

Análise de Sentimento no *Twitter* e sua Influência no Mercado Financeiro: Um Estudo de Caso do Índice IBOVESPA

Elvira Helena Oliveira de Medeiros* Diego Pitta de Jesus†
Lucas Lucio Godeiro‡ Andressa Lemes Proque§

Resumo

Este artigo tem o objetivo de verificar o potencial de dados do *twitter* na previsão dos retornos do índice IBOVESPA. Assim foi construído um índice de sentimento de *tweets* que mencionam o IBOVESPA para captar a polaridade de tais publicações. O desempenho do índice de sentimento foi comparado com estratégias provenientes da Análise Técnica e com o retorno histórico do índice representativo do mercado acionário brasileiro (*benchmark*). Além disso, foi também calculado o equivalente certeza (ganho de utilidade) de um investidor individual de média variância associado a cada modelo de previsão. Os dados são de frequência diária, iniciando em 02 janeiro de 2007 a 31 dezembro de 2022. Os resultados mostraram que o modelo baseado no índice de sentimento em *twitter* supera a maioria dos modelos que dependem de indicadores técnicos e o modelo de referência. Outro resultado bem surpreendente é que os preditores baseados no indicador OBV tiveram um desempenho superior aos demais indicadores técnicos, bem como ao preditor baseado nas informações do *twitter*.

Palavras-chave: Previsão do IBOVESPA; Análise de Sentimentos; *twitter*; Indicadores Técnicos.

Abstract

This paper aims to verify the potential of twitter data in predicting the returns of the IBOVESPA index. Thus, a sentiment index of tweets that mention the IBOVESPA was built to capture the polarity of such publications. The performance of the sentiment index was compared with strategies derived from Technical Analysis and with the historical return of the representative index of the Brazilian stock market benchmark). In addition, the certainty equivalent (utility gain) of an individual average variance investor associated with each forecast model was also calculated. Data are of daily frequency, starting from January 2, 2007 to December 31, 2022. The results showed that the model based on the sentiment index in twitter outperforms most models that rely on technical indicators and the model of reference. Another very surprising result is that the predictors based on the OBV indicator outperformed the other technical indicators, as well as the predictor based on information from twitter.

Keywords: IBOVESPA forecast; Sentiment Analysis; Twitter; Technical Indicators.

JEL Classification: C01, C22.

Área Temática - Econometrics and Numerical Methods

*Doutoranda em Economia Aplicada - UFJF e Mestre em Economia Aplicada - UERN ravilelenna@yahoo.com.br - Financiada por meio de bolsa CAPES.

†Doutor em Economia Aplicada - UFPB e Professor Substituto da UFPB diegopitta13@hotmail.com

‡Doutor em Economia Aplicada - UFPB e Professor da Universidade Federal Ruraldo Semi Árido - UFERSA lucasgodeiro@ufersa.edu.br

§Doutora em Economia Aplicada - UFJF e Professora na Universidade Federal de São João del-Rei - UFSJ e PhD in Economics, Federal University of Juiz de Fora (UFJF) alesproque@gmail.com

1 Introdução

A previsão precisa dos retornos do mercado de ações não apenas ajuda os investidores do mercado a ajustar suas carteiras no tempo, mas também permite que os tomadores de decisão entendam os riscos do mercado e formulem políticas correspondentes de acordo com a situação do mercado. De acordo com [Rapach e Zhou \(2013\)](#) do ponto de vista dos profissionais de finanças, a alocação de ativos requer previsões em tempo real dos retornos das ações, e previsões aprimoradas de retorno das ações prometem melhorar o desempenho do investimento. Portanto, não surpreende que os profissionais de finanças empreguem uma infinidade de variáveis na tentativa de prever o retorno das ações. Acadêmicos em finanças também estão profundamente interessados em previsões de retorno de ações, uma vez que a capacidade de prever retornos tem implicações importantes para testes de eficiência de mercado; de maneira mais geral, entender a natureza da previsibilidade do retorno das ações ajuda os pesquisadores a produzir modelos de precificação de ativos mais realistas que explicam melhor os dados.

Esse tema tem sido objeto de muitos estudos na literatura financeira. Dentre os principais trabalhos, destacam-se aqueles que propuseram modelos e técnicas para prever os retornos do mercado de ações, como o modelo de [Black e Scholes \(1973\)](#) para precificação de opções, a metodologia ARIMA proposta por [Box e Jenkins \(1976\)](#) para previsão de séries temporais, o modelo de três fatores proposto por [Fama e French \(1993\)](#) para explicar os retornos das ações, e a estratégia de momentum proposta por [Jegadeesh \(1992\)](#). Além disso, estudos mais recentes têm explorado abordagens mais sofisticadas, como modelos de aprendizado de máquina, redes neurais e inteligência artificial, para prever os retornos do mercado de ações com maior precisão.

A maioria desses estudos utiliza dados tradicionais, como preços de ações e indicadores macroeconômicos. Entretanto, nos últimos anos, com o aperfeiçoamento da Linguagem de Processamento Natural (LPN), surgiram diversas técnicas voltadas para a análise de informações não numéricas, possibilitando, assim, a otimização das investigações empíricas que utilizam variáveis qualitativas. A LPN tem favorecido a realização de estudos que fazem uso de medidas qualitativas, como as pesquisas que tratam sobre narrativas ([Beattie \(2014\)](#); [Fisher, Garnsey e Hughes \(2016\)](#)). Assim alguns estudos começaram a explorar o potencial dos tweets e das notícias como fontes de dados para a previsão dos retornos de mercado (ver os trabalhos de Vários estudos relacionados sugerem que o sentimento do investidor afeta significativamente o mercado financeiro (ver os trabalhos de [Brown \(1999\)](#); [Baker e Wurgler \(2006\)](#); [Garcia \(2013\)](#); [Huang, Zang e Zheng \(2014\)](#); [Han e Li \(2017\)](#); [Jiang et al. \(2019\)](#); [Liang et al. \(2020\)](#); [Shapiro, Wilson et al. \(2017\)](#); [Huynh et al. \(2021\)](#); [Obaid e Pukthuanthong \(2022\)](#))).

No âmbito financeiro, parte das informações qualitativas pode ser descrita como informações linguísticas, expressas mediante textos ou por meio do discurso. No que diz respeito às informações textuais, as principais características analisadas nos estudos, geralmente, são: (a) legibilidade, definida como a complexidade sintática da comunicação, ou seja, nível de compreensão do texto (([Lehavy, Li e Merkley, 2011](#)); ([Beattie, 2014](#))); e (b) tom ou sentimento textual que faz referência à semântica do processo de comunicação, quanto à interpretação da informação, podendo esta apresentar um sentido otimista ou pessimista, conforme as palavras empregadas no texto (([Beattie, 2014](#)); ([Huang, Zang e Zheng, 2014](#))).

Tais dados são usados para a criação de indicadores que possam ajudar na melhoria da acurácia da previsão dos retornos. O objetivo dessa abordagem é extrair informações úteis sobre o comportamento dos investidores a partir de textos não estruturados, como notícias e posts em redes sociais, e utilizá-las como indicadores de tendência para prever os movimentos futuros do mercado. Os indicadores de sentimento são gerados a partir da análise de sentimentos expressos em textos. Essa análise é realizada utilizando técnicas de processamento de linguagem natural, que permitem identificar e classificar as emoções expressas em um texto, tais como positividade, negatividade e neutralidade.

