estudos incide nos Estados Unidos (57,7%), o que denota uma clara preferência por países desenvolvidos em detrimento dos países emergentes.

Outros trabalhos brasileiros recentes contribuem para a literatura sobre sentimento do investidor, tal como Miranda (2023) e Ribeiro (2022), porém com respaldo na metodologia de Baker e Wurgler (2006), reforçando a contribuição deste trabalho na utilização de notícias do mercado e aplicação da abordagem léxica.

O quadro 1 apresenta uma síntese dos principais estudos empíricos identificados na revisão da literatura, relacionados com análise de sentimento:

Quadro 1 - Resumo trabalhos relacionados

Autor	Objetivo	País	Período Estudado	Fonte informações	Dicionário / Proxy sentimento	Principais resultados / Conclusões
Baker e Wurgler (2006)	Estudar o modo como o sentimento do investidor afeta a <i>cross-section</i> dos retornos das ações.	EUA	1962 e 2001	Mercado	Baker e Wurgler (2006)	* Sentimento tem efeitos transversais significativos no retorno das ações; * Observada relação inversa entre nível de sentimento (alto ou baixo) e retorno subsequente (alto ou baixo) para determinados tipos de ações.
Tetlock (2007)	Investigar a interação entre os media e o mercado de ações.	EUA	1984 a 1999	Notícias de jornais	<i>Harvard-IV</i>	* Elevado pessimismo nas notícias está relacionado a uma queda no preço das ações; * Perspetivas pessimistas altas quanto baixas nas notícias publicadas estavam relacionadas ao alto volume de negociação de ações.
Pagliarussi et al. (2016)	Extrair o tom de relatórios de administração de empresas brasileiras, relacionando-as com variáveis de mercado.	Brasil	1997 a 2009	<i>Reports</i> de empresas	Desenvolveram o próprio dicionário	* Fraca associação entre a métrica do tom da notícia e volume de negociações e da volatilidade dos preços das ações; * Lista de palavras positivas é positivamente associada com retorno anormal e negativamente associada com volatilidade do retorno das ações.
Silva (2017)	Analisar o impacto das notícias sobre a volatilidade dos preços, considerando um cenário de incertezas econômicas.	Brasil	25/07/2011 a 30/06/2017	Notícias de jornais	Desenvolveram o próprio dicionário	* O sentimento textual pode ser considerado como preditor da volatilidade; * Sentimento de pessimismo reduzem o retorno em períodos de incerteza económica.

Galdi e Gonçalves (2018)	Investigar se existe associação entre a rentabilidade e volatilidade dos índices de mercado com o teor positivo ou negativo extraído das notícias divulgadas pelos media.	Brasil	02/01/2012 a 30/12/2016	Notícias de jornais	Utilizam o dicionário de Pagliarussi et al. (2016)	* Mercado avalia com maior peso palavras de incerteza e negativas; * Termos do tipo “incerteza” tem relação negativa com a rentabilidade, e “negativas” tem relação com a volatilidade.
Piccoli et al. (2018)	Investigar a influência do sentimento do investidor na relação risco-retorno no mercado de ações Brasileiro.	Brasil	01/2002 a 10/2015	Mercado	Índice de Confiança do Consumidor (ICC)	* Relação positiva entre variância condicional e retorno das ações em períodos de baixo sentimento, exceto para pequenas ações; * Deterioração da relação positiva entre risco e retorno, quando o sentimento é alto, resultado do aumento de investidores menos sofisticados.
Mello (2019)	Investigar se o sentimento do investidor afeta os retornos da bolsa de valores brasileira.	Brasil	10/2003 a 09/2018	Mercado	Adaptação da metodologia de Baker e Wurgler (2006)	* Sentimento do investidor explica o retorno das ações com maior volatilidade; * Ações mais especulativas são mais difíceis de arbitrar, estando mais sujeitas ao sentimento dos investidores irracionais.
Silva e Machado (2019)	Analisar se o risco sistemático das ações preferenciais da Vale (VALE5) sofreu influência do tom e do volume de notícias atrelado ao acidente ocasionado pela Samarco.	Brasil	05/11/2015 a 16/02/2016	Notícias de jornais	Desenvolveram o próprio dicionário	* Volume e o tom das notícias veiculadas nos media influenciam o beta da ação; * Índícios de que o risco sistemático apresenta conexão com as divulgações de notícias pelos media, nos períodos de maior incerteza.

Reis e Pinho (2020a)	Investigar o impacto da racionalidade ou irracionalidade nos retornos das ações de diferentes países e setores.	<i>USA, UK, France, Italy, Spain e Germany</i>	2015 até 04/2020	<i>Google Trends / Notícias</i>	Baseado no método de Schatteman and Waymire (2017) e Gao, Ren, and Zhang (2018) - <i>Merriam-Webster</i>	* Os retornos dos EUA são mais sensíveis ao sentimento, e por consequência, mais propensos a fatores irracionais, quando comparados com os retornos na Europa; * O fator “país” influencia os retornos de forma diferente.
Igarashi et al. (2021)	Análise da correlação do comportamento variante do preço de ativos da bolsa de valores, conjuntamente com polaridade das notícias.	Brasil	18/11/2019 a 02/12/2019	Notícias de jornais	<i>Machine Learning - Amazon Comprehend</i>	* Correlação existente entre os sentimentos e a movimentação do preço da ação é, predominantemente, moderada.
Reis e Pinho (2020b)	Apresentam um novo índice de sentimento do investidor europeu, com base em novas proxies de sentimento individuais, bem como a existência de repercussão entre Estados Unidos e Europa.	EUA e Europa	01/1973 a 07/2019	Mercado	CCI, economic sentiment indicator, VSTOXX, gold bullion, spread between 10- and others	* Forte co-movimento entre os mercados de ações dos Estados Unidos e Europa; * Otimismo e boas notícias tem um impacto mais forte na volatilidade do que o pessimismo e más notícias.
Souza e Martins (2020)	Investigar a relação entre as informações divulgadas no <i>Twitter</i> e o retorno e volume de negócios no mercado brasileiro.	Brasil	01/01/2017 a 31/12/2018	<i>Twitter</i>	Sentimento mensurado pelo <i>Google Cloud Natural Language API</i>	* Associação significativa entre o volume diário de tweets e o volume de negócios realizados; * Associação entre as informações divulgadas em redes sociais, nomeadamente o <i>Twitter</i> , e os movimentos de mercado na realidade brasileira, principalmente em períodos de pessimismo.

Reis e Pinho (2020c)	Analisar a adequação de 13 <i>proxies</i> de sentimento do investidor, normalmente utilizadas como indutoras dos retornos mensais de ações do índice europeu S&P 350 ao longo de 45 anos.	16 países europeus e 362 empresas	01/1973 a 07/2019	Mercado	13 <i>proxies</i> de sentimento do investidor	* Diferentes medidas de sentimento do investidor explicam o retorno das ações; * Algumas medidas apresentaram efeito significativo na previsão de comportamentos irracionais de investidores para determinar o retorno das ações.
Ferreira et al. (2021)	Analisar o efeito do sentimento do investidor sobre a volatilidade do mercado acionário brasileiro.	Brasil	01/2006 a 12/2017	Mercado	Utilizaram o Índice de Miranda (2018)	* Sentimento tem relação negativa e significativa com a volatilidade do mercado acionário brasileiro, bem como evidenciam um comportamento assimétrico, mais forte em períodos pessimistas; * Capacidade explicativa do sentimento é sensível às características das empresas.
Costa (2021)	Estudar o papel da complexidade textual das notícias financeiras em relação ao processo de transmissão de informação dentro do mercado financeiro.	Brasil	10/2012 a 03/2020	Notícias de jornais	Utilizaram o dicionário de Silva (2017)	* Quanto menor a complexidade textual, mais intenso é o processo de transmissão de informação aos preços; * Quanto menos complexas forem as notícias, maiores serão os níveis de investimento posteriores das empresas.
Nogueira et al. (2021)	Construir um índice de sentimento do investidor para o mercado brasileiro.	Brasil	01/01/2008 a 31/12/2017	Mercado	Adaptação da metodologia de Baker e Wurgler (2006, 2007) e Yoshinaga (2009)	* Correlação positiva e estatisticamente significativa entre o índice de sentimento e o Ibovespa; * Evolução do índice de sentimento revelou-se condizente com os momentos econômicos do país.

Valcareggi (2022)	Esclarecer se notícias boas ou ruins impactam no Índice Ibovespa.	Brasil	2004 a 2020	Notícias de jornais	Utilizaram o dicionário de Shapiro et al. (2020), que uniram os dicionários VADER, LM e HL	<ul style="list-style-type: none"> * Em períodos de estabilidade, os índices reagem proporcionalmente na mesma direção; * Em períodos económicos favoráveis, notícias pessimistas não influenciaram o Ibovespa. * Em períodos económicos desfavoráveis, notícias otimistas não impulsionaram o Ibovespa.
Santos (2022)	Análise de sentimento nas notícias económicas e financeiras buscando compreender seu impacto sobre a economia e agentes económicos.	Brasil	2011 a 2022	Notícias no <i>Twitter</i>	VADER e LM-SA-2020	<ul style="list-style-type: none"> * Dicionário LM-SA-2020 obteve melhores resultados do que o dicionário VADER; * Com o dicionário LM-SA-2020, observou-se que altas na atividade coincidiam com elevações dos preços e juros no longo prazo.
Duarte (2023)	Elaborar o índice de sentimento do investidor para o mercado brasileiro de ações com base em Baker e Wurgler (2006), e propor a construção de um índice alternativo por meio de outras <i>proxies</i> .	Brasil	2010 a junho/2022	Mercado	Adaptação da metodologia de Baker e Wurgler (2006)	<ul style="list-style-type: none"> * Em períodos de pré-crise o sentimento é mais exacerbado; * A proxy denominada "captação líquida de fundos de ações" é capaz de capturar mais tempestivamente mudanças no sentimento; * Resultados demonstraram a existência de indicadores eficazes para avaliar o sentimento do investidor.
Castro (2023)	Investigar como os sentimentos dos investidores influenciam o preço do fechamento do Bitcoin, diante das publicações na rede social <i>Twitter</i> .	Brasil	10/02/2021 a 06/01/2023	Notícias no <i>Twitter</i>	VADER	<ul style="list-style-type: none"> * Os resultados indicaram que os tweets influenciam o preço da bitcoin.

3. Dados e metodologia

Neste capítulo são apresentadas as fontes de informação para a construção das variáveis do modelo e, os testes estatísticos que garantem a qualidade e a representatividade das informações analisadas. São ainda descritas as técnicas e ferramentas utilizadas para tratar a base das notícias e a mensuração do sentimento do investidor.

3.1. Variáveis do modelo

O índice Ibovespa será utilizado como *proxy* do retorno de mercado (variável dependente) e, as variáveis explicativas (independentes) serão o sentimento do investidor – avaliado a partir do conteúdo das notícias, a taxa de câmbio (USD/BRL), o retorno de uma *commodity*, o ouro, o retorno de uma criptomoeda, o *Bitcoin*, o “risco país” do Brasil, avaliado pelo índice *Credit Default Swap* (CDS) do Brasil e, a frequência diária de notícias publicadas.

