

Previsibilidade do Retorno das Ações pela Análise do Sentimento Textual: uma revisão

Renan Silva de Carvalho

<https://orcid.org/0000-0001-5475-3233>

Joyce Menezes da Fonseca Tonin

<https://orcid.org/0000-0003-1371-1314>

Simone Leticia Raimundini Sanches

<https://orcid.org/0000-0002-7363-2573>

Resumo

Objetivo: Analisar a evolução das pesquisas com a abordagem de análise textual como ferramenta de predição do retorno das ações no mercado de capitais.

Método: Revisão sistemática auxiliada pelo software VosViewer, em que foram analisados 78 artigos empíricos da temática, publicados em inglês e indexados na Web of Science.

Resultados: Os resultados aglutinam evidências de que o sentimento textual é capaz de prever o retorno das ações, sendo capturado em diversas fontes de informação. Na análise, emergiram quatro categorias que correspondem às fontes de informação para a análise textual: notícias financeiras (31), divulgações corporativas (29), mídias sociais (16) e outros documentos (8).

Contribuições: Este trabalho contribui com a academia ao demonstrar os principais achados de pesquisas elaboradas sobre o assunto, além de sugerir tópicos para trabalhos futuros. No âmbito prático, as evidências transmitem aos investidores que informações textuais acerca das empresas também podem gerar reações no mercado.

Palavras-Chave: Análise Textual; Sentimento Textual; Retorno das Ações; Revisão Sistemática.

Editado em Português e Inglês. Versão original em Português.

Rodada 1: Recebido em 25/5/2022. Pedido de revisão em 22/11/2022. Rodada 2: Resubmetido em 21/12/2022. Pedido de revisão em 3/2/2023. Rodada 3: Resubmetido em 4/3/2023. Aceito em 1/5/2023 por Vinicius Gomes Martins, Doutor (Editor assistente) e por Gerlando Augusto Sampaio Franco de Lima, Doutor (Editor). Publicado em 30/6/2023. Organização responsável pelo periódico: Abracicon.

1. Introdução

A Hipótese de Mercado Eficiente (HME) prevê que os preços dos ativos financeiros refletem todo o conteúdo de informações relevantes divulgadas (Fama, 1970). Apesar das críticas existentes acerca desta teoria, como nos trabalhos de Kahneman (1994) e Lo (2004), assume-se que o mercado é, no mínimo, parcialmente eficiente (Markiel, 2003). Uma das possíveis razões para essa parcialidade é o grande volume de informações que são divulgadas nos relatórios corporativos, em comunicados à imprensa e em notícias, o que dificulta a análise manual e minuciosa dessas informações (Xing, 2017).

Nesse sentido, é do interesse dos investidores buscar formas de automatizar esse processo de análise de informação, considerando que eles não são capazes de absorver todas as informações relevantes, visto que a racionalidade é limitada (Kahneman, 1994) e, com a automatização desse processo, eles poderiam adquirir algum nível de vantagem competitiva. No entanto, várias das informações divulgadas são qualitativas, o que, em um primeiro momento, dificultaria uma análise feita sem que exigisse uma atenção individual do investidor. Nesse sentido, surge a análise textual que, conforme indicam Loughran e McDonald (2016), permite extrair o sentimento das informações divulgadas, transformando-o em dados quantitativos para tratá-los com o objetivo de, dentre outros, prever as variações nos preços das ações.

Pesquisas que buscaram traçar a relação entre o sentimento ou tom textual e o retorno das ações encontraram evidências de que: uma mudança no tom textual, em relatórios anuais, altera significativamente as variações nos preços das ações em uma janela de evento de dois dias depois da divulgação (Feldman, Govindaraj, Livnat, & Segal, 2010); o tom linguístico em teleconferências de resultado é um indicador significativo de retornos anormais e volume de negócios posteriores (Price, Doran, Peterson, & Bliss, 2012); e altos níveis de incerteza nos textos de formulários registrados na *Securities and Exchange Commission* (SEC) antes do processo de Oferta Pública Inicial, ou *Initial Public Offering* (IPO), têm retornos mais elevados no primeiro dia de negociação na bolsa de valores (Loughran & McDonald, 2013).

Além de divulgações financeiras realizadas pelas companhias, encontram-se na literatura análises textuais desenvolvidas para capturar o sentimento de notícias e mídias sociais. Os achados da pesquisa de Tetlock (2007) revelaram que altos níveis de palavras pessimistas na coluna do *Wall Street Journal* proporcionam retornos mais baixos no mercado no dia seguinte. Da mesma forma, Chen, De, Hu e Hwang (2014) demonstraram que uma análise textual de opiniões dos investidores transmitidas por meio das mídias sociais é capaz de prever o retorno das ações das empresas, bem como surpresas de lucro, que ocorrem quando uma empresa relata lucros que são significativamente maiores ou menores do que o que era esperado.

Diante da ampla possibilidade de aplicação da análise textual em finanças e contabilidade, a mudança da forma como os textos são analisados com os avanços da tecnologia suscita a necessidade de revisitar a literatura e organizar as principais contribuições de estudos e lacunas deixadas, que podem ser exploradas em pesquisas futuras. Nesse sentido, Kearney e Liu (2014), Xing (2017) e Loughran e McDonald (2020) realizaram revisões acerca da temática; os últimos autores se concentram no uso de análise textual em mídia social, preconceito político e detecção de fraude.

No entanto, diferentemente dos estudos citados, o presente estudo explora, especificamente, a análise textual como forma de prever o retorno das ações. Isso posto, o objetivo deste artigo é analisar a evolução das pesquisas com a abordagem de análise textual como ferramenta de predição do retorno das ações no mercado de capitais. Foi realizada uma revisão sistemática da literatura, auxiliada pelo software VosViewer, em que foram analisados 78 artigos empíricos da temática, publicados em inglês e indexados na base de dados *Web of Science*. A pesquisa não teve recorte temporal de início da janela e limitou-se aos artigos publicados até o fim de 2021.

A revisão possibilitou a identificação de quatro principais fontes de informação para a análise textual: notícias financeiras (31), divulgações corporativas (29), mídias sociais (16) e a quarta fonte, que representa outros documentos (8). A pesquisa contribui com a literatura ao compilar as principais evidências encontradas em cada uma dessas categorias, além de auxiliar novas pesquisas, indicando possíveis direções futuras. Além disso, há contribuições no âmbito prático, ao demonstrar aos *stakeholders* como as informações qualitativas sobre as empresas podem afetar os preços praticados no mercado.

Este artigo apresenta, além desta introdução ao tema de estudo, quatro outras seções: na segunda seção, é apresentada a revisão de literatura, contemplando a hipótese de mercado eficiente, suas críticas e o surgimento da análise textual em finanças; na terceira seção, são apresentados os aspectos metodológicos; na quarta seção, são evidenciados os resultados da pesquisa; e, na quinta e última seção, estão as considerações finais do trabalho.

2. Referencial teórico

2.1 Hipótese de Mercado Eficiente, Hipótese de Mercados Adaptativos e Análise Textual

A eficiência de mercado é um ponto-chave na literatura de finanças, pois a aceitação ou não dessa eficiência é decisiva na elaboração de modelos financeiros em pesquisas na área de mercado de capitais. A discussão acerca da eficiência de mercado surge com Fama (1970), que apresenta a Hipótese de Mercado Eficiente (HME), na qual se assume que os preços dos ativos financeiros refletem todo o conteúdo de informações relevantes divulgadas.

O autor separa a eficiência de mercado em três tipos: eficiência fraca, eficiência semiforte e eficiência forte, nas quais, posteriormente, Fama (1991) realiza uma mudança em suas nomenclaturas para os testes: I) teste de previsibilidade de retorno; II) teste para estudos de evento; e III) teste de informações privadas, respectivamente. A alteração na nomenclatura não alterou a essência de cada um dos testes; houve apenas uma mudança na forma fraca, que passou a incorporar a previsão dos retornos.

O mercado eficiente, na sua forma fraca, prevê que o preço das ações reflete completamente todas as informações passadas, em especial aquelas referentes às cotações dos ativos e, a partir do estudo de Fama (1991), o rendimento de dividendos e a relação preço/lucro. Um mercado eficiente semiforte presume que os preços se ajustam de forma instantânea a outras informações públicas como relatórios anuais e desdobramento de ações, por exemplo. Os testes para eficiência forte preocupam-se com o fato de os investidores individuais, ou em grupo, terem acesso a informações privilegiadas, as quais são relevantes para a formação de preços, e como essas informações estão incorporadas aos preços.

Após a publicação do estudo de Fama (1970), pesquisadores têm contestado a HME, apresentando evidências contrárias a essa teoria, demonstrando algumas anomalias de mercado e questionando a existência de um mercado eficiente. Entre as evidências contrárias, tem-se, por exemplo, o efeito janeiro, em que os retornos nos meses de janeiro são superiores aos obtidos nos outros meses do ano (Rozeff & Kinney, 1976); o efeito índice Preço/Lucro (P/L), no qual ações com menor P/L trariam retornos superiores ao mercado (Ball, 1992); a obtenção de retornos anormais por meio da aplicação de médias móveis (Brock, Lakonishok, & Lebaron, 1992), dentre outros.