Desse modo, muitas informações textuais têm sido o foco das investigações empíricas, tais como:

jornais (ver os trabalhos de Tetlock (2010); Griffin, Hirschey e Kelly (2011); Dougal et al. (2012); Garcia (2013); Liu, McConnell e Xu (2017); Hendershott, Livdan e Schürhoff (2015)); relatórios da administração ou relatórios específicos (Li (2008); Loughran e McDonald (2011); Huang, Zang e Zheng (2014)); mensagens de chat, *twitter*, Facebook e mensagens de spam relacionadas aos ativos (Antweiler e Frank (2004); Bollen, Mao e Zeng (2011); Hu, McInish e Zeng (2010); Karabulut (2013)); e previsão dos analistas (Twedt e Rees (2012)).

A literatura que trata sobre o conteúdo informacional das notícias tem investigado se o mercado acionário reage ao volume e ao sentimento das notícias. Para a realização dessas investigações, são utilizadas várias hipóteses, que fazem alusão à suposição de que os investidores seriam sensíveis ao conteúdo informacional das notícias (Garcia (2013); Liu e McConnell (2013)). Evidências Empíricas mostram indícios da sensibilidade dos investidores às notícias, ao fazerem inferências relacionadas à dinâmica do mercado, constatando que o sentimento das notícias da mídia tem valor preditivo sobre os retornos das ações, na volatilidade e sobre os volumes de negociações (Tetlock (2010); Tetlock, Saar-Tsechansky e Macskassy (2008); Garcia (2013); Liu (2014); Ferguson et al. (2015)).

De acordo com Phan, Nguyen e Hwang (2022) a partir do desenvolvimento da mídia social online apresentada na forma de conteúdo de texto, a forma como os investidores, produtores e consumidores entendem as tendências do mercado mudou tremendamente. A análise de texto tornou-se um novo método para atender aos enormes dados gerados em diferentes serviços de mídia social (Hu e Liu (2012)). Em particular, o compartilhamento de conteúdo pelos usuários nas redes sociais carrega pontos de vista e sentimentos, que estão em alta demanda para extração e captura no campo da pesquisa ativa

Essa literatura é relativamente recente no Brasil, mas já existem alguns trabalhos importantes nessa linha de pesquisa. Por exemplo, um estudo realizado por de Passos et al. (2019) analisou o impacto da análise de sentimentos em tweets no mercado de ações brasileiro, e descobriu que as informações contidas nos tweets têm poder preditivo para os retornos das ações. Além disso, outro estudo de Junqueira e Fernandes (2018) propôs um modelo para previsão de retornos do mercado de ações brasileiro utilizando técnicas de análise de sentimento em notícias, e verificou-se que essa abordagem é capaz de superar os modelos tradicionais de previsão de séries temporais.

Nesse contexto, o artigo pretende verificar e explorar o potencial de tweets como indicadores de previsão dos retornos do índice IBOVESPA. Assim, o objetivo do artigo é construir um índice de sentimento a partir de dados do *twitter* e testar o seu poder de predição dos retornos do índice IBOVESPA. O desempenho do índice de sentimento será comparado com estratégias provenientes da análise técnica e com o retorno histórico do índice IBOVESPA. Além disso, também pretende-se calcular o equivalente certeza (ganho de utilidade) de um investidor individual de média variância associado a cada modelo de previsão. Os dados são de frequência diária, iniciando em 02 janeiro de 2007 a 31 dezembro de 2022. Sendo assim, um dos diferenciais do presente artigo é o uso de alta frequência nos dados.

Além desta introdução, o artigo está dividido em mais três seções: a seguir será apresentada de forma detalhada a metodologia. Posteriormente, serão apresentados resultados e discussões e por fim, na última seção são apresentadas as principais conclusões.

2 Revisão da Literatura

A literatura sobre a previsão do retorno de ações e índices acionários é vasta, de modo que existem uma variedade de métodos e de preditores. Serão apresentados trabalhos da literatura de previsão de retornos de ações e de índices de ações que usaram as notícias de jornais e os tweets para prever os retornos.

Uma das primeiras pesquisas a discorrer sobre a interação entre o sentimento textual no mercado acionário, utilizando análise de conteúdo automatizada, foi o estudo de Tetlock (2007), no qual foi analisada a coluna financeira *Abreast of the Market* presente no *Wall Street Journal*, no período de

1984 a 1999. Segundo o autor, a coluna financeira estaria dentre as mais populares dos Estados Unidos, visto que nela seriam encontradas informações sobre as movimentações do mercado, assim como as previsões dos analistas e o posicionamento de corretoras, explicando o que ocorreu no mercado e projetando comportamentos futuros. No estudo, investigou-se a correlação entre o conteúdo da coluna diária do WSJ e as atividades do mercado de ações, com o intuito de verificar se o pessimismo da mídia poderia prever os movimentos no mercado, em termos de volume das negociações e de retornos anormais. Os principais resultados mostraram que níveis elevados de pessimismo na mídia preveem uma pressão sobre os preços de mercado, os quais reagem com quedas, mas que são revertidas rapidamente.

Tetlock, Saar-Tsechansky e Macskassy (2008) analisaram em que medida a linguagem negativa explicava lucros contábeis e o retorno das ações. Assim, analisaram os impactos das informações apresentadas no WSJ e no Dow Jones News Service (DJNS), quanto às notícias sobre as empresas integrantes do índice S&P500, no período de 1980 a 2004. Os autores partiram da suposição de que palavras negativas apresentavam uma relação forte com o retorno dos ativos financeiros, pois elas provocariam um maior impacto nos investidores. Desse modo, o estudo restringiu-se à análise do impacto das palavras negativas utilizadas nos textos das notícias que faziam referências às empresas integrantes da amostra, fazendo uso do General Inquirer's Harvard IV-4 para a classificação e mensuração da frequência das palavras. Os resultados evidenciaram que as palavras negativas utilizadas nas notícias analisadas transmitem informações negativas sobre os ganhos e que poderiam predizer ganhos além do que seria possível captar com a previsão dos analistas e com as medidas tradicionais de avaliação dos preços.

Kothari, Li e Short (2009) estudaram o efeito sistemático do risco com a disseminação de informações financeiras, no contexto do mercado acionário dos Estados Unidos. Para isso, foram adotadas como proxies de risco: o custo do capital, a variação da previsão dos analistas e a volatilidade do retorno. Além disso, realizaram uma análise de conteúdo para classificação 50 das informações favoráveis (otimistas) e desfavoráveis (pessimistas), por meio do General Inquirer's Harvard, em 326.357 documentos textuais, relacionados à: (a) disclosures das empresas (MD&A); (b) textos com previsão dos analistas; e (c) notícias financeiras, partindo da suposição de que as informações textuais podem afetar o mercado, e modo que, as informações desfavoráveis (favoráveis) tendem a aumentar (reduzir) o risco do fluxo de caixa. Foi constatado que tons de otimismo não afetaram significativamente o custo de capital da empresa, sugerindo que divulgações otimistas da gestão não gerariam credibilidade.

Engelberg (2009) analisaram o processo de negociação de derivativos, investigando as vantagens que as informações de domínio público poderiam possibilitar nesse mercado. Os autores partem da suposição de que a vantagem das negociações em derivativos encontrava-se na capacidade de analisar informações publicamente disponíveis. Com isso, é realizado um cruzamento entre dados relacionados às negociações diárias de derivativos na NYSE com notícias específicas sobre as empresas, chamadas de comunicados de imprensa, pertencentes à Dow Jones News Service e ao WSJ, no período de 2005 a 2007, considerando para o tratamento dos textos os dicionários Harvard-IV e L&M. As principais evidências mostraram que existia um aumento nas negociações após a divulgação de notícias, sugerindo que estratégias baseadas em notícias de domínio público poderiam gerar vantagens competitivas.