O índice Ibovespa é frequentemente identificado em trabalhos anteriores (ex.: Galdi e Gonçalves, 2018; Valcareggi, 2022; Igarashi et al., 2021; Marschner et al., 2021) como uma *proxy* do retorno de mercado. A variável sentimento do investidor pode ser identificada nos trabalhos de Galdi e Gonçalves (2018), Silva (2017), Silva e Machado (2019), Costa (2021), entre outros. A variável taxa de câmbio é utilizada na literatura, como por exemplo no trabalho de Marschner et al. (2021) e Paraboni et al. (2016). A variável ouro, representando uma *commodity*, e a variável *Bitcoin*, também são utilizadas na literatura, como por exemplo nos trabalhos de Karameshinova (2023), Vartanian (2022), Castro (2023) e Pereira e Arevalo (2024). A variável “risco país” do Brasil, o CDS, foi utilizada por Sing (2019) e Azevedo Neto (2019). Por fim, a frequência diária de notícias foi uma *proxy* utilizada por Silva e Machado (2019), Ferguson et al. (2015) e Faustino (2013).

Para a variável dependente Ibovespa, foi considerado o retorno logarítmico dos preços de fecho, de acordo com Santana (2018). O autor destaca que a vantagem do retorno logarítmico reside no fato de as propriedades estatísticas melhorarem, sendo adequado para substituir a variação relativa de preços. O cálculo foi realizado da seguinte forma:

$$\ln_ibov = \ln\left(\frac{ibov_t}{ibov_{t-1}}\right)$$

A cotação das variáveis independentes, nomeadamente, *Bitcoin*, ouro, câmbio e “risco país” do Brasil foram extraídas do site “Investing.com”, para o período de novembro de 2021 a agosto de 2024. Consideraram-se as variações diárias com base na informação de fecho diário, sendo calculado da seguinte forma:

$$var = \frac{P_t}{P_{t-1}} - 1$$

Os dados são analisados com recurso à regressão linear múltipla em séries temporais. A equação a ser testada no *Stata BE 18.5* é expressa da seguinte forma:

$$\begin{aligned} \ln_ibov_n = & \beta_0 + \beta_1 qtd_news_n + \beta_2 sent_VADER_n + \beta_3 sent_LM_n + \beta_4 sent_HIV_n \\ & + \beta_5 Bitcoin_n + \beta_6 Ouro_n + \beta_7 USD_BRL_n + \beta_8 CDS_Brasil_n + \varepsilon_n \end{aligned}$$

(equação 1)

Onde, \ln_ibov_n representa o logaritmo do retorno do índice de mercado Ibovespa no período n ; β_0 representa o intercepto da regressão; qtd_news_n representa a frequência diária de notícias publicadas no período n ; $sent_VADER_n$ representa o sentimento do investidor mensurado pelo dicionário VADER no período n ; $sent_LM_n$ representa o sentimento do investidor mensurado pelo dicionário de *Loughran e McDonald* no período n ; $sent_HIV_n$ representa o sentimento do investidor mensurado pelo dicionário *Harvard-IV* no período n ; $Bitcoin_n$ representa o retorno do Bitcoin no período n ; $Ouro_n$ representa o retorno do Ouro no período n ; USD_BRL_n representa a variação do câmbio Dólar/Real no período n ; CDS_Brasil_n representa a variação do *Credit Default Swap* do Brasil no período n ; ε_n representa o termo de erro no período n . Os coeficientes β_1 a β_8 indicam a magnitude e a direção do impacto de cada variável explicativa sobre a variável dependente, isto é, o índice de mercado Ibovespa (\ln_ibov_n).

3.2. Validação do modelo -Testes estatísticos

Para validação do modelo proposto, serão realizados pré-testes à estacionariedade das variáveis, multicolinearidade, autocorrelação dos resíduos e heterocedasticidade.

O pressuposto da estacionariedade das variáveis será verificado pelo teste *Dickey–Fuller*, desenvolvido por Dickey e Fuller (1979). Este teste verifica a existência ou não de uma raiz unitária, isto é, se choques externos causam efeito permanente na série. A hipótese nula (H_0) é de que a série possui raiz unitária e, portanto, não é estacionária. A hipótese alternativa (H_1), a série é estacionária. Portanto, se uma série não possui raiz unitária será considerada estacionária, indicando que há uma tendência de retorno à média após certo tempo. Espera-se que a hipótese nula seja rejeitada (Dawid, 2004).

O pressuposto da multicolinearidade será testado pelo *Variance Inflation Factor* (VIF). Este teste investiga se há correlações elevadas entre as variáveis do modelo. A presença de multicolinearidade implica uma estimativa imprecisa dos coeficientes de regressão. A estatística VIF superior a 10 indica problemas de multicolinearidade, sendo ideal então que o resultado do teste seja inferior a 10. A matriz de correlações também será verificada, para ferir o grau de correlação entre as variáveis do modelo (Fávero, 2015; Akinwande et al., 2015).

O pressuposto da autocorrelação dos resíduos será verificado pelo teste de *Breusch–Godfrey LM*, desenvolvido por Breusch (1978) e Godfrey (1978). Este teste verifica a hipótese de aleatoriedade e independência dos termos de erro. A hipótese nula (H_0) é de que não existe autocorrelação dos resíduos (os resíduos são independentes) e a hipótese alternativa (H_1) identifica autocorrelação nos resíduos (os erros não são independentes). A hipótese nula não deve ser rejeitada.

O pressuposto da heterocedasticidade será verificado pelo teste *Breusch–Pagan*, desenvolvido por Breusch e Pagan (1979). Este teste investiga se a variância dos resíduos é constante ao longo da variável explicativa. A hipótese nula (H_0) é de que a variância dos termos é constante, indicando homoscedasticidade. A hipótese alternativa (H_1) é de que a variância dos termos não é constante, indicando então heterocedasticidade. Espera-se que a hipótese nula não seja rejeitada.

3.3. Processamento das notícias e o sentimento do investidor

Considerando que o sentimento não é diretamente observável, torna-se então necessário construir *proxies* para o avaliar e realizar testes para aferir a sua capacidade de prever o comportamento dos preços dos ativos. Reis e Pinho (2020c) destacam que as medidas mais utilizadas incluem dados de mercado, dados de inquéritos e dados coletadas pela análise de textos.

Na construção do índice de sentimento do investidor, foi tido em conta um conjunto de informações, nomeadamente: o título e o conteúdo das notícias sobre aspetos financeiros, criados por especialistas e, divulgadas para o público institucional sujeito a licenciamento. Assim, o público em geral tem acesso apenas ao título das notícias e, à informação completa mediante um pagamento. Trabalhos anteriores (ex.: Igarashi et al., 2021; Galdi e Gonçalves, 2018) também utilizaram como fonte de notícias *sites* especializados, mas apenas o conteúdo gratuito das notícias. Neste sentido o presente estudo, pretende ser mais exaustivo na abordagem que realiza ao conteúdo informativo das notícias divulgadas.

A base de dados, com as notícias utilizadas no âmbito deste trabalho, foi adquirida à empresa *Valor Económico*, do serviço “Valor +News”, que constitui uma referência na divulgação de notícias de índole financeira do Brasil e do mundo, com periodicidade diária e, informação disponível para o horizonte temporal de novembro de 2021 a agosto de 2024. O arquivo original, em formato *JavaScript Object Notation* (JSON), conta com 9.583 notícias.

Bases de dados do *Valor Económico* foram utilizadas como fonte de informação financeira nos trabalhos realizados por Silva e Machado (2019), Galdi e Gonçalves (2018) e Silva (2017), por conter dados relevantes sobre a evolução do mercado brasileiro, pelo que, pode ser considerada uma fonte de informação importante para construir uma *proxy* do sentimento do investidor, utilizando as notícias de mercado. Igarashi et al. (2021) destacam a importância de utilização de fontes de informações fiáveis, mitigando assim os efeitos de ambiguidades e de incertezas relacionadas com o conhecimento do autor da informação.

A vantagem de utilizar métodos computacionais no tratamento de grandes bases de dados reside na velocidade de processamento, produzindo resultados consistentes, mitigando o efeito de enviesamento por opiniões ou predileções individuais (Liu e Zhang, 2012).

Assim, a base com as notícias foi tratada com recurso à ferramenta *NotePad++*, de modo que a estrutura JSON fosse interpretada corretamente pelo código *Python* desenvolvido especificamente no âmbito deste trabalho. Foram excluídas as *TAGs HTML* por meio da identificação de expressões regulares. Tal procedimento foi realizado de modo a não contaminar a base de notícias.

Durante o processamento do arquivo com as notícias no formato original, foi criado um arquivo em formato “.csv” com 04 (quatro) colunas, a saber: “*issued*” a data de publicação; “*id*” o código de sequência alfanumérica como identificação única de cada notícia; “*title*” o título da notícia; e “*text*”, o conteúdo da notícia. Para esta última informação, foi necessário concatenar

todas as chaves (*keys*) do tipo “*text*”, tendo em vista que cada chave representava um parágrafo da notícia.

A partir deste novo arquivo com as notícias estruturadas, foi utilizado o “*Google Translator*” da *library* “*deep-translator*” *Python* para traduzir as notícias para o idioma inglês. Este procedimento foi necessário, porque os dicionários de léxicos possuem palavras em inglês. Além disso, optou-se por garantir a eficiência na apuração do sentimento (tom) da notícia, isto é, não realizar a tradução durante o processamento. Tal procedimento também foi útil para novos processamentos somente do sentimento (tom), não sendo necessário traduzir a mesmas notícias várias vezes.

A biblioteca “*deep-translator*” realizou a tradução para as notícias até 5.000 caracteres. Para as notícias com um número superior de caracteres, foi necessário utilizar a funcionalidade de “*Translate files*” do site *Deepl*. Após tradução de todas as notícias, foi criado um arquivo “.csv” com o “*id*” e o conteúdo da notícia, constituindo este o *input* final para análise de sentimento.

O trabalho de Reis et al. (2015) conclui que a tradução de texto para outras línguas não afeta significativamente a acurácia ou abrangência dos métodos testados na análise de sentimentos. Sendo assim, a utilização de ferramentas de tradução não constitui um problema, desde que utilizado o idioma original do método. Santos (2022) também recorreu a estratégia de tradução para inglês das notícias em português extraídas no *Twitter*, dado que, este é o idioma dos léxicos mais utilizados nesta área de estudos.

Os arquivos de notícias foram processados separadamente para cada um dos 3 (três) códigos *Python* associados aos dicionários de léxicos: *Harvard-IV*, *Loughran e McDonald* (Loughran e McDonald, 2011) e *VADER* (Hutto e Gilbert, 2014). Para os léxicos *Harvard-IV* e *Loughran e McDonald* foi utilizado a biblioteca “*pysentiment2 0.1.1*” (DeRobertis, n.d.), e para o léxico *VADER* foi utilizado a biblioteca “*vaderSentiment 3.3.2*” (Hutto, n.d.). Ambos os dicionários fornecem funções específicas para mensuração do sentimento por meio da abordagem léxica, com as respectivas implementações em código *Python* para calcular a polaridade do texto. Os *outputs* com a indicação do sentimento (tom) de cada notícia foram guardados num novo arquivo “.csv”, contendo novamente o “*id*” e o respetivo sentimento.

