



## A Influência da Mídia Social no Valor da Petrobras

**MILENE DIAS ALMEIDA**

*Universidade Federal do Rio de Janeiro*

**VINICIUS MOTHÉ MAIA**

*Universidade Federal do Rio de Janeiro*

**ROBERTO TOMMASETTI**

*Universidade Federal do Rio de Janeiro*

### Resumo

A presente pesquisa busca contribuir com a discussão do sentimento textual sobre a organização. O objetivo do trabalho foi analisar a relação entre o sentimento exposto no Twitter e o retorno da Petrobras. Desse modo, *tweets* sobre a Petrobras, no período de 2010 até 2020, foram analisados a fim de se verificar se o sentimento textual dessas mensagens impacta no retorno. O método escolhido para classificar as palavras foi o mapeamento perceptual conjuntamente com a média e desvio padrão da frequência dos termos em dias positivos ou negativos. Três grupos de regressões foram criadas. O primeiro analisou o impacto do sentimento no retorno, o segundo verificou o relacionamento do sentimento positivo e negativo no retorno e o terceiro analisou se o sentimento amplifica ou reduz o movimento do mercado em relação ao retorno. Os resultados apontaram que a variação do sentimento total e a variação do sentimento negativo tem relação com o retorno. A variação do sentimento é significativa quando o mercado está pessimista, demonstrando que a variação negativa do sentimento intensifica o movimento pessimista do mercado e ameniza a queda do mercado quando a variação do sentimento é positiva. Os achados corroboram com a teoria de finanças comportamentais de que o sentimento possui relacionamento com o retorno, no caso do estudo, o sentimento textual. Por fim, foram analisadas diferentes amostras com a mesma metodologia. Com isso, foi possível verificar que os indivíduos falam mais sobre a estatal, de modo geral, do que sobre os *tickers* da ação ou com a própria empresa na sua página oficial.

**Palavras-chave:** Sentimento Textual, Mídias Sociais, Petrobras.



São Paulo 28 a 30 de julho 2021.

## 1. INTRODUÇÃO

As noções de valor, fama e reputação de pessoas, antes da imprensa, eram muito vagas. Com o surgimento da imprensa, as ideias se tornam mais difundidas e claras, se tornando um poderoso agente de imitação. A publicação de cópias de julgamentos individuais semelhantes sobre mudanças no valor de mercado é chamada de opinião. Esse fator pode determinar a glória de um homem ou de um bem (Tarde, 1902). As emoções e humores dos indivíduos afetam seu processo de tomada de decisão, levando a uma relação entre “sentimento público” e comportamento do mercado acionário (Nofsinger, 2010).

No contexto da era da internet, os consumidores estão cada vez mais olhando para as mídias sociais de uma determinada marca para formar julgamento sobre a mesma (Naylor, Lamberton & West, 2012). As mídias sociais são um meio de compartilhar essas emoções públicas que têm impacto na opinião pública geral e nos mercados. Pesquisas anteriores relacionadas a análise de sentimento por meio das mídias sociais se dividem em duas categorias. A primeira, classifica os dados gerais de forma lexical e a segunda, por meio de instrumentos automatizados (Birmingham & Smeaton, 2010; Go, Bhayani & Huang, 2009).

Na primeira categoria, os sentimentos são classificados manualmente, como o léxico de sentimento Loughran – McDonald Financial Dictionary. Esse dicionário é uma abordagem específica para conteúdo financeiro e foi feito manualmente por Loughran e McDonald (2011). A segunda categoria possui o objetivo de classificar os sentimentos por meio de *machine learning*. Nessa categoria, um dos métodos mais usados é o chamado Classificadores de Bayes, onde os dados são rotulados e valores negativos, positivos ou neutros são atribuídos à opinião das pessoas (Alkubaisi, Kamaruddin & Husni, 2018; Rao & Srivastava, 2012). Outro método bastante utilizado é o OpinionFinder (Bollen & Mao, 2011; Oliveira, Cortez & Areal, 2013), SentiWordNet, SenticNet (Li, Wu & Wang, 2020).

Dicionário de sentimento específico de finanças modela melhor os novos sentimentos do que outros (Li, Wu & Wang, 2020). Uma lista de palavras gerais não é aplicável a textos financeiros. Loughran e McDonald (2011) verificaram que 73,8% da contagem de palavras negativas na lista de palavras sobre emoções de domínio aberto Harvard-IV-4 não são consideradas negativas no contexto financeiro. É importante que o dicionário de palavras seja na mesma língua que o texto analisado, pois a simples tradução da lista em inglês poderia levar a um resultado errôneo da análise de sentimento (Bannier, Pauls & Walter, 2019).

A relação entre sentimento do investidor e sentimento textual é complexa. O último pode incluir o primeiro, mas também demonstra as condições dentro das empresas e mercado. O presente estudo buscou medir o sentimento textual, demonstrando o efeito positivo ou negativo das palavras. Com isso, surge o questionamento: *Qual o efeito do sentimento expresso via Twitter sobre os retornos no mercado acionário brasileiro?*

O objetivo do estudo é analisar a relação entre o sentimento exposto no Twitter e o retorno das ações. Para alcançar o objetivo geral, foram estabelecidos os seguintes objetivos específicos: (i) criar uma lista de termos/palavras a partir das publicações no Twitter, a fim de construir um dicionário de palavras associadas ao sentimento do investidor; (ii) estimar um índice de sentimento baseado no dicionário criado anteriormente pelo método do mapa perceptual; (iii) verificar se este índice de sentimento está relacionado ao comportamento dos preços no mercado acionário brasileiro; e, (iv) um teste de robustez do modelo analisando se likes e compartilhamentos afetam o resultado.

Pesquisadores utilizaram vários classificadores para realizar análises de sentimento no mercado de ações com o Twitter como fonte de dados. Este estudo realiza a análise de sentimento com o método de mapeamento perceptual para classificar os *tweets*, verificando



São Paulo 28 a 30 de julho 2021.

quais palavras foram classificadas como positivas ou negativas. O modelo é uma maneira gráfica de representar como os indivíduos analisam os produtos ou marcas do mercado analisado (Hair Jr et al., 2009), no caso, a empresa analisada foi a Petrobrás.

A instituição é uma empresa de referência nacional e está entre as marcas que mais foi percebida por alguma publicidade, notícias ou propaganda “boca a boca” na internet no ano de 2019 em relação ao ano anterior. A Petrobrás foi a empresa que teve o maior salto de popularidade (“Buzz Rankings de 2019: Brasil | YouGov - BrandIndex”, 2020). A empresa teve uma liderança isolada por 4 anos, de 2016 até 2019, no Top of Mind na categoria Marca que Representa o Brasil (Pinto, 2019).

A escolha também foi feita pelo fato do nome “Petrobrás” não ser um nome homônimo, o que auxilia na obtenção dos dados, pois se fosse outra empresa como a Oi, por exemplo, todos os tweets com a palavra “oi” seriam coletadas e poderiam surgir mensagens que não são relacionadas com a empresa, e sim, com um cumprimento. Por esses motivos, popularidade, relevância e não ser um homônimo, foi realizada a escolha da empresa no presente estudo. Desse modo, o trabalho buscou entender o sentimento em relação à Petrobrás e como se relaciona com o seu retorno.

Dessa forma, espera-se identificar padrões de comportamento atrelados aos movimentos de oscilação do mercado, em relação ao retorno, e com isso, contribuir com o gerenciamento de risco e a otimização da tomada de decisão no mercado acionário do Brasil. O diferencial do estudo é criação de uma lista de termos financeiros com base no idioma oficial do país a partir de uma nova metodologia. A escolha de *tweets* como amostra foi feita pois reduz problemas de metodologias qualitativas como *survey*. Visto que ao ser questionado, o entrevistado pode omitir dados ao acreditar que seria inadequado dizer algo que o entrevistador não perguntou diretamente por exemplo (Bauer & Gaskell, 2003). Na mídia social o indivíduo se expressa sem estímulo, comenta o que pensava sem influência de um entrevistador.