Além das anomalias, surgiram também outras teorias buscando explicar o funcionamento do mercado de capitais. Contrariamente a uma das principais premissas da HME, que é a racionalidade dos investidores, as finanças comportamentais surgem tendo Kahneman e Tversky (1979) como precursores. Os autores defendem que os investidores são irracionais e que as emoções interferem em sua tomada de decisão. Nesse mesmo sentido, a Hipótese de Mercados Adaptativos (HMA) também contrapõe a HME e defende que o mercado funciona de forma cíclica e, em alguns momentos, fatores institucionais e aspectos comportamentais mudam as condições do mercado (Lo, 2004).

Dessa forma, Lo (2004), explica que a HMA se baseia em uma abordagem que visa conciliar a estrutura neoclássica das interações econômicas da HME com os fatores psicológicos dos investidores que, por vezes, resultam em um comportamento irracional que, conseqüentemente, afeta o mercado. Sob a HMA, no mercado existem tendências que podem gerar bolhas, pânico e ciclos de alta ou baixa, que são fenômenos rotineiramente testemunhados em ecologias naturais de mercado (Lo, 2004).

Uma implicação prática da HMA é que o uso de estratégias de investimentos para obter retornos acima do mercado pode funcionar durante determinadas épocas, mas, a partir do momento em que o ciclo de mercado se altera, a estratégia pode não ter o desempenho esperado e retornar quando as condições se tornarem mais propícias (Lo, 2004).

Assim, mesmo em um mercado de capitais competitivo, oportunidades de arbitragem surgem de vez em quando em um “Mercado Adaptável”. Isso porque, conforme apontam Grossman e Stiglitz (1980), sem essas oportunidades, não haveria incentivo para a coleta e a análise de informações, e o aspecto de descoberta de oportunidades dos mercados financeiros não existiria, fazendo assim com que este entrasse em colapso.

Uma das premissas da HMA é que a velocidade na qual as informações são absorvidas pelos participantes do mercado é diferente (Xing, 2017). Isso causa certos desvios de eficiência que podem persistir por curtos períodos, possibilitando ganhos adicionais, se identificados antecipadamente (Markiel, 2003).

Considerando a HMA, e diante de tantas informações disponíveis sobre as empresas e sobre o mercado, os investidores interessam-se em agilizar o processo de análise das informações de modo a sair na frente em sua tomada de decisão e, conseqüentemente, obter retornos acima do mercado. Dessa forma, surge a análise textual, permitindo realizar estudos empíricos que buscam, dentre outras possibilidades, verificar se o sentimento textual de informações públicas sobre o mercado de capitais é capaz de prever as variações nos preços das ações (Loughran & McDonald, 2016).

As notícias financeiras transmitidas pelas mídias tradicionais (textos jornalísticos) é um dos principais objetos da análise textual na literatura de finanças, que tem o trabalho seminal de Tetlock (2007) como um dos precursores. Nessa mesma perspectiva, a popularidade crescente das mídias sociais fez com que surgissem espaços para que investidores compartilhassem suas opiniões acerca de empresas de capital aberto e suas visões de mercado. Assim, tornou-se um campo promissor para capturar o sentimento do investidor por meio de suas publicações e relacioná-lo com o retorno das ações.

Divulgações corporativas realizadas pelas companhias também podem conter informações que permitam identificar quais as perspectivas futuras para a empresa, e essas expectativas sobre o desempenho futuro são de grande interesse dos investidores, que sempre buscam se antecipar aos acontecimentos. As divulgações que podem ter seu conteúdo textual analisado vão desde documentos fornecidos ao órgão regulador, antes mesmo da abertura de capital, passando por outras divulgações obrigatórias, como relatórios anuais, até as divulgações voluntárias, como as teleconferências de resultado.

3. Procedimentos metodológicos

Esta pesquisa trata-se de uma revisão sistemática da literatura. A realização da revisão se deu por se tratar de uma forma de pesquisa que utiliza como fonte de dados a literatura, buscando sintetizar e relatar as evidências acerca de uma temática, com o emprego de técnicas de busca e organização de dados de forma sistematizada, com o intuito de diminuir o viés da pesquisa. O método selecionado, portanto, está em conformidade com o objetivo proposto.

A pesquisa possui uma abordagem qualitativa e descritiva, devido à descrição das características de uma amostra e à sua base teórica estabelecida. Por sua vez, classifica-se como documental, pois a coleta de dados ocorreu na própria literatura, a partir da seleção de artigos em base de dados (Creswell, 2010).

Para a coleta de dados, foram realizadas buscas de artigos, no *Google Scholar*, para identificar terminologias comumente utilizadas nos estudos sobre esta temática, de modo a auxiliar na definição dos termos pesquisados. Nessa busca preliminar, foram utilizadas as palavras-chave das pesquisas de Price et al. (2012) e de Jiang, Lee, Martin e Zhou (2019).

A partir dos resultados do *Google Scholar* sobre as palavras-chave dos artigos citados, foram selecionadas as que obtiveram o maior número de ocorrências – uma associada à análise textual e outra às reações de mercado. Com isso, definiram-se os termos: *Textual Analysis* e *Stock Return*, de modo que se selecion os artigos que continham ambos os termos em seu título, resumo ou palavras-chave.

Com os termos definidos, para realização das buscas, optou-se por utilizar a base de dados *Web of Science*, como forma de certificar de que as publicações foram analisadas por pares (Ahmad & Omar, 2016), e por ser uma base de dados multidisciplinar considerada relevante (Wang & Waltman, 2016).

A busca retornou 114 resultados na *Web of Science*, considerando os artigos publicados até o fim de 2021 e sem recorte temporal de início da janela. Do total de 114 artigos encontrados, realizou-se a leitura dos resumos dos artigos, para selecionar quais fizeram, de alguma forma, a associação entre a análise textual e o retorno das ações empiricamente. Nesta fase foram excluídos 36 artigos que não atendiam a esse critério. As fases da revisão sistemática estão resumidas na Figura 1:

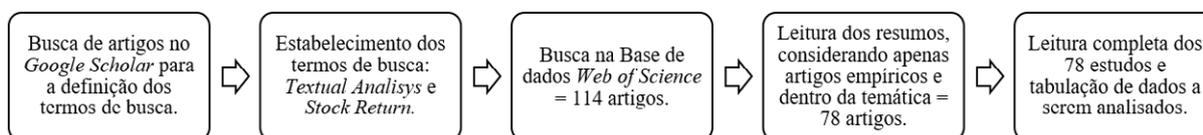


Figura 1. Fases para a revisão sistemática

Fonte: elaborado pelo autor, 2022.

Após o processo de filtragem, a amostra final foi composta por 78 artigos, os quais estão expostos na Figura 2, em que está demonstrada a evolução do número de publicações da temática no decorrer dos anos.



Figura 2. Número de artigos publicados por ano

Fonte: dados da pesquisa, 2022.

Antes mesmo de iniciar o processo de análise dos artigos, com a tabulação dos estudos, é possível observar que, no ano de 2016, houve um crescimento expressivo na quantidade de publicações que relacionam o sentimento textual com o retorno das ações, e, a partir desse ano, um crescimento contínuo até 2020. No ano de 2021, novamente há um salto no número de artigos publicados na temática.

Os artigos foram analisados inicialmente por meio do software VosViewer®. Este *software* possibilitou a visualização de um mapa com as palavras-chave mais utilizadas nos estudos e como estas se conectam, e também auxiliou na divisão em categorias, que representam as fontes de informação para a análise textual. Para a completa divisão em categorias, foi necessária a leitura dos artigos, para a identificação do documento em que foi realizada a análise textual, isto é, a fonte de informação.

4. Resultados e discussões

Esta seção está dividida em dois principais tópicos. No primeiro, são apresentados dados como: visão geral dos artigos, em quais periódicos foram publicados, quais países foram investigados, como as pesquisas evoluíram no decorrer dos anos e quais os principais métodos adotados. O tópico seguinte dispõe sobre as principais contribuições dos estudos, além de lacunas de pesquisa identificadas nas análises. Cada um dos subtópicos representa os documentos que foram objeto da análise textual nos artigos da amostra.

4.1 Visão Geral dos Artigos Analisados

A primeira parte da investigação identificou os periódicos em que os artigos da amostra estavam publicados e o fator de impacto de cada um. Observa-se, na Tabela 1, uma notável variedade de publicações sobre o tema, totalizando 50 periódicos para os 78 artigos da amostra, na área de contabilidade, finanças, administração e tecnologia da informação.