O sentimento (tom) de cada notícia foi representado pela sua polaridade, calculado com base na quantidade de palavras positivas e negativa para os dicionários *Loughran e McDonald* e *Harvard-IV*, e com base na valência (intensidade emocional) das palavras para o dicionário

VADER. O *output* dos códigos é um valor normalizado entre “-1” e “+1”, sendo “-1” indicativo de um sentimento máximo negativo (pessimismo), e “+1” indicativo de um sentimento máximo positivo (otimismo) (Silva e Machado, 2019). Esta faixa de valores foi a utilizada pelos 3 (três) dicionários léxicos, o que garantiu a uniformidade para efeitos de comparação.

A base final de notícias, com o *id* e o respetivo sentimento mensurado de forma quantitativa, foi construída no *Excel*. Foi calculada a média aritmética simples do sentimento para os dias com mais de uma notícia, em linha com o trabalho de Silva e Machado (2019). Esta etapa foi essencial para a contagem de notícias por dia, sendo esta uma variável explicativa que irá incorporar o modelo. A base final contou com 9.515 notícias, agrupadas em 697 observações diárias com a média do sentimento para os dias que tiveram variação do índice de mercado Ibovespa.

Em síntese, o código *Python* desenvolvido neste trabalho utilizou técnicas de leitura, escrita e manipulação de arquivos, além de integrar bibliotecas de terceiros para tradução e análise de sentimento (tom). A combinação dessas abordagens possibilitou a automação do processamento de um elevado volume de notícias, aumentando a eficiência na extração e organização de dados textuais. Isso permitiu a geração de *insights* significativos, servindo como base para análises subsequentes. Os códigos foram executados no *Google Colab*, e podem ser consultados no Apêndice A.

Ao realizar este trabalho com o recurso a um algoritmo na análise de dados, esperamos reforçar a importância crescente que a tecnologia do processamento de grande volume de dados tem e, fornecer contribuições valiosas para o estudo de questões cada vez mais complexas.

Este estudo inova ao utilizar notícias exclusivas de uma base de notícias técnicas, de acesso restrito, evitando enviesamento de notícias de fontes diversas, bem como ao utilizar 3 (três) dicionários de palavras para modelar os resultados, procurando identificar qual deles tem maior poder preditivo na explicação das variações do mercado de ações.

4. Análise dos resultados

Neste seção são apresentados os resultados obtidos a partir da análise dos dados coletados. Inicialmente, serão discutidas as estatísticas descritivas, com uma visão geral das variáveis do estudo. De seguida, serão analisados os pressupostos da regressão linear múltipla, com o objetivo de verificar se os dados atendem aos requisitos fundamentais para a aplicação deste modelo. Por fim, serão apresentados os resultados da regressão linear múltipla, com a interpretação dos coeficientes, significância estatística e o poder explicativo do modelo.

4.1. Estatística descritiva

A primeira análise incide sobre a estatística descritiva das variáveis do modelo. Esta análise permite explorar as principais características dos dados, tais como distribuição dos valores, medidas de dispersão e de tendência central. Selecionamos o número de observações, média, os percentis a 50% (mediana) e 95%, desvio padrão e valores máximo e mínimo. Os resultados podem ser analisados na Tabela 1 abaixo:

Tabela 1 - Estatística descritiva das variáveis do modelo

Variable	Obs	Mean	Percentiles		Std. dev.	Min	Max
			50%	95%			
ln_ibov	697	.000424	.0004274	.0184775	.0111302	-.034488	.053928
qtd_news	697	13.65136	14	20	3.845486	2	31
sent_VADER	697	.5254248	.5506	.8323067	.2191688	-.57755	.96885
sent_LM	697	-.2838067	-.2888985	-.0352045	.1423894	-.7456858	.1867071
sent_HIV	697	.2433882	.2429599	.4057619	.1004818	-.0863323	.5678228
Bitcoin	697	.0000403	-.0013	.0497	.0319629	-.1563	.1194
Ouro	697	.0004073	.0003	.0159	.0093175	-.0281	.0314
USD_BRL	697	.0000568	0	.0165	.0092238	-.0458	.0371
CDS_Brasil	697	-.0002139	-.0002	.0373	.0366329	-.2847	.4219

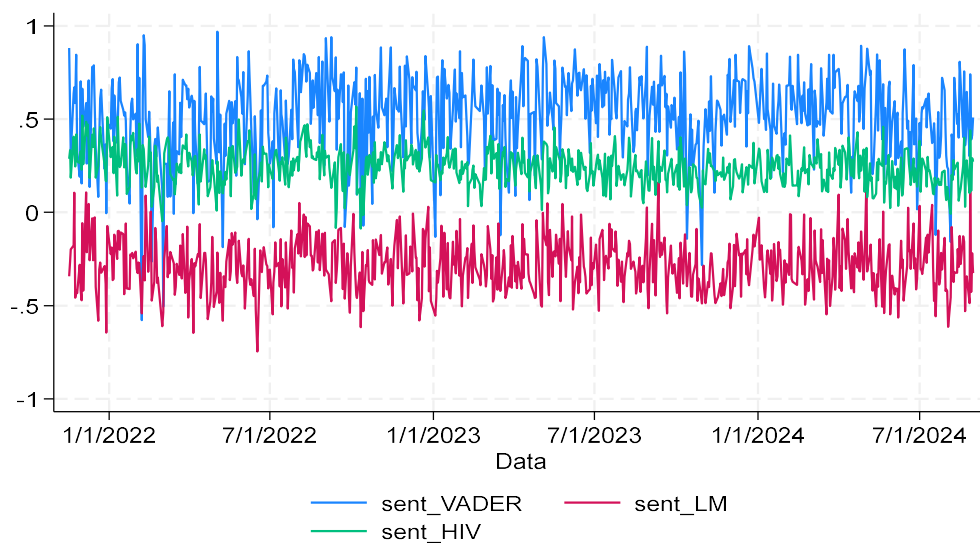
Em relação ao Ibovespa (ln_ibov) 50% das variações foram iguais ou inferiores a 0,04274% e 95% das variações foram iguais ou inferiores a 1,84%, sendo a média aritmética de 0,0424%. O desvio padrão foi de aproximadamente 1,1%, sendo a variação mínima negativa de aproximadamente 3,4% e variação máxima positiva de aproximadamente 5,3%. Silva (2017) também identificou um valor médio positivo para o logaritmo do retorno do Ibovespa, embora tenha deflacionado os valores pelo Índice Nacional de Preços ao Consumidor Amplo (IPCA), à inflação oficial do Brasil. Porém, a mediana da variável indicou um valor negativo no período

estudado, sugerindo a presença de *outliers*. Souza e Martins (2020) identificaram valores positivos para média e mediana do Ibovespa, tendo sido mensurada pelo retorno dos *Exchange-Traded Funds* (ETFs).

Em relação a quantidade de notícias (*qtd_news*) 50% do número diários de notícias foi igual ou inferior a 14 e 95% do número diário de notícias foi igual ou inferior a 20, sendo a média aritmética de 13,6 notícias por dia. O desvio padrão foi de aproximadamente 384%, com frequência mínima diária de aproximadamente 2 notícias e frequência máxima de 31 notícias.

Em relação as variáveis que medem o sentimento do investidor (*sent_VADER*, *sent_LM*, *sent_HIV*), embora o intervalo possível seja de -1 a +1, os resultados são diferentes para cada variável. O gráfico 1 abaixo evidencia o comportamento das 3 (três) variáveis de sentimento do investidor ao longo do período estudado. Embora o sentimento mensurado tenha ocorrido em faixas distintas para cada variável, é possível verificar uma tendência (padrão) de comportamento similar no período estudado:

Gráfico 1 - Comportamento sentimento do investidor de acordo com diferentes indicadores



Em relação ao sentimento obtido pelo dicionário VADER (*sent_VADER*), 50% da média do sentimento diário foi igual ou inferior 0,55 e 95% da média do sentimento diário foi igual ou inferior a 0,83, com uma média aritmética de 0,52. O desvio padrão foi de aproximadamente 22%, com valor mínimo diário de -0,57 e valor máximo diário de 0,96. O valor mínimo não se aproximou de -1, indicando que, em média, não foram observadas notícias predominantemente negativas, mas sim com tendências negativas. O valor máximo indica que, para alguns dias, observou-se uma média de notícias muito otimistas.

Em relação ao sentimento avaliado pelo dicionário *Loughran e McDonald*, 50% da média do sentimento diário foi igual ou inferior a -0.288 e 95% da média do sentimento diário foi igual ou inferior a -0.035, com uma média aritmética de -0.28. O desvio padrão foi de aproximadamente 14%, com valor mínimo diário de -0.74 e valor máximo diário de 0.18. Para este dicionário léxico, observa-se que não foram mensurados sentimentos positivos superiores a 0.18, isto é, houve uma tendência de classificação de notícias negativas, o que é evidenciado pelo percentil de 95%.

Em relação ao sentimento aferido pelo dicionário *Harvard-IV*, 50% da média do sentimento diário foi igual ou inferior a 0.242 e 95% da média do sentimento diário foi igual ou inferior a 0.405, com a média aritmética de 0.243. O desvio padrão foi de aproximadamente 10%, com valor mínimo diário de -0.086 e valor máximo diário de 0.56. Para este dicionário léxico, observa-se uma tendência de classificação de notícias positivas, com no mínimo 50% dos dados positivas e um valor mínimo médio pouco expressivo para negatividade.

Silva (2017) destaca que há uma tendência de concentração de termos negativos nos dicionários léxicos no contexto financeiro, diante do conteúdo informacional que palavras negativas apresentam para os investidores. Esse comportamento também foi observado por Loughran e McDonald (2011).