O estudo se mostra relevante, pois as palavras são um componente importante dos sistemas de linguagem natural. O modelo semântico usa as palavras para prever sentimentos. A análise de sentimentos oferece às organizações uma maneira rápida de monitorar os sentimentos do público em relação à marca, negócios, diretores, etc. Isso auxiliará não apenas a atrair novos investidores ou clientes, mas também a manter os atuais.

A criação de mapas perceptuais pode permitir prever ganhos ou perdas futuras no valor da empresa, visto que uma crise na imagem custa caro. A desconfiança do mercado sobre algum aspecto pode diminuir o valor das empresas. Com a construção desses mapas pode surgir um poderoso instrumento de gerenciamento de marketing e análise do sentimento do investidor sobre a organização.

## 2. REFERENCIAL TEÓRICO

Um dos primeiros estudos que analisa a relação entre Twitter e o mercado de ações foi o de Bollen, Mao e Zeng (2011). O objetivo do estudo foi verificar se o humor coletivo exposto no Twitter estava correlacionado com o valor da Média Industrial Dow Jones. O *OpinionFinder* e o perfil do Google dos estados de humor foram usados para classificar os textos da mídia social e o modelo estatístico foi a causalidade de Granger. Foi escolhido dois períodos para fazer os testes da resposta do público: eleições presidenciais e dia de ação de graças de 2008. Como resultado, não foi vista correlação pelo método *OpinionFinder*, mas foi visto pela dimensão “calma” do Google. Assim, “calma” é um preditivo dos valores da bolsa. Os autores também buscaram verificar se a mídia social, Twitter, exerce influência nos retornos financeiros. Nesse caso, os autores aumentaram a amostra e incluíram mais variáveis no modelo



São Paulo 28 a 30 de julho 2021.

de causalidade de Granger, notícias e consultas de pesquisas de opinião. Como resultado, foi verificado que pesquisas tradicionais não tem poder preditivo do mercado. Contudo, o Twitter e a frequência de termos financeiros no Twitter nos dois dias anteriores foram considerados estatisticamente preditores do retorno diário das ações.

Bartov, Faurel e Moharam (2018) analisaram se a opinião dos indivíduos expostos no Twitter entre 2009 e 2012 podem prever seus ganhos e retornos. Os autores utilizaram o modelo bayesiano como forma de mensurar os *tweets* e duas regressões como metodologia para analisar se a opinião impactava os ganhos e os retornos das ações. Além disso, foi feita uma regressão para analisar as empresas que buscavam transmitir informações sólidas por meio dos seus canais de transmissão, como cobertura da imprensa, relatórios de analistas e comunicados à imprensa. Como resultado, foi verificado que o sentimento do investido no Twitter foi significativo nas regressões, demonstrando assim, que é importante analisar o sentimento do investidor expostos no Twitter para tentar prever o valor futuro das ações das empresas, mesmo quando em ambiente de assimetria informacional.

A pesquisa de Pengnate e Riggins (2020) analisou se indicadores de características de qualidade textual cognitiva em descrições de empréstimos - número de palavras, legibilidade, erros de ortografia e erros gramaticais- influência no sucesso do financiamento. Os dados foram obtidos por meio de uma ferramenta de análise de texto online (Readable) e depois foi feita a classificação manual do número de palavras, legibilidade e erros ortográficos a fim de verificar se a ferramenta era confiável. Para detectar as emoções foi utilizado o aplicativo Warson NLU disponível na IBM Cloud Computing PaaS. O aplicativo também classificou as mensagens em positivas, neutras e negativas. Foi verificado que na ausência da taxa de juros, as empresas podem aumentar o sucesso do financiamento melhorando a qualidade textual, sendo assim, mais atraentes para os credores.

McGurk, Nowak e Hall (2020) e Seng e Yang (2017) também fizeram uma análise subjetiva e manual para classificação textual, sendo o objetivo verificar se os textos financeiros influenciam nos retornos das ações. O dicionário criado teve efeito positivo e significativo, consistente com outros modelos estudados. Outro método é a análise conjunta utilizando a semântica de notícias, relatórios ou cartas do CEO e a automatizada (Yadav et al., 2020).

Ito et al. (2017) apresentaram uma nova forma de desenvolver o indicador de sentimento textual denominado GAP, esse que realiza a análise morfológica das palavras. Os autores mostraram que os preços das ações têm relação com as postagens do Yahoo Finance e o GAP demonstrou mais fortemente essa relação do que outros indicadores. Também é possível verificar um impacto do sentimento textual a partir da análise léxica das palavras (Uhr, Zenkert & Fathi, 2014). Mao et al. (2014) também realizaram uma análise baseada em características linguísticas para verificar se a notícia é um fator importante para o movimento do mercado. As etapas de classificação do sentimento são de acordo com a classe gramatical: substantivo, adjetivo e adverbio. O sentimento textual teve efeito significativo sobre o movimento do mercado.

Li, Wu e Wang (2020) e Mangee (2018) analisaram informações financeiras e o impacto no retorno das ações. Diferentes dicionários de palavras financeiras foram comparados. Entre os dicionários de sentimento, o dicionário de sentimento específico do domínio de finanças (Loughran-McDonald Financial Dictionary) modela melhor os novos sentimentos, possui uma relação mais forte com o comportamento do preço das ações. Loughran e McDonald (2015) verificaram que o dicionário Loughran-McDonald (2011) é melhor para a captura de tons de textos sobre negócios do que o Diction, que é uma plataforma da literatura de finanças.

Outras listas foram criadas pelo dicionário Loughran-McDonald Financial, sendo que adaptando a lista ao idioma do texto, como fez Bannier, Pauls e Walter (2019) e outros



São Paulo 28 a 30 de julho 2021.

trabalhos, em língua inglesa, aplicaram o mesmo dicionário mas para amostras diferentes, como documentos financeiros (Gandhi, Loughran & McDonald, 2019; Hájek, 2018), mensagens de mídias sociais (Houlihan & Creamer, 2017) e até em documentos que a análise textual foi capaz de prever falência de bancos (Gupta, Simaan & Zaki, 2017).

O aumento do risco, regulação crise financeira e características das empresas possuem relação com os atributos textuais dos relatórios financeiros. Esse resultado foi obtido pela classificação dos termos em clichê, subjetividade, redundância, especificidade, nebulosidade e fator de risco (Li et al., 2020)

A pontuação do sentimento textual pela análise de rede neural de documentos financeiros (Ito et al., 2019; Ahmadi et al., 2019; Ito et al., 2017), notícias financeiras (Sarigiannidis et al., 2018; Seng & Yan, 2017) e mídia social (Wang, 2017; Mai et al., 2018). Análise do sentimento deve ser feita pela união dos métodos (Darena et al., 2018).

O estudo de Ranco et al. (2015) constatou que o sentimento agregado expresso em *tweets* financeiros transmite informações sobre a avaliação das empresas. Só foram utilizados os *tweets* com *tag* de dinheiro sobre essas companhias. Os *tweets* foram rotulados com três sentimentos (negativo, neutro ou positivo) com base no *Support Vector Machine* que é um algoritmo de aprendizado supervisionado. Com o mesmo algoritmo, Fang et al. (2017) verificaram que os *tweets* possuem capacidade de previsão e Ab. Rahman, Abdul-Rahman e Mutalib (2017) verificaram que as notícias financeiras publicadas recentemente também possuem capacidade de previsão do retorno dos preços.

O estudo de Pagolu et al. (2016) concluiu que as mudanças nos preços das ações de uma empresa possuem relacionamento com as opiniões públicas expressas em *tweets*. Os métodos *word2vec* e *N-grama* foram utilizados para analisar o sentimento dos *tweets*. A conclusão é que postagens positivas no Twitter sobre uma empresa incentivaram as pessoas a investir nas ações dessa empresa e conseqüentemente, os preços das ações aumentaram. Ito et al. (2016) também utilizaram o algoritmo *word2vec* para avaliar dados textuais do Yahoo! e como resultado, o índice teve significância para a previsão dos preços das ações.