Tabela 1

Periódicos com artigos analisados

Periódico	Fator de impacto	Quant. de artigos	Periódico	Fator de impacto	Quant. de artigos
Journal of Banking and Finance	3,539	5	ACM Transactions on Information Systems	4,657	1
Journal of Behavioral Finance	1,798	5	Decision Sciences	4,551	1
Review of Financial Studies	8,414	4	Journal of Financial and Quantitative Analysis	4,337	1
Pacific-Basin Finance Journal	3,239	4	The Quarterly Review of Economics and Finance	4,324	1
Management Science	6,172	3	Journal of Corporate Finance	4,249	1
Review of Accounting Studies	4,011	3	Accounting, organizations and society	4,114	1
Applied Economics	1,916	3	Journal of Financial Markets	3,095	1
Journal of Financial Economics	8,238	2	Journal of Financial Research	2,811	1
International Review of Financial Analysis	8,235	2	Journal of Business Finance & Accounting	2,709	1
Knowledge-Based Systems	8,139	2	Multimedia Tools and Applications	2,577	1
Decision Support Systems	6,969	2	Accounting & Finance	2,473	1
International Review of Economics & Finance	3,399	2	Kybernetes	2,352	1
European Financial Management	2,295	2	Financial Analysts Journal	2,345	1

Periódico	Fator de impacto	Quant. de artigos	Periódico	Fator de impacto	Quant. de artigos
Computational Economics	1,741	2	Amfiteatru Economic	2,304	1
Journal of Intelligent & Fuzzy Systems	1,737	2	A Journal of Accounting, Finance, and Business Studies	2,060	1
The Quarterly Journal of Economics	15,563	1	Discrete Dynamics in Nature and Society	1,457	1
International Journal of Production Economics	11,251	1	Applied Economics Letters	1,287	1
Finance Research Letters	9,846	1	Revista de Administração de Empresas	1,100	1
Expert Systems with Applications	8,665	1	Lecture Notes in Computer Science	0,402	1
The Journal of Finance	7,870	1	Review of Behavioral Finance	N/A	1
Information Processing & Management	7,466	1	Intelligent Systems in Accounting, Finance and Management	N/A	1
Journal of Business Ethics	6,331	1	Global Business Review	N/A	1
Neural Computing and Applications	5,102	1	Journal of Advanced Computational Intelligence and Intelligent Informatics	N/A	1
Review of Finance	5,059	1	Qualitative Research in Financial Markets	N/A	1
Emerging Markets Finance and Trade	4,859	1	Review of Quantitative Finance and Accounting	N/A	1

Fonte: dados da pesquisa, 2022.

A Tabela 2 identifica em quais ambientes as pesquisas foram desenvolvidas e demonstra os mercados investigados nos estudos analisados.

Tabela 2
Mercados investigados nos artigos

Mercado investigado	Quant. de artigos	Percentual (%)
Estados Unidos	45	57,7
China	13	16,7
Mundo	8	10,3
Taiwan	3	3,8
Alemanha	2	2,6
Brasil	2	2,6
Índia	2	2,6
Inglaterra	1	1,3
Irã	1	1,3
Japão	1	1,3
Total	78	100

Fonte: dados da pesquisa, 2022.

É possível observar a predominância dos Estados Unidos como mercado investigado nos estudos analisados, representando cerca de 58% dos artigos da amostra; em seguida, consta a China, com 17%. Observa-se, ainda, que as pesquisas que envolvem mercados emergentes representam uma pequena parcela dos artigos analisados, demonstrando a possibilidade de replicação dos estudos realizados nos EUA nos mercados emergentes, para fins de comparação.

Oito estudos foram classificados como “Mundo”, pois se referem a estudos que utilizaram o mercado de vários países, sob critérios diferentes. Glasserman e Mamayski (2019), por exemplo, investigaram os 50 principais bancos, seguradoras e imobiliárias globais por capitalização de mercado, dessa forma, abrangendo vários países do mundo. Já Maragoudakis e Serpanos (2016) analisaram os principais mercados de ações europeus, asiáticos e americanos. E a pesquisa de Anand, Basu, Pathak e Thampy (2021) teve como foco os índices de ações de países da União Europeia.

Na próxima etapa da pesquisa, foi criado no software VosViewer® um mapa com as palavras-chave mais utilizadas nos 78 estudos que fizeram parte da amostra (Figura 3), o que auxilia na análise preliminar dos artigos, identificando como se dividem os assuntos dentro dessa temática e como os assuntos foram evoluindo no decorrer dos anos. Para sua elaboração, foram consideradas as palavras-chave que apareceram em mais de dez estudos.

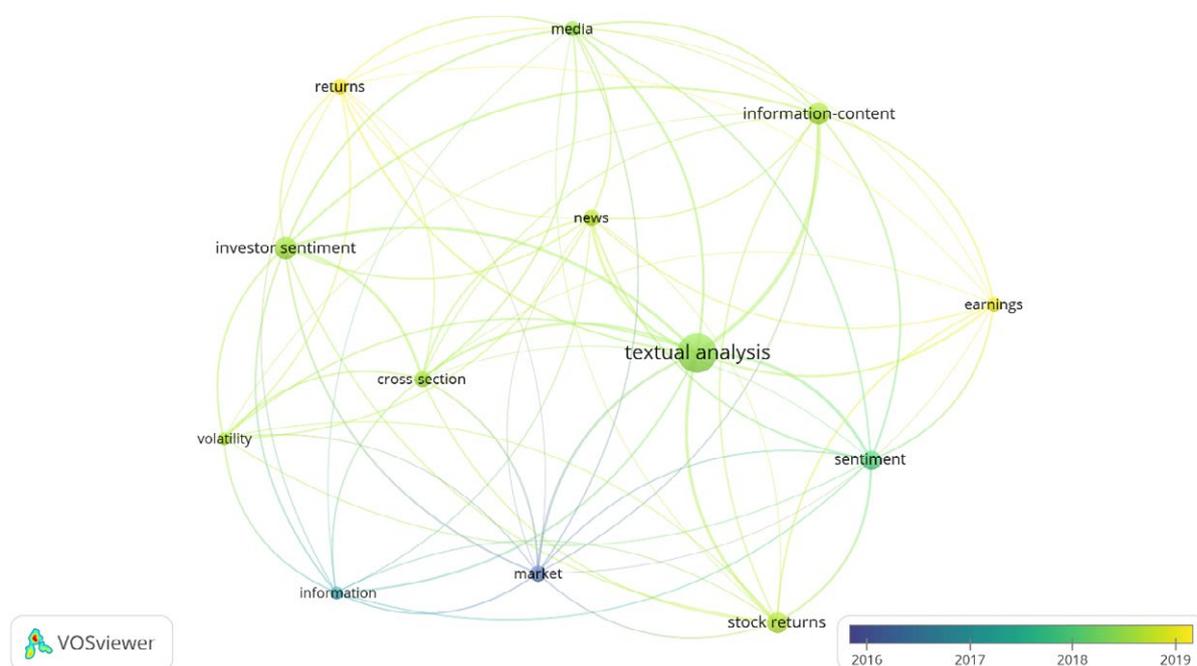


Figura 3. Evolução das palavras-chave dos artigos

Fonte: dados da pesquisa, 2022.

Conforme observado na Figura 3, o termo “*Textual analysis*” está próximo ao centro do mapa, sendo a palavra-chave encontrada com maior frequência nos artigos, com 45 ocorrências, além de possuir conexão com todas as demais palavras. É possível observar outras palavras-chave que também buscam demonstrar o foco do artigo em capturar o sentimento de informações textuais, como “*Sentiment*”, “*Information-content*” e “*Investor sentiment*”, além da presença de “*Returns*”, que se refere à parte do retorno das ações.

Nota-se, ainda, algumas palavras-chave relacionadas ao objeto da análise textual: “*Media*”, “*News*”, “*Earnings*”; a última se refere ao sentimento na divulgação dos resultados realizado pela companhia. No entanto, tais palavras demonstram as diferentes possibilidades de aplicação da análise textual, de modo que a previsão do retorno das ações não se restringe apenas a documentos corporativos divulgados pela companhia, mas também abrangem as notícias financeiras; logo, o sentimento dos investidores pelas mídias sociais pode ter valor preditivo.

Quanto à evolução das palavras-chave no tempo, verifica-se que a maior parte das palavras encontra-se em tons de verde, o que representa que estão bem distribuídas ao decorrer dos anos em que foram publicadas. Ao analisar as palavras que se referem às pesquisas mais recentes, destaca-se a palavra “*Earnings*”, que pode significar que as pesquisas publicadas a partir de 2018 se voltaram para analisar o sentimento na divulgação dos resultados. Vale destacar que a legenda com a descrição dos anos foi formada automaticamente pelo software, de forma que é possível perceber que as pesquisas nessa temática começam a ganhar forma a partir de 2016.

Posteriormente, foi realizada a leitura dos artigos, para identificar quais eram os documentos abrangidos pela análise textual. Além dos que puderam ser observados na Figura 3, destaca-se a teleconferência de resultado, que foi o objeto de análise em 10 artigos. Os documentos utilizados para análise textual, bem como a frequência de seu uso nos artigos da amostra, podem ser visualizados na Tabela 3.

Tabela 3

Objetos da análise textual

Objeto de análise	Quantidade de artigos
Notícias financeiras	31
Divulgações corporativas	29
Mídias sociais	16
Outros documentos	8

Fonte: dados da pesquisa, 2022.

Vale ressaltar que a soma dos números da Tabela 3 não resulta nos 78 artigos que compuseram a amostra. Isso se deve ao fato de que alguns estudos tiveram múltiplas fontes de dados, analisando as notícias financeiras em conjunto com as mídias sociais (Sun, Najand & Shen, 2016; Maragoudakis, 2016; Li, Wang, Wang, Li, Liu, & Chen, 2017; Griffith, Najand, & Shen, 2019; Gan, Alexeev, Bird, & Yeung, 2020), ou as notícias financeiras junto às divulgações corporativas (Eachempati & Srivastava, 2021).

