UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

MILENE DIAS ALMEIDA

A INFLUÊNCIA DA MÍDIA SOCIAL NO VALOR DA PETROBRAS

RIO DE JANEIRO

2021

MILENE DIAS ALMEIDA

A INFLUÊNCIA DA MÍDIA SOCIAL NO VALOR DA PETROBRAS

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciências Contábeis da Universidade Federal do Rio de Janeiro como requisito parcial à obtenção do título de Mestre em Ciências Contábeis.

Orientador: Prof. Dr. Vinicius Mothé Maia

Coorientador: Dr. Roberto Tommasetti

RIO DE JANEIRO

2021

FICHA CATALOGRÁFICA

A447 Almeida, Milene Dias.

A influência da mídia social no valor da Petrobras / Milene Dias Almeida. - 2021.

56 f.; 31 cm.

Orientador: Vinicius Mothé Maia.

Coorientador: Roberto Tommasetti.

Dissertação (mestrado) – Universidade Federal do Rio de Janeiro, Faculdade de Administração e Ciências Contábeis, Programa de Pós-Graduação em Ciências Contábeis, 2021.

Bibliografia: f. 44 – 49.

Ficha catalográfica elaborada pelo bibliotecário: Luiza Hiromi Arao CRB 7 – 6787 Biblioteca Eugênio Gudin/CCJE/UFRJ

Milene Dias Almeida

A INFLUÊNCIA DA MÍDIA SOCIAL NO VALOR DA PETROBRAS

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciências Contábeis da Universidade Federal do Rio de Janeiro como requisito parcial à obtenção do título de Mestre em Ciências Contábeis.

Aprovada em

_____

(Vinicius Mothé Maia, Dr. em Administração e UFRJ)

_____

(Marcelo Álvaro da Silva Macedo, Dr. em Engenharia de Produção e UFRJ)

_____

(Marcio André Veras Machado, Dr. em Administração e UFPB)

AGRADECIMENTOS

Primeiramente, agradeço a Deus que me deu força durante essa caminhada.

Aos meus pais, Antonio Valter Dias Almeida e Lucilene Aguiar de Almeida, pelo amor e pelo incentivo nas horas de desânimo e cansaço, além do apoio emocional.

Aos professores, Vinicius Mothé Maia e Roberto Tommasetti, pela orientação, paciência e apoio na elaboração deste trabalho e no meu desenvolvimento.

Aos amigos que fiz na UFRJ, pelos momentos divertidos nos almoços. O carinho de todos tornou mais fácil as tarefas da trajetória acadêmica.

RESUMO

ALMEIDA, Milene Dias. **A INFLUÊNCIA DA MÍDIA SOCIAL NO VALOR DA PETROBRAS**. Rio de Janeiro, 2021. Dissertação (Mestrado em Ciências Contábeis) – Faculdade de Administração e Ciências Contábeis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2021.

A presente pesquisa busca contribuir com a discussão do sentimento textual sobre a organização. O objetivo do estudo 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 a 2020, foram analisados, a fim de 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, tanto em dias positivos ou negativos. Para tal, três grupos de regressões foram criados: o primeiro analisou o impacto do sentimento no retorno; o segundo verificou o relacionamento do sentimento positivo e do sentimento negativo com o retorno; o terceiro analisou o impacto incremental do sentimento nos dias em que o mercado está positivo ou negativo. Os resultados apontaram que a variação do sentimento total e a variação do sentimento negativo têm relação com o retorno. A variação do sentimento total é significativa, quando o mercado está pessimista, demonstrando que essa variação negativa do sentimento intensifica o movimento pessimista do mercado. Por outro lado, quando a variação do sentimento é positiva, ameniza a queda do mercado. 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. Esses resultados são importantes para entender a opinião pública sobre a empresa Petrobras e, assim, contribuir para que gestores e agentes de política pública possam vir a ter uma ferramenta que auxilia a tomada de decisão. Para as pesquisas futuras, sugere-se incluir uma ponderação dos termos, mudar a amostra para outras empresas de capital aberto e analisar como o movimento da mídia social Twitter ou outras se relaciona com o movimento de mercado.

Palavras-chave: Sentimento Textual; Mídias Sociais; Petrobras.

ABSTRACT

ALMEIDA, Milene Dias. **A INFLUÊNCIA DA MÍDIA SOCIAL NO VALOR DA PETROBRAS**. Rio de Janeiro, 2021. Dissertação (Mestrado em Ciências Contábeis) – Faculdade de Administração e Ciências Contábeis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2021.