Já Bollen, Mao e Zeng (2011) analisaram os estados emocionais do público por meio de *feeds* do Twitter para prever tendências. Os *tweets* foram classificados com positivos e negativos com o buscador de opinião e pelo algoritmo Google-Profile of Mood States, que classifica sentimento em seis classes: calmo, alerta, seguro, vigoroso, amistoso e feliz. O resultado do estudo demonstrou que a incorporação do sentimento do Twitter melhorou significativamente a precisão da previsão. Esses modelos possuem o diferencial de apresentar dimensões de humor que são, na maioria dos casos, ignorados pelos métodos de classificação tradicional.

Twitter de pessoas importantes para a sociedade também foram investigados. Foi verificado que os *tweets* do Presidente dos Estados Unidos possuem influência no mercado acionário americano, aumentando o volume e volatilidade dos preços além de atrair atenção do investidor. Contudo, esse efeito se dissipa em poucos dias (Ge, Kurov & Wolfe, 2018).

No Brasil, a análise de sentimento por meio das redes começou a ser abordado de forma lexical. Um dos primeiros estudos, analisou o sentimento nas opiniões de produtos escritos em português no Brasil. O método que se mostrou mais eficiente foi o Sentilex (Avanço & Nunes, 2014). Posteriormente, os *tweets* foram analisados junto com as notícias sobre o mercado financeiro em diversos sites. O objetivo era criar uma técnica de classificação das palavras por meio de um método já usado no idioma inglês, SentiPipe, adaptando para a língua portuguesa. Foi constatado que o método SentiPipe gerou resultados melhores do que a adaptação do método de língua inglesa (Martins, Pereira & Benevenuto, 2015).



São Paulo 28 a 30 de julho 2021.

O estudo de Souza, Lucena e Queiroz (2019) avaliou o sentimento do investidor por meio do Twitter sobre o mercado brasileiro durante o período eleitoral de 2018. Os dados analisados foram o retorno e volume negociado do IBOVESPA e os tweets postados durante o período eleitoral. Os *tweets* foram escolhidos pela língua, portuguesa ou inglesa, e que tivessem o nome de alguma das ações que fazem parte do índice. Os autores decidiram utilizar a biblioteca chamada Tweepy para classificar os posts. O resultado apontou a influência positiva dos tweets sobre os retornos do IBOVESPA e o volume negociado.

As técnicas de Machine Learning para análise de sentimento (Bayes, Suport Vector Machine, Entropia Máxima e multicamadas Perceptron) foram comparadas nos tweets postados durante o período eleitoral brasileiro. Os resultados indicam que a multicamada Perceptron é o melhor algoritmo para estabelecer relações entre sentimentos em português via *twitter* e o mercado de ações brasileiro. (Carosia, Coelho & Silva, 2019).

No Brasil, foi investigado o mesmo efeito no setor bancário para o Presidente Bolsonaro. Nesse caso, como resultado foi verificado que as postagens do Presidente no Twitter não possuem impacto no retorno ou no volume negociados das ações de empresas do setor bancário. Os autores do estudo argumentam que esse resultado foi provavelmente devido ao fato do setor bancário ser muito concentrado, pouco tempo de governo do Presidente e o conteúdo das postagens que podem ser interpretadas mais voltadas para o meio político do que ao mercado (Tommasetti, Maia & Macedo, 2020).

Além dos estudos anteriores, para analisar o sentimento do investidor via Twitter sobre mercado acionário brasileiro, foram utilizados outros métodos, como, o SentiPipe, com adaptação para a língua portuguesa (Martins, Pereira & Benevenuto, 2015) e o *machine learning*, Perceptron (Carosia, Coelho & Silva, 2019). Todos esses estudos observaram influências significativas.

A presente pesquisa procurou criar uma lista de termos/palavras a partir das publicações no Twitter, a fim de construir um dicionário de palavras em língua portuguesa associadas ao sentimento do investidor. Dados do mercado de capitais e da mídia social foram coletados para entender melhor o comportamento dos investidores e o que influencia em sua tomada de decisão. O modelo geral do estudo analisou a influência do sentimento textual do Twitter no retorno da Petrobrás, tendo em vista que o retorno do ativo possui influência do retorno do mercado presente e do sentimento textual presente como foi visto nos estudos citados anteriormente. Dessa forma, tem-se as hipóteses de pesquisa:

H1: O sentimento textual se relaciona com o retorno;

H2: O sentimento textual positivo e negativo se relacionam com o retorno; e,

H3: O sentimento textual amplifica ou reduz o movimento de mercado em relação ao retorno.

### 3. METODOLOGIA

A presente pesquisa buscou avaliar o efeito do sentimento expresso via Twitter sobre os retornos no mercado acionário brasileiro. Para alcançar esse objetivo, foram coletados os dados sobre o mercado acionário brasileiro e as mensagens expostas no Twitter entre 2010 e 2020. Para o referencial teórico, realizou-se uma revisão sistemática da literatura mediante busca na base de dados Scopus utilizando os termos “Sentimento textual” e “finanças” em inglês. Foram encontrados 57 artigos até o momento da pesquisa. Outros estudos também foram acrescentados para incrementar a revisão.



São Paulo 28 a 30 de julho 2021.

Para analisar o comportamento das ações da Petrobras, foi calculado o log retorno dos preços das ações da organização. O histórico da série diária de preços da Petrobras foi coletado na Bloomberg.

Para realizar a análise do sentimento exposto no Twitter foi necessário coletar as postagens acerca da Petrobras em pré-processamento dos dados. De acordo com Godsay (2015), para realizar essa análise é necessário fazer as seguintes etapas:

- i) definir o objetivo;
- ii) determinar a fonte de onde está tomando os dados da web. Coletar os dados e eliminar palavras indesejadas. Além disso, observou-se que para expressar sentimentos fortes, as letras maiúsculas do alfabeto são usadas;
- iii) analisar o conteúdo e classificar os termos de acordo com uma polaridade desejada;
- iv) Para garantir o refinamento, retirar as *StopWords*;
- v) Analisar e pontuar os termos. Nesse processo é onde a intensidade do sentimento é analisado.

Os *tweets* foram coletadas por meio da biblioteca *GetOldTweets3* no Python. Foram selecionadas as postagens que estão em língua portuguesa e continham o nome da instituição (Petrobras), o ticker da ação (PETR3 e PETR4) e as respostas as postagens da empresa na sua conta oficial (@petrobras) entre 2010 e 2020. A escolha do banco de dados foi feita com o propósito de avaliar qual informação seria mais importante para o investidor: sobre a Petrobrás, no geral, sobre a ação da instituição ou sobre as respostas direcionadas à empresa por meio de sua conta oficial no Twitter. Esse parâmetro para captar determinados *tweets* também foi utilizado nos estudos de Oliveira, Cortez e Areal (2013) e Souza, Lucena e Queiroz (2019), onde coletaram postagens em português ou inglês que tinham algum nome das instituições participantes do índice IBOVESPA e/ou com termos escolhidos pelos autores: IBOVESPA, IBOV, Brasil, Bolsa, Balcão e Bolsa de Valores Brasileira.

Depois da coleta dos *tweets*, cada palavra foi separada como uma *string* no R. Realizou-se uma limpeza nessa lista, tirando as *StopWords*, que são palavras que não possuem valor semântico. Por esse motivo, essas palavras foram removidas da amostra. Essa mineração de dados foi construída com algoritmos em Python contendo esse tratamento de limpeza nos dados.

Para a classificação das palavras que continuaram na amostra, depois da limpeza de dados, foi realizada uma análise léxica de cada palavra, indicando a polaridade. Se o sentimento foi positivo (1), negativo (-1) ou neutro (0). Um conjunto de palavras que expressam sentimento é denominado de léxico de sentimentos. Essas palavras podem pertencer a diferentes classes gramaticais (adjetivos, nomes, verbos, advérbios) (Avanço, 2015).