Garcia (2013) verificou o efeito do sentimento textual das notícias sobre o preço dos ativos, no período de 1905 a 2005, no mercado acionário dos Estados Unidos. O estudo faz inferências ao comportamento dos investidores, com embasamento na literatura de psicologia e economia, sugerindo que os investidores estão mais sensíveis às notícias em períodos de recessão. Assim, investigou o poder informacional do sentimento das notícias no mercado acionário, ao longo dos ciclos econômicos. Foram encontrados padrões de previsibilidade no retorno dos ativos financeiros atrelados ao conteúdo das notícias, principalmente nos períodos de recessão.

Liu e McConnell (2013) investigaram a sensibilidade dos gestores das empresas de capital aberto norte-americanas à mídia pública, nos momentos de decisão de alocação de capital, que foram

acompanhados por uma reação negativa do mercado. Assim, foram analisados 636 anúncios referentes às fusões e às propostas de aquisições, com transações no valor mínimo de US\$100 milhões, entre 01 de janeiro de 1990 a 31 de dezembro de 2010. Os resultados mostraram que as propostas de aquisição estão mais propensas de serem abandonadas quando o mercado reage com preços negativos ao anúncio das negociações.

Liu (2014) investigou a relação existente entre o sentimento de pessimismo das notícias, a concentração/volume de notícias e os spreads dos rendimentos dos *sovereign bond* (títulos do governo), ao longo dos anos de 2009 a 2012, período marcado por uma crise na Europa e por uma ampla divulgação de notícias, fazendo referência a situação de créditos dos países. Nesse contexto, a análise do estudo foi restrita aos seguintes países: (a) Grécia; (b) Irlanda; (c) Itália; (d) Portugal; e (e) Espanha. O autor procurou analisar em que medida o sentimento textual se aplicava aos mercados de dívidas, que apresentavam os títulos de longo prazo, *sovereign bond*. Os resultados das regressões lineares indicaram que, à medida que o pessimismo e o número de notícias aumentavam, os preços tendiam a cair, enquanto, os resultados das regressões logísticas mostraram que a probabilidade de expansão *spreads* dos rendimentos dos títulos aumentavam em 21%, quando o pessimismo da mídia se encontrava acima da média.

Ferguson et al. (2015) examinam o poder informacional da combinação do tom (positivo/negativo) e do volume de notícias específicas sobre as empresas, analisando a capacidade de previsão dos retornos futuros das ações. O estudo restringiu-se à análise das notícias financeiras do Reino Unido, publicadas nos jornais The Financial Times, The Times, The Guardian e Mirror, referentes às empresas integrantes do índice FTSE 100, no período de 1981 a 2010, totalizando uma amostra de 264.647 notícias. Os resultados mostraram indícios das notícias positivas predizerem retornos futuros maiores, enquanto notícias negativas indicavam retornos menores.

Strauß, Vliegthart e Verhoeven (2016) investigaram a emoção apresentada nas matérias de jornais holandeses com grande circulação, analisando os seus efeitos no preço de algumas ações listadas no índice Amsterdam Exchange (AEX), entre os anos de 2002 a 2013, com o objetivo de avaliar o papel dos meios de comunicação no processo de transmissão de emoções ao mercado, por meio de uma análise dos efeitos recíprocos existentes entre o tom dos artigos e o preço de abertura de ações do índice AEX. As evidências sugerem que a emoção positiva das notícias apresenta efeitos no preço de abertura das ações em 12 empresas da amostra, dentre as quais prevaleceu um efeito negativo sobre os preços dos ativos.

Após as eleições de 2008, as mídias sociais, principalmente o *twitter*, tornaram-se uma ferramenta estratégica crucial nas campanhas políticas principalmente pela facilidade de disseminação de informações em todo o mundo. Além disso, é um “lugar livre”, onde os políticos se auto-isolam de fóruns de mídia rigidamente controlados com regras e disciplina, que os obrigam a enviar mensagens com cautela (Newkirk II, 2016). O *twitter* foi um fator chave na campanha eleitoral de Donald Trump e, de fato, dotou-o de mais poder de mídia (Stolee e Caton (2018)). Além disso, a conta do *twitter* de Trump, em contraste com a de qualquer presidente anterior ou agências estatais, conquistou um vasto número de usuários com extensos debates acalorados, o que lhe conferiu o poder de uma pessoa influente (Cornfield (2017)). Uma análise de pesquisa considerou os tweets como eventos acadêmicos que afetaram momentaneamente o mercado, feito diretamente no Stata (Ullah et al. (2021)).

Ao analisar os tweets de Donald Trump, McGranahan (2019) constatou que as pessoas se concentravam nos últimos e mais recentes tweets do que era visto como a “crise” da época, e o nível de sua reação dependia da mensagem que os tweets transmitiam. Simultaneamente, o autor levantou preocupações sobre a “mentira” dos políticos por meio de investigações de contexto em debates na mídia e palavras em tweets. Estudos mostraram que a linguagem de Donald Trump é mais “autêntica” porque ele é percebido como um orador inocente e franco que fica feliz em dizer o que pensa, mesmo que seja incorreto. Ao analisar declarações públicas, Ott (2017) observou que Trump costuma usar linguagem mais dura e violenta, que são mais facilmente vistas como insultos. No entanto, a fama de Trump pode explicar a popularidade de sua conta no *twitter*; ele é visto como um homem da

era da nova tecnologia. Durante a crise da pandemia, as guerras da mídia e a aversão contínua do presidente Donald Trump ao conselho federal impactaram negativamente as respostas às políticas governamentais e a compreensão das pessoas sobre o COVID-19, causando uma séria perda de vidas americanas porque o número de infecções e mortalidade atingiu o pico dentro 2 meses.

Burggraf, Fendel e Huynh (2020) analisou mais de 3.200 tweets do presidente dos EUA, Donald Trump, e descobriu que tweets com conteúdo criticando ou incluindo palavras negativas relacionadas à guerra comercial EUA-China afetaram adversamente o mercado de ações e muitas indústrias dos EUA que dependiam da intensidade do comércio dos EUA com a China. Durante os estágios iniciais da pandemia de COVID-19, Yaqub (2020) usou o modelo VADER (Valence Aware Dictionary and sEntiment Reasoner) para estudar as emoções dos tweets do presidente Donald Trump e revelou uma correlação entre o tom de seus tweets e o número de a pandemia e as vidas e economias da América; ele notou ainda negatividade nas mensagens de Trump ao mencionar a China e o COVID-19. No entanto, de acordo com Ajjoub, Walker e Zhao (2021), analisando cada um dos tweets de Trump como um acontecimento mostra a influência e o controle das notícias sobre os fatores macroeconômicos do mercado. Em particular, as variáveis de controle são determinadas com base no nível de influência para 49 indústrias nos Estados Unidos.

Wei, Mao e Wang (2016) mostram que há relação entre o pico de volume no *twitter*, ou seja, entre um alto número de tweets, e o preço das ações. Também foi visto que a volatilidade das ações costuma ser maior antes de um alto volume de tweets, e que a volatilidade implícita aumenta acentuadamente antes de um pico no volume do *twitter*, diminuindo rapidamente depois disso. Nota-se ainda que após um pico de volume no *twitter*, as ações parecem continuar superfaturadas, o que, segundo os autores, deve ser utilizado como uma estratégia para venda. Ao usar essa estratégia, os autores conseguiram obter um retorno de 34,3%, sendo esse retorno superior ao retorno do S&P 500 no mesmo período, que foi de 12,8%.

Quanto à relação entre o sentimento e o retorno das ações, Oliveira, Cortez e Areal (2017) buscaram identificar se o sentimento tem impacto no retorno das ações por mais de uma maneira. Os autores utilizaram, para além de um indicador de sentimento formado por pesquisa de opinião, um indicador tendo por base o *twitter*. Ambos os indicadores mostraram ter impacto sobre o preço das ações, demonstrando que elas podem ser alternativas complementares na predição do preço das ações.