Para capturar o sentimento textual dos documentos, na maioria dos artigos, optou-se pela utilização de um dicionário, classificando as palavras em tom positivo, negativo ou outras categorias preestabelecidas. O dicionário mais utilizado foi o de Loughran e McDonald (2011), aplicado em 36 dos 78 artigos analisados. A justificativa pelo seu uso é descrita como sendo preferível o uso de um dicionário criado especificamente para capturar o sentimento no âmbito financeiro (Myšková, Hájek, & Olej, 2018).

Nesse sentido, pesquisas que usaram a classificação do tom mediante o dicionário de Loughran e McDonald (2011) e o dicionário de Harvard evidenciaram que a obra criada para uso específico em finanças é mais precisa do que a publicação de Harvard (Price et al., 2012; Ferris, Hao, & Liao, 2013; Li, Wu, & Wang, 2020).

Dentre os pesquisadores que optaram por não utilizar o dicionário de Loughran e McDonald (2011), alguns criaram a própria lista de palavras e fizeram a classificação do tom ou sentimento de cada uma, adaptando-a para seu objetivo. Brau, Cicon e McQueen (2016), por exemplo, argumentam que criaram sua própria lista de palavras, devido à especificidade do objetivo, que foi analisar o tom estratégico em documentos de registro de IPOs.

Outros pesquisadores não usaram lista de palavras. É o caso de Ibriyomova, Kogan, Salganik-Shoshan e Stolin (2017), que utilizaram a impressão digital semântica para medir a similaridade entre as empresas, com base em suas descrições, e demonstrar que o grau de similaridade auxilia na previsão da correlação dos retornos das ações dessas empresas. Na amostra, também foram encontrados métodos mais sofisticados para analisar o sentimento textual, a exemplo de Schumaker e Chen (2009), que constroem um sistema para prever as variações de mercado aplicando três técnicas de análise textual por meio da aprendizagem de máquina: I) saco de palavras; II) sintagmas nominais; e III) entidades nomeadas. Destaca-se também a pesquisa de Zhang, Wang, Zhu, Wang e Ghei (2019), que propõe um modelo de rede neural híbrida para classificação e computação dos sentimentos.

Em relação à *proxy* para medir o retorno das ações, é predominante o uso do retorno bruto para a “empresa x” no ano/mês/dia/hora t após determinado evento. Destaca-se também o uso de outras medidas como o *Cumulative Abnormal Return* (CAR), considerando a diferença entre o retorno bruto da ação e o retorno de um índice de mercado (Kiesel, 2021), o *Buy-and-Hold Return* (Jiang, Lee, Martin & Zhou, 2019) e o retorno anormal estimado pelo modelo de mercado (Campbell, Cecchini, Cianci, Ehinger, & Werner, 2019; Wu & Lin, 2017).

Foram encontrados, também, estudos que utilizaram o excesso de retorno obtido por meio do modelo de fatores de risco de Carhart (1997), como a pesquisa de Hillert, Jacobs e Müller (2018), ou pelos três fatores de risco de Fama e French (1993) e cinco fatores de risco de Fama e French (2015), nas pesquisas de Houlihan e Creamer (2017) e Buehlmaier e Whited (2018), respectivamente. A abordagem utilizada na pesquisa de Doukas, Guo, Lam e Xiao (2016) foi o retorno anormal ajustado pelo mercado para análise de curto prazo e o *Buy-and-Hold Abnormal Return* (BHAR) para o longo prazo; isso porque, com o BHAR, é possível calcular o retorno anormal de ações com volatilidade diferente e menor viés no longo prazo do que o CAR (Yan, Xiong, Meng, & Zou, 2019).

4.2 Principais Contribuições e Possibilidades de Pesquisa

4.2.1 Notícias Financeiras

As notícias financeiras foram a principal fonte de dados para a análise textual nos estudos. O artigo mais antigo da amostra é o de Schumaker e Chen (2009) que, por meio de uma pesquisa experimental, usando um sistema baseado em aprendizagem de máquina para analisar as notícias, mostra a correlação entre os preços futuros das ações de companhias e as notícias relacionadas a essas companhias.

Outras pesquisas utilizaram algoritmo de previsão, com o auxílio de redes neurais, para conectar o sentimento textual das notícias obtido pela análise textual a um modelo predominantemente numérico, baseado em indicadores de análise técnica, para prever o retorno das ações intradiário (Geva & Zahavi, 2014; Li et al., 2020). Os resultados demonstraram que a integração de dados de mercado com dados textuais aumenta o poder preditivo dos modelos, conforme testes realizados com 72 empresas do S&P 500 (Geva & Zahavi, 2014) e também na Bolsa de Valores de Hong Kong (Li et al., 2020).

Ahmad, Han, Hutson, Kearney e Liu (2016) analisaram as notícias sobre 20 grandes empresas dos Estados Unidos da América (EUA) publicadas em um período de 10 anos, utilizando o dicionário de Loughran e McDonald (2011) com as palavras que são tipicamente negativas em contextos de finanças. Os autores estimaram um modelo com vetores autorregressivos e constataram que o tom negativo extraído das notícias tem efeito contraproducente sobre o retorno – que, quando significativo, tende a perdurar por longos períodos, em vez de reverter no curto prazo.

Em artigos mais recentes, Glasserman e Mamayski (2019) demonstraram que notícias consideradas “incomuns” e com sentimento negativo aumentam a volatilidade no mercado de ações; já notícias “incomuns” com sentimento positivo preveem uma menor volatilidade. Na mesma perspectiva, Gan et al. (2020) revelaram que a ligação entre volatilidade e sentimento das notícias é mais persistente do que entre retornos e sentimento. Os autores também encontraram evidências de que, a partir de 2016, o sentimento textual das redes sociais passou a ser um melhor preditor das variações do mercado do que as notícias.

A pesquisa de Debata, Ghatge e Renganathan (2021) analisou o sentimento no cenário da pandemia de Covid-19, especificamente no ano de 2020. Os autores constaram que um índice baseado na intensidade do volume de pesquisa do Google é uma *proxy* mais representativa do sentimento público do que um índice baseado em manchetes de jornal. No entanto, esse artigo foi o único da amostra a estudar o cenário de pandemia e tem como principal limitação o estudo apenas do mercado indiano – deixando, assim, lacunas de pesquisa significativas no que se refere à relação entre o sentimento pandêmico e o mercado de ações.

No geral, as pesquisas que capturam o sentimento textual das notícias financeiras possuem uma grande variedade, não apenas de métodos, como o emprego de dicionários, redes neurais e aprendizagem de máquina, como também no foco dado em cada pesquisa. É consenso na literatura que as notícias podem afetar os preços praticados no mercado de ações e constituem a fonte de informações em que mais se aplica a análise textual em finanças.

4.2.2 Divulgações corporativas

As divulgações corporativas correspondem aos documentos textuais que são elaborados e divulgados pelas próprias companhias. Neste tópico, as divulgações corporativas são divididas em 3 categorias: Relatórios 10-K e 8-K (14), Teleconferências de Resultado (11) e Documentos pré-IPO (6). Novamente, a soma dos artigos não resulta nos 29 artigos que trataram de divulgações corporativas, visto que alguns deles investigaram mais de um tipo de documento.

Ao analisar individualmente as *proxies* de retorno utilizadas quando a análise textual foi empregada nesses documentos, verificou-se que a medida predominante é o CAR, seguida pelo *Buy-and-Hold Return*, que, juntos, compreendem a parcela de 59% dos artigos. Um possível motivo desse resultado é que diversos artigos que focaram notícias financeiras e mídias sociais foram publicados em periódicos da área de tecnologia da informação (TI), especificamente na captura do sentimento textual com o uso da tecnologia; já as pesquisas que utilizam as divulgações corporativas se concentram na área de contabilidade e finanças, na qual há maior familiaridade com medidas de retorno mais sofisticadas.

Os documentos 8-K tratam de eventos específicos e fornecem detalhes e exposições sobre o que vem ocorrendo com a empresa e é registrado na SEC. As pesquisas de Feuerriegel & Prölloch (2021) e de Filip, Lobo, Paugam e Stolowy (2021) investigaram como os preços das ações variam em resposta à divulgação de fusões e aquisições – o primeiro, com uma abordagem de mineração de dados, mais especificamente a *Latent Dirichlet allocation (LDA)*; e o segundo, com uma lista própria de palavras. Evidências empíricas encontradas em ambos os estudos demonstraram um efeito negativo e estatisticamente significativo no retorno anormal em resposta à divulgação de fusões e aquisições em intervalos de 1 e 3 dias após a divulgação.

Os efeitos gerados pela divulgação dos relatórios não se limitam às reações no curto prazo. Cohen, Malloy e Nguyen (2020) revelaram, em seu estudo, que as mudanças na linguagem e na construção de relatórios financeiros têm fortes implicações para retornos e operações futuras das empresas. Tais mudanças são gradualmente incorporadas aos preços dos ativos, no período de 12 a 18 meses após a alteração do relatório, e não têm efeitos a curto prazo depois do anúncio.