This research seeks to contribute to the discussion of the textual sentiment about the organization. The study's objective was to analyze the relationship between the sentiment exposed on Twitter and Petrobras' return. In this way, in the period from 2010 to 2020, tweets about Petrobras were analyzed to verify if the textual sentiment of these messages impacts the return. The method used to classify words was perceptual mapping and the mean and standard deviation of the terms' frequency on positive or negative days. Three groups of regressions were created. The first analyzed the impact of sentiment on the return. The second verified the relationship of the positive and negative sentiment with the return. The last analyzed whether the sentiment impact the market's return. The results showed that the total and negative sentiment variation are related to the return. The variation in sentiment is significant when the market is pessimistic, demonstrating that the negative variation in sentiment intensifies the market's pessimistic movement and mitigates the fall in the market when the variation in sentiment is positive. The findings corroborate the behavioral finance theory that the sentiment has a relationship with the return, in the study's case, the textual feeling. Finally, different samples were analyzed using the same methodology. With that, it was possible to verify that the individuals talk more about the state-owned company, in general, than about the stock tickers or with the company itself on its official page. These results are important to understand public opinion about the company Petrobras and, thus, managers and public policy agents have a tool that helps decision making. For future research, it is suggested to move the sample to other publicly traded companies and to analyze how the social media movement, Twitter or others, relates to the market movement.

Keywords: Textual Sentiment; Social Media; Petrobras.

LISTA DE ABREVIATURAS

MQO: Mínimos Quadrados Ordinários

VAR: Modelo de Autorregressão Vetorial

LISTA DE ILUSTRAÇÃO

Figura 1. Mecanismos de transmissão – demanda de informações.....21
Figura 2. Etapas para o Processamento da Linguagem Natural usados no estudo.....29
Figura 3. Oscilações das variáveis ao longo do tempo.....36

LISTA DE TABELAS

Tabela 1. Parte da amostra dos *tweets* que continham "PETR3" e/ou "PETR4".....34
Tabela 2. Quantidade de *Tweets*.....35
Tabela 3. Estatística Descritiva do sentimento das variáveis do estudo.....37
Tabela 4. Teste de estacionariedade das séries.....38
Tabela 5. Correlação Das Variáveis Independentes Com A Dependente.....38
Tabela 6. Impacto da variação do sentimento *total* no retorno da Petrobras (2,550 Observações).....38
Tabela 7. Impacto da variação do sentimento *médio* no retorno da Petrobras (2,550 Observações).....40

SUMÁRIO

1. INTRODUÇÃO 11
2. REVISÃO DA LITERATURA 14
2.1 Teoria de Finanças Comportamentais 14
2.2 Sentimento do Investidor 17
2.3 Sentimento Textual 20
2.4 Estudos anteriores sobre o sentimento textual em finanças 24
3. METODOLOGIA 29
4. ANÁLISE DE RESULTADOS 34
4.1 Análise das Amostras 34
4.2 Construção do sentimento 36
4.3 Análise Das Regressões 38
5. CONCLUSÃO 42
REFERÊNCIAS 44
ANEXO A – EXEMPLOS DE TWEETS PRESENTES EM PERÍODOS DE PICOS DE NÚMERO DE OBSERVAÇÕES 50
ANEXO B – FREQUÊNCIA, MÉDIA E DESVIO PADRÃO DAS PALAVRAS MAIS CITADAS NOS TWEETS 51
ANEXO C – LISTA DE TERMOS POSITIVOS E NEGATIVOS DO ESTUDO 56

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 tornaram mais difundidas e claras, formando 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 seus processos de tomadas de decisão, levando-os a uma relação entre "sentimento" 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 ela (NAYLOR; LAMBERTON; WEST, 2012). A mídia social é um meio de compartilhar essas emoções públicas e impactar na opinião pública geral e nos mercados. Pesquisas anteriores relacionadas à análise de sentimento por meio das mídias sociais, se dividem em duas categorias: a primeira classifica os dados gerais de forma lexical; a segunda, por meio de instrumentos automatizados (BERMINGHAM; SMEATON, 2010; GO; BHAYANI; HUANG, 2009). Nesta 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 os valores negativos, positivos ou neutros são atribuídos à opinião das pessoas (ALKUBAISI; KAMARUDDIN; HUSNI, 2018; RAO; SRIVASTAVA, 2012). Outros métodos utilizados, por exemplo, são: *OpinionFinder*, (BOLLEN; MAO, 2011; OLIVEIRA; CORTEZ; AREAL, 2013), *SentiWordNet*, *SenticNet* (LI; WU; WANG, 2020).