Sob a perspectiva de tentar identificar se havia relação entre sentimento extraído do *twitter* e o retorno das ações, Shen, Liu e Zhang (2018) provaram que essa relação existe, trazendo evidências com base em 26 índices internacionais. De acordo com os autores, há uma estreita relação entre a dinâmica dos sentimentos on-line e o desempenho do mercado de ações. Das et al. (2018) também apresentam provas de que os *feeds* do *twitter* tem potencial para serem usados na previsão do movimento dos preços das ações do mercado.

Shen, Liu e Zhang (2018) investigaram como o sentimento diário de felicidade do *twitter* pode afetar os retornos das ações. Os autores perceberam que o sentimento e o retorno estão ligados, tanto ao utilizar modelos lineares como ao utilizar modelos não lineares. Isso mostra que há interdependência entre as atividades on-line e o mercado de ações. Todavia, (Ruan, Durresti e Alfantoukh, 2018) mostram que essa interdependência pode ser ainda maior quando se pondera quem foi o autor dos tweets. Ou seja, a relação entre o indicador de sentimento via *twitter* e o mercado acionário é ainda mais forte a depender do poder ou reputação do usuário na comunidade.

Visando identificar como o *twitter* impacta o mercado em termos de retorno e de volume de negociações, e se ele poderia ser empregado como ferramenta na previsão das ações, Nisar e Yeung (2018) desenvolveram um estudo tendo por foco as eleições para presidente do Reino Unido em 2016. Foi utilizado um período de 6 dias, que englobava o período antes, durante e depois das eleições. Os resultados encontrados mostraram haver potencial de uma forte relação de casualidade entre o sentimento e o retorno, mostrando correlação entre as variáveis, e uma significância estatística ao nível de 10%. Todavia, não foram encontradas fortes relações entre as variáveis relacionadas ao volume de tweets e o volume de negociação, ou ao volume de tweets e o retorno das ações, nem mesmo quando

a quantidade das mensagens foi destrinchada em positivas e negativas. Também foi visto que, em média, o sentimento no dia t só impactará o retorno do mercado em $t+3$. Assim, os autores acreditam que há um potencial para que o *twitter* seja utilizado como ferramenta de previsão do mercado.

3 Metodologia e Dados

3.1 Construindo o Índice de Sentimento com Dados do *Twitter*

Para a construção do índice de sentimento aplicado ao retorno do mercado acionário brasileiro foram utilizados *tweets* como *proxy* para análise de sentimento.

O *twitter* foi escolhido por ser uma fonte de informação com a qual os usuários podem postar e transmitir informação de maneira instantânea, e da qual as informações podem ser obtidas facilmente. Para atingir o propósito do estudo, utilizou-se o IBOVESPA como *proxy* para conhecer o comportamento do mercado acionário brasileiro. O IBOVESPA foi escolhido para representar o mercado, porque ele é formado pelas ações de maior liquidez e maior volume de negociação da Bolsa, Brasil, Balcão (B3), assim, ele é o índice que melhor representa o mercado como um todo e que é mais suscetível a ser afetado pelo sentimento dos investidores.

No *twitter* usamos os termos "IBOV", "IBOVESPA" e "B3" para selecionar as notícias e os *tweets*. Antes de executar a análise lexicográfica nos *tweets*, realizamos uma série de transformações no texto original. O texto é primeiro dividido em uma sequência de *substrings* (*tokens*) cujos caracteres são todos transformados em minúsculas. Removemos *stop words* em inglês e limpamos o texto usando o pacote *tolower* do R.

Posteriormente, após baixar os dados salvaremos em um “corpus”, segundo Godeiro et al. (2018) “um corpus é uma coleção de textos escritos, ou seja, um conjunto de notícias financeiras ou termos financeiros”. Porém é sabido que no *twitter* além desses dados de notícias sobre esses termos também existem outros *tweets* que contêm outros tipos de termos. Dessa forma, com o uso do *Software R-Stúdio*, foi aplicada uma filtragem de todos os textos buscando os termos com referência apenas financeiras.

Conforme Shapiro, Sudhof e Wilson (2020) há uma breve literatura que apresenta fontes metodológicas para quantificar os sentimentos contidos em textos ou *tweets*. Dentre dessas, destaca-se a lexical. Esta técnica, baseia-se em listas de textos cujo teor de sentimento já são predefinidas por palavras. O sentimento encontrado nesses conteúdos obtêm valor 1 quando o teor das notícias são positivas, -1 quando são negativas e 0 quando elas possuem sentimento neutro. A contagem de palavras contidas em cada notícia são determinadas com base na predominância de palavras positivas *versus* palavras negativas.

Dentre esse tipo de método destaca-se o dicionário criado por Loughran e McDonald (2011) (L&M). Os autores aperfeiçoaram o dicionário *Harvard General Inquirer* (GI/Harvard) que havia sido criado para a psicologia, porém muitas das palavras utilizadas foram ajustada ao mercado acionário. O dicionário (L&M) foi construído contendo listas de palavras positivas e negativas. Estas listas são selecionadas com notícias relacionadas ao mercado financeiro. Nele há 2.355 palavras negativas e 354 palavras positivas. Desta forma, a quando da construção de índices de sentimentos via procedimento de dicionários fixos, o dicionário (L&M) é o mais utilizado.

Por fim, qualquer tipo de dicionário, calcula o índice de sentimento pela diferença entre palavras positivas e negativas, dividida pela soma de palavras positivas e negativas, como foi proposto por Hubert e Labondance (2018):

$$S_t = \frac{\text{PositiveWords}_t - \text{NegativeWords}_t}{\text{PositiveWords}_t + \text{NegativeWords}_t} \quad (1)$$

Portanto, obtemos a medida de sentimentos, S , que varia entre -1 e 1 (Jesus e Besarria,). Para o exercício de previsão utilizamos o dicionário (L&M) para detectar o sentimento contido no teor dos

twitters de notícias diárias, e desta forma, detectar com base em palavras já pré definidas quais são mais representativas para esse tipo de notícias. Após essa etapa é feita a descrição dos preditores técnicos.

3.2 Preditores Técnicos

Normalmente a comparação da previsão do retorno acionário de qualquer economia é feita por observações cuja características pertencem a dados macroeconômicos. Contudo, há uma literatura que destaca essa relação com variáveis técnicas. Estas variáveis devido ao seu retorno econômico e estatístico, normalmente são utilizadas por agentes do mercado financeiro, destacam-as como bons preditores, (Medeiros, Godeiro e Miranda,).

Os estudiosos dessa técnica denotam que o seu desenvolvimento advêm inicialmente por Charles Down (1851-1902), pioneiro da análise técnica, tinha como principal pressuposto analisar comportamentos e padrões dos níveis de preços. Tais procedimentos por sua vez, eram observados por meio de índices como indicativo de mercado, conforme ainda visto nos dias atuais. Segundo Neely et al. (2014) esse indicador toma como base padrões de preços e volumes anteriores visando identificar persistência no futuro.

É sabido que na análise técnica tais comportamentos ocorrem por meio de demonstrações gráficas que são comumente observadas por tendências. Significativamente, essas tendências possuem características próprias, dentre as quais podemos destacar: a Média Móvel e Volume Financeiro. A primeira, caracterizase por meio de linhas gráficas que são sucessivamente as médias dos níveis de preços de um determinado período, sendo estas calculadas a partir de intervalos específicos de tempos. Pequenos intervalos são conhecidos por média móveis curtas e intervalos maiores são denotados por médias móveis longas.