Sobre as teleconferências de resultado, Price et al. (2012) realizaram uma pesquisa com o objetivo de capturar o tom das chamadas, dividindo a parte de apresentação realizada pela administração e a parte de perguntas e respostas. Os achados revelaram que a parte de perguntas e respostas da chamada tem capacidade preditiva significativa para o CAR de reação inicial, e apresenta poder explicativo durante os 60 dias de negociação após a teleconferência. Resultados adicionais demonstraram, ainda, que o tom passa a ter mais relevância em empresas que não pagam dividendos.

O sentimento do gerente analisado nas teleconferências de resultado é considerado complementar ao sentimento do investidor na previsão dos retornos das ações, o que implica que o sentimento do gerente tem um impacto diferente na avaliação em relação ao sentimento do investidor (Jiang et al., 2019). Esse achado possibilita a interação de ambos os sentimentos, obtidos por meio de diferentes fontes textuais para aumentar o poder preditivo de retornos anormais com base em sentimentos, o que não foi encontrado nos estudos analisados nesta revisão.

Os elogios feitos por analistas durante as chamadas também demonstraram possuir poder explicativo dos retornos das ações. A pesquisa de Milian e Smith (2017) examinou 16.609 teleconferências de empresas S&P 500 e constatou que o número de elogios, como “ótimo ano” e “bom trimestre” prevê retornos anormais das ações no trimestre seguinte. O tom de analistas e de representantes das empresas pode ser mais explorado na literatura, investigando se são complementares ou se possuem poder explicativo próximo.

Informações divulgadas ao órgão regulador, caso da SEC nos EUA, antes da abertura de capital de uma empresa possibilitam a análise textual como ferramenta para prever retornos pós-IPO. A pesquisa de Loughran e McDonald (2013) demonstrou que o tom do primeiro registro na SEC no processo de IPO é capaz de prever retornos no primeiro dia de negociação do ativo. Os achados demonstraram que altos níveis de texto incerto têm retornos mais elevados no primeiro dia de negociação.

No mesmo sentido, Ferris, Hao e Liao (2013) revelaram que o elemento do conservadorismo no texto de prospectos de IPO de empresas que não são da área de tecnologia está inversamente relacionado ao retorno anormal das ações da empresa pós-IPO. Esses resultados demonstram que tanto o tom do relatório S-1 utilizado na pesquisa de Loughran e McDonald (2013) quanto das informações prospectivas têm poder explicativo no retorno das ações pós-IPO.

É notável que as pesquisas que envolvem divulgações corporativas, sejam elas por meio dos relatórios 10-K e 8-K, teleconferências de resultado e os documentos pré-IPO, não possuem uma literatura evoluída como a análise textual nas mídias tradicionais e sociais, de modo que não há uma variedade nos métodos, no mercado investigados e nos objetivos propostos. Isso possibilita um leque de oportunidades de pesquisa nessa temática, investigando o mercado de capitais de outros países, com diferentes recortes temporais e objetivos.

Especificamente sobre as pesquisas que envolvem as teleconferências de resultado, existem evidências de que os investidores estão atentos, também, às divulgações voluntárias realizadas pelas empresas, considerada a influência que esse evento corporativo causa no mercado. Por representar uma área em expansão na literatura, é possível identificar lacunas de pesquisa nos aspectos metodológicos, principalmente na medição do sentimento textual, que pode ser quantificado de formas distintas às já documentadas nas pesquisas que investigam as teleconferências de resultado.

4.2.3 Mídias Sociais

Uma forma de captar o sentimento e as opiniões dos investidores é por meio das mídias sociais, em que investidores, analistas e demais interessados no mercado de capitais podem compartilhar suas análises de investimento e/ou comentar as opiniões de outros investidores. Além disso, as opiniões do investidor transmitidas nas mídias sociais podem se referir a um ativo específico, como também ser uma opinião mais ampla acerca da direção do mercado. Isto é, tanto as opiniões específicas de empresas quanto as mais amplas de mercado são capazes de prever as variações nos preços em uma janela de cinco minutos, conforme constatado por Broadstock e Zhang (2019).

Chen et al. (2014), por meio do dicionário de palavras de Loughran e McDonald (2011), buscaram entender se as opiniões compartilhadas por investidores eram capazes de prever as variações no mercado de capitais. Os resultados demonstraram que a fração de palavras negativas em artigos e comentários em uma rede social popular entre investidores nos Estados Unidos é capaz de prever negativamente os retornos das ações nos três meses seguintes, mesmo controlando os efeitos de recomendações de analistas e mídias de notícias tradicionais.

A linguagem adotada em mídias sociais é, por vezes, diferente da usual formalidade dos relatórios corporativos emitidos pelas empresas e das notícias publicadas pelas mídias tradicionais. Isso ocorre devido ao fato de que publicações e comentários nas mídias sociais são realizados pelos mais diversos tipos de usuários, em linguagem informal; enquanto isso, nas mídias tradicionais, as notícias são publicadas por jornalistas que apenas comunicam os fatos com formalidade. Diante disso, Renault (2017) construiu um léxico de palavras comumente usadas por investidores nas mídias sociais. Esse léxico teve poder preditivo significativamente superior aos dicionários mais utilizados pela literatura. O autor acrescentou algoritmos de aprendizagem de máquina mais complexos, e os resultados encontrados continuaram competitivos.

Já na pesquisa de Xu, Pang e Han (2021), os autores optaram por uma alternativa diferente da lista de palavras para medir o sentimento textual. O estudo utilizou um índice derivado de postagens do Twitter por meio de análise textual como *proxy* do sentimento on-line, que é o *Daily Happiness Index (DHS)*. Os achados revelaram que existe uma relação de correlação cruzada de lei de potência entre o mercado financeiro e o sentimento on-line em alguns países desenvolvidos e em todos os países em desenvolvimento analisados na amostra, que foi representado pelo Brasil e pela Índia.

As mídias sociais nas quais o sentimento dos usuários foi capturado estão expostas na Tabela 4, a seguir. Alguns estudos optaram por selecionar apenas uma das mídias (Broadstock & Zhang, 2019; Chen et al., 2014), outros utilizaram mais de uma (Griffith Najand & Shen, 2020; Gan et al., 2020); a escolha da mídia se justifica por se tratar das mais populares entre os investidores nos EUA.

Tabela 4
Mídias Sociais utilizadas

Mídia Social	
Seeking Alpha	Yahoo!Finance
StockTwits	Sina
Twitter	EastMoney

Fonte: dados da pesquisa, 2022.

As pesquisas que buscam analisar o sentimento dos investidores expresso nas mídias sociais são mais recentes em comparação àquelas que têm como fonte de informações as mídias tradicionais. Apesar de as pesquisas com as mídias sociais representarem cerca de 21% dos artigos analisados nesta revisão, é preciso considerar a crescente expansão deste tipo de comunicação, o que proporciona novas oportunidades de pesquisa, analisando, por exemplo, outras mídias utilizadas pelos investidores.

4.2.4 Outros Documentos

Dentre as pesquisas que investigaram outros documentos, destacam-se os discursos do Banco Central, que foram objeto de pesquisa em dois artigos de Anand, Basu, Pathak e Thampy (2021) e no artigo de Möller e Reichmann (2021). Foi examinada a linguagem usada pelos funcionários do Banco Central em conferências de imprensa públicas e como esse fator influencia os retornos das ações na área do euro, utilizando o dicionário de Loughran e McDonald (2011). Os resultados mostraram que o sentimento de incerteza tem um efeito positivo significativo nos retornos intradiários do índice das 50 principais empresas da zona do euro antes da crise financeira global, ou seja, a incerteza é interpretada como um sinal de acomodação política futura, fazendo com que o mercado reaja positivamente (Möller & Reichmann, 2021).

O tom textual das informações qualitativas dos relatórios de analistas também demonstrou poder explicativo sobre os retornos anormais de 12 meses, evidenciado na pesquisa de Caylor, Cecchini e Winchel (2017). Da mesma forma, a divulgação de informações qualitativas de risco de empresas que são clientes pode afetar negativamente o retorno das ações das empresas que são os fornecedores em estudo realizado mercado chinês (Li, He, Chan, 2021), corroborando com a ideia de que há influência do sentimento textual no retorno das ações em documentos externos à empresa.

Outro documento cujo tom textual pode interferir nas variações nos preços das ações são os relatórios de *rating* de crédito. Agarwal, Chen e Zhang (2016) identificaram que o tom linguístico líquido (tom negativo menos o tom positivo) nos relatórios está significativa e negativamente relacionado a retornos anormais das ações, além de ser capaz de prever mudanças futuras no *rating*. Por outro lado, a evidência empírica encontrada na pesquisa de Kiesel (2021), que foi realizada com empresas listadas nos Estados Unidos e na Europa, não traz conclusões claras sobre se as atualizações de *rating* realmente são capazes de afetar o retorno das ações.

Dessa forma, no que se refere aos relatórios de *rating* de crédito, novas pesquisas poderiam explorar a medição do sentimento textual de formas alternativas, buscando resolver esse impasse sobre a possível influência do tom deste relatório nos preços das ações. Além disso, todas as pesquisas que investigaram os relatórios de *rating* de crédito, os relatórios de analistas e os discursos do Banco Central estão concentradas no mercado norte-americano ou europeu. Isso faz com que pesquisas futuras possam obter evidências empíricas sobre a relação entre o tom textual e o retorno das ações nos documentos citados em mercados emergentes, para fins de comparação.