O 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. Além disso, a lista deve estar no mesmo idioma do texto analisado. Uma simples tradução de uma lista de um conteúdo específico, 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 inclui o primeiro e demonstra as condições dentro das empresas e mercado. O presente estudo buscou medir o sentimento textual, apresentando 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 foi de 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; (ii) estimar um índice de sentimento, baseado no dicionário criado anteriormente pelo método do mapa perceptual; e, (iii) verificar se este índice de sentimento está relacionado ao comportamento dos preços no mercado acionário brasileiro.

Pesquisadores utilizaram vários classificadores para realizar análises de sentimento no mercado de ações tendo o Twitter como fonte de dados. Para classificar os *tweets,* o estudo realiza a análise de sentimento com o método de mapeamento perceptual, verificando quais palavras foram qualificadas como positivas ou negativas. O modelo é uma maneira gráfica de representar como os indivíduos analisam os produtos ou marcas do mercado avaliado (GOWER; LEGENDRE, 1986).

Optou-se, então, pela Petrobras para participar do estudo, porque é uma empresa de referência nacional e está entre as marcas que mais foi percebida por alguma publicidade, por notícias ou propagandas "boca a boca" na internet, no ano de 2019, em relação ao ano anterior. A instituição foi a que teve o maior salto de popularidade ("Buzz Rankings de 2019: Brasil | YouGov - BrandIndex", 2020). Além disso, a empresa teve uma liderança isolada por quatro anos, de 2016 até 2019, no *Top of Mind*, na categoria "Marca que representa o Brasil" (PINTO, 2019).

A seleção da empresa Petrobras no presente estudo também se deu pela sua relevância e popularidade. Além do fato de que o nome "Petrobras" não é homônimo, o que auxilia na obtenção dos dados. Caso a escolha fosse por outra empresa, como a Oi, todos os *tweets* com a palavra "oi" seriam coletadas. Com isso, poderiam surgir mensagens que não são relacionadas com a operadora, e sim, com uma saudação.

O trabalho apresentou dois diferenciais. O primeiro é a personalização do dicionário para organizações individuais. A Petrobras foi escolhida para o estudo de caso. Os dicionários existentes são genéricos e, portanto, com menor qualidade. Dessa forma, o indicador de sentimento particular pode mitigar ou melhorar o movimento de mercado dessa empresa. O

segundo é a 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 e comenta o que pensa sem influência de um entrevistador.

A revisão da literatura foi apresentada em quatro tópicos. O capítulo 3 trata-se da metodologia utilizada para a análise, a fonte e a natureza dos dados utilizados. No capítulo 4, são apresentados os resultados obtidos. Está subdividido em três partes: a primeira é a análise das amostras; a segunda, construção do dicionário; a terceira, a aplicação desse dicionário para verificar se o sentimento público exposto no Twitter alterou os valores das ações. Por último, o capítulo 5, apresenta as conclusões do presente trabalho e as sugestões de pesquisas futuras relacionadas ao tema.

2. REVISÃO DA LITERATURA

2.1 Teoria de Finanças Comportamentais

A teoria da economia comportamental surgiu em 1902, com a publicação do livro *La Psychologie Économique*, pelo jurista francês Gabriel Tarde. Nessa obra, é abordada a concepção geral da sociedade, no aspecto social, e como isso impacta o lado econômico. Os indivíduos interagem uns com os outros em sociedade, mas não de uma forma pré-definida. A teoria da imitação explica que existem ações mais frequentes do que outras. Se um grupo de mentes concebem uma ideia e se essa, pelo menos, parecer lógica para as pessoas que estão à volta do grupo, a tendência é que essa ideia se propague, de acordo com um tipo de progressão geométrica. Essa teoria também foi chamada de radiação imitativa.

"*Aussi bien que l'Utilité, la Vérité et la Beauté sont filles del'Opinion, de l'opinion de la masse en lutte ou en accord constant avec la raison d'une élite qui influe sur elle*" (TARDE, 1902, p.52). Ou seja, para haver uma noção de verdade, é necessário que as opiniões sobre ideias, informações ou até bens tangíveis sejam unificadas.

No final dos anos 40, o estudo de Herbert Simon discutiu sobre a ideia de racionalidade limitada, demonstrando o processo de tomada de decisão em condições de incerteza. A qualquer momento, existem inúmeras alternativas possíveis que um indivíduo pode escolher, e ele precisaria analisar todas essas alternativas e suas consequências, se tivesse alto grau de racionalidade, além de ter um sistema de valores como critério para escolher uma alternativa dentre todas as demais. Sendo que o ser humano não possui informação completa das condições que cercam sua ação e, mesmo se tivesse, seria incapaz de absorver toda a informação disponível, o que o obrigaria a fazer um recorte subjetivo da situação e assumir premissas de alto valor subjetivo (SIMON, 1947). "*This is impossible for the behavior of a single, isolated individual to reach any high degree of rationality*" (SIMON, 1947, p. 79).