Já o segundo por sua vez, foi inicialmente introduzido por Granville (1960) demonstrando se o Volume Financeiro de um determinado ativo cresce ou diminui no decorrer do tempo, Gonçalves (2018). Quando era observado um aumento havia divulgação de que o volume era alto, caracterizando possivelmente em aumento dos níveis de preços. Já quando o volume era menor, os níveis de preços diminuam, originando em uma menor transação.

Sendo assim, o artigo comparou o Índice de Sentimento proposto, com a técnica adotada no trabalho de Neely et al. (2014), onde os mesmos constroem quatorze Indicadores Técnicos a partir de três categorias: Média Móvel (MA), Momento (MOM) e Volume (VOL). Para isso, foram utilizados dados do Índice IBOVESPA considerado o principal indicador de desempenho do mercado Brasileiro. Nesse sentido, foi feito junto ao site do *Yahoo Finance* uma busca por tais indicadores¹.

A Média Móvel² (MA) tem como objetivo dá o valor de compra e venda de uma determinada ação S_i no tempo $t(0, 1)$ conforme apresentado na equação abaixo.

$$S_{i,t} = \begin{cases} 1 & \text{se } MA_{s,t} \geq MA_{l,t} \\ 0 & \text{se } MA_{s,t} \leq MA_{l,t} \end{cases}$$

Quando

$$MA_{i,j} = \frac{1}{j} \sum_{i=0}^{j-1} P_{t-i} \text{ para } j = s, l$$

Onde: P_t é o nível de preço de uma determinada ação S_i ; a Média Móvel (MA) é representada como longa (l) e curta (s), sendo ($s < l$) denotando que uma MA_s será mais sensível ao movimento de preço do que MA_l . Sendo assim, quando um determinado preço começar a subir a MA_s será mais sensível que a MA_l caracterizando sinal de compra que é representada pelo valor 1 na equação acima

¹ A estimação dos ocorreu com o Software *R- Stúdio* através dos pacotes “quantmod”, “xts”, “BETS” e “CARET”.

² As combinações para cada classificação dos Indicadores Técnicos (Média Móvel, Momento e Volume) são apresentados no Apêndice 1 deste artigo.

e um valor de venda quando o resultado for 0 . Os valores atribuídos as médias móveis (MA) são: $MA_s(1, 2, 3)$ e $MA_l(9, 12)$.

Já os Indicadores de Momentos (MOM) são os valores de compra e venda de uma ação, S_i . Automaticamente, quando o preço de uma ação S_i no período atual (t) é superior aos preços ($t - m$) indica que o momento (t) é positivo, gerando um sinal de compra da ação, S_i . Assim, o modelo de Momento (MOM) é representado pela seguinte equação:

$$S_{i,t} = \begin{cases} 1 & \text{se } P_t \geq P_{t-m} \\ 0 & \text{se } P_t \leq P_{t-m} \end{cases}$$

Já com relação ao Volume (VOL), segundo (Neely et al., 2014) especialistas da análise técnica constantemente utilizam o volume financeiro em grupo com os preços dos períodos anteriores visando identificar tendência no mercado. Assim, nossa estratégia termina acrescentando esse Volume (OBV) da seguinte forma:

$$OBV_t = \sum_{k=1}^t VOL_k D_k$$

Onde: OBV_t é o retorno do Índice IBOVESPA no momento t ; VOL_k é o volume no período k e D_k é um binário com valores P_k e $P_k \leq k$ e -1 caso contrário. Assim, conforme o que foi descrito será apresentado um sinal de negociação para o o Índice IBOVESPA, OBV_t da seguinte forma:

$$S_{i,t} = \begin{cases} 1 & \text{se } MA_{s,t}^{OBV} \geq MA_{l,t}^{OBV} \\ 0 & \text{se } MA_{s,t}^{OBV} \leq MA_{l,t}^{OBV} \end{cases},$$

Quando

$$MA_{i,j}^{OBV} = \frac{1}{j} \sum_{i=0}^{j-1} OBV_{t-i}, \text{ para } j = s, l.$$

Se a média móvel MA_s do IBOVESPA for maior que a média móvel MA_l indicará momento favorável de compra, caso contrário representará sinal de venda.

3.3 Prevendo Passos a Frente - Fora da Amostra

Para se fazer essa análise com previsões dentro e fora da amostra foi feito uma divisão dos dados. Seguimos a estratégia adotada no trabalho (Medeiros, Godeiro e Miranda,) para a divisão dos dados. Dividimos o total de observações em dentro da amostra com uma proporção de 60% das observações diárias e fora da amostra com proporção de 40% do total.

Após a divisão dos dados, realizamos uma regressão Δs_{t+1} sobre a constante, e partir disso é gerado um preditor x_s^t para $s = 1, \dots, t - 1$, e posteriormente computa-se a previsão como sendo $\Delta s_{t+1} = \hat{a}_t + \hat{\theta}_t x_t^i$, de forma que \hat{a}_t e $\hat{\theta}_t$ especifica os Mínimos Quadrados Ordinários (MQO) e x_t^i reporta as modelagens adotadas: Dados de *Tweets* e Preditores Técnicos.

Após essa estratégia inicial, foi utilizado o R_{OS}^2 fora da amostra como método de avaliação. Este procedimento ocorre com a comparação do modelo de previsão $\Delta s_{t+1} = \hat{a}_t + \hat{\theta}_t x_t^i$, em relação ao benchmark³ $\bar{\Delta} s_{t+1} = \frac{1}{t} \sum_{s=1}^t \Delta s_t$. Assim, o R_{OS}^2 é especificado da seguinte forma:

$$R_{OS}^2 = 1 - \frac{\sum_{t=R}^{T-1} (\Delta s_{t+1} - \hat{\Delta} s_{t+1})^2}{\sum_{t=R}^{T-1} (\Delta s_{t+1} - \bar{\Delta} s_{t+1})^2}$$

Diante da equação pode-se observar as seguintes especificações: se a previsão $\hat{\Delta} s_{t+1}$ bate a previsão do *benchmark* então $R_{a,s,2}^2 \geq 0$. Desta forma, o (MSPE) dos preditores em R_{OS}^2 especificará a relação com o modelo de referência (média histórica do índice IBOVESPA).

³ Nosso modelo benchmark é representado pela média histórica dos retornos do IBOVESPA.

Além disso, é avaliado a significância dos modelos de previsões, e esta por sua vez, é feito por meio da estatística de [Diebold e Mariano \(1999\)](#). No entanto, devido as características dos modelos (aninhados) tal técnica torna-se inadequada. Visando corrigir possíveis erros de especificações, [Clark e West \(2007\)](#) aprimoraram a técnica de [Diebold e Mariano \(1999\)](#), na qual ficou conhecida na literatura como Erro Quadrático Médio de Previsão Ajustado (MSPE). Tal melhoramento tem como objetivo testar se a hipótese nula é significativa frente a hipótese alternativa. A sua especificação é dada da seguinte forma:

$$\Delta s_{t+1} = \left(\Delta s_{t+1} \bar{\Delta} s_{t+1} \right)^2 - \left[\left(\Delta s_{t+1} - \hat{\Delta} s_{t+1} \right)^2 - \left(\bar{\Delta} s_{t+1} - \hat{\Delta} s_{t+1} \right)^2 \right]$$

Diante da equação acima, o MSPE é encontrado regredindo $\{\Delta s_{t+1}\}_{t=1}^{T-1}$ sobre o intercepto e posteriormente calcula-se a estatística t . Após esse passo, é possível averiguar se as previsões baseadas nos modelos de dados do *tweets* possui um MSPE superior ao modelo dos retorno histórico do índice representativo do mercado acionário brasileiro (*benchmark*).⁴

Por fim, é calculado o equivalente certeza (ganho de utilidade) de um investidor de média variância. Este ganho pode ser observado como o retorno que cada modelo de previsão pode propor a esse investidor, ou seja, se eles seriam capazes de repassar mais informações do mercado em relação ao modelo de referência (*benchmark*) e desta forma, esse agente conseguir tomar suas melhores estratégias. Aqui foi empregado a mesma abordagem adotada em diversos estudos para calcular o ganho de utilidade. Um desses trabalhos estão [Campbell e Thompson \(2008\)](#) e [Rapach e Strauss \(2010\)](#), sendo esses os pioneiros nessa abordagem.