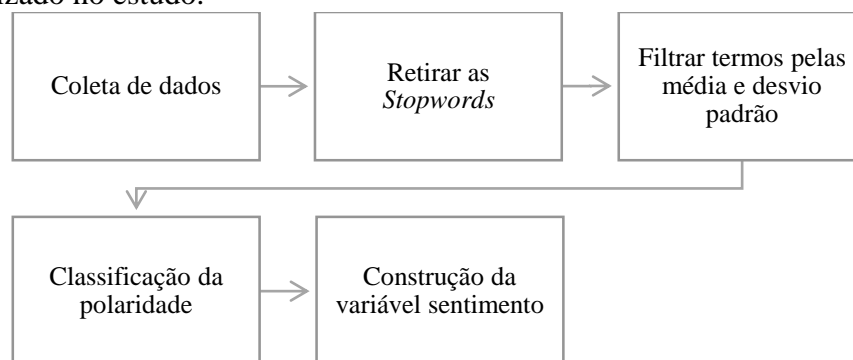
A fim de confirmar a análise da lista de palavras, paralelamente, cada palavra recebeu um valor e esse é exposto graficamente no mapa perceptual (Hair Jr et al., 2009). Na presente pesquisa, esse valor dependerá da frequência das palavras em dias de retornos positivos e em dias negativos em relação ao total das palavras. Optou-se por filtrar as palavras que tinham média e desvio padrão da frequência das palavras nos dias positivos ou negativos acima de 20%. Essa escolha foi realizada com o objetivo de não elevar artificialmente o peso de cada termo, assim, se estivesse próximo de zero, receberia o valor neutro. Assim, foram descartadas as palavras que apareceram todos os dias na mesma frequência independente do retorno. Essa opção foi feita para refinar as palavras, pois o número era muito grande.

Para construir um mapa perceptual, primeiro é necessário reunir as medidas de similaridade ou preferência em todo o conjunto de objetos a serem analisados. Segundo, usa-se as técnicas de modelagem de escalonamento multidimensional para estimar a posição relativa de cada objeto em dimensões multidimensionais de espaço. E por último, identificar e

interpretar os eixos do espaço dimensional em termos de percepção e / ou objetivo dos atributos (Hair Jr et al., 2009).

No contexto do presente trabalho, suponha que os termos “A” e “B” da amostra receberão à valores mais semelhantes do que outros possíveis objetos (“C”, “D” e “E”). O escalonamento multidimensional posiciona os objetos “A” e “B”, de forma que a distância euclidiana entre eles no espaço multidimensional seja menor que a distância entre os outros objetos.

A partir da presente etapa foi realizado o mapa perceptual com os tweets e a partir do resultado, verificou-se quais palavras foram classificadas como positivas, negativas ou neutras e receberam escores numérico de acordo com sua polaridade. O mapa perceptivo demonstrou o posicionamento relativo de todas as palavras. O objetivo foi analisar como as palavras se comportam próximo do retorno positivo e do retorno negativo. A Figura 1 ilustra o fluxo de trabalho utilizado no estudo.



**Figura 1: Etapas para o Processamento da Linguagem Natural usados no estudo**

Fonte: Elaboração Própria.

Depois do mapa perceptual, foi estimado o valor do sentimento de um *tweet* com base no peso das palavras positivas e negativas. O primeiro indicador de sentimento diário será o sentimento total diário dos *tweets*, de acordo com a equação 1.

$$STO_t = \sum PP_t + \sum PN_t \quad (1)$$

Onde:

$STO_t$ : Sentimento Total dos Tweets j;

$PP_t$ : Quantidade de palavras positivas;

$PN_t$ : Quantidade de palavras negativas;

A segunda variável de sentimento diário é a média de sentimento dos *tweets* em determinado dia. A medida é o total das palavras positivas menos as negativas, dividido pelo total de palavras. O mesmo indicador foi utilizado por Feuerriegel, Wolff e Neumann (2015). Essa rotulação foi realizada a partir do resultado da Equação 2.

$$SME_t = \frac{\sum PP_t + \sum PN_t}{\sum PT_t} \quad (2)$$

Onde:

$SME_t$ : Sentimento Médio do Tweet j;

$PT_t$ : Quantidade de palavras totais.





São Paulo 28 a 30 de julho 2021.

Para calcular o retorno observado no período da Petrobrás, será utilizada a função logarítmica que analisa o retorno sob o preço da ação no dia presente e no dia anterior (Fama, 1995). A partir do modelo de mercado desenvolvido por Sharpe (1963), foi construída uma equação para verificar se o retorno da Petrobrás é relacionado com o sentimento textual presente em *tweets*, além do movimento do mercado. O retorno do mercado foi o índice IBOVESPA. Com base nisso, a Equação 3 foi estimada para analisar se o sentimento textual se relaciona com o retorno do ativo.

$$Ret_t = \alpha + \beta_1 R_{m,t} + \beta_2 ST_t + \varepsilon_t \quad (3)$$

$Ret_t$ : Retorno da Petrobrás;

$R_{m,t}$ : Retorno do mercado;

$ST_t$ : Sentimento Textual;

A fim de aprofundar o tema, a variável sentimento foi dividida em sentimento positivo e negativo. Optou-se por esse caminho para analisar se existe alguma diferença entre o efeito do sentimento separadamente como está presente na equação 4.

$$Ret_t = \alpha + \beta_1 R_{m,t} + \beta_2 STP_t + \beta_3 STN_t + \varepsilon_t \quad (4)$$

Onde:

$STP_t$ : Sentimento Textual Positivo;

$STN_t$ : Sentimento Textual Negativo;

O comportamento do sentimento textual sobre o movimento do mercado foi dividido em duas variáveis pois se espera comportamentos diferentes em cada uma delas. Esperou-se encontrar significância e sinal positivo no  $\beta_2$ , sentimento positivo aumenta os retornos, e o sentimento negativo diminui os retornos, no  $\beta_3$ .

E por último, um teste para aprofundar o modelo. Com o objetivo de verificar o relacionamento entre o sentimento com o movimento do mercado, foram introduzidas as variáveis: retorno positivo do mercado e retorno negativo do mercado. Conforme, pode ser visualizado na Equação 5.

$$Ret_t = \alpha + \beta_1 R_{m,t} + \beta_2 ST_t + \beta_3 RP_{m,t} ST_t + \beta_4 RN_{m,t} ST_t + \varepsilon_t \quad (5)$$

Onde:

$RP_{m,t}$ : Retorno Positivo do mercado;

$RN_{m,t}$ : Retorno Negativo do mercado;

Dessa forma, espera-se investigar se o sentimento amplifica ou diminui o retorno positivo e negativo do mercado. Buscou-se analisar se quando o mercado está caindo, negativo, o sentimento positivo modera o efeito no retorno ou se o sentimento negativo amplifica o efeito no retorno. E quando o mercado for otimista, positivo, analisar se o sentimento positivo amplifica esse efeito ou se o sentimento negativo modera. Por isso, o sinal esperado das variáveis  $\beta_3$  é positivo e  $\beta_4$  é negativo.

Para testar as hipóteses de pesquisa, foi utilizado um modelo de regressão linear para séries temporais, com a estimação por Métodos Generalizado dos Momentos (GMM). Pois os retornos do mercado possuem grandes oscilações e esse método é uma forma de estimação que



São Paulo 28 a 30 de julho 2021.

não faz suposições sobre ausência de autocorrelação dos erros, ausência de multicolinearidade ou ausência de heterocedasticidade.

#### 4. ANÁLISE DE RESULTADOS

Com o objetivo de avaliar o efeito do sentimento expresso via Twitter sobre os retornos da Petrobrás, buscou-se, nesta seção, apresentar as estatísticas descritivas, testar as hipóteses de pesquisa e apresentar uma lista de termos presentes na rede social que exercem influência no retorno da ação.

A frequência diária dos termos mais usados foi relacionada com o retorno diário da ação da Petrobrás. Se as palavras foram citadas em dias de retornos positivos, suas frequências eram multiplicadas por +1 e em dias negativos, por -1. Essa classificação de polaridade (positiva ou negativa) foi realizada a fim de se verificar se o peso dos sentimentos, afeta o retorno das ações.

Depois de verificar a polaridade, foram escolhidas as palavras que possuem variação da média (Positiva/Negativa) e do desvio padrão (Positivo/Negativo) superior a 20%. Esse critério foi feito para excluir as palavras que aparecem nos dois momentos e poderiam gerar resultados positivos e negativos, mas que não agregariam tanto para análise de resultado quanto palavras citadas nos extremos. Nomes de políticos foram retirados pois são autoridades que tem mandato temporário e não agregaria na lista de palavras.