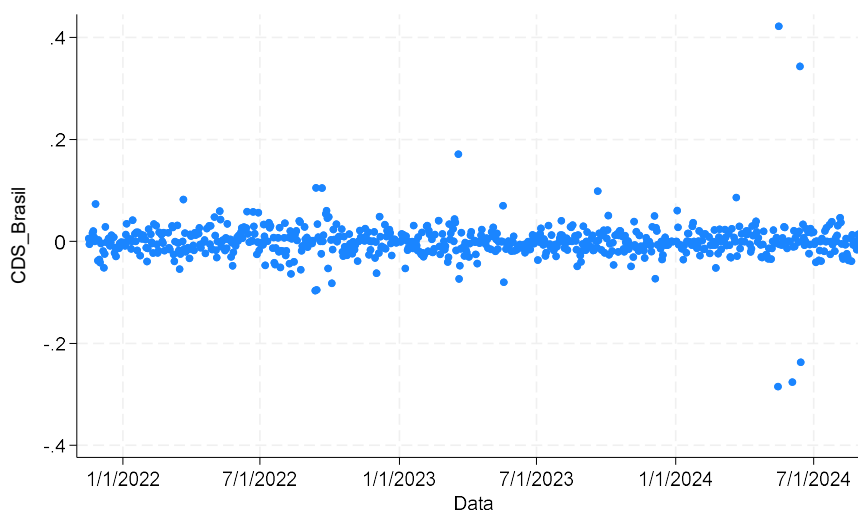
Além disso, é importante considerar o contexto e o propósito para os quais cada dicionário léxico foi desenvolvido, pois esses fatores influenciam diretamente o conjunto de palavras, que pode variar significativamente em termos de foco, tom (conteúdo) e aplicação. Por exemplo, o dicionário de *Loughran e McDonald* foi desenvolvido num contexto de análise de relatórios corporativos e comunicados de empresas. A lista de palavras negativas é maior tendo em vista que textos financeiros frequentemente incluem termos com maior ênfase em negatividade, tal como riscos, perda e incertezas. O dicionário VADER foi desenvolvido num contexto de análise de sentimento nos media sociais, com tendência de linguagem informal. Logo, há uma tendência de balanceamento entre palavras positivas e negativas. Por fim, o dicionário *Harvard-IV*, mais balanceado do que o dicionário de *Loughran e McDonald* em termos de palavras negativas e positivas, foi desenvolvido num contexto de estudos psicológicos, sendo aplicável a diversos campos de estudo de uma forma genérica.

Em relação ao risco Brasil (CDS_Brasil), 50% das variações foram iguais ou inferiores a -0,02% e 95% das variações foram iguais ou inferiores a 3,73%, sendo a média aritmética de

-0,0213%. O desvio padrão foi de aproximadamente 3,66%, com variação mínima negativa de aproximadamente -28,4% e variação máxima positiva de aproximadamente 42,1%.

Os *outliers* do risco Brasil (-28,4% e 42,1%) ocorreram aproximadamente nos meses de maio e junho do ano de 2024, conforme evidenciado no gráfico 2 de dispersão abaixo:

Gráfico 2 - Gráfico de dispersão da variável “risco brasil”



Fatores como a revisão das metas de superavit pelo governo federal, a mudança nas presidências da Petrobras e Banco Central e as dificuldades em compor a receita são explicações para aumento do risco Brasil no período, indicando uma deterioração da percepção do risco pelos investidores (Azevedo, 2024; Ferrari, 2024; Benevides, 2024).

No que diz respeito à alteração da presidência do Banco Central, o receio do mercado é de que haja interferência na indicação pelo governo federal, pressionando uma queda brusca da taxa básica de juros da economia, a Taxa Selic. Consequentemente, poderia ocorrer um aumento da inflação e redução do juro real pago pelos títulos brasileiros (Azevedo, 2024; Benevides, 2024)

Por sua vez, o aumento dos gastos públicos transmite ao mercado a percepção de que o governo federal não tem capacidade de honrar os compromissos. A deterioração fiscal tende a impulsionar as taxas de juros, com investidores exigindo maiores retornos para assumir maiores riscos. Taxas de juros mais altas, por si mesmas, resultam em maiores despesas, devido ao impacto do pagamento dos juros da dívida (Azevedo, 2024; Ferrari, 2024).

4.2. Pressupostos do modelo de regressão

O passo seguinte, incorpora um conjunto de testes para garantir a validade e a robustez dos resultados da regressão, nomeadamente ao nível da estacionaridade, heterocedasticidade, autocorrelação dos resíduos e da multicolinearidade.

O pressuposto da estacionaridade da série foi confirmado pelo teste de *Dickey-Fuller* em todas as variáveis do modelo. O *p-value* (0.000) inferior ao nível de significância indica que a hipótese nula deve ser rejeitada, evidenciando a ausência de raiz unitária. Os resultados podem ser verificados na Tabela 2:

Tabela 2 - Resultados do teste *Dickey-Fuller*

<i>Dickey-Fuller test for unit root</i>					
<i>Variable</i>	<i>Test statistic</i>	<i>critical value</i>			<i>p-value Z(t)</i>
		1%	5%	10%	
ln_ibov	-23.228				0.0000
qtd_news	-13.349				0.0000
sent_VADER	-20.195				0.0000
sent_LM	-22.981				0.0000
sent_HIV	-17.173	-3.430	-2.860	-2.570	0.0000
Bitcoin	-25.707				0.0000
Ouro	-23.097				0.0000
USD_BRL	-23.999				0.0000
CDS Brasil	-31.225				0.0000

A pressuposição da heterocedasticidade foi verificado com recurso ao teste *Breusch-Pagan*. Considerando que o *p-value* (0.3785) foi superior ao nível de significância, logo não se rejeita a hipótese de homoscedasticidade, indicando que a variância dos erros é constante e não varia em função das variáveis independentes. Os resultados podem ser verificados na Tabela 3:

Tabela 3 - Resultados teste *Breusch-Pagan*

<i>Breusch-Pagan/Cook-Weisberg test for heteroskedasticity</i>	
<i>Assumption: Normal error terms</i>	
<i>Variable: Fitted values of ln_ibov</i>	
H0: <i>Constant variance</i>	
chi2(1) = 0.78	
Prob > chi2 = 0.3785	

O pressuposto da autocorrelação dos resíduos foi verificado pelo teste *Breusch–Godfrey LM*. Considerando que o *p-value* (0.6422) foi superior ao nível de significância, logo não se rejeita a hipótese nula de inexistência de autocorrelação. Os resultados podem ser verificados na Tabela 4:

Tabela 4 - Resultados teste *Breusch–Godfrey LM*

<i>Breusch–Godfrey for autocorrelation</i>			
lags(p)	chi2	df	Prob > chi2
1	0.216	1	0.6422
<i>H0: no serial correlation</i>			

O pressuposto da ausência de multicolinearidade foi verificado pelo teste *Variance Inflation Factor (VIF)*. A média foi de 1.21, significativamente inferior a 10, indicando que não existe correlação elevada entre as variáveis explicativas do modelo. Os resultados podem ser visualizados na Tabela 5:

Tabela 5 - Resultados teste VIF

<i>Variable</i>	<i>VIF</i>	<i>1/VIF</i>
sent_LM	1.50	0.668065
sent_HIV	1.43	0.697472
sent_VADER	1.42	0.705601
USD_BRL	1.18	0.850373
Ouro	1.06	0.945840
Bitcoin	1.05	0.948244
qtd_news	1.04	0.959235
CDS_Brasil	1.02	0.976322
Mean VIF	1.21	

4.3. Matriz de correlação

Os resultados da matriz de correlação entre as variáveis do modelo podem ser visualizados na Tabela 6:

Tabela 6 - Matriz de correlações entre as variáveis do modelo

	ln_ibov	qtd_news	sent_VADER	sent_LM	sent_HIV	Bitcoin	Ouro	USD_BRL	CDS_Brasil
ln_ibov	1.0000								
qtd_news	-0.0353	1.0000							
sent_VADER	0.3212*	0.1533*	1.0000						
sent_LM	0.2876*	0.0285	0.4497*	1.0000					
sent_HIV	0.2455*	0.0738	0.4044*	0.5061*	1.0000				
Bitcoin	0.2246*	0.0422	0.1002*	0.0824*	0.0846*	1.0000			
Ouro	0.0920*	-0.0327	0.0537	0.0727	0.0925*	0.0916*	1.0000		
USD_BRL	-0.6051*	0.0575	-0.2785*	-0.2026*	-0.2037*	-0.1955*	-0.1853*	1.0000	
CDS_Brasil	-0.0710	0.0126	-0.0621	-0.0327	-0.0436	-0.0815*	-0.1198*	0.0155	1.0000

A matriz de correlação evidencia um número considerável de correlações estatisticamente significativas ao nível de 5%. As variáveis relacionadas com o sentimento (sent_VADER, sent_LM, sent_HIV) apresentaram correlação positiva com a variação do Ibovespa, indicando uma tendência de as variáveis moverem-se na mesma direção. Logo, um sentimento otimista está relacionado com o aumento do índice de mercado Ibovespa, representando a carteira de ações.

A quantidade de notícias (qtd_news), embora não tenha apresentado correlação estatisticamente significativa ao nível de 5%, indicou uma relação negativa com a rentabilidade do Ibovespa. Estes resultados sugerem que um aumento da frequência diária de notícias implicaria uma tendência de queda nas ações. Faustino (2013) também identificou que as correlações apresentadas entre quantidade de notícias e retorno do Ibovespa são foram estatisticamente significativas, embora o coeficiente tenha sido positivo. A diferença encontrada entre coeficientes (positivo e negativo), considerando a presente investigação e o trabalho de Faustino (2013), podem ser explicadas, em parte, pelo período estudado, já que Faustino trabalhou com o período entre 2010 e 2012.

O retorno do Bitcoin apresentou uma correlação positiva e significativa com as variáveis relacionadas ao sentimento do investidor (sent_VADER, sent_LM, sent_HIV). Este comportamento foi observado por Castro (2023) com a aplicação do modelo *Autoregressive Distributed Lag* (ARDL), embora a base de dados tenha sido composta por *tweets* em detrimento de notícias. O autor utilizou o dicionário VADER para mensuração do sentimento.

O risco (CDS_Brasil), embora não significativo ao nível de 5%, e o dólar (USD_BRL), significativo ao nível de 5%, apresentaram um comportamento negativo em relação ao

Ibovespa. Os resultados sugerem que, com o aumento do risco do país e uma valorização do Dólar em relação ao Real, se verifica uma tendência de queda na rendibilidade das ações.

Mello (2019) também identificou uma relação negativa do sentimento, mensurado por uma metodologia adaptada de Baker e Wurgler (2006), e o risco, representado pela variável *Emerging Markets Bond Index* (EMBI). Sing (2019) identificou uma correlação negativa entre a segunda componente, representando o risco CDS dos países estudados, e o índice S&P500. O autor verificou que, tanto para os períodos pré-crise e pós-crise, pode-se notar uma correlação significativa dos primeiros componentes principais com as variáveis representativas do mercado acionista americano. Marschner et al. (2021) identificaram uma relação negativa entre a taxa de câmbio e as medidas de sentimento, representadas pelo Índice de Confiança do Consumidor (ICC) e pelo Índice de Confiança Empresarial (ICE).

Por fim, as variáveis Bitcoin e Ouro estabelecem uma relação positiva e significativa com o Ibovespa. Logo, um aumento da cotação destes valores está relacionados com o aumento do Ibovespa. Mello (2019) identificou uma relação positiva entre sentimento e *commodity*, representado pelo índice calculado pelo Fundo Monetário Internacional (FMI). Karameshinova (2023) também identificou uma correlação positiva entre Bitcoin e Ouro para com o índice de mercado S&P500. Complementarmente, Vartanian et al. (2022) e Karameshinova (2023) identificaram uma relação positiva entre Ouro e Bitcoin, indicando uma tendência de as variáveis caminharem no mesmo sentido.

4.4. Regressão linear múltipla

Embora os pressupostos para aplicar a regressão linear múltipla tenham sido todos verificados, nomeadamente ausência de multicolinearidade, heterocedasticidade e autocorrelação dos resíduos, bem como a presença de comportamento estacionário, optou-se por estimar o modelo com recurso à regressão linear múltipla robusta.