Este método tem como principio de que este agente é avesso ao risco e busca sempre alocar seu rendimento numa cesta de ativos que contenham a média do mercado, ou seja, ativos com risco e sem risco no tempo t . Essa relação é baseada na taxa sem risco e na previsão do retorno do índice IBOVESPA para o período posterior, Δs_{t+1} . Na literatura tem observado um valor de peso para o ativo arriscado. Este peso é dado por:

$$\omega_t = \frac{1}{\gamma} \frac{\hat{\Delta} s_{t+1}}{\sigma_{t+1}^2} \quad (2)$$

De forma que: γ é apresentado como sendo o parâmetro de aversão ao risco e σ_{t+1}^2 representa a variância do retorno do índice IBOVESPA. Desta forma, o retorno um passo a frente em $t + 1$ é $R_{t+1} = \omega_t \Delta s_{t+1} + (1 - \omega_t) r_{t+1}^f$ onde é imposto a ressalva de que $\omega_t \in (-1, 1)$.

Diante disso, para o passo a frente $t + 1$, este agente (investidor) perceberá que seu nível de utilidade é dado por:

$$U = \hat{\mu} - \frac{1}{2} \gamma \hat{\sigma}^2$$

Onde $\hat{\mu} = \frac{1}{p} \sum_t R_t \hat{\sigma}_p^2 = \text{Var} \left(R_t = \frac{1}{p} \sum_t (R_t - \hat{\mu}_t) \right)^2$ é a quantidade total de observações fora da amostra.

O ganho de utilidade é dado como a diferença da utilidade que é obtida entre os modelos de previsões. Este método é adotado por ([Godeiro et al., 2018](#)); ([Rapach e Strauss, 2010](#)); ([Medeiros, Godeiro e Miranda,](#)) dentre outros.

4 Resultados

Na tabela 1, estão dispostos os resultados do teste de ([Clark e West, 2007](#)) e o R_{OOS}^2 , que servem para avaliar a capacidade preditiva dos indicadores técnicos e do índice de sentimento baseado nos *tweets*. Vale ressaltar que um R_{OOS}^2 maior do que zero indica que o modelo supera o *benchmark*. A

⁴ Esse passo testa a hipótese nula de que $R_{OOS}^2 \leq 0$ contra a hipótese alternativa de que $R_{OOS}^2 > 0$.

tabela também traz o ganho de utilidade, que serve para avaliar o ganho econômico obtido pelo investidor, ou seja, o retorno econômica que o investidor obtém ao utilizar o modelo proposto ao invés do *benchmark*.

Tabela 1 – Resultado das Previsões Fora da Amostra

Modelo	$R^2_{OOS}\%$	CW	Utilidade
ma_01_09_BOVA11.SA	-0,06	0,29	0,12
ma_01_12_BOVA11.SA	-0,23	0,92	-0,38
ma_02_09_BOVA11.SA	-0,23	0,75	-0,33
ma_02_12_BOVA11.SA	-0,25	0,89	-0,33
ma_03_09_BOVA11.SA	-0,22	0,86	-0,3
ma_03_12_BOVA11.SA	-0,23	0,97	-0,33
mom_09_BOVA11.SA	-0,23	0,93	-0,37
mom_12_BOVA11.SA	-0,26	0,87	-0,25
obv_01_09_BOVA11.SA	-0,13	0,24	0,12
obv_01_12_BOVA11.SA	-0,02	0,1	0,53
obv_02_09_BOVA11.SA	-0,14	0,32	0,13
obv_02_12_BOVA11.SA	-0,26	0,73	-0,27
obv_03_09_BOVA11.SA	-0,23	0,8	-0,21
obv_03_12_BOVA11.SA	-0,25	0,99	-0,41
Sentimento_Twitter	-0,17	0,95	-0,21

Fonte: Elaboração dos Autores.

No que diz respeito à análise do R^2_{OOS} , nota-se que nenhum dos modelo consegue superar a média histórica, dado que a mesma é um modelo com baixa variância. Outro resultado interessante é que o modelo baseado no índice de sentimento *twitter* supera a maioria dos modelos que dependem de indicadores técnicos. O modelo baseado nos *tweets* é pior do que 5 modelos de indicadores técnicos e melhor que 9. Nesse sentido, nota-se que o sentimento do *twitter*, mesmo sendo calculado com base na lista fixa de palavras de (Loughran e McDonald, 2011), pode trazer alguma melhora na previsão do retorno diário do IBOVESPA. Assim, sugere-se que, para futuras pesquisas a melhora no cálculo do sentimento, utilizando técnicas que dependem do dicionário variante no tempo, como em (Lima, Godeiro e Mohsin, 2019).

Com relação ao ganho de utilidade, nota-se que os modelos baseados nos preditores (*obv_01_12*), (*obv_02_09*), (*ma_01_09*) e (*obv_01_09*) superaram a média histórica, com ganhos de utilidade de 0.53%, 0.13%, 0.12% e 0.12 %, respectivamente. Assim, tem-se que para o investidor que adotou algum dos modelos acima, obteve um ganho econômico considerável para dados na frequência diária. Outro resultado bem surpreendente é que os preditores baseado no indicador OBV tiveram um desempenho superior aos demais indicadores técnicos, bem como ao preditor baseado nas informações do *twitter*. Dado a dificuldade de prever retornos financeiros em frequência diária, sugere-se que tais preditores podem ser utilizados em estratégias quantitativas de investimento. Ressalta-se ainda que a presente pesquisa inova em dois sentidos: utilização de uma larga amostra de *tweets* e previsão de retorno na frequência diária.

5 Conclusões

Este artigo tem o objetivo de verificar o potencial de dados do *twitter* na previsão dos retornos do índice IBOVESPA. Assim foi construído um índice de sentimento de *tweets* que mencionam o IBOVESPA para captar a polaridade de tais publicações. O desempenho do índice de sentimento foi comparado com estratégias provenientes da Análise Técnica e com o retorno histórico do índice representativo do mercado acionário brasileiro (*benchmark*).

O sentimento do *twitter* foi calculado com base no dicionário fixo de Loughran e McDonald (2011). Os resultados mostraram que o modelo baseado no índice de sentimento conseguiu obter um bom desempenho em comparação com o *benchmark* e em relação aos indicadores técnicos. Entretanto, o modelo com índice de sentimento obteve uma acurácia inferior quando comparado a alguns modelos de indicadores técnicos. Além disso, foi constatado que o investidor que usou esses melhores modelos de indicadores técnicos conseguiu ter ganhos de utilidade.

Por fim, diante da dificuldade de prever os retornos do IBOVESPA em alta frequência, sugere-se que tais preditores podem ser utilizados em estratégias quantitativas de investimento. Ressalta-se ainda que a presente pesquisa inova em dois sentidos: utilização de uma larga amostra de *tweets* e previsão de retorno na frequência diária. Além disso, também é possível o uso de algoritmos de aprendizado de máquina para melhorar a acurácia com a adição de mais preditores.