No geral, todas os termos se mostraram frequentes tanto em dias positivos quanto nos negativos. As palavras blog e petróleo, por exemplo, aparecem tanto nos dias positivos quanto nos dias negativos, como pode ser visto pela variação da frequência, média e desvio padrão que é próximo a 1. Nesses casos, os termos não tiveram peso na variação dos retornos. Logo, o mapa perceptual foi criado sem essas palavras.

Foi realizado o mapeamento perceptual das palavras depois da limpeza dos dados, tanto das *stopwords* quanto dos termos que tiveram desvio padrão e média relativa menor que 20%. A alocação de cada termo foi calculada de acordo com a distância euclidiana desse em relação a todos os outros termos.

Foram calculados 2 diferentes índices de sentimento para verificar qual teria mais impacto no retorno. Para cada variável, foi testada a estacionaridade, pois, só é possível fazer a regressão se todas as séries foram estacionárias. A tabela 1 apresenta os resultados do teste Kwiatkowski-Phillips-Schmidt-Shin (KPSS) que verifica a estacionaridade da série.

**Tabela 1:** Teste de estacionaridade das séries

Teste KPSS	Ret	Ibov	STO	SME
<b>P-value</b>	0.01	0.01	0.01	0.01

Fonte: Elaborado pelos autores a partir de dados da pesquisa.

O p-valor de todos os indicadores de sentimento foi menor que a significância do estudo, que é de 5%. Logo, todas as séries possuíam problemas com estacionaridade. Para resolver esse problema, optou-se por tomar as primeiras diferenças da série temporal para transformar a regressão espúria em estacionária (Arellano & Bond, 1991). Na tabela 2 foi feito o mesmo teste anterior para verificar se as novas variáveis passam no teste de estacionaridade.

**Tabela 2:** Teste de estacionaridade das séries

Teste KPSS	Ret	Ibov	STO	SME
<b>P-value</b>	0.1	0.1	0.1	0.1



São Paulo 28 a 30 de julho 2021.

Fonte: Elaborado pelos autores a partir de dados da pesquisa.

Com a mudança, os p-valores de todos os indicadores de sentimento foram maiores que a significância do estudo e aceito a hipótese nula do teste que diz que a série é estacionária. Contudo, é importante ressaltar que apesar dessa resolução ter a vantagem de remover a tendência temporal da série, teve a desvantagem da mudança da variável analisada. Não será mais analisado se o sentimento impacta o retorno, e sim, será verificado se a variação do sentimento impacta o retorno.

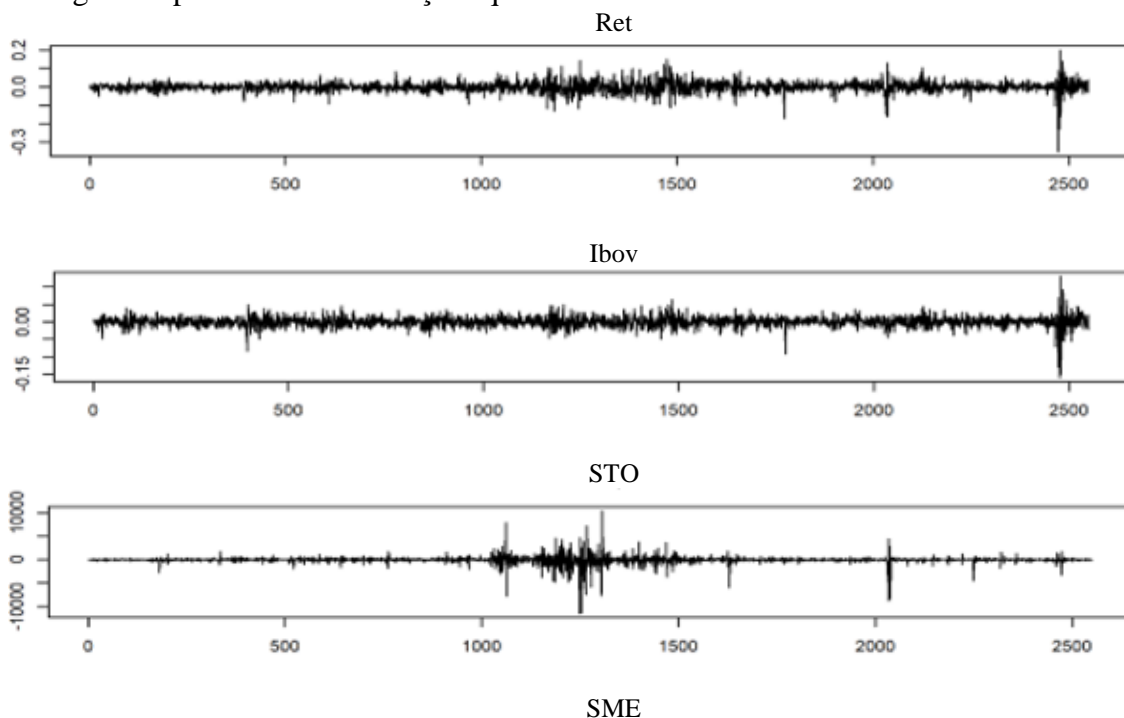
Os testes de autocorrelação e autocorrelação parcial encontram-se no Anexo A. Nenhuma das variáveis apresentaram tendência ou sazonalidade. Na tabela 3 estão dispostas as estatísticas descritivas de cada variável que foi testada no estudo.

**Tabela 3:** Estatística Descritiva do sentimento das variáveis do estudo

Variável	Média	Desvio Padrão	Mínimo	Máximo
<b>STO</b>	-55.401	891.69	-11491	10366
<b>SME</b>	-0.00033	0.01774	-0.110922	0.07260
<b>Ret</b>	-0.00021	0.03022	-0.35236	0.20067
<b>Ibov</b>	0.00011	0.01612	-0.15993	0.13022

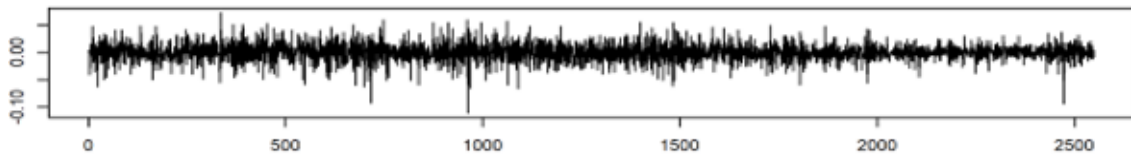
Fonte: Elaborado pelos autores a partir de dados da pesquisa.

Nota-se que o valor do retorno do IBOVESPA e da ação da Petrobrás tiveram média , respectivamente, negativa e positiva. Apesar do mercado se manter positivo durante o período, as ações da Petrobras tiveram, em média durante o período em análise, retornos negativos. As duas variáveis tiveram valor mínimo negativo e um valor máximo de positivo, o que demonstra grandes oscilações no retorno do mercado, principalmente no retorno da Petrobras. Os gráficos da Figura 2 apresentam as oscilações que ocorreram nas variáveis em análise de 2010 até 2020.





São Paulo 28 a 30 de julho 2021.

**Figura 2:** Oscilações das variáveis ao longo do tempo

Fonte: Elaborado pelos autores a partir de dados da pesquisa.

Devido a essas oscilações presentes no mercado acionário, é difícil obter normalidade e homocedasticidade. Dessa forma, optou-se pelo Métodos Generalizado dos Momentos (GMM). Visto que é uma forma de estimação que não faz suposições sobre ausência de autocorrelação dos erros, ausência de multicolinearidade ou ausência de heterocedasticidade.

Antes da regressão, foram visualizadas as correlações entre o retorno da Petrobras, retorno do mercado e os indicadores de sentimento. Os resultados estão dispostos na tabela 4.

**Tabela 4:** Correlação Das Variáveis Independentes Com A Dependente

	Ibov	STO	SME
Ret	0.78	0.21	0.14

Fonte: Elaborado pelos autores a partir de dados da pesquisa.

O retorno do mercado apresentou alta correlação com o retorno da Petrobrás (78%). Todos os dois indicadores de sentimento tiveram correlação com o retorno da estatal (21%, e 14%). O resultado demonstra que existe uma correlação linear entre as variáveis independentes e a dependente.