A regressão linear múltipla robusta em séries temporais é uma técnica utilizada para modelar as relações entre uma variável dependente e várias variáveis explicativas em contextos em que as séries temporais apresentam características que podem comprometer a validade do modelo, como heterocedasticidade a autocorrelação dos resíduos ou a presença de *outliers* (Jajo, 2005). Yu e Yao (2017) destacam que a estimativa *Ordinary Least Squares* (OLS) é sensível aos *outliers*, sendo que a presença de uma única observação por ter grande efeito na estimativa OLS.

Verificados os pressupostos que garantem a validade e robustez dos resultados, foi possível a partir do modelo expresso na Equação 1, obter os valores estimados expressos na Tabela 7:

Tabela 7 - Regressão Linear Múltipla

<i>Linear regression</i>		<i>Number of obs =</i>		697		
		<i>F(8, 688) =</i>		50.14		
		<i>Prob > F =</i>		0.0000		
		<i>R-squared =</i>		0.4186		
		<i>Root MSE =</i>		.00854		
<i>ln_ibov</i>	<i>Coefficient</i>	<i>Robust std. err.</i>	<i>t</i>	<i>P>t</i>	<i>[95% conf. interval]</i>	
qtd_news	-.00009	.0000851	-1.06	0.291	-.0002572	.0000771
sent_VADER	.0052522	.0016387	3.21	0.001	.0020347	.0084696
sent_LM	.0085515	.0027541	3.11	0.002	.003144	.0139589
sent_HIV	.0039104	.0037577	1.04	0.298	-.0034675	.0112883
Bitcoin	.0345588	.010782	3.21	0.001	.0133892	.0557285
Ouro	-.0467565	.0360065	-1.30	0.195	-.1174524	.0239393
USD_BRL	-.6423121	.0433709	-14.81	0.000	-.7274672	-.557157
CDS_Brasil	-.0144245	.0077406	-1.86	0.063	-.0296225	.0007734
_cons	.0004199	.0019075	0.22	0.826	-.0033253	.0041652

O modelo contém 697 observações de dados diários. O *R-squared* indica a capacidade do modelo em explicar 41,86% da variação da variável dependente (*ln_ibov*). Logo, 58,14% da variação é explicado por variáveis que não estão no modelo. Percentagem semelhante a esta pode também ser identificado na literatura. Marschner et al. (2021) identificaram um poder explicativo de 42% para o conjunto de variáveis do modelo 1, tendo o Índice de Confiança do Consumidor (ICC) como variável dependente, e 49% para o conjunto de variáveis do modelo 2, tendo o Índice de Confiança Empresarial (ICE) como variável dependente. Os resultados indicaram que quase metade da variação do sentimento do investidor foi explicado pelos fundamentos do mercado.

Os resultados evidenciam que a quantidade de notícias (*qtd_news*) não foi estatisticamente significativa para explicar a variação do Ibovespa, com um *p-value* (0.291) superior ao nível de significância estabelecido (0.05). Tal resultado é coerente, na medida em que a qualidade da notícia é mais relevante do que a quantidade divulgada diariamente. Faustino (2013) também identificou que a quantidade e o conteúdo semântico das notícias não tem efeito significativo sobre o retorno do Ibovespa. Complementarmente, Silva e Machado (2019) encontraram evidências estatisticamente significativas da relação entre volume de notícias e o risco sistemático, evidenciando que os investidores são sensíveis ao sentimento das notícias.

Em relação as variáveis que representam o sentimento do investidor, apenas as variáveis associadas ao dicionário *VADER* (*sent_VADER*) e *Loughran e McDonald* (*sent_LM*) revelaram significância estatística, na medida em que os respectivos *p-values* (0.001 e 0.002) foram inferiores ao nível de significância estabelecida (0.05). Logo, há evidência de que o sentimento do investidor, aferido por estes indicadores, contribui para explicar a variação do Ibovespa. Os coeficientes, sendo positivos, indicam que há uma relação significativa entre as notícias otimistas e a valorização das ações no mercado.

A relação expressa entre as variáveis acima também foi corroborada por outros autores. Silva (2017) identificou evidências estatisticamente significativas da relação positiva e persistente no mercado em relação as notícias divulgadas nos media. Valcareggi (2022) identificou que, em períodos de estabilidade, o índice de sentimento reage proporcionalmente e na mesma direção do Ibovespa, reagindo desproporcionalmente nos demais períodos. Galdi e Gonçalves (2018) encontraram evidências estatisticamente significativas entre palavras classificadas como “incerteza” e “negativas” com o retorno e volatilidade do Ibovespa, respectivamente. Os resultados indicaram que palavras classificadas nestas categorias tendem a implicar em pessimismo no mercado. Piccoli et al. (2018), ao analisarem a relação entre risco-retorno e volatilidade de ativos financeiros, encontraram evidências de que sentimentos de pessimismo e otimismo podem influenciar o comportamento do mercado, indicando que a dinâmica emocional dos investidores tem consequências no preço dos ativos.

Além disso, há indícios de que o sentimento foi refletido no mercado no mesmo dia, já que as variáveis que representam o sentimento do investidor (*sent_VADER* e *sent_LM*) são medidas no mesmo período n em que a variável dependente (\ln_ibov), correspondendo assim ao mesmo período temporal. Isso deve-se, possivelmente, à ação dos *Brokers* de atuarem de forma rápida, indicando uma tendência de absorção imediata das notícias divulgadas nos media. Souza e Martins (2020) também encontraram evidências do reflexo no mesmo dia, do sentimento no retorno do mercado. Complementarmente, Costa (2021) identificou que, quanto menor a complexidade textual – sendo esta uma *proxy* para o custo da informação, mais intenso é o processo de incorporação das informações nos preços.

A variável que representa o sentimento do investidor com base no léxico *Harvard-IV* não se revela estatisticamente significativo, dado que o respectivo *p-value* (0.298), é superior ao nível de significância estabelecido no estudo (0.05). A ausência de significância deste dicionário está de acordo com os argumentos de Loughran e McDonald (2011), de que a utilização de um dicionário com conotações que não pertencem ao domínio da pesquisa em

estudo podem culminar numa classificação pouco relevante. Esta constatação levou os autores a desenvolverem o próprio dicionário, que tem sido recorrentemente utilizado na literatura por diferentes autores, inclusivamente nesta pesquisa. Por outro lado, os resultados obtidos por Tetlock (2007), utilizando o dicionário *Harvard-IV*, apontaram para uma correlação entre o conteúdo das notícias e o comportamento do mercado, concluindo que o alto pessimismo nas notícias estava relacionado a uma queda no preço das ações.

A variável risco (CDS_Brasil) embora tenha o *p-value* (0.063) ligeiramente superior a 0.05, mas inferiores a 0.1, pode ser considerada estatisticamente como moderadamente significativa. O coeficiente (-0.014), sendo negativo, indica que o aumento do risco está relacionado a uma queda do Ibovespa. Este comportamento faz sentido na medida que os investidores tendem a ter atitudes conservadoras e de receio na presença de incertezas. Azevedo Neto (2019) também identificou uma relação negativa e estatisticamente significativa entre o CDS e o Ibovespa.

A variável câmbio (USD_BRL) foi estatisticamente significativa, na medida em que o *p-value* (0.000) foi inferior ao nível de significância. Deste modo, há evidências de que a taxa de câmbio Dólar/Real contribui para explicar a variação do Ibovespa. O coeficiente (-0.64), sendo negativo, indica que a valorização do Dólar e a conseqüente desvalorização do Real dá origem à queda na rentabilidade do Ibovespa. Este comportamento será o esperado, na medida em que, com a valorização do dólar, os investidores procuram ativos mais seguros, reduzindo a sua exposição a mercados emergentes, como o Brasil, refletindo assim uma queda no valor das ações. Logo, na presença de incerteza, aumenta-se a demanda por moedas mais estáveis, como o dólar, provocando uma desvalorização da moeda local e refletindo-se numa taxa de câmbio mais alta.

Pereira e Arevalo (2024) também encontram uma relação negativa e estatisticamente significativa entre o Ibovespa e a taxa de câmbio, sem considerar as defasagens. Os autores argumentam que este comportamento é esperado e perceptível, uma vez que o crescimento da taxa de câmbio denota escassez de moeda estrangeira pela saída de capitais. Marschner et al. (2021) e Paraboni et al. (2016) identificaram uma relação negativa e estatisticamente significativa entre a taxa de câmbio e o sentimento mensurado pelo Índice de Confiança do Consumidor (ICC). Para o sentimento mensurado pelo Índice de Confiança Empresarial (ICE), Marschner et al. (2021) não obtiveram resultados significativos. Sing (2019) identificou uma relação negativa, mas não significativa, entre câmbio (USD/BRL) e os índices de mercado brasileiro (MSCI Brasil) e americano (S&P500).

Complementarmente, Sing (2019) e Azevedo Neto (2019) identificaram uma relação estatisticamente significativa e positiva entre o CDS Brasil e o câmbio (USD/BRL), sendo a variável CDS dependente na regressão. Sing (2019) concluiu ainda que, de um modo geral, parte do risco de crédito dos países é influenciado por variáveis financeiras globais, com destaque para o índice acionista americano S&P 500, enquanto as variáveis locais (câmbio e mercado acionista) apresentam maior capacidade de explicar o comportamento dos CDS soberanos. De modo similar, Azevedo Neto (2019) concluiu que a maioria dos países estudados é mais sensível a fatores locais (ex.: taxa de câmbio), seguido por fatores globais (ex.: S&P500) e por último fator regional (ex. Ibovespa).

A variável Bitcoin foi estatisticamente significativa, com um *p-value* (0.001) inferior ao nível de significância. Os resultados evidenciam que a variação do Bitcoin contribui para explicar a variação do Ibovespa. O coeficiente (0.034), sendo positivo, indica que a valorização do Bitcoin no mercado está associada a valorização da rentabilidade do Ibovespa. Este comportamento pode ser explicado pelo comportamento otimista do mercado, em cenários económicos favoráveis, refletido na procura por ativos de risco pelos investidores. Karameshinova (2023), à luz da metodologia aplicada na pesquisa, não identificou resultados estatisticamente significativos entre o Bitcoin, enquanto variável dependente, e o índice de mercado S&P500.

A variável ouro não foi estatisticamente significativa, com o *p-value* (0.195) superior ao nível de significância. Logo, a hipótese nula de que a variação do ouro não influencia o comportamento do Ibovespa não pode ser rejeitada. Contudo, Pereira e Arevalo (2024) encontraram uma relação significativa e positiva entre o Ibovespa e o Ouro (LnOuro), sem considerar as defasagens. Nesta perspectiva, os autores argumentam que o investimento em ouro não oferece benefícios de *hedge* (cobertura) e diversificação para investidores. Complementarmente, Karameshinova (2023) identificou que a variável ouro, de entre as selecionadas no estudo, foi a segunda mais representativa no impacto ao Bitcoin, com influência positiva.