References

- AJJOUR, C.; WALKER, T.; ZHAO, Y. Social media posts and stock returns: The trump factor. *International Journal of Managerial Finance*, Emerald Publishing Limited, v. 17, n. 2, p. 185–213, 2021.
- ANTWEILER, W.; FRANK, M. Z. Is all that talk just noise? the information content of internet stock message boards. *The Journal of finance*, Wiley Online Library, v. 59, n. 3, p. 1259–1294, 2004.
- BAKER, M.; WURGLER, J. Investor sentiment and the cross-section of stock returns. *The journal of Finance*, Wiley Online Library, v. 61, n. 4, p. 1645–1680, 2006.
- BEATTIE, V. Accounting narratives and the narrative turn in accounting research: Issues, theory, methodology, methods and a research framework. *The British Accounting Review*, Elsevier, v. 46, n. 2, p. 111–134, 2014.
- BLACK, F.; SCHOLES, M. The pricing of options and corporate liabilities. *Journal of political economy*, The University of Chicago Press, v. 81, n. 3, p. 637–654, 1973.
- BOLLEN, J.; MAO, H.; ZENG, X. Twitter mood predicts the stock market. *Journal of computational science*, Elsevier, v. 2, n. 1, p. 1–8, 2011.
- BOX, G.; JENKINS, G. Time series analysis: forecasting and control. gep. 1976.
- BROWN, G. W. Volatility, sentiment, and noise traders. *Financial Analysts Journal*, Taylor & Francis, v. 55, n. 2, p. 82–90, 1999.
- BURGGRAF, T.; FENDEL, R.; HUYNH, T. L. D. Political news and stock prices: evidence from trump's trade war. *Applied Economics Letters*, Taylor & Francis, v. 27, n. 18, p. 1485–1488, 2020.
- CAMPBELL, J. Y.; THOMPSON, S. B. Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, Society for Financial Studies, v. 21, n. 4, p. 1509–1531, 2008.
- CLARK, T. E.; WEST, K. D. Approximately normal tests for equal predictive accuracy in nested models. *Journal of econometrics*, Elsevier, v. 138, n. 1, p. 291–311, 2007.
- CORNFIELD, M. Empowering the party-crasher: Donald j. trump, the first 2016 gop presidential debate, and the twitter marketplace for political campaigns. *Journal of Political Marketing*, Taylor & Francis, n. just-accepted, 2017.

- DAS, S. et al. Real-time sentiment analysis of twitter streaming data for stock prediction. *Procedia computer science*, Elsevier, v. 132, p. 956–964, 2018.
- DIEBOLD, F.; MARIANO, R. Comparing predictive accuracy', journal of business and economic statistics, 13 (3), july, 253-63. *INTERNATIONAL LIBRARY OF CRITICAL WRITINGS IN ECONOMICS*, EDWARD ELGAR PUBLISHING LTD, v. 108, p. 263–273, 1999.
- DOUGAL, C. et al. Journalists and the stock market. *The Review of Financial Studies*, Oxford University Press, v. 25, n. 3, p. 639–679, 2012.
- ENGELBERG, J. Costly information processing: evidence from earnings announcements (january 18, 2008). In: *AFA 2009 San Francisco meetings paper*. SSRN: <http://ssrn.com/abstract>. [S.l.: s.n.], 2009. v. 1107998.
- FAMA, E. F.; FRENCH, K. R. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, Elsevier, v. 33, n. 1, p. 3–56, 1993.
- FERGUSON, N. J. et al. Media content and stock returns: The predictive power of press. *Multinational Finance Journal*, v. 19, n. 1, p. 1–31, 2015.
- FISHER, I. E.; GARNSEY, M. R.; HUGHES, M. E. Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research. *Intelligent Systems in Accounting, Finance and Management*, Wiley Online Library, v. 23, n. 3, p. 157–214, 2016.
- GARCIA, D. Sentiment during recessions. *The journal of finance*, Wiley Online Library, v. 68, n. 3, p. 1267–1300, 2013.
- GODEIRO, L. L. et al. Ensaios sobre modelos de previsão econômica. Universidade Federal da Paraíba, 2018.
- GONÇALVES, L. I. O uso de indicadores técnicos como suporte à tomada de decisões no mercado financeiro. 2018.
- GRANVILLE, J. E. *A strategy of daily stock market timing for maximum profit*. [S.l.]: Prentice-Hall, 1960.
- GRIFFIN, J. M.; HIRSCHEY, N. H.; KELLY, P. J. How important is the financial media in global markets? *The Review of Financial Studies*, Oxford University Press, v. 24, n. 12, p. 3941–3992, 2011.
- HAN, X.; LI, Y. Can investor sentiment be a momentum time-series predictor? evidence from china. *Journal of Empirical Finance*, Elsevier, v. 42, p. 212–239, 2017.
- HENDERSHOTT, T.; LIVDAN, D.; SCHÜRHOFF, N. Are institutions informed about news? *Journal of financial economics*, Elsevier, v. 117, n. 2, p. 249–287, 2015.
- HU, B.; MCINISH, T.; ZENG, L. Gambling in penny stocks: the case of stock spam e-mails. *International Journal of Cyber Criminology*, International Journal of Cyber Criminology, v. 4, n. 1/2, p. 610, 2010.
- HU, X.; LIU, H. Text analytics in social media. *Mining text data*, Springer, p. 385–414, 2012.
- HUANG, A. H.; ZANG, A. Y.; ZHENG, R. Evidence on the information content of text in analyst reports. *The Accounting Review*, American Accounting Association, v. 89, n. 6, p. 2151–2180, 2014.
- HUBERT, P.; LABONDANCE, F. Central bank sentiment. URL: <https://www.nbp.pl/badania/seminaria/14xi2018.pdf>. Working paper, 2018.

- HUYNH, T. L. D. et al. Feverish sentiment and global equity markets during the covid-19 pandemic. *Journal of Economic Behavior & Organization*, Elsevier, v. 188, p. 1088–1108, 2021.
- JEGADEESH, N. Does market risk really explain the size effect? *Journal of Financial and Quantitative Analysis*, Cambridge University Press, v. 27, n. 3, p. 337–351, 1992.
- JESUS, D. P. de; BESARRIA, C. da N. Narrativas do banco central e previsões macroeconômicas: Usando análise textual de machine learning.
- JIANG, F. et al. Manager sentiment and stock returns. *Journal of Financial Economics*, Elsevier, v. 132, n. 1, p. 126–149, 2019.
- JUNQUEIRA, K. T.; FERNANDES, A. M. da R. Análise de sentimento em redes sociais no idioma português com base em mensagens do twitter. *Anais do Computer on the Beach*, p. 681–690, 2018.
- KARABULUT, Y. Can facebook predict stock market activity? In: *AFA 2013 San Diego meetings paper*. [S.l.: s.n.], 2013.
- KOTHARI, S. P.; LI, X.; SHORT, J. E. The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. *The Accounting Review*, v. 84, n. 5, p. 1639–1670, 2009.
- LEHAVY, R.; LI, F.; MERKLEY, K. The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review*, v. 86, n. 3, p. 1087–1115, 2011.
- LI, F. The determinants and information content of the forward-looking statements in corporate filings-a naive bayesian machine learning approach. *Journal of Accounting Research*, v. 1001, 2008.
- LIANG, C. et al. Which sentiment index is more informative to forecast stock market volatility? evidence from china. *International Review of Financial Analysis*, Elsevier, v. 71, p. 101552, 2020.
- LIMA, L. R.; GODEIRO, L.; MOHSIN, M. Time-varying dictionary and the predictive power of fed minutes. *Available at SSRN 3312483*, 2019.
- LIU, B.; MCCONNELL, J. J. The role of the media in corporate governance: Do the media influence managers' capital allocation decisions? *Journal of Financial Economics*, Elsevier, v. 110, n. 1, p. 1–17, 2013.
- LIU, B.; MCCONNELL, J. J.; XU, W. The power of the pen reconsidered: The media, ceo human capital, and corporate governance. *Journal of Banking & Finance*, Elsevier, v. 76, p. 175–188, 2017.
- LIU, S. The impact of textual sentiment on sovereign bond yield spreads: Evidence from the eurozone crisis. *Multinational Finance Journal*, v. 18, n. 3/4, p. 215–248, 2014.
- LOUGHRAN, T.; MCDONALD, B. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, Wiley Online Library, v. 66, n. 1, p. 35–65, 2011.
- MCGRANAHAN, C. Extreme speech| a presidential archive of lies: Racism, twitter, and a history of the present. *International Journal of communication*, v. 13, p. 19, 2019.
- MEDEIROS, E. H. O. de; GODEIRO, L. L.; MIRANDA, K. F. Previsão do prêmio de risco no mercado de ações brasileiro utilizando notícias financeiras, fatores de risco, indicadores técnicos e aprendizado de máquina supervisionado.
- NEELY, C. J. et al. Forecasting the equity risk premium: the role of technical indicators. *Management science*, INFORMS, v. 60, n. 7, p. 1772–1791, 2014.