O trabalho critica a teoria clássica da racionalidade quando conclui que, nessa teoria, os indivíduos são limitados por seus hábitos, reflexos, valores, concepções de propósito, conhecimento e informação. O que significa que cada decisão envolve a seleção de uma meta (elemento de valor) e de um comportamento (elemento de fato) com ela relacionada. Toda decisão é composta por dois tipos de elementos, chamados de elementos de "fato" e de "valor". O indivíduo até pode ser racional, se tiver informações completas e perfeitas sobre as condições que cercam sua ação, e se tiver um pensamento correto do objetivo da ação, mas essa racionalidade não determina o comportamento. O problema é como analisar os limites da

racionalidade e a maneira que a organização afeta esses limites para que a pessoa tome uma decisão (SIMON, 1947).

Em oposição à teoria do homem puramente racional (*homo economicus*), Simon apresenta o conceito do Homem Administrativo. Nessa teoria, os indivíduos realizam escolhas que os satisfazem em vez das decisões otimizadas. A decisão satisfatória é a melhor diante de um determinado contexto, não necessariamente ótima (SIMON, 1947).

O efeito ancoragem diz que as pessoas podem ser influenciadas por informações completamente hipotéticas. Como exemplo, o pesquisador girou uma roda com números aleatórios e mostrou o número sorteado aos participantes da experiência. Em seguida, perguntou a eles se o número dos países africanos membros das Nações Unidas era maior ou menor que um número tirado na roda. Depois que os participantes responderam, foi pedido que estimassem o número de países africanos que participam das Nações Unidas. Apesar do número da roleta ter sido aleatório, exerceu influência sobre o valor das estimativas dos participantes (TVERSKY; KAHNEMAN, 1974).

Quanto mais opções de escolha o indivíduo tiver, melhor será para o processo de tomada de decisão. Contudo, o excesso de opções também pode ser um problema. Foi realizado um estudo no supermercado, onde havia duas marcas de geleias de frutas. Uma marca tinha 24 sabores de geleia e a outra, apenas seis. A marca com mais opções de sabores chamou maior atenção dos clientes. Contudo, a segunda marca vendeu mais geleias do que a primeira. Ter mais opções foi mais atraente para o consumidor, mas por outro lado, dificultou seu processo de decisão (IYENGAR; LEPPER, 2000).

Emoções e erros cognitivos influenciam na tomada de decisão e o mesmo problema pode ser analisador de diversas formas diferentes. No entanto, o estudo de Halfeld e Torres (2001) aponta que existem padrões de comportamentos, em alguns casos, em que as anomalias no comportamento das pessoas são previsíveis, como aversão à perda, excesso de confiança e otimismo em exagero.

A linha das finanças comportamentais surgiu no final da década de 1970. A nova vertente defende que os retornos podem ser influenciados por aspectos comportamentais e psicológicos dos investidores. Os agentes atuam de maneira irracional e cometem erros esperados de previsão (NOFSINGER, 2010). O homem (*homo economicus*), do modelo de finanças comportamentais, não é perfeitamente irracional e nem racional. No processo de tomada de decisão, o agente analisa as informações disponíveis e considera todos os meios de solucionar o problema. A racionalidade não é o centro do pensamento para tomada de decisão. (HALFELD; TORRES, 2001).

Há uma assimetria entre ganhos e perdas. A função de utilidade, nesse modelo, demonstra que uma perda gera mais impacto do que um ganho na mesma magnitude. Os indivíduos têm aversão ao risco, não aversão à perda (KAHNEMAN; TVERSKY, 1979).

Um dos primeiros estudos publicados sobre finança comportamental foi o trabalho de Kahneman e Tverski (1990). Os pesquisadores analisaram o processo de tomada de decisão dos investidores em situações de risco. Nesse estudo, foi verificado que indivíduos com riscos e benefícios (ganho ou perda) iguais, têm comportamentos diferentes nos processos de tomadas de decisão. O conceito de aversão à perda também surgiu por esse estudo. Os indivíduos sentem muito mais a dor de uma perda do que o prazer de um ganho.

O sentimento cria um ruído no mercado e gera um risco no preço do ativo. Quando há muitos torcedores, é provável que o preço do ativo seja maior do que seria com indivíduos racionais e ausência de risco. O risco provocado por anomalias no mercado financeiro causa excesso de volatilidade, subestimação dos fundos e reversão média de retorno das ações (DE LONG et al., 1990).