Concluída a análise dos resultados do modelo de regressão linear múltipla apresentados na tabela 7, e procurando melhorar o ajuste do modelo, realizou-se uma nova estimação com a primeira diferença da variável do sentimento do investidor, mensurada pelo léxico *Harvard-IV* (sent_HIV), que não tinha revelado significância na variável original e, assim, não foram inicialmente detetados como estatisticamente significativos. A primeira diferença refere-se a uma transformação em série temporal tomando a diferença de valores de períodos adjacentes

para determinada variável, onde o valor do período anterior (X_{t-1}) é subtraído valor do período posterior (X_t) (Wooldridge, 2013). Os resultados da nova regressão linear múltipla podem ser verificados na Tabela 8 abaixo:

Tabela 8 - Regressão Linear Múltipla – 1º Diferença para o léxico *Harvard-IV* (sent_HIV)

<i>Linear regression</i>		<i>Number of obs =</i>		538		
		<i>F(8, 529) =</i>		32.98		
		<i>Prob > F =</i>		0.0000		
		<i>R-squared =</i>		0.3712		
		<i>Root MSE =</i>		.00864		
<i>ln_ibov</i>	<i>Coefficient</i>	<i>Robust std. err.</i>	<i>t</i>	<i>P>t</i>	<i>[95% conf. interval]</i>	
qtd_news	-.0001265	.000098	-1.29	0.197	-.000319	.0000659
sent_VADER	.0051236	.0019789	2.59	0.010	.0012363	.009011
sent_LM	.0094587	.0031417	3.01	0.003	.0032869	.0156305
sent_HIV D1.	.0066181	.0031509	2.10	0.036	.0004283	.012808
Bitcoin	.0452514	.0139452	3.24	0.001	.0178566	.0726463
Ouro	-.0618394	.0428264	-1.44	0.149	-.1459701	.0222913
USD_BRL	-.5719692	.0512493	-11.16	0.000	-.6726464	-.471292
CDS_Brasil	-.0082856	.0079744	-1.04	0.299	-.0239509	.0073797
cons	.0025399	.0020947	1.21	0.226	-.0015752	.0066549

Neste cenário, a variável sentimento do investidor (sent_HIV D1.), mensurado a partir do dicionário léxico *Harvard-IV*, passa a ser estatisticamente significativa, com *p-value* de 0.036. Verifica-se, assim, que uma mudança no sentimento medido por esta variável ainda que não conotada diretamente com a área financeira (por exemplo, aumento no otimismo geral), produz um impacto positivo no retorno acionista. Desde que o panorama geral da economia mude favoravelmente, este indicador capta o otimismo e transmite-o, ainda que lentamente, aos investidores o que, por sua vez, provoca um aumento do retorno do Ibovespa. Todas as restantes variáveis explicativas não alteram o seu poder explicativo, se comparados com o primeiro modelo regressivo (Tabela 7), à exceção da variável que representa o “risco Brasil” (CDS_Brasil) que perdeu a sua relevância. A capacidade explicativa do modelo também reduziu para 37,12%, face à percentagem do modelo anterior de 41,86%.

Conclusão

O presente estudo analisou a relação entre o sentimento do investidor e o comportamento do mercado acionista brasileiro, representado pelo índice Ibovespa, partindo da hipótese de que o tom (conteúdo) implícito nos textos divulgados pelos media especializada tem impacto no comportamento dos investidores.

Assim, foi analisado o sentimento expresso no conteúdo integral das notícias do serviço “Valor +News” do jornal *Valor Econômico*, por meio da abordagem léxica, com o recurso a 3 (três) dicionários utilizados na literatura (*Harvard IV*, *Loughran e McDonald* e *VADER*). Os códigos foram escritos na linguagem *Python*, na plataforma *Google Colab*. Os dados foram modelados por meio da regressão linear múltipla, e complementarmente, realizada uma nova regressão com a primeira diferença da variável *Harvard IV*, na procura de significância estatística. Os testes de robustez e validade das inferências do modelo foram efetuados.

De um modo geral, a presente pesquisa forneceu evidências da relevância informacional das notícias divulgadas nos media especializados do Brasil. O sentimento do investidor, mensurado pelos dicionários *Loughran e McDonald* e *VADER*, bem como *Harvard IV*, na primeira diferença, foram estatisticamente significativos na explicação do comportamento do mercado, representado pelo índice Ibovespa, revelando uma relação positiva. Ou seja, o sentimento otimista do investidor está relacionado com a valorização das ações das empresas no mercado.

A variável quantidade de notícias não foi estatisticamente significativa na explicação do comportamento do mercado, sugerindo que o mercado avalia com maior intensidade a qualidade (conteúdo informacional) do que a quantidade de notícias divulgadas. É importante destacar que este estudo foi desenvolvido com acesso ao conteúdo integral das notícias, o que permite uma análise mais detalhada e robusta da qualidade das informações. Este resultado, alinhado com a significância estatística das variáveis de sentimento, sugere a importância de fontes credíveis de informações nas decisões dos investidores.

Os resultados indicaram também uma tendência de absorção imediata das notícias divulgadas nos media. Este facto pode revelar-se um fator positivo para os investidores no que diz respeito ao acompanhamento das tendências do mercado, constituindo um importante instrumento de auxílio no processo de tomada de decisão e no planeamento de investimentos.

As demais variáveis do modelo, nomeadamente taxa de câmbio, *Bitcoin* e risco Brasil revelaram-se estatisticamente significativas. Assim, um aumento do risco e uma valorização do

câmbio dão origem a uma queda do Ibovespa, enquanto que, uma valorização do *Bitcoin* impacta positivamente o Ibovespa. A variável ouro não apresentou significância estatística.

A presente investigação apresenta limitações no que diz respeito ao período estudado, embora seja este o horizonte temporal para o qual há informação disponível na base de dados selecionada para análise. Foram analisadas apenas notícias do tópicos “Finanças”, do serviço “Valor+News”, não sendo avaliados, por exemplo, notícias do caderno Política ou Mundo. Além disso, é de considerar a evolução de questões específicas relacionadas aos aspetos gramaticais da língua, implicando, por consequência, em avanços constantes nas técnicas de análise de sentimento.

Para estudos futuros, recomenda-se a unificação dos diversos dicionários já utilizados na literatura, incluindo tanto os tradicionais utilizados em pesquisas na área de finanças quanto aqueles propostos pelos autores em trabalhos anteriores. Essa consolidação, após a eliminação de palavras redundantes, permitiria a criação de um dicionário único que reúna todas as palavras previamente mapeadas. Dessa forma, seria possível melhorar a capacidade explicativa do modelo para identificar o tom (conteúdo) dos textos e, conseqüentemente, analisar de modo mais preciso o sentimento do investidor.

Técnicas de *machine learning* podem ser implementadas para realizar a leitura em tempo real de notícias divulgadas em diferentes media relevantes. Além disso, é possível investigar o intervalo de tempo (em minutos, horas, etc.) entre a publicação das notícias e o impacto de seu conteúdo informacional no mercado.

Embora haja uma tendência crescente de publicação de trabalhos relacionadas com a análise de sentimento, seja através da abordagem léxica – com a reutilização ou criação de novos dicionários, ou da abordagem baseada em aprendizagem de máquina – algoritmos relacionados à inteligência artificial, observa-se uma ampla gama de possibilidades para combinar técnicas de análise e abstrair a complexidade dos textos. Esse cenário ressalta a importância de pesquisas futuras que investiguem os efeitos da racionalidade limitada dos investidores no mercado financeiro.

Referências bibliográficas

- Akinwande, M. O., Dikko, H. G., & Samson, A. (2015). Variance inflation factor: as a condition for the inclusion of suppressor variable (s) in regression analysis. *Open journal of statistics*, 5(7), 754. <https://doi.org/10.4236/ojs.2015.57075>
- Azevedo Neto, G. A. D. (2019). Fundamentos macroeconômicos do credit default swaps nos mercados emergentes [Dissertação de Mestrado, Fundação Getúlio Vargas]. Repositório da Fundação Getúlio Vargas. <https://hdl.handle.net/10438/28084>
- Azevedo, V. (2024, junho 26). *Risco Brasil flerta com maior patamar em 8 meses, com fiscal e falas de Lula*. InfoMoney. <https://www.infomoney.com.br/mercados/risco-brasil-flerta-com-maior-patamar-em-8-meses-com-fiscal-e-falas-de-lula/>
- Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, 61(4), 1645-1680. <https://doi.org/10.1007/s11156-018-0756-z>
- Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2), 129-151. <https://doi.org/10.1257/jep.21.2.129>
- Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial economics*, 9(1), 3-18. [https://doi.org/10.1016/0304-405X\(81\)90018-0](https://doi.org/10.1016/0304-405X(81)90018-0)
- Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3), 307-343. [https://doi.org/10.1016/S0304-405X\(98\)00027-0](https://doi.org/10.1016/S0304-405X(98)00027-0)
- Benevides, G. (2024). *Risco-país do Brasil tem a 2ª maior alta do G20 em 2024*. Poder360. <https://www.poder360.com.br/economia/risco-pais-do-brasil-tem-a-2a-maior-alta-do-g20-em-2024/>
- Black, F., & Scholes, M. (1972). The valuation of option contracts and a test of market efficiency. *The journal of finance*, 27(2), 399-417. <https://doi.org/10.2307/2978484>
- Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, 17(31), 334-355. <https://doi.org/10.1111/j.1467-8454.1978.tb00635.x>
- Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47(5), 1287-1294. <https://doi.org/10.2307/1911963>
- Carosia, A. E. O., Coelho, G. P., & Silva, A. E. A. (2019). Analyzing the Brazilian Financial Market through Portuguese Sentiment Analysis in Social Media. *Applied Artificial Intelligence*, 34(1), 1-19. <https://doi.org/10.1080/08839514.2019.1673037>
- Carvalho, R. S. de., Tonin, J. M. da F., & Sanches, S. L. R. (2023). Previsibilidade do Retorno das Ações pela Análise do Sentimento Textual: uma revisão. *Revista De Educação E Pesquisa Em Contabilidade (REPeC)*, 17(2). <https://doi.org/10.17524/repec.v17i2.3111>