Após o tratamento dos dados e das variáveis, buscou-se testar as hipóteses do estudo por meio dos modelos de série temporal. Para cada indicador de sentimento, foi realizado o cálculo de um modelo. Com isso, foi possível a comparação dos modelos e determinou-se qual foi o melhor. Na primeira regressão, a hipótese de que a variação sentimento expresso via Twitter impacta o retorno da Petrobras foi testada. Os resultados estão demonstrados na Tabela 5.

**Tabela 5:** Regressão do impacto da variação do sentimento sobre o retorno da Petrobrás

Ret	STO	SME
<i>Ibov</i>	1.4466e+00*** ( $< 2.2e-16$ )	1.45758099*** ( $< 2.2e-16$ )
<i>ST<sub>i</sub></i>	4.7366e-06*** (3.182e-06)	0.10441803*** (9.815e-05)
Const	-1.2957e-04 (0.717)	-0.00035801 (0.3329)
R <sup>2</sup>	0,6334	0,6177
Teste de Normalidade (Jarquebera)	$< 2.2e-16$	$< 2.2e-16$
Teste de Heterocedasticidade (Breusch-Pagan)		
chi <sup>2</sup> (1)	$< 2,2e-16$	$< 1.716e-06$

Fonte: elaboração própria com base no programa Rstudio. Nota: 1) Método GMM; 2) entre parênteses encontram-se os valores dos coeficientes; 3) Níveis de significância: “\*\*\*”: Significativo a 0,01%; “\*\*”: Significativo a 1%; “\*”: Significativo a 5%.



São Paulo 28 a 30 de julho 2021.

Os resultados demonstraram estar de acordo com a teoria. O retorno do mercado foi significativo e positivo, da mesma forma que explicitada que o modelo de Sharpe (1963). Todas as três variáveis de sentimento tiveram significância e foram positivas em relação ao retorno. Esses achados corroboram com os resultados dos estudos de Bartov, Faurel e Moharam (2018), Carosia, Coelho e Silva (2019) e McGurk, Nowak e Hall (2020) que também verificaram influências do sentimento via Twitter com o retorno. A primeira hipótese da pesquisa foi confirmada, a variação do sentimento textual se relacionou com o retorno.

Na segunda regressão, disposta na Tabela 6, o objetivo foi analisar a segunda hipótese de pesquisa. Buscou-se verificar se a variação dos sentimentos negativos e positivos, separadamente, possuem impacto no retorno e qual seria a direção desses relacionamentos.

**Tabela 6:** Regressão do impacto da variação do sentimento positivo e do negativo sobre o retorno da Petrobrás

Ret	STO	SME
<i>Ibov</i>	1.4435e+00*** ( $< 2.2e-16$ )	1.45334652*** ( $< 2.2e-16$ )
<i>STP<sub>i</sub></i>	3.4063e-07 (0.722851)	-0.02407614 (0.5706119)
<i>STN<sub>i</sub></i>	7.1590e-06*** (1.439e-07)	0.21895770*** (0.0002617)
Const	1.1165e-03** (0.006874)	0.00125524 . (0.0601886)
<b>R<sup>2</sup></b>	0,6407	0,6198
<b>Teste de Normalidade (Jarquebera)</b>	$< 2.2e-16$	$< 2.2e-16$
<b>Teste de Heterocedasticidade (Breusch-Pagan)</b>		
<b>chi<sup>2</sup> (1)</b>	$< 2,2e-16$	8.618e-07

Fonte: elaboração própria com base no programa Rstudio. Nota: 1) Método GMM; 2) entre parênteses encontram-se os valores dos coeficientes; 3) Níveis de significância: “\*\*\*”: Significativo a 0,01%; “\*\*”: Significativo a 1%; “\*”: Significativo a 5%.

Nesse caso, somente a variação dos sentimentos negativos dos três indicadores de sentimento possuíram impacto negativo no retorno. Esse resultado confirma a teoria de finanças comportamentais que afirma que há uma assimetria entre ganhos e perdas (Kahneman & Tversky, 1979). No presente estudo foi possível verificar que a variação negativa do sentimento possuiu maior impacto nos indivíduos do que a variação positiva do sentimento. As pessoas se sentem mais sensibilizadas pela perda do que pelo ganho, mesmo que seja na mesma magnitude. É um comportamento de defesa ao seu patrimônio. Os mesmos achados foram encontrados no trabalho de Silva (2017). A segunda hipótese de pesquisa foi parcialmente confirmada, somente a variação do sentimento textual negativo se relacionou com o retorno.

Na Tabela 7 foram apresentados os resultados da terceira regressão. Nesses modelos, foram analisados se a variação do sentimento exposto nos *tweets* amplificam ou diminuem o movimento do mercado.

**Tabela 7:** Regressão do impacto da variação do sentimento no movimento de mercado sobre o retorno da Petrobrás

Ret	STO	SME
<i>Ibov</i>	1.3840e+00 *** ( $< 2e-16$ )	1.4169e+00 *** ( $< 2.2e-16$ )



São Paulo 28 a 30 de julho 2021.

<b><math>ST_i</math>*Ibov positivo</b>	1.2564e-05 (0.8688)	-1.5893e+00 (0.501020)
<b><math>ST_i</math>*Ibov negativo</b>	-4.1118e-04*** ( $<2e-16$ )	-1.1218e+01** (0.002098)
<b>Const</b>	-1.2957e-04 (0.717)	-1.5146e-04 (0.680044)
<b>R<sup>2</sup></b>	0,6463	0,6275
<b>Teste de Normalidade (Jarquebera)</b>	$< 2.2e-16$	$< 2.2e-16$
<b>Teste de Heterocedasticidade (Breusch-Pagan)</b>		
<b>chi<sup>2</sup> (1)</b>	7.901e-09	0.0315

Fonte: elaboração própria com base no programa Rstudio. Nota: 1) Método GMM; 2) entre parênteses encontram-se os valores dos coeficientes; 3) Níveis de significância: “\*\*\*”: Significativo a 0,01%; “\*\*”: Significativo a 1%; “\*”: Significativo a 5%.

Na linha dos resultados da segunda regressão, a variação do sentimento textual teve efeito somente com o mercado negativo. A variação do sentimento obteve significância e com efeito negativo em relação ao retorno. Esse achado demonstra que no cenário de o mercado estar mal, com retornos negativos, e se a variação do sentimento também for negativa, o impacto no retorno da Petrobras será intensificado negativamente. No mesmo cenário, caso a variação do sentimento na rede social for positiva, o impacto no retorno da estatal será amaciado. Esses resultados estão de acordo com a linha teórica de finanças comportamentais. O sentimento promoveu desvios no valor do ativo. A terceira hipótese da presente pesquisa foi parcialmente confirmada, o sentimento textual amplificou ou reduziu o movimento de mercado em relação ao retorno, porém, somente quando o mercado está negativo.

A fim de aprofundar a análise, optou-se verificar se o mesmo impacto ocorre se as postagens forem coletadas diretamente do seu perfil no Twitter (@Petrobras) e se forem citadas apenas o *ticker* da ação (PETR3 e/ou PETR4). Com isso, a coleta da amostra seria mais facilitada e menos custosa, visto que são menos *tweets*, para obter o mesmo resultado. Por isso, foi realizada a mesma metodologia, porém, com amostras diferentes.

Tiveram 139.530 *tweets* na amostra que tinha como busca a página oficial da Petrobras e 210.881 *tweets* com a citação dos *tickers* da ação. É possível verificar que os usuários falam mais *sobre* do que *com* a empresa, no caso, a página oficial no Twitter. Aplicando a mesma metodologia para essas amostras, não foi obtido um resultado satisfatório.

No caso da amostra da página oficial da Petrobras, poucos termos apresentaram relevância. As palavras foram muito frequentes, apareciam tanto em dias positivos quanto em negativos e não apresentaram grandes desvios padrões. Com isso, não seria interessante ter a conta oficial como filtro. Apenas 19 termos seriam utilizados como lista. Com isso, optou-se por utilizar a amostra com todos os *tweets* que apresentaram a palavra Petrobras.