O modelo também diverge da teoria dos curtos irrecuperáveis. Nessa teoria, o valor pago em um item não pode ser recuperado depois de comprado. Com isso, os custos irrecuperáveis não influenciam o comportamento futuro. Porém, isso não ocorre na realidade. As pessoas tendem a dar importância ao que já pagaram por um bem (GENESOVE; MAYER, 2001).

Lakonishok, Shleifer e Vishny (1992) estudaram o efeito manada, onde as ações são compradas ou vendidas em massa. Os autores verificaram que o mercado possui esse efeito apenas onde as negociações apresentaram *feedback* positivo nas ações menores. O mesmo não aconteceu nas maiores ações. Dessa forma, o resultado não exclui o efeito rebanho como um todo e não analisa a possibilidade de demandas altamente inelásticas, que não altera a quantidade de rebanho e, por consequência, não sofre grandes impactos sobre os preços das ações.

A estratégia denominada “momentum” é onde se compra ações que tiveram alta no período anterior e vendem ações que tiveram queda no período anterior. Essa estratégia gera um retorno positivo. A interpretação desse resultado é que os retornos adquiridos no primeiro ano, após a formação do portfólio, se dissipam nos dois anos seguintes. Contudo, isso não se deve a reações a fatores comuns ou ao risco sistemático (JEGADEESH; TITMAN, 1993).

Jin e Kothari (2008) analisaram os determinantes do comportamento da decisão dos gestores de vender suas ações. Os CEOs venderam grandes montantes de seus patrimônios, ao contrário do que espera o senso comum, que evita por medo de uma reação adversa do mercado

de ações. Como resultado, verificou-se que os impostos diminuem a confiança dos gestores de venderem seus patrimônios.

A partir de 2009, cresceu o número de pesquisas de finanças comportamentais. Um possível motivo é o aumento de anomalias ocorridas no mercado financeiro, depois da crise de 2007, devido à ineficiência desses mercados (VIANEZ; MARTÍNEZ; ROMÁN, 2020).

Graham, Harvey e Huang (2009) concluíram que os investidores que acreditam que são mais competentes, habilidosos ou com conhecimento em tomar decisões, estão mais dispostos a negociar com mais frequência. O aspecto psicológico é intrínseco a esse achado. O estudo mostra que investidores homens, com carteiras maiores e com mais educação, são mais propensos a se considerarem mais competentes do que investidores mulheres, com carteiras menores ou com menos educação.

Em momento de incerteza, os investidores tendem a agir com determinados padrões psicológicos para determinadas circunstâncias. Esses vieses cognitivos são relacionados à ineficiência informacional, emoções e sentimentos. Na época, a crise financeira americana foi caracterizada por um colapso de confiança nos agentes e nos preços dos ativos. O futuro era incerto, e a ansiedade que surgiu com a crise era ruim para o sistema financeiro, porque poderia provocar um desespero nos bancos para não falhar, o que também implicaria a uma questão moral. Uma solução racionalizada não teria efeito. É necessário analisar a economia com base na teoria do comportamento humano (DOW, 2011).

Quando os investidores prestam menos atenção no mercado financeiro, alocam mais atenção nas informações do mercado como um todo que nas informações específicas da firma. A atenção é escassa. No estudo, com intuito de testar a teoria, foi utilizado como evento os dias das loterias nacionais de Taiwan como choques exógenos, que atraem a atenção dos investidores para longe do mercado de ações. Foi descoberto que os dias de loteria estão associados a menor rotatividade de ações e de volume de pesquisa sobre empresas, mas com maior volume de pesquisa de loterias no Google. Logo, os investidores podem passar o tempo pesquisando sobre a loteria, assistindo à mídia a falar sobre o evento, fazendo filas para comprar os bilhetes, ou até mesmo conversando com os amigos sobre o jogo, mas ainda assim, não dispensam a atenção para o mercado de ações (HUANG; HUANG; LIN, 2019).

2.2 Sentimento do Investidor

O sentimento do investidor está embutido nas opiniões e nos discursos não verbais. Nesse sentido, é o sentimento que dá o tom mais otimista (positivo) ou pessimista (negativo) a um texto opinativo. A análise linguística das palavras pode captar o sentimento otimista ou

pessimista de quem as escreveu. Nesse sentido, a avaliação do sentimento textual está atrelada a aspectos léxicos, de acordo com o contexto que a palavra está inserida, e não com o significado ou sintaxe (BEATTIE, 2014).