- Castro, P. A. C. (2023). Análise de sentimento dos tweets e sua influência sobre o preço do bitcoin [Dissertação de Mestrado, Universidade Federal do Ceará]. Repositório da Universidade Federal do Ceará. <http://repositorio.ufc.br/handle/riufc/75404>
- Colombo, M. E. , Varela, U. do N., & Sabino, V. J. B. (2023). Jornalismo e internet: evolução e perspectivas dos processos de circulação de notícias. *Lumina*, v. 17, n. 2, 108–124. <https://doi.org/10.34019/1981-4070.2023.v17.39413>
- Costa, R. R. A. (2021). *Notícias financeiras e o custo da informação: a relação entre complexidade textual e a difusão de informação no mercado financeiro* [Tese de Doutorado, Universidade Federal do Rio Grande do Norte]. Repositório da Universidade Federal do Rio Grande do Norte. <https://repositorio.ufrn.br/handle/123456789/48506>
- Dawid, P. E. (2004). Estacionariedade em séries temporais com quebras estruturais [Dissertação de Mestrado, Universidade de São Paulo]. Instituto de Matemática e Estatística. <https://doi.org/10.11606/D.45.2004.tde-20210729-134638>
- DeRobertis, N. (n.d.). *PySentiment* [Software]. Acessado em 8 de outubro de 2024 de <https://nickderobertis.github.io/pysentiment/index.html>
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74(366a), 427–431. <https://doi.org/10.1080/01621459.1979.10482531>
- Duarte, C. (2023). Sentimento do investidor no mercado de ações brasileiro [Dissertação de Mestrado, Universidade de São Paulo]. Repositório da Universidade de São Paulo. <https://doi.org/10.11606/D.96.2023.tde-12072023-145002>
- Esuli, A., & Sebastiani, F. (2005). Determining the semantic orientation of terms through gloss classification. In *Proceedings of the 14th ACM international conference on Information and knowledge management* (pp. 617-624). <https://doi.org/10.1145/1099554.1099713>
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of finance*, 25(2), 383-417. <https://doi.org/10.2307/2325486>
- Famá, R., Cioffi, P. L. de M., & Coelho, P. A. R. (2008). Contexto das finanças comportamentais: anomalias e eficiência do mercado de capitais brasileiro. *Revista de Gestão USP*, 15(2), 65-78. <https://repositorio.usp.br/item/001710071>
- Faustino, C. C. R. (2013). O impacto das notícias no mercado financeiro brasileiro [Dissertação de Mestrado, Fundação Getúlio Vargas]. Repositório da Fundação Getúlio Vargas. <https://hdl.handle.net/10438/10532>
- Fávero, L. P. (2015). Análise de Dados: Modelos de Regressão com Excel®, Stata® e SPSS®
- Ferguson, N. J., Philip, D., Lam, H., & Guo, J. M. (2015). Media content and stock returns: The predictive power of press. *Multinational Finance Journal*, 19(1), 1-31. <https://ssrn.com/abstract=2611046>

- Ferrari, H. (2024). *Incerteza fiscal empurra risco Brasil com Lula a 147 pontos*. Poder360. <https://www.poder360.com.br/economia/incerteza-fiscal-empurra-risco-brasil-com-lula-a-147-pontos/>
- Ferreira, T. S. V., Machado, M. A. V., & Silva, P. Z. P. (2021). Impacto assimétrico do sentimento do investidor na volatilidade do mercado acionário brasileiro. *Revista de Administração Mackenzie*, 22(4), 1–29. <https://doi.org/10.1590/1678-6971/eRAMF210208>
- Galdi, F. C., & Gonçalves, A. M. (2018). Pessimismo e incerteza das notícias e o comportamento dos investidores no Brasil. *RAE-Revista de Administração de Empresas*, 58 (2), 130-148. <https://doi.org/10.1590/S0034-759020180203>
- Gao, Z., Ren, H., & Zhang, B. (2020). Googling investor sentiment around the world. *Journal of Financial and Quantitative Analysis*, 55(2), 549-580. <https://doi.org/10.1017/S0022109019000061>
- García, D. (2013) Sentiment during recessions. *The Journal of Finance*, v. 68, n. 3, p. 1267-1300. <https://doi.org/10.1111/jofi.12027>
- Godfrey, L. G. (1978). Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables. *Econometrica*, 46(6), 1293-1301. <https://doi.org/10.2307/1913829>
- Haugen, R. A. (2000). *Os Segredos da Bolsa*. São Paulo, Ed. Pearson Educação.
- Haugen, R. A., & Jorion, P. (1996). The January effect: Still there after all these years. *Financial Analysts Journal*, 52(1), 27-31. <https://doi.org/10.2469/faj.v52.n1.1963>
- Hutto, C. J. (n.d.). *vaderSentiment* [Software]. Acessado em 8 de outubro de 2024 de <https://github.com/cjhutto/vaderSentiment>
- Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216-225. <https://doi.org/10.1609/icwsm.v8i1.14550>
- Igarashi, W., Valdevieso, G. S., & Igarashi, D. C. C. (2021). Análise de sentimentos e indicadores técnicos: uma análise da correlação dos preços de ativos com a polaridade de notícias do mercado de ações. *Brazilian Journals of Business*, 3 (1), 470-486. <https://doi.org/10.34140/bjbv3n1-029>
- Indurkha, N., & Damerau, F. J. (2010). *Handbook of natural language processing*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781420085938>
- Jaffe, J., Keim, D. B., & Westerfield, R. (1989). Earnings yields, market values, and stock returns. *The Journal of Finance*, 44(1), 135-148. <https://doi.org/10.1111/j.1540-6261.1989.tb02408.x>

- Jajo, N. K. (2005). A Review of Robust Regression and Diagnostic Procedures in Linear Regression. *Acta Mathematicae Applicatae Sinica, English Series* 21, 209–224. <https://doi.org/10.1007/s10255-005-0230-2>
- Jurafsky, D. & Martin, J. H. (2024). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*, 3rd edition. Online manuscript released August 20, 2024. <https://web.stanford.edu/~jurafsky/slp3>.
- Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, Princeton, 47(2), 263–291. <https://doi.org/10.2307/1914185>
- Kamps, J., Marx, M., Mokken, R. J., & De Rijke, M. (2004). Using WordNet to measure semantic orientations of adjectives. In *Lrec*, 4, 1115-1118.
- Karameshinova, E. A. (2023). *Análise de sensibilidade: Fatores que influenciam o preço das Criptomoedas* [Dissertação de Mestrado, Universidade de Lisboa]. Repositório do Instituto Universitário de Lisboa. <http://hdl.handle.net/10071/30911>
- Kimura, H. (2003). Aspectos comportamentais associados às reações do mercado de capitais. *Revista de Administração de Empresas*, 2 (1). <https://doi.org/10.1590/S1676-56482003000100007>
- Klibanoff, P., Owen, L., & Thierry, W. (1998). Investor Reaction to Salient News in Country Closed-End Funds, *Journal of Finance*, 53, 673-699. <https://www.jstor.org/stable/117365>
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Springer. <https://doi.org/10.1007/978-3-031-02145-9>
- Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In C. Aggarwal & C. Zhai (Eds.) *Mining text data* (pp. 415-463). Boston, USA: Springer. https://doi.org/10.1007/978-1-4614-3223-4_13
- Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35-65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>
- Marschner, P. F., Ceretta, P. S., & Ambrozini, M. A. (2021, julho, 28-30). Impactos dos Sentimentos Racionais e Irracionais dos Investidores sobre o Retorno e a Volatilidade do Mercado Brasileiro. 21º USP International Conference in Accounting, São Paulo. <https://congressosp.fipecafi.org/anais/21UspInternational/ArtigosDownload/3323.pdf>
- Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, v. 7(1), 77-91. <https://doi.org/10.2307/2975974>
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <http://doi.org/10.1016/j.asej.2014.04.011>

- Mello, P. H. S. (2019). O sentimento do investidor no mercado de ações brasileiro [Dissertação de Mestrado, Fundação Getúlio Vargas]. Repositório da Fundação Getúlio Vargas. <https://hdl.handle.net/10438/28827>
- Milanez, D. Y. (2003). Finanças comportamentais no Brasil [Dissertação de Mestrado, Universidade de São Paulo]. Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo. <https://doi.org/10.11606/D.12.2003.tde-09022004-130012>
- Miller, M. H., & Modigliani, F. (1961). Dividend policy, growth, and the valuation of shares. *the Journal of Business*, 34(4), 411-433. <https://www.jstor.org/stable/2351143>
- Miranda, K. F. (2018). Sentimento do investidor e a influência do horizonte de investimento em decisões corporativas: Evidências baseadas na Teoria de Catering [Tese de Doutorado, Universidade Federal da Paraíba, João Pessoa, PB, Brasil]. Repositório Institucional da UFPB. <https://repositorio.ufpb.br/jspui/handle/123456789/15201>
- Miranda, M. C. de. (2023). Efeito dos fatores conjunturais e empresariais no sentimento do investidor: evidências do mercado brasileiro [Dissertação de Mestrado, Universidade Federal Rural do Semi-Árido]. Programa de Pós-Graduação em Administração, Mossoró. <https://repositorio.ufersa.edu.br/handle/prefix/11039>
- Mitra, G., & Mitra, L. (Eds.). (2011). *The handbook of news analytics in finance*. Hoboken, USA: John Wiley & Sons. <https://doi.org/10.1002/9781118467411.fmatter>
- Modigliani, F., & Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *The American economic review*, 48(3), 261-297. <https://www.jstor.org/stable/1809766>
- Mussa, A., Yang, E., Trovão, R., & Famá, R. (2008). Hipótese de mercados eficientes e finanças comportamentais: as discussões persistem. *FACEF PESQUISA*, 11 (1). <http://periodicos.unifacef.com.br/facefpesquisa/article/view/109/173>
- Naik, P. K., & Padhi, P. (2016). Investor Sentiment, Stock Market Returns and Volatility: Evidence from National Stock Exchange of India. *Int. J. Management Practice*, vol 9, no 3, 213-237. <http://dx.doi.org/10.1504/IJMP.2016.077816>
- Nasukawa, T., & Yi, J. (2003). Sentiment Analysis: Capturing Favorability Using Natural Language Processing. *Proceedings of the 2nd International Conference on Knowledge Capture*, Florida, 70-77. <https://doi.org/10.1145/945645.945658>
- Nogueira, B. T. B., Avelino, B. C., Colares, A. C. V., & Reis, D. E. A. dos. (2021). Índice de sentimento do investidor no mercado de ações brasileiro. *Revista de Contabilidade e Controladoria*, 12(3), 46-71. <http://dx.doi.org/10.5380/rcc.v12i3.71338>
- Pagliarussi, M. S., Aguiar, M. O., & Galdi, F. C. (2016). Sentiment analysis in annual reports from Brazilian companies listed at the BM&FBovespa. *BASE-Revista de Administração e Contabilidade da Unisinos*, 13(1), 53-64. <https://revistas.unisinos.br/index.php/base/article/view/base.2016.131.04>