- NISAR, T. M.; YEUNG, M. Twitter as a tool for forecasting stock market movements: A short-window event study. *The journal of finance and data science*, Elsevier, v. 4, n. 2, p. 101–119, 2018.
- OBAID, K.; PUKTHUANTHONG, K. A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news. *Journal of Financial Economics*, Elsevier, v. 144, n. 1, p. 273–297, 2022.
- OLIVEIRA, N.; CORTEZ, P.; AREAL, N. The impact of microblogging data for stock market prediction: Using twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Systems with applications*, Elsevier, v. 73, p. 125–144, 2017.
- OTT, B. L. The age of twitter: Donald j. trump and the politics of debasement. *Critical studies in media communication*, Taylor & Francis, v. 34, n. 1, p. 59–68, 2017.
- PASSOS, L. F. C. et al. Twittadas presidenciais: Análise dos tweets do atual e alguns ex presidentes do brasil. *Revista do Seminário Internacional de Estatística com R*, v. 4, n. 2, 2019.
- PHAN, H. T.; NGUYEN, N. T.; HWANG, D. Convolutional attention neural network over graph structures for improving the performance of aspect-level sentiment analysis. *Information Sciences*, Elsevier, v. 589, p. 416–439, 2022.
- RAPACH, D.; ZHOU, G. Forecasting stock returns. In: *Handbook of economic forecasting*. [S.l.]: Elsevier, 2013. v. 2, p. 328–383.
- RAPACH, D. E.; STRAUSS, J. K. Bagging or combining (or both)? an analysis based on forecasting us employment growth. *Econometric Reviews*, Taylor & Francis, v. 29, n. 5-6, p. 511–533, 2010.
- RUAN, Y.; DURRESI, A.; ALFANTOUKH, L. Using twitter trust network for stock market analysis. *Knowledge-Based Systems*, Elsevier, v. 145, p. 207–218, 2018.
- SHAPIRO, A. H.; SUDHOF, M.; WILSON, D. J. Measuring news sentiment. *Journal of Econometrics*, Elsevier, 2020.
- SHAPIRO, A. H.; WILSON, D. J. et al. What’s in the news? a new economic indicator. *FRBSF Economic Letter*, Federal Reserve Bank of San Francisco, v. 10, 2017.
- SHEN, D.; LIU, L.; ZHANG, Y. Quantifying the cross-sectional relationship between online sentiment and the skewness of stock returns. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 490, p. 928–934, 2018.
- STOLEE, G.; CATON, S. Twitter, trump, and the base: A shift to a new form of presidential talk? *Signs and society*, University of Chicago Press Chicago, IL, v. 6, n. 1, p. 147–165, 2018.
- STRAUSS, N.; Vliegenthart, R.; VERHOEVEN, P. Lagging behind? emotions in newspaper articles and stock market prices in the netherlands. *Public Relations Review*, Elsevier, v. 42, n. 4, p. 548–555, 2016.
- TETLOCK, P. C. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, Wiley Online Library, v. 62, n. 3, p. 1139–1168, 2007.
- TETLOCK, P. C. Does public financial news resolve asymmetric information? *The Review of Financial Studies*, Oxford University Press, v. 23, n. 9, p. 3520–3557, 2010.
- TETLOCK, P. C.; SAAR-TSECHANSKY, M.; MACSKASSY, S. More than words: Quantifying language to measure firms’ fundamentals. *The journal of finance*, Wiley Online Library, v. 63, n. 3, p. 1437–1467, 2008.

TWEDT, B.; REES, L. Reading between the lines: An empirical examination of qualitative attributes of financial analysts' reports. *Journal of Accounting and Public Policy*, Elsevier, v. 31, n. 1, p. 1–21, 2012.

ULLAH, I. et al. Rweetminer: Automatic identification and categorization of help requests on twitter during disasters. *Expert Systems with Applications*, Elsevier, v. 176, p. 114787, 2021.

WEI, W.; MAO, Y.; WANG, B. Twitter volume spikes and stock options pricing. *Computer Communications*, Elsevier, v. 73, p. 271–281, 2016.

YAQUB, U. Tweeting during the covid-19 pandemic: Sentiment analysis of twitter messages by president trump. *Digital Government: Research and Practice*, ACM New York, NY, USA, v. 2, n. 1, p. 1–7, 2020.

Apêndice 1: Combinações dos Indicadores Técnicos

Tabela 2 – Combinações de Pares dos Indicadores Técnicos

Preditores	Descrição dos Pares conforme o Indicador Técnico
(MA_01_09_BOVA11.SA)	Média Móvel Curta do Primeiro Período com a Média Móvel Longa do Nono Período
(MA_01_12_BOVA11.SA)	Média Móvel Curta do Primeiro Período com a Média Móvel Longa do Décimo Segundo Período
(MA_02_09_BOVA11.SA)	Média Móvel Curta do Segundo Período com a Média Móvel Longa do Nono Período
(MA_02_12_BOVA11.SA)	Média Móvel Curta do Segundo Período com a Média Móvel Longa do Décimo Segundo Período
(MA_03_09_BOVA11.SA)	Média Móvel Curta do Terceiro Período com e Média Móvel Longa do Nono Período
(MA_03_12_BOVA11.SA)	Média Móvel Curta do Terceiro período com a Média Móvel Longa do Décimo Segundo Período
(MOM_01_09_BOVA11.SA)	Retorno dos Níveis de Preços no Primeiro Período com Retorno dos Níveis de Preços do Nono Período
(MOM_01_12_BOVA11.SA)	Retornos dos Níveis de Preços no Primeiro Período com os Retornos dos Níveis do Décimo Segundo período
(OBV_01_09_BOVA11.SA)	Saldo de Volume do Primeiro Período com o Saldo de volume do Nono Período
(OBV_01_12_BOVA11.SA)	Saldo de Volume do Primeiro Período com o Saldo de Volume do Décimo Segundo Período
(OBV_02_09_BOVA11.SA)	Saldo de Volume do Segundo Período com o Saldo de Volume do Nono Período
(OBV_02_12_BOVA11.SA)	Saldo de Volume do Segundo Período com o Saldo de Volume do Décimo Segundo Período
(OBV_03_09_BOVA11.SA)	Saldo de Volume do Terceiro Período com o Saldo de Volume do Nono Período
(OBV_03_12_BOVA11.SA)	Saldo de Volume do Terceiro Período com o Saldo de Volume do Décimo Segundo Período

A tabela 1 retrata das combinações de pares para cada Indicadores Técnicos conforme seus períodos: seis pares para a Média Móvel (*MA_BOVA11.SA*); dois pares o Momento (*MOM_BOVA11.SA*) e seis pares para o Volume (*OBV_BOVA11.SA*).