O sentimento dos investidores pode promover desvios no valor de um ativo que, em teoria, deveria ser igual ao seu valor intrínseco. A teoria analisa os sentimentos que levam os investidores a tomar decisões distantes das decisões ótimas, de acordo com a teoria de mercados eficientes. Uma mudança no sentimento do investidor pode gerar consequências de eventos, como novos padrões de compras. Com base nessa premissa, diversos pesquisadores buscaram analisar qual é a relação entre o sentimento e o mercado (BAKER; WURGLER, 2006).

Os investidores formam suas expectativas sobre os retornos dos ativos, com base em sentimentos que não possuem natureza contábil e nem fundamento clássico (KAWAKATSU; MOREY, 1999). Barberis, Shleifer e Vichny (1998) apresentaram modelo sobre como os investidores formam crenças que impactam suas tomadas de decisão. A cada período, o investidor observa os ganhos ou as perdas e usa essa informação para atualizar suas crenças. Nessa atualização, o investidor é bayesiano. Se houver uma alta frequência de retornos positivos, a probabilidade do investidor de se manter no regime, aumenta. Por outro lado, se houver uma surpresa negativa, aumenta a probabilidade dele sair do regime que está.

Os modelos propostos foram baseados no conservadorismo, que afirma que as pessoas demoram para mudar suas crenças frente a novos achados, e na heurística da representatividade, onde as pessoas observam padrões dentro da aleatoriedade, ao comparar uma amostra a uma população que é considerada padrão, como quando uma ação tem o histórico de crescimento, e o passado se torna um potencial de crescimento e supervalorizam os ganhos (*overreact*), mesmo que o crescimento tenha sido por sorte (BARBERIS, SHLEIFER E VICHNY, 1998).

O modelo do conservadorismo analisou se o anúncio de ganho não impacta muito na previsão de preços. Já no segundo modelo, observa-se que acontece ao contrário. No estudo, o modelo de previsão não muda, só a transição que se altera. O peso e a força das notícias também impactam de formas variadas nas tomadas de decisão dos agentes. Notícia com mais força gera maior reação nos investidores (BARBERIS; SHLEIFER; VISHNY, 1998).

O componente emocional “confiança” foi objeto de estudo de Barber e Odean (2001). O resultado encontrado foi de que o excesso de confiança leva aos agentes a aumentarem o número de negociações, gerando assim, maior rotatividade nos ativos do portfólio. Contudo, Ko e Huang (2007) notaram que o excesso de confiança e a arrogância só aumentam a volatilidade e o retorno se houver uma quantidade de informação privada alta. Tirando esse caso, os mercados seriam menos eficientes e criariam preços incorretos.

As pesquisas de opinião dos investidores deveriam ser as medidas mais importantes e que sintetizassem o sentimento. Porém, foi observado pela comunidade acadêmica que, o que os entrevistados respondem, não é visto na realidade. Outros problemas também são vistos nessas pesquisas, como amostra selecionada, quantidade de respondentes e formulação de perguntas (YOSHINAGA, 2009).

Diante disso, Domingues (2008) criou um índice de sentimento do investidor, entrevistando diariamente os analistas sobre as expectativas de desempenho no dia seguinte. Foram feitas regressões lineares para testar a previsibilidade com o índice IBOVESPA. Como achado, foi observado que o índice criado conseguiu antecipar as oscilações futuras do mercado acionário brasileiro.

A literatura em Psicologia descobriu que a luz do sol pode melhorar o humor e a emoção de um indivíduo. Pessoas com bom humor possuem comportamento mais otimista. Sabendo disso, foi testado se a luz do sol afeta os retornos das ações. Não é analisada a notícia da previsão do clima, mas sim a relação entre a nebulosidade diária e o retorno diário de 26 bolsas de valores ao redor do mundo. Foi verificado que a luz do sol possui relação positiva com os retornos. Os investidores ficam mais otimistas com o mercado. Também foi testado para os climas de chuva e neve, porém, não obtiveram significância (HIRSHLEIFER; SHUMWAY, 2005).

Eventos, como férias, criam um sentimento positivo nos indivíduos, enquanto guerras ou desastres naturais, criam um sentimento negativo. Com base nisso, foi analisado qual seria o efeito se houvesse uma guerra durante o feriado. Era esperado que um efeito positivo anulasse o outro, negativo, gerando uma influência neutra. Porém, foi verificado que os dois efeitos estavam presentes, mas que o sentimento negativo da guerra dominava o sentimento positivo do feriado. Portanto, existem níveis de impacto entre efeitos positivos e negativos e é necessário que se tenha uma interpretação de cada situação em específico (KAPLANSKI; LEVY, 2012).