- Paraboni, A. L., dos Santos Fraga, L., Bender Filho, R., Vieira, K. M., & Coronel, D. A. (2016). Impacto de variáveis macroeconômicas e corrupção na Petrobrás no Sentimento de Mercado. *Revista ESPACIOS*, 37(31).
<https://www.revistaespacios.com/a16v37n31/16373120.html>
- Pedro, E. C. (2018). Efeito manada no Brasil: o impacto dos sentimentos dos investidores [Dissertação de Mestrado, Universidade Federal de Santa Catarina]. Centro Socioeconômico, Programa de Pós-Graduação em Economia, Florianópolis.
<https://repositorio.ufsc.br/handle/123456789/198760>
- Pereira, J. R. G., Jr., & Arevalo, J. L. S. (2024). Relations Between The Price Of Gold, Oil, and The Exchange Rate With The IBOVESPA: An Application of The ARDL Model. *Revista de Gestão Social e Ambiental*, 18(10), e08688.
<https://doi.org/10.24857/rgsa.v18n10-011>
- Piccoli, P., da Costa, N. C. A., da Silva, W. V., & Cruz, J. A. W. (2018). Investor sentiment and the risk-return tradeoff in the Brazilian market. *Accounting & Finance*, 58, 599–618. <https://doi.org/10.1111/acfi.12342>
- Reis, J., Gonçalves, P., Araújo, M., Pereira, A., & Benevenuto, F. (2015). Uma Abordagem Multilíngue para Análise de Sentimentos. In *Anais do IV Brazilian Workshop on Social Network Analysis and Mining*, Porto Alegre.
<https://doi.org/10.5753/brasnam.2015.6767>
- Reis, P. M. N., & Pinho, C. (2020a). COVID-19 and investor sentiment influence on the US and European countries sector returns. *Investment Management and Financial Innovations*, v. 17, n. 3, 373-386. <http://dx.doi.org/10.2139/ssrn.3647577>
- Reis, P. M. N., & Pinho, C. (2020b). A new European investor sentiment index (EURsent) and its return and volatility predictability. *Journal of Behavioral and Experimental Finance*, v. 27. <https://doi.org/10.1016/j.jbef.2020.100373>
- Reis, P. M. N., & Pinho, C. (2020c). A reappraisal of the causal relationship between sentiment proxies and stock returns. *Journal of Behavioral Finance*, 22(4), 420-442.
<https://doi.org/10.1080/15427560.2020.1792910>
- Ribeiro, M. A. S. (2022). O Impacto do sentimento do investidor na cross-section de retornos de ações brasileiras: uma análise da iliquidez como limite à arbitragem. [Trabalho de Conclusão de Curso (Bacharelado em Ciências Econômicas)]. Universidade de Brasília, Faculdade de Economia, Administração e Contabilidade.
<https://bdm.unb.br/handle/10483/33166>
- Santana, H. N. (2018). Eleições presidenciais brasileiras e a volatilidade do IBOVESPA: relações com variáveis conjunturais e risco político [Dissertação de Mestrado, Universidade Federal de Minas Gerais]. Centro de Pós-Graduação e Pesquisas em Administração da Faculdade de Ciências Econômica. <http://hdl.handle.net/1843/30298>
- Santos, L. da S. (2022). Aplicações de técnicas de análise de sentimentos às notícias econômicas e sua relação com variáveis macroeconômicas. [Trabalho de Conclusão de

- Curso (Graduação em Ciências Econômicas]. Faculdade de Economia, Universidade Federal Fluminense, Niterói. <http://app.uff.br/riuff/handle/1/32843>
- Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2020). Measuring news sentiment. *Journal of Econometrics*, 228, 221–243. <http://www.frbsf.org/economic-research/publications/working-papers/2017/01/>
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3), 425-442. <https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>
- Schatteman, A. M., & Waymire, T. R. (2017). The state of nonprofit finance research across disciplines. *Nonprofit management and leadership*, 28(1), 125-137. <https://doi.org/10.1002/nml.21269>
- Shefrin, H. (2015). The behavioral paradigm shift. *Revista de Administração de Empresas*, 55(1), 95-98. <https://doi.org/10.1590/S0034-759020150109>
- Shiller, R. (2000). *Exuberância Irracional*. São Paulo, Makron Books.
- Shleifer, A. & Summers, L. (1990). The Noise Trader Approach to Finance. *Journal of Economic Perspectives*, vol. 4, n. 2, 19–33. <https://doi.org/10.1257/jep.4.2.19>
- Silva, M. D. de O. P. da. (2017). O efeito do sentimento das notícias sobre o comportamento dos preços no mercado acionário brasileiro. [Tese de Doutorado, Universidade Federal da Paraíba]. Repositório Institucional da UFPB. <https://repositorio.ufpb.br/jspui/handle/123456789/12575>
- Silva, M. D. de O. P. da, & Machado, M. A. V. (2019). Índice de sentimento textual: uma análise empírica do impacto das notícias sobre risco sistemático. *Revista Contemporânea De Contabilidade*, v. 16, n. 40, 24–42. <https://doi.org/10.5007/2175-8069.2019v16n40p24>
- Sing, A. M. K. (2019). Análise dos Credit Default Swap (CDS) de títulos soberanos [Dissertação de Mestrado,]. Insper Instituto de Ensino e Pesquisa, São Paulo. <https://repositorio.insper.edu.br/handle/11224/2604>
- Souza, I. M. O. de, Macedo, A. M., & Lucena, W. G. L. (2023). Finanças comportamentais, anomalias de mercado e bolhas especulativas: Uma análise de notícias. *Teoria E Prática Em Administração*, 13(2). <https://doi.org/10.22478/ufpb.2238-104X.2023v13n2.63112>
- Souza, D. M. S. de, & Martins, O. S. (2020). Sentimento do investidor via twitter e desempenho do mercado acionário brasileiro. XIV Congresso ANPCONT, Foz do Iguaçu.
- Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). *General Inquirer*. Cambridge, MA: MIT Press.

- Tahmasebi, N., Borin, L., Jatowt, A., Xu, Y., & Hengchen, S. (2021). *Computational approaches to semantic change*. Language Science Press. <https://doi.org/10.5281/zenodo.5040241>
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3), 1139-1168. <https://doi.org/10.1111/j.1540-6261.2007.01232.x>
- Tetlock, P. C. (2014). Information transmission in finance. *Annu. Rev. Financ. Econ.*, vol 6, n. 1, 365-384. <https://doi.org/10.1146/annurev-financial-110613-034449>
- Terblanche, M., & Marivate, V. (2021). Loughran McDonald-SA-2020 Sentiment Word List. <https://doi.org/10.25403/UPresearchdata.14401178.v1>.
- Turney, P. D., & Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *acm Transactions on Information Systems (tois)*, 21(4), 315-346. <https://doi.org/10.1145/944012.94401>
- Valcareggi, S. M. (2022). Análise de sentimentos para o mercado brasileiro [Dissertação de Mestrado, Universidade de São Paulo]. Repositório da Universidade de São Paulo. <https://doi.org/10.11606/D.96.2022.tde-24012023-165235>
- Valor Econômico. (n.d.). Venda de Conteúdo. Acessado em 15 de setembro de 2024 de <https://valor.globo.com/venda-de-conteudo/>
- Vartanian, P. R., Jr., Moura, A. A., Racy, J. C., & Simioni, R. (2022). Preços dos non-fungible tokens (NFTS), criptomoedas, taxa de juros e ouro: uma análise econométrica (jan. De 2019 – ago. De 2022). In: *Anais do Fórum Mackenzie de Liberdade Econômica*, São Paulo, Universidade Presbiteriana Mackenzie. [https://www.even3.com.br/anais/libecon2022/564426-precos-dos-non-fungibletokens-\(nfts\)-criptomoedas-taxa-de-juros-e-ouro--uma-analise-econometrica-\(jan-de-2019/](https://www.even3.com.br/anais/libecon2022/564426-precos-dos-non-fungibletokens-(nfts)-criptomoedas-taxa-de-juros-e-ouro--uma-analise-econometrica-(jan-de-2019/)
- Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). Boston, MA: South-Western Cengage Learning.
- Yoshinaga, C. E. (2009). A relação entre índice de sentimento de mercado e as taxas de retorno das ações: uma análise com dados em painel. [Tese de Doutorado, Universidade de São Paulo]. Faculdade de Economia, Administração e Contabilidade, São Paulo. <https://doi.org/10.11606/T.12.2009.tde-17122009-105238>
- Yu, C., & Yao, W. (2017). Robust linear regression: A review and comparison. *Communications in Statistics-Simulation and Computation*, 46(8), 6261-6282. <https://doi.org/10.1080/03610918.2016.1202271>

Apêndices

Apêndice A – Código Python

```
!pip install deep-translator

#Lendo o arquivo JSON das notícias e criando um novo arquivo .csv estruturado.

import json

def import_json():

    file_path = r"xxxxxxxxxx.json"
    with open(file_path, encoding="utf8") as file:
        data = json.load(file)
    arquivo = open("xxxxxxxxxx.csv", "w", encoding='UTF-8')

    for i in data:
        arquivo.write(i['issued']+i['_id']['$oid']+i['title']+i['bodyData']['blocks'])
        a = i['bodyData']['blocks']
        b = ""
        for j in a:
            if(j['text']!= ""):
                b = b + j['text'] + '!'
            textoSemQuebra = b.replace("\n", "")
            arquivo.write(textoSemQuebra+'\n')
        arquivo.close()

import_json()

# Criando um novo arquivo com as notícias traduzidas para o inglês.
import csv
from deep_translator import GoogleTranslator

file_path = r"xxxxxxxxxx.csv"
tradutor = GoogleTranslator(source="portuguese", target="english")
arquivo = open("xxxxxxxxxx.csv", "w", encoding='UTF-8')

with open(file_path, encoding="utf8", mode='r') as file:
    leitor_csv = csv.reader(file, delimiter='|')
    a = 0
    for linha in leitor_csv:
        if(len(linha[3])<=5000 ): # Verifica se o texto tem no máximo 5.000 caracteres, que é o
limite para tradução.
            arquivo.write(linha[1] + '|')
            traducao = tradutor.translate(linha[3])
            arquivo.write(traducao + '\n')
            a = a +1
arquivo.close()
!pip install vaderSentiment
```

```

#Dicionário Vader Sentiment
import csv
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

file_path = r"xxxxxxxxxx.csv"
analyzer = SentimentIntensityAnalyzer()
arquivo = open("xxxxxxxxxx.csv", "w", encoding='UTF-8')
with open(file_path, encoding="utf-8", mode='r') as file:
    leitor_csv = csv.reader(file, delimiter='|')
    for linha in leitor_csv:
        arquivo.write(linha[0] + '|')
        vs = analyzer.polarity_scores(linha[1])
        arquivo.write(str(vs) + '|' + str(vs['compound']) + '\n')

arquivo.close()
!pip install pysentiment2

#Dicionário Loughran and McDonald
import pysentiment2 as ps

file_path = r"xxxxxxxxxx.csv"
arquivo = open("xxxxxxxxxx.csv", "w", encoding='UTF-8')
lm = ps.LM()
with open(file_path, encoding="utf8", mode='r') as file:
    leitor_csv = csv.reader(file, delimiter='|')

    for linha in leitor_csv:
        arquivo.write(linha[0] + '|')
        tokensLm = lm.tokenize(linha[1])
        scoreLm = lm.get_score(tokensLm)
        arquivo.write(str(scoreLm) + '|' + str(scoreLm['Polarity']) + '\n')

arquivo.close()

#Dicionário Harvard
import pysentiment2 as ps

file_path = r"xxxxxxxxxx.csv"
arquivo = open("xxxxxxxxxx.csv", "w", encoding='UTF-8')
hiv4 = ps.HIV4()
with open(file_path, encoding="utf8", mode='r') as file:
    leitor_csv = csv.reader(file, delimiter='|')

    for linha in leitor_csv:
        arquivo.write(linha[0] + '|')
        tokensHiv4 = hiv4.tokenize(linha[1])
        scoreHiv4 = hiv4.get_score(tokensHiv4)
        arquivo.write(str(scoreHiv4) + '|' + str(scoreHiv4['Polarity']) + '\n')

arquivo.close()

```