5. Considerações finais

Esta pesquisa de revisão sistemática fornece uma visão geral da evolução das pesquisas sobre a abordagem de análise textual como ferramenta de predição do retorno das ações no mercado de capitais. Para tanto, 78 estudos da temática publicados em inglês, divididos em 50 periódicos, foram analisados por meio do software VosViewer® e classificados em categorias que representam os documentos nos quais a análise textual foi aplicada.

Por meio da classificação, identificaram-se quatro principais fontes de informação para a realização da análise textual: notícias financeiras (31), divulgações corporativas (29), mídias sociais (16) e a quarta fonte, que representa outros documentos (8). A partir das categorias, foram compiladas as principais evidências encontradas em cada uma delas, e indicadas possíveis direções futuras a serem seguidas em novas pesquisas.

Ao analisar a amostra, constatou-se que, para capturar o sentimento textual dos documentos, a maioria dos artigos optou pela utilização de um dicionário classificando as palavras de acordo com o tom; o dicionário de Loughran e McDonald (2011) foi preponderante, encontrado em 36 dos 78 artigos analisados. Em se tratando da *proxy* aplicada para capturar as reações do mercado, a mais utilizada foi o retorno bruto da ação. Os estudos incluídos nesta revisão foram realizados principalmente no mercado de capitais norte-americano, representando cerca de 58% dos artigos analisados.

As pesquisas que se encontram classificadas em notícias financeiras e mídias sociais são as que possuem maior representatividade na amostra (47 artigos) e ainda apresentam maior variedade de métodos, amostras e particularidades, quando comparadas às demais categorias. Dessa forma, as notícias financeiras juntamente com as mídias sociais representam as principais fontes de informações para a aplicação da análise textual em finanças, quando o interesse é associar o tom textual com as reações do mercado.

Quanto às demais categorias, há os documentos que envolvem as divulgações realizadas pelas próprias companhias, podendo ser voluntárias (teleconferências de resultado), ou obrigatórias (relatórios 10-K, 8-K e documentos pré-IPO). Nestes casos, é visível algumas diferenças desses estudos, demonstrando que mais pesquisas são necessárias para explorar as diversas lacunas deixadas. Além disso, investigar países em desenvolvimento é uma oportunidade para estudos futuros, bem como estudos comparativos entre países desenvolvidos e em desenvolvimento.

Há, ainda, as pesquisas que foram classificadas em outros documentos (8), que se referem a documentos que não foram divulgados pelas próprias companhias e nem se referem às notícias financeiras e mídias sociais. Nessa categoria, destaca-se o sentimento dos discursos do Banco Central, que se mostrou relacionado com o retorno das ações. No entanto, apenas 3 pesquisas foram realizadas, todas no ambiente europeu, deixando a possibilidade de investigações adicionais acerca da relação dessa divulgação com o preço das ações.

As evidências encontradas nos artigos analisados demonstram que, no geral, o sentimento textual é capaz de prever o retorno das ações, sendo capturado das mais diversas fontes de informação. Diante disso, a HME não é capaz de explicar tais resultados, considerando que esta teoria tem como premissa a racionalidade dos investidores. No entanto, a HME ainda pode ser testada em condições diferentes às utilizadas nos estudos da amostra.

Esta pesquisa contribui para a disseminação da técnica de revisão sistemática em pesquisas desenvolvidas na área de finanças e contabilidade, pois, embora exista um crescimento das publicações, a aplicação da técnica não é disseminada nesta área do conhecimento. No âmbito acadêmico, a pesquisa contribui para sistematizar o estado da arte na temática e pode ser útil àqueles que pretendem realizar estudos semelhantes futuramente. Como implicações para a prática, investidores podem se informar sobre como o sentimento textual pode afetar seus investimentos em ações.

Como limitações desta pesquisa, em primeiro lugar, destaca-se o uso de artigos publicados na língua inglesa; sendo assim, não foram inclusos estudos em outros idiomas. Em segundo lugar, o uso exclusivo da *Web of Science* como base de dados pode ter omitido alguma pesquisa relevante que não tenha sido indexada nessa plataforma. Consequentemente, recomenda-se a realização de pesquisas adicionais para investigar outras bases de dados e contemplar também publicações em outro idioma. Além disso, novos estudos poderiam explorar esta temática realizando a categorização das pesquisas, considerando as regiões geográficas, o que poderia trazer novos *insights* e enriquecer o debate sobre a influência do sentimento textual no retorno das ações em diferentes contextos regionais.

Referências

- Agarwal, S., Chen, V. Y. S., & Zhang, W. (2016). The information value of credit rating action reports: a textual analysis. *Management Science*, 62(8), 2218-2240. Recuperado de <https://doi.org/10.1287/mnsc.2015.2243>
- Ahmad, K., Han, J., Hutson, E., Kearney, C., & Liu, S. (2016). Media-expressed negative tone and firm-level stock returns. *Journal of Corporate Finance*, 37, 152-172. Recuperado de <https://doi.org/10.1016/j.jcorpfin.2015.12.014>
- Ahmad, S., & Omar, R. (2016). Basic corporate governance models: a systematic review. *International Journal of Law and Management*, 58(1), 73-107. Recuperado de <https://doi.org/10.1108/ijlma-10-2014-0057>
- Anand, A., Basu, S., Pathak, J., & Thampy, A. (2021). The impact of sentiment on emerging stock markets. *International Review of Economics & Finance*, 75(1), 161-177. Recuperado de <https://doi.org/10.1016/j.iref.2021.04.005>
- Anand, A., Basu, S., Pathak, J., & Thampy, A. (2021). Whose speeches impact European markets: ECB's or the national central banks? *European Financial Management*, 28(5), 1413-1476. Recuperado de <https://doi.org/10.1111/eufm.12334>
- Ball, R. (1992). The earnings-price anomaly. *Journal of Accounting and Economics*, 15(2-3), 319-345. Recuperado de [https://doi.org/10.1016/0165-4101\(92\)90023-u](https://doi.org/10.1016/0165-4101(92)90023-u)
- Bian, S., Jia, D., Li, R., Sun, W., Yan, Z., & Zheng, Y. (2021). Can management tone predict IPO performance? – Evidence from mandatory online roadshows in China. *Pacific-Basin Finance Journal*, 68(1), 101588. Recuperado de <https://doi.org/10.1016/j.pacfin.2021.101588>
- Boudoukh, J., Feldman, R., Kogan, S., & Richardson, M. (2018). Information, trading, and volatility: evidence from firm-specific news. *The Review of Financial Studies*, 32(3), 992-1033. Recuperado de <https://doi.org/10.1093/rfs/hhy083>
- Box, T. (2018). Qualitative similarity and stock price comovement. *Journal of Banking & Finance*, 91, 49-69. Recuperado de <https://doi.org/10.1016/j.jbankfin.2018.04.010>
- Brau, J. C., Cicon, J., & McQueen, G. (2016). Soft strategic information and IPO underpricing. *Journal of Behavioral Finance*, 17(1), 1-17. Recuperado de <https://doi.org/10.1080/15427560.2016.1133619>
- Broadstock, D. C., & Zhang, D. (2019). Social-media and intraday stock returns: the pricing power of sentiment. *Finance Research Letters*, 30, 116-123. Recuperado de <https://doi.org/10.1016/j.frl.2019.03.030>
- Brochet, F., Kolev, K., & Lerman, A. (2018). Information transfer and conference calls. *Review of Accounting Studies*, 23(3), 907-957. Recuperado de <https://doi.org/10.1007/s11142-018-9444-4>
- Brock, W., Lakonishok, J., & Lebaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *The Journal of Finance*, 47(5), 1731-1764. Recuperado de <https://doi.org/10.1111/j.1540-6261.1992.tb04681.x>
- Brockman, P., Li, X., & Price, S. M. (2017). Conference Call tone and stock returns: evidence from the stock exchange of Hong Kong. *Asia-Pacific Journal of Financial Studies*, 46(5), 667-685. Recuperado de <https://doi.org/10.1111/ajfs.12186>
- Buehlmaier, M. M. M., & Whited, T. M. (2018). Are financial constraints priced? Evidence from textual analysis. *The Review of Financial Studies*, 31(7), 2693-2728. Recuperado de <https://doi.org/10.1093/rfs/hhy007>