A religião também exerce influência no sentimento do investidor. O mês sagrado do Ramadã exerce influência positiva no sentimento do investidor em países de maioria muçulmana. Contudo, esse efeito diminui gradualmente com o passar dos anos, pois os investidores começaram a conhecer essa anomalia. Com isso, o impacto do Ramadã foi revertido ao longo do tempo, deixando de gerar um aumento, para acontecer uma diminuição nas flutuações dos preços das ações (BIAŁKOWSKI et al., 2013).

Silva (2017) investigou o efeito do sentimento textual das notícias dos jornais sobre o comportamento dos preços no mercado acionário brasileiro. Foi criada uma lista de termos financeiros, utilizando a técnica do *vector space model*, a partir das reportagens, e estimado um índice, analisando a frequência das palavras e ponderando-as. Como resultado, foi observado

que as notícias de jornais, no Brasil, têm influência no retorno e na volatilidade do mercado acionário. Em momentos de incerteza econômica e crises, esses impactos são mais expressivos, chegando a prever queda de -76.21 pontos-base sobre os menores retornos do mercado.

A análise de sentimento também é uma forma de pesquisa de *big data*. É uma maneira de extrair emoções e sentimentos de informações textuais dentro do conteúdo da *web* de determinadas páginas. Diariamente, as pessoas escrevem posts, comentários, críticas e *tweets* sobre variados tópicos. Com isso, é possível verificar se produtos, marcas e pessoas são visualizados de forma positiva ou negativa na internet. (GODSAY, 2015).

2.3 Sentimento Textual

As pesquisas relacionadas ao mercado de capitais utilizam, em sua maioria, os dados do Twitter e Facebook. A escolha está relacionada à disponibilidade de dados, número de usuários ou alta popularidade no mundo ocidental, onde se examina os mercados de capitais dos Estados Unidos e Europa (BUKOVINA, 2016). A mídia social utilizada na pesquisa é o Twitter. Essa escolha foi realizada com análise em diversas razões. A mídia é popular. O estudo, realizado pelo site Statista (CLEMENT, 2020), demonstrou que o Brasil está na lista dos dez países com mais usuários na mídia social, com 12,15 milhões de usuários ativos. Outro motivo é que as postagens são públicas. As postagens e os seguidores das contas do Twitter são visíveis e de fácil acesso do público.

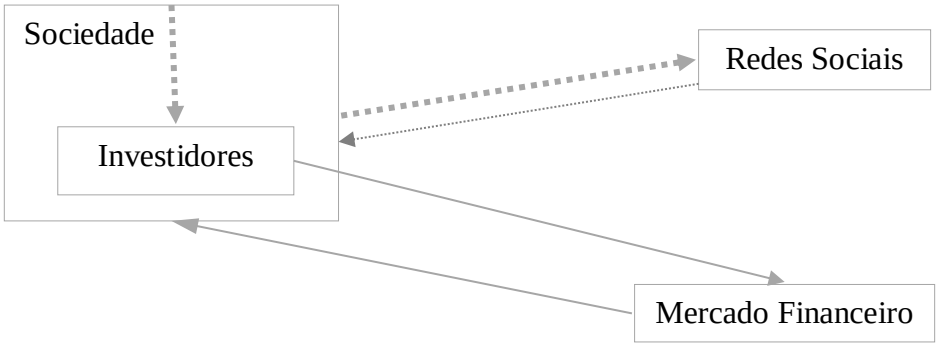
A publicação no Twitter, conhecida como *tweet*, é composta por 280 caracteres ou menos. Os usuários podem postar para o público, de forma imediata, seus pensamentos, curiosidades e opiniões sobre qualquer tópico. Além de ser uma mídia aberta ao público, os dados do Twitter são valiosos pela velocidade da exposição do pensamento e pela disponibilidade de consumo em tempo real dos conteúdos.

O texto é comparado com a lista de palavras. O processo da análise de sentimento tem três etapas: baixar os dados da mídia social, transformar o conteúdo da web em texto e transformar a natureza qualitativa para quantitativa, por meio do aprendizado da máquina ou método lexical. Em geral, existem dois métodos para análise de sentimentos em mídias sociais. O primeiro é o modo lexical, em que é utilizado uma lista de palavras pré-definidas, sendo que cada palavra transmite um sentimento específico. O segundo é o método de aprendizagem da máquina, em que o programa aprende os dados. Esse programa tem um modelo pré-definido, que classifica o texto, como sentimento positivo ou sentimento negativo (GODSAY, 2015).

Sprenger et al. (2014) estudou os fóruns de investimento do Twitter como fonte de informação alternativa do varejo e criou um mecanismo que exemplifica como funciona a

demanda de informações. O mecanismo está exposto na Figura 1, que foi reelaborada a partir desse estudo citado anteriormente com o trabalho de Santos, Laender e Pereira (2015).