No caso da amostra dos *tweets* que tiveram alguma citação dos *tickers* da ação, os *tweets* não apresentaram sentimento. As palavras apareceram tanto em dias positivos como negativos e também não apresentaram desvios padrões. Para demonstrar o que ocorreu, parte dos *tweets* foram expostos na tabela 8.

**Tabela 8:** Parte da amostra dos *tweets* que continham “PETR3” e/ou “PETR4”

<b>Dias</b>	<b>Tweets</b>
24/09/2010	\$PETR4: Pedidos d reserva dos investidores Ñ institucionais e do FIA Petrobras foram atendidos proporcionalmente sendo rateados em 45,77% \$\$

06/09/2012	\$PETR4 - Petrobras e Repsol descobrem reserva de gás no Peru <a href="http://t.co/4sp7US4u">http://t.co/4sp7US4u</a>
13/10/2015	\$PETR3 - Relatório sobre os resultados operacionais e financeiros da Petrobras (PETR3, PETR4, PBR, PBR.A, XPBR) <a href="http://br.advfn.com/noticias/PAPERBR/2015/artigo/68864874?adw=1126416">http://br.advfn.com/noticias/PAPERBR/2015/artigo/68864874?adw=1126416</a>
23/08/2016	\$PETR3 - Bovespa: Índice Ibovespa fecha em alta de +0,41% nesta terça-feira, 23 de agosto de 2016 <a href="http://br.advfn.com/noticias/PAPERBR/2016/artigo/72274612?adw=1126416">http://br.advfn.com/noticias/PAPERBR/2016/artigo/72274612?adw=1126416</a>
30/12/2019	\$PETR3 - PETROBRAS (PETR-N2) - Regimento Comite Auditoria Estatutario - 18/12/19 <a href="http://br.advfn.com/noticias/BOV/2019/artigo/81437435?adw=1126416">http://br.advfn.com/noticias/BOV/2019/artigo/81437435?adw=1126416</a>

Fonte: Elaborado pelos autores a partir de dados da pesquisa

Na tabela 9 foi possível verificar que os *tweets* apresentaram conteúdo informacional, atualizando os valores diários da ação. O conteúdo da mensagem não variou substancialmente ao longo dos dias. Também foi visto que diferentes usuários copiam o conteúdo e repostam em seus perfis. Importante lembrar que *retweets* foram retirados das amostras, mas mensagens copiadas não foi motivo de filtragem da amostra.

Das três amostras colhidas para o estudo, apenas a que tinha o filtro “petrobras” gerou uma grande lista de termos. Analisando a amostra, foi possível verificar que essa teve mais *tweets* em relação as outras, 3.826.463 *tweets*, e também apresentou um conteúdo mais diversificado, com mais opiniões e destaques diários.

## 5. CONSIDERAÇÕES FINAIS

O objetivo deste trabalho foi analisar a relação entre o sentimento exposto no Twitter e o retorno das ações. Desse modo, *tweets* sobre a Petrobras foram analisados a fim de se verificar se o sentimento textual dessas mensagens impacta no valor do retorno. Os retornos diários foram coletados da base Bloomberg. Para estimar o sentimento, foram criados três diferentes indicadores para verificar qual seria o melhor. Por fim, foi testada a mesma metodologia para diferentes amostras - conta oficial no Twitter e PETR3 e/ou PETR4 - para analisar qual explicaria melhor os retornos diários.

A pesquisa partiu da premissa que dentro do conteúdo da *web* de determinadas páginas podem ser extraídas informações que denotam emoções e sentimentos. As pessoas escrevem posts, comentários, críticas e *tweets* sobre variados tópicos diariamente (Godsay, 2015). Com isso, é possível verificar se a Petrobras é visualizada de forma positiva ou negativa na internet.

As regressões foram divididas em três grupos de acordo com os objetivos específicos do estudo. O primeiro grupo era para analisar se o índice de sentimento impacta no retorno, o segundo avaliou se sentimentos positivos e negativos impactam o retorno e o terceiro, verificou se o sentimento influencia no movimento do mercado sobre os retornos. Para cada grupo, foi testado os dois indicadores de sentimento, totalizando seis regressões.

Todos os indicadores de variação do sentimento do primeiro grupo tiveram resultados positivos e significativos. A variação do sentimento textual dos *tweets* tem relação direta com o retorno da Petrobras. O resultado corrobora com os estudos de Bartov, Faurel e Moharam (2018), Carosia, Coelho e Silva (2019) e McGurk, Nowak e Hall (2020).

No segundo grupo, somente os indicadores de variação do sentimento negativo tiveram resultados significativos. A variação negativa do sentimento possuiu maior impacto nos



São Paulo 28 a 30 de julho 2021.

indivíduos do que a variação positiva do sentimento. As pessoas possuem comportamento avesso ao risco, sentem mais perda do que pelo ganho, mesmo que seja na mesma magnitude. Os mesmos achados foram encontrados no trabalho de Silva (2017).

No último grupo de regressões, foram avaliadas as variações do sentimento sobre o movimento de mercado. Somente o sentimento negativo teve resultado negativo e significativo. Desse modo, o sentimento negativo potencializa o movimento negativo do mercado e ameniza o movimento positivo do mercado. Esse resultado corrobora a teoria das finanças comportamentais que afirma que há uma assimetria entre ganhos e perdas. Uma perda gera mais impacto do que um ganho na mesma magnitude (Kahneman & Tversky, 1979).

A fim de verificar se haveria resultados diferentes com outras amostras, optou-se por analisar as postagens da conta oficial da Petrobras e *tweets* que continha algum dos *tickers* das ações – PETR3 e/ou PETR4. As duas amostras apresentaram menos de 10% do tamanho da amostra principal. A lista das palavras mais frequentes da conta oficial da Petrobras foi pequena. As pessoas falam mais da estatal Petrobras do que com a Petrobras. As mensagens que continham PETR3 e/ou PETR4, em sua maioria, eram informativas e apresentavam apenas as cotações diárias.

Dessa forma, a amostra que foi tida como mais sentida, diversidade de palavras e quantidade de *tweets* foi a primeira amostra, a que continha apenas a palavra “petrobras” como filtro de seleção. Apesar de mais custosa, em relação ao tempo de coleta de dados e ao uso de programas para salvar a base de dados, foi a amostra que apresentou uma lista de termos que apresentam desvios relacionados aos retornos diários. Por isso, o presente estudo optou por se aprofundar na lista de termos formada por essa amostra.

Esses resultados são importantes para os gestores e agentes de política pública, uma vez que é importante entender a opinião pública sobre a empresa Petrobras para realizar uma tomada de decisão. Assim, o estudo pode criar um instrumento de gerenciamento de marketing e análise do sentimento do investidor sobre a organização. Para pesquisas futuras, sugere-se mudar a amostra para outras empresas de capital aberto e analisar como o movimento da mídia social se relaciona com o movimento de mercado.

## REFERÊNCIAS

- Ab. Rahman, A. S., Abdul-Rahman, S. & Mutalib, S. (2017). Mining textual terms for stock market prediction analysis using financial news. *International Conference on Soft Computing in Data Science*.
- Ahmadi, Z. et al. (2019). Towards bankruptcy prediction: Deep sentiment mining to detect financial distress from business management reports. *Proceedings - 2018 IEEE 5th International Conference on Data Science and Advanced Analytics*.
- Alkubaisi, G. A. A. J., Kamaruddin, S. S. & Husni, H. (2018). Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naive Bayes Classifiers. *Computer and Information Science*, 11 (1).
- Arellano, M. & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58 (1).
- Avanço, L. V. (2015). Sobre normalização e classificação de polaridade de textos opinativos na web. (Dissertação de Mestrado) *Instituto de Ciências Matemáticas e de Computação - ICMC-USP*.





São Paulo 28 a 30 de julho 2021.