- Campbell, J. L., Cecchini, M., Cianci, A. M., Ehinger, A. C., & Werner, E. M. (2018). Tax-related mandatory risk factor disclosures, future profitability, and stock returns. *Review of Accounting Studies*, 24(1), 264-308. Recuperado de <https://doi.org/10.1007/s11142-018-9474-y>
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1), 57-82. Recuperado de <https://doi.org/10.1111/j.1540-6261.1997.tb03808.x>
- Carlini, F., Cucinelli, D., Previtali, D., & Soana, M. G. (2020). Don't talk too bad! stock market reactions to bank corporate governance news. *Journal of Banking & Finance* (Vol. 121, p. 105962). Recuperado de <https://doi.org/10.1016/j.jbankfin.2020.105962>
- Caylor, M., Cecchini, M., & Winchel, J. (2017). Analysts' qualitative statements and the profitability of favorable investment recommendations. *Accounting, organizations and society*, 57, 33-51. Recuperado de <https://doi.org/10.1016/j.aos.2017.03.005>
- Chen, H., De, P., Hu, Y. (Jeffrey), & Hwang, B.-H. (2014). Wisdom of crowds: the value of stock opinions transmitted through social media. *Review of Financial Studies*, 27(5), 1367-1403. Recuperado de <https://doi.org/10.1093/rfs/hhu001>
- Chen, J., Demers, E., & Lev, B. (2018). Oh what a beautiful morning! Diurnal influences on executives and analysts: evidence from conference calls. *Management Science*, 64(12), 5899-5924. Recuperado de <https://doi.org/10.1287/mnsc.2017.2888>
- Chen, S., Liu, T., Peng, Q., & Zhao, Y. (2021). Manager Sentiment Bias and Stock Returns: Evidence from China. *Emerging Markets Finance and Trade*, 58(3), 823-836. Recuperado de <https://doi.org/10.1080/1540496x.2021.1918543>
- Cohen, L., Malloy, C., & Nguyen, Q. (2020). Lazy Prices. *The Journal of Finance*, 75(3), 1371-1415. Recuperado de <https://doi.org/10.1111/jofi.12885>
- Debata, B., Ghate, K., & Renganathan, J. (2021). COVID-19 pandemic sentiment and stock market behavior: evidence from an emerging market. *Review of Behavioral Finance*. Recuperado de <https://doi.org/10.1108/rbf-05-2021-008>
- Doukas, A. J., Guo, J. M., Lam, H. Y. T., & Xiao, S. H. (2014). Media endorsements of new product announcements: a new marketing strategy. *European Financial Management*, 22(3), 394-426. Recuperado de <https://doi.org/10.1111/eufm.12057>
- Eachempati, P., & Srivastava, P. R. (2021). Accounting for investor sentiment in news and disclosures. *Qualitative Research in Financial Markets*, 14(1), 53-75. Recuperado de <https://doi.org/10.1108/qrfm-01-2020-0006>
- Fama, E. F. (1970). Efficient capital markets: a review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417. Recuperado de <https://doi.org/10.1111/j.1540-6261.1970.tb00518.x>
- Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3-56. Recuperado de [https://doi.org/10.1016/0304-405x\(93\)90023-5](https://doi.org/10.1016/0304-405x(93)90023-5)
- Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1-22. Recuperado de <https://doi.org/10.1016/j.jfineco.2014.10.01>
- Feldman, R., Govindaraj, S., Livnat, J., & Segal, B. (2009). Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies*, 15(4), 915-953. Recuperado de <https://doi.org/10.1007/s11142-009-9111-x>
- Ferris, S. P., Hao, (Grace) Qing, & Liao, (Stella) Min-Yu. (2012). The effect of issuer conservatism on IPO pricing and performance. *Review of Finance*, 17(3), 993-1027. Recuperado de <https://doi.org/10.1093/rof/rfs018>

- Feuerriegel, S., & Prendinger, H. (2016). News-based trading strategies. *Decision Support Systems*, 90, 65–74. Recuperado de <https://doi.org/10.1016/j.dss.2016.06.020>
- Feuerriegel, S., & Pröllochs, N. (2018). Investor Reaction to Financial Disclosures across Topics: An Application of Latent Dirichlet Allocation. *Decision Sciences*, 52(3), 608–628. Recuperado de <https://doi.org/10.1111/dec.12346>
- Filip, A., Lobo, G. J., Paugam, L., & Stolowy, H. (2021). Disclosures About Key Value Drivers in M&A Announcement Press Releases: An Exploratory Study. *Abacus*, 58(1), 62–104. Recuperado de <https://doi.org/10.1111/abac.12234>
- Galdi, F. C., & Gonçalves, A. M. (2018). Pessimism and uncertainty of the news and investor's behavior in Brazil. *Revista de Administração de Empresas*, 58(2), 130-148. Recuperado de <https://doi.org/10.1590/s0034-759020180203>
- Gan, B., Alexeev, V., Bird, R., & Yeung, D. (2020). Sensitivity to sentiment: news vs social media. *International Review of Financial Analysis*, 67, 101390. Recuperado de <https://doi.org/10.1016/j.irfa.2019.101390>
- Garcia-Lopez, F. J., Batyrshin, I., & Gelbukh, A. (2018). Analysis of relationships between tweets and stock market trends [JB]. *Journal of Intelligent & Fuzzy Systems*, 34(5), 3337-3347. Recuperado de <https://doi.org/10.3233/JIFS-169515>
- Geva, T., & Zahavi, J. (2014). Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news. *Decision Support Systems*, 57, 212–223. Recuperado de <https://doi.org/10.1016/j.dss.2013.09.013>
- Glasserman, P., & Mamaysky, H. (2019). Does Unusual news forecast market stress? *Journal of Financial and Quantitative Analysis*, 54(5), 1937-1974. Recuperado de <https://doi.org/10.1017/s0022109019000127>
- Griffith, J., Najand, M., & Shen, J. (2019). Emotions in the stock market. *Journal of Behavioral Finance*, 21(1), 42-56. Recuperado de <https://doi.org/10.1080/15427560.2019.1588275>
- Gu, W., Zhang, L., Xi, H., & Zheng, S. (2021). Stock Prediction Based on News Text Analysis. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 25(5), 581–591. Recuperado de <https://doi.org/10.20965/jaciii.2021.p058>
- Hájek, P. (2017). Combining bag-of-words and sentiment features of annual reports to predict abnormal stock returns. *Neural Computing and Applications*, 29(7), 343-358. Recuperado de <https://doi.org/10.1007/s00521-017-3194-2>
- Hassan, T. A., Hollander, S., van Lent, L., & Tahoun, A. (2019). Firm-level political risk: measurement and effects. *The Quarterly Journal of Economics*, 134(4), 2135-2202. Recuperado de <https://doi.org/10.1093/qje/qjz021>
- Hatefi Ghahfarrokhi, A., & Shamsfard, M. (2020). Tehran stock exchange prediction using sentiment analysis of online textual opinions. *Intelligent Systems in Accounting, Finance and Management*, 27(1), 22-37. Recuperado de <https://doi.org/10.1002/isaf.1465>
- He, Y., Qu, L., Wei, R., & Zhao, X. (2021). Media-based investor sentiment and stock returns: a textual analysis based on newspapers. *Applied Economics*, 54(7), 774–792. Recuperado de <https://doi.org/10.1080/00036846.2021.1966369>
- Hendricks, B. E., Lang, M., & Merkley, K. (2021). Through the eyes of the founder: CEO characteristics and firms' regulatory filings. *Journal of Business Finance & Accounting*, 49(3–4), 383–422. Recuperado de <https://doi.org/10.1111/jbfa.12572>
- Hillert, A., Jacobs, H., & Müller, S. (2018). Journalist disagreement. *Journal of Financial Markets*, 41, 57-76. Recuperado de <https://doi.org/10.1016/j.finmar.2018.09.002>

- Houlihan, P., & Creamer, G. G. (2017). Can sentiment analysis and options volume anticipate future returns? *Computational Economics*, 50(4), 669-685. Recuperado de <https://doi.org/10.1007/s10614-017-9694-4>
- Huang, A. G., Tan, H., & Wermers, R. (2020). Institutional Trading around Corporate News: Evidence from Textual Analysis. *The Review of Financial*, 33(10), 4627-4675. Recuperado de <https://doi.org/10.1093/rfs/hhz136>
- Ibriyamova, F., Kogan, S., Salganik-Shoshan, G., & Stolin, D. (2016). Using semantic fingerprinting in finance. *Applied Economics*, 49(28), 2719-2735. Recuperado de <https://doi.org/10.1080/00036846.2016.1245844>
- Ibriyamova, F., Kogan, S., Salganik-Shoshan, G., & Stolin, D. (2018). Predicting stock return correlations with brief company descriptions. *Applied Economics*, 51(1), 88-102. Recuperado de <https://doi.org/10.1080/00036846.2018.1494377>
- Jiang, F., Lee, J., Martin, X., & Zhou, G. (2019). Manager sentiment and stock returns. *Journal of Financial Economics*, 132(1), 126-149. Recuperado de <https://doi.org/10.1016/j.jfineco.2018.10.001>
- Kahneman, D. (1994). New Challenges to the Rationality Assumption. *Journal of Institutional and Theoretical Economics (JITE)*, 150(1), 18-36. Recuperado de <http://www.jstor.org/stable/40753012>
- Kearney, C., & Liu, S. (2014). Textual sentiment in finance: a survey of methods and models. *International Review of Financial Analysis*, 33, 171-185. Recuperado de <https://doi.org/10.1016/j.irfa.2014.02.006>
- Kiesel, F. (2021). It's the tone, stupid! Soft information in credit rating reports and financial markets. *Journal of Financial Research*, 44(3), 553-585. Recuperado de <https://doi.org/10.1111/jfir.12250>
- Li, J., Li, J., & Zhu, X. (2020). Risk dependence between energy corporations: a text-based measurement approach. *International Review of Economics & Finance*, 68, 33-46. Recuperado de [10.1016/j.iref.2020.02.009](https://doi.org/10.1016/j.iref.2020.02.009)
- Li, Q., Wang, J., Wang, F., Li, P., Liu, L., & Chen, Y. (2016). The role of social sentiment in stock markets: a view from joint effects of multiple information sources. *Multimedia Tools and Applications*, 76(10), 12315-12345. Recuperado de <https://doi.org/10.1007/s11042-016-3643-4>
- Li, S., Ning, K., & Zhang, T. (2021). Sentiment-aware jump forecasting. *Knowledge-Based Systems*, 228, 107292. Recuperado de <https://doi.org/10.1016/j.knosys.2021.107292>
- Li, X., Wu, P., & Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: a case of Hong Kong. *Information Processing & Management*, 57(5), 102212. Recuperado de <https://doi.org/10.1016/j.ipm.2020.102212>
- Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, 69, 14-23. Recuperado de <https://doi.org/10.1016/j.knosys.2014.04.022>
- Li, Y., He, J., & Chan, K. C. (2021). Information transmission along supply chains: Stock price reaction of suppliers upon a customer's release of qualitative risk information. *International Journal of Production Economics*, 239, 108189. Recuperado de <https://doi.org/10.1016/j.ijpe.2021.108189>
- Liu, S., & Han, J. (2020). Media tone and expected stock returns. *International Review of Financial Analysis*, 70, 101522. Recuperado de <https://doi.org/10.1016/j.irfa.2020.101522>
- Lo, A. W. (2004). The adaptive markets hypothesis. *The Journal of Portfolio Management*, 30(5), 15-29. Recuperado de <https://doi.org/10.3905/jpm.2004.442611>
- 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. Recuperado de <https://doi.org/10.1111/j.1540-6261.2010.01625.x>