- Avanço, L. V. & Nunes, M. D. G. V. (2014). Lexicon-based sentiment analysis for reviews of products in Brazilian Portuguese. *Proceedings - 2014 Brazilian Conference on Intelligent Systems, BRACIS*, 277–281.
- Bannier, C., Pauls, T. & Walter, A. (2019). Content analysis of business communication: introducing a German dictionary. *Journal of Business Economics*, 89 (1), 79–123.
- Bartov, E., Faurel, L. & Mohanram, P. S. (2018). Can twitter help predict firm-level earnings and stock returns? *The Accounting Review*, 93 (3), 25–57.
- Bauer, M. W. & Gaskell, G. (2003). *Pesquisa qualitativa com texto, imagem e som: um manual prático*. 2<sup>o</sup> ed. Petrópolis: Vozes.
- Bermingham, A. & Smeaton, A. (2010). Classifying sentiment in microblogs: Is brevity an advantage? *International Conference on Information and Knowledge Management, Proceedings*.
- Bollen, J. & Mao, H. (2011). Twitter mood as a stock market predictor. *Journal of Computational Science*, 2(1).
- Bollen, J., Mao, H. & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8.
- Carosia, A. E. O., Coelho, G. P. & Silva, A. E. A. (2019) Analyzing the Brazilian Financial Market through Portuguese Sentiment Analysis in Social Media. *Applied Artificial Intelligence*, 34(1), 1–19.
- Dařena, F. et al. (2018). Machine learning-based analysis of the association between online texts and stock price movements. *Inteligencia Artificial*, 21(61), 95–110.
- Fama, E. F. (1995). Random Walks in Stock Market Prices. *Financial Analysts Journal*, 51(1), 75–80.
- Fang, B. et al. (2017). Research on sentiment analysis of financial texts. *Proceedings of 2017 IEEE 7th International Conference on Electronics Information and Emergency Communication, ICEIEC 2017*.
- Feuerriegel, S., Wolff, G. & Neumann, D. (2015). Information processing of foreign exchange news: Extending the overshooting model to include qualitative information from news sentiment. *2015 International Conference on Information Systems: Exploring the Information Frontier, ICIS 2015*.
- Gandhi, P., Loughran, T. & Mcdonald, B. (2019). Using Annual Report Sentiment as a Proxy for Financial Distress in U.S. Banks. *Journal of Behavioral Finance*, 20 (4), 424–436.
- Ge, Q., Kurov, A. & Wolfe, M. H. (2018). Stock Market Reactions to Presidential Statements: Evidence from Company-Specific Tweets. *SSRN Electronic Journal*.
- Go, A., Bhayani, R. & Huang, L. (2009). Twitter Sentiment Classification using Distant Supervision. *Processing*, 150, 1–6.
- Godsay, M. (2015). The Process of Sentiment Analysis: A Study. *International Journal of*



São Paulo 28 a 30 de julho 2021.

*Computer Applications*, 126 (7), 26–30.

Gupta, A., Simaan, M. & Zaki, M. J. (2017). Investigating bank failures using text mining. *2016 IEEE Symposium Series on Computational Intelligence, SSCI 2016*.

Hair Jr, J. F. et al. (2009). *Multivariate Data Analysis*. 7. ed. Prentice Hall.

Houlihan, P. & Creamer, G. G. (2017). Can Sentiment Analysis and Options Volume Anticipate Future Returns? *Computational Economics*, 50 (4), 669–685.

Ito, T. et al. (2017). Development of an interpretable neural network model for creation of polarity concept dictionaries. *IEEE International Conference on Data Mining Workshops, ICDMW*.

Ito, T. et al. (2019). Word-level Sentiment Visualizer for Financial Documents. *CIFEr 2019 - IEEE Conference on Computational Intelligence for Financial Engineering and Economics*.

Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Journal of the Econometrica society*, 47 (2), 263–292..

Li, X., Wu, P. & Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. *Information Processing and Management*, 57 (5).

Loughran, T. & McDonald, B. (2011). Internet appendix for “When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks”. *The Journal of Finance*, 66 (1), 35–65.

Loughran, T. & McDonald, B. (2015) The Use of Word Lists in Textual Analysis. *Journal of Behavioral Finance*, 16, (1), 1–11.

Mai, F. et al. (2018). How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis. *Journal of Management Information Systems*, 35 (1), 19–52.

Mangee, N. (2018) Stock Returns and the Tone of Marketplace Information: Does Context Matter? *Journal of Behavioral Finance*, 19 (4), 396–406.

Mao, H. et al. (2014). An automatic news analysis and opinion sharing system for exchange rate analysis. *Proceedings - 11th IEEE International Conference on E-Business Engineering, ICEBE 2014 - Including 10th Workshop on Service-Oriented Applications, Integration and Collaboration, SOAIC 2014 and 1st Workshop on E-Commerce Engineering, ECE 2014*.

Mao, H., Counts, S. & Bollen, J. (2011). Predicting Financial Markets: Comparing Survey, News, Twitter and Search Engine Data. *Economics, Computer Science, Physics - ArXiv*, 1–10.

Martins, R., Pereira, A. & Benevenuto, F. (2015). An approach to sentiment analysis of web applications in Portuguese. *WebMedia 2015 - Proceedings of the 21st Brazilian Symposium on Multimedia and the Web*, 105–112.

Mcgurk, Z., Nowak, A. & Hall, J. C. (2020). Stock returns and investor sentiment: textual analysis and social media. *Journal of Economics and Finance*, 44 (3), 458–485.

Naylor, R. W., Lamberton, C. P. & West, P. M. (2012). Beyond the “like” button: The impact of mere virtual presence on brand evaluations and purchase intentions in social media settings. *Journal of Marketing*, 76 (6), 105–120.

- Nofsinger, J. (2010). *The Psychology of Investing*. 4. ed. New Jersey: Prentice Hall.
- Oliveira, N., Cortez, P. & Areal, N. (2013). Some experiments on modeling stock market behavior using investor sentiment analysis and posting volume from twitter. *ACM International Conference Proceeding Series*.
- Pengnate, S. F. & Riggins, F. J. (2020). The role of emotion in P2P microfinance funding: A sentiment analysis approach. *International Journal of Information Management*, 54.
- Pinto, A. E. De S. (2019). *Pelo 4<sup>o</sup> ano seguido, Petrobras é a Marca que Representa o Brasil*. Recuperado a partir de <https://top-of-mind.folha.uol.com.br/2019/10/pelo-4o-ano-seguido-petrobras-e-a-marca-que-representa-o-brasil.shtml>.
- Ranco, G. et al. (2015). The effects of twitter sentiment on stock price returns. *PLoS ONE*, 10 (9), 1–21.
- Sarigiannidis, A. et al. (2018). A Novel Lexicon-Based Approach in Determining Sentiment in Financial Data Using Learning Automata. *LNCS*, 10750.
- 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.
- Sharpe, W. F. (1963) A Simplified Model For Portfolio Analysis. *Magazine Science*, 9 (2), 277–293.
- Silva, M. D. De O. P. Da. (2017). *O Efeito do Sentimento das Notícias sobre o Comportamento dos Preços no Mercado Acionário Brasileiro*. (Tese de Doutorado) Programa Multi-institucional e Inter-regional de Pós-graduação em Ciências Contábeis da Universidade de Brasília, Universidade Federal da Paraíba e Universidade Federal do Rio Grande do Norte.
- Souza, D. M. S. De, Lucena, W. G. L. & Queiroz, D. B. De. (2019). O Efeito do Sentimento do Investidor Expresso via Twitter sobre o Comportamento do Mercado Acionário Brasileiro Durante o Período Eleitoral. *XIX Congresso USP - FIEPECAFI*, 1–19.
- Tarde, G. (1902). *Psychologie Économique*. Paris: Bibliothèque de philosophie contemporaine.
- Tommasetti, R.; Maia, V. M.; Macedo, M. Á. Da S. (2020). Twitter as ‘bully pulpit’: Brazilian banking sector case. *Applied Economics Letters*, 1–4.
- Uhr, P., Zenkert, J. & Fathi, M. (2014). Sentiment analysis in financial markets: A framework to utilize the human ability of word association for analyzing stock market news reports. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*.
- Wang, Y. (2017). Stock market forecasting with financial micro-blog based on sentiment and time series analysis. *Journal of Shanghai Jiaotong University (Science)*, 22 (2), 173–179.
- Yadav, A. et al. (2020). Sentiment analysis of financial news using unsupervised approach. *Procedia Computer Science*.