- Loughran, T., & McDonald, B. (2011b). Barron's Red Flags: Do They Actually Work? *Journal of Behavioral Finance*, 12(2), 90-97. Recuperado de <https://doi.org/10.1080/15427560.2011.575971>
- Loughran, T., & McDonald, B. (2013). IPO first-day returns, offer price revisions, volatility, and form S-1 language. *Journal of Financial Economics*, 109(2), 307-326. Recuperado de <https://doi.org/10.1016/j.jfineco.2013.02.017>
- Loughran, T., & McDonald, B. (2020). Textual Analysis in Finance. *Annual Review of Financial Economics*, 12(1), 357-375. Recuperado de <https://doi.org/10.1146/annurev-financial-012820-032249>
- Lutz, B., Pröllochs, N., & Neumann, D. (2020). Predicting sentence-level polarity labels of financial news using abnormal stock returns. *Expert Systems with Applications*, 148, 113223. Recuperado de <https://doi.org/10.1016/j.eswa.2020.113223>
- Ly, Y., Piao, J., Li, B., & Yang, M. (2021). Does online investor sentiment impact stock returns? Evidence from the Chinese stock market. *Applied Economics Letters*, 29(15), 1434-1438. Recuperado de <https://doi.org/10.1080/13504851.2021.1937490>
- Malaquias, R. F., & Júnior, D. M. B. (2021). Positive Tone in Management Reports and Volatility of Stock Returns. *Global Business Review*, 097215092110542. Recuperado de <https://doi.org/10.1177/09721509211054279>
- Mangee, N. (2018). Stock Returns and the Tone of Marketplace Information: does Context matter? *Journal of Behavioral Finance*, 19(4), 396-406. Recuperado de <https://doi.org/10.1080/15427560.2018.1405268>
- Maragoudakis, M., & Serpanos, D. (2015). Exploiting financial news and social media opinions for stock market analysis using MCMC Bayesian Inference. *Computational Economics*, 47(4), 589-622. Recuperado de <https://doi.org/10.1007/s10614-015-9492-9>
- Marquez-Illescas, G., Zebedee, A. A., & Zhou, L. (2018). Hear Me Write: does CEO narcissism affect disclosure? *Journal of Business Ethics*, 159(2), 401-417. Recuperado de <https://doi.org/10.1007/s10551-018-3796-3>
- Milian, J. A., & Smith, A. L. (2017). An investigation of analysts' praise of management during earnings conference calls. *Journal of Behavioral Finance*, 18(1), 65-77. Recuperado de <https://doi.org/10.1080/15427560.2017.1276068>
- Möller, R., & Reichmann, D. (2021). ECB language and stock returns – A textual analysis of ECB press conferences. *The Quarterly Review of Economics and Finance*, 80, 590-604. Recuperado de <https://doi.org/10.1016/j.qref.2021.04.003>
- Myšková, R., & Hájek, P. The effect of managerial sentiment on market to-book ratio. *Transformations in Business & Economics*, 15(2A), 498-513. Recuperado de <https://dk.upce.cz/handle/10195/67653>
- Myšková, R., Hájek, P., & Olej, V. (2017) Predicting abnormal stock return volatility using textual analysis of news – A meta-learning approach. *Amfiteatru Economic*, 20(47), 185-201. Recuperado de <https://www.ceeol.com/search/article-detail?id=678584>
- Price, S. M., Doran, J. S., Peterson, D. R., & Bliss, B. A. (2012). Earnings conference calls and stock returns: the incremental informativeness of textual tone. *Journal of Banking & Finance*, 36(4), 992-1011. Recuperado de <https://doi.org/10.1016/j.jbankfin.2011.10.013>
- Renault, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking & Finance*, 84, 25-40. Recuperado de <https://doi.org/10.1016/j.jbankfin.2017.07.002>
- Rozeff, M. S., & Kinney, W. R., Jr. (1976). Capital market seasonality: the case of stock returns. *Journal of Financial Economics*, 3(4), 379-402. Recuperado de [https://doi.org/10.1016/0304-405x\(76\)90028-3](https://doi.org/10.1016/0304-405x(76)90028-3)

- Sampaio, R.F., & Mancini, M.C. (2007). Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. *Revista Brasileira de Fisioterapia*, 11(1),83-89. Recuperado de <https://www.redalyc.org/articulo.oa?id=2350/235016477013>
- Sampath, V. S., O'Connor, A. J., & Legister, C. (2021). Moral leadership and investor attention: An empirical assessment of the potus's tweets on firms' market returns. *Review of Quantitative Finance and Accounting*, 58(3), 881–910. Recuperado de <https://doi.org/10.1007/s11156-021-01012-0>
- Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news. *ACM Transactions on Information Systems*, 27(2), 1-19. Recuperado de <https://doi.org/10.1145/1462198.1462204>
- Seng, J.-L., & Yang, H.-F. (2017). The association between stock price volatility and financial news – a sentiment analysis approach. *Kybernetes*, 46(8), 1341-1365. Recuperado de <https://doi.org/10.1108/k-11-2016-0307>
- Sun, L., Najand, M., & Shen, J. (2016). Stock return predictability and investor sentiment: a high-frequency perspective. *Journal of Banking & Finance*, 73, 147–164. Recuperado de <https://doi.org/10.1016/j.jbankfin.2016.09.010>
- Suslava, K. (2021). “Stiff Business Headwinds and Uncharted Economic Waters”: The Use of Euphemisms in Earnings Conference Calls. *Management Science*, 67(11), 7184–7213. Recuperado de <https://doi.org/10.1287/mnsc.2020.3826>
- Takahashi, S., Takahashi, M., Takahashi, H., & Tsuda, K. (2006, october). Analysis of stock price return using textual data and numerical data through text mining. In *Lecture Notes in Computer Science*. Recuperado de https://doi.org/10.1007/11893004_40
- Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of science and scopus. *Journal of Informetrics*, 10(2), 347-364. Recuperado de <https://doi.org/10.1016/j.joi.2016.02.003>
- Wu, C.-H., & Lin, C.-J. (2017). The impact of media coverage on investor trading behavior and stock returns. *Pacific-Basin Finance Journal*, 43, 151-172. Recuperado de <https://doi.org/10.1016/j.pacfin.2017.04.001>
- Xu, K., Pang, Y., & Han, J. (2021). Dynamic Cross-Correlation between Online Sentiment and Stock Market Performance: A Global View. *Discrete Dynamics in Nature and Society*, 2021, 1–11. Recuperado de <https://doi.org/10.1155/2021/6674379>
- Yan, Y., Xiong, X., Meng, J. G., & Zou, G. (2019). Uncertainty and IPO initial returns: evidence from the tone analysis of China's IPO prospectuses. *Pacific-Basin Finance Journal*, 57, 101075. Recuperado de <https://doi.org/10.1016/j.pacfin.2018.10.004>
- Yang, A. S. (2020). Misinformation corrections of corporate news: corporate clarification announcements. *Pacific-Basin Finance Journal*, 61, 101315. Recuperado de <https://doi.org/10.1016/j.pacfin.2020.101315>
- Zhang, W., Wang, M., Zhu, Y., Wang, J., & Ghei, N. (2019). A hybrid neural network approach for fine-grained emotion classification and computing [JB]. *Journal of Intelligent & Fuzzy Systems*, 37(3), 3081-3091. Recuperado de <https://doi.org/10.3233/JIFS-179111>
- Zhao, W., Zhang, W., Xiong, X., & Zou, G. (2019). Share pledges, tone of earnings communication conferences, and market reaction: evidence from China. *Accounting & Finance*, 59(5), 2817-2853. Recuperado de <https://doi.org/10.1111/acfi.12585>