Fig. 1. Mecanismos de transmissão – demanda de informações



Fonte: Elaboração própria baseada nos artigos de Sprenger et al. (2014) e de Santos, Laender e Pereira (2015).

É possível perceber, pela Figura 1, que os fóruns de investimento presentes nas mídias sociais se conectam com os investidores com uma discussão sobre ações e mercados. O mecanismo começa com os investidores e a sociedade postando suas emoções nas mídias sociais. Os investidores buscam informações para fazer sua tomada de decisão, no mecanismo de pesquisa do Google ou em fóruns de investimento em mídias sociais. Como a mídia social não é a única fonte de informação dos investidores, pode-se presumir, pelo menos, um impacto parcial da demanda de informações. Isso é indicado pela seta pontilhada dos investidores às mídias sociais mutuamente. O mercado define o valor do preço das ações e índice pela negociação dos investidores, de forma agregada. Isso é indicado pela seta sólida. Por fim, o sentimento adquirido influencia na tomada de decisão do investidor, que se reflete no preço das ações (SPRENGER et al., 2014). Contudo, os movimentos das ações, eventos e notícias também são capazes de impactar o Twitter e de gerar picos de postagens, e a frequência das postagens acompanha o início das negociações e se mantém por três horas (SANTOS; LAENDER; PEREIRA, 2015).

Na literatura de aprendizado da máquina e previsão do mercado de ações são utilizados, por exemplo, os seguintes classificadores: i) o *Opinion Finder*, que é dicotômico e classifica o sentimento como negativo ou positivo; ii) GPOMS, que é o perfil de estados de humor do Google, onde cada palavra do texto analisado é ponderado, de acordo com sua relação com o humor desejado (calmo, alerta, claro, animado, gentil e feliz); iii) Árvores de decisão reforçadas, que são classificadores bayesianos; iv) Mecanismos de classificação de idioma natural, como o *Google Trends*; v) ANEW, que são as normas para palavras inglesas; vi) LC, usado para sentimentos dicotômicos, útil para grandes textos; vii) *WorldTracker*, onde se usa

palavras-chaves de pesquisa para analisar os termos mais usados, a frequência de uso das palavras e número de pesquisas realizadas; e, viii) *SentiWN*, baseado no *WordNet*, que mede a polaridade e a intensidade (NARDO; PETRACCO-GIUDICI; NALTSIDIS, 2015).

O sentimento da mídia social é importante preditor na determinação da avaliação do indivíduo, mas nem todas as mensagens de mídia social têm o mesmo impacto. O grupo de ativos no Twitter é a minoria. Entretanto, acontece uma influência desproporcional na população. Ainda assim, a minoria posta cada vez mais, pois assim possui maior probabilidade de se tornar um líder. No Twitter, as informações são diárias, visíveis e acessíveis, e os investidores são mais propensos a ler e a responder a essa mídia social. Esse efeito também acontece no mercado financeiro, que pode levar ao comportamento de manada, em que os indivíduos do grupo reagem todos da mesma forma (MAI et al., 2018).

O sentimento é capturado rastreando variações na frequência de palavras relacionadas às emoções nas mídias sociais. Especificamente, o humor positivo ou negativo, em relação às ações, é calculado em termos da fração de palavras positivas ou negativas e ajustado de acordo com a importância da postagem e data de publicação. O poder das postagens é seu poder de influência (LI et al, 2017).

A maioria dos estudos utiliza a ponderação proporcional dos termos e considera todas as palavras da lista igualmente importantes. O estudo de Harvard, por exemplo, calcula frequências simples para palavras que aparecem no texto que se enquadram em cada categoria. Contudo, existem estudos que ponderam cada termo de forma diferente. Loughran e McDonald (2011) usaram dois esquemas de ponderação: uma ponderação proporcional simples e outra que pondera cada palavra proporcionalmente ao inverso simples e a frequência dela no documento. Assim, quanto mais recorrente foi o termo, menor foi sua ponderação. Mao et al. (2014) calcularam o valor baseado em dois níveis: i) nível de palavra, em que a palavra recebe um valor fixo, independentemente do contexto da frase em que a palavra está; e, ii) nível da frase, onde a influência do advérbio determina um valor para a frase. O valor final é a soma do valor das sentenças, e o valor relativo é o valor somado dividido pelo número de frases.

Na abordagem baseada em dicionário as medidas mais comuns do sentimento textual são calculadas a partir do número de palavras de determinada categoria de sentimento, em relação ao número total de palavras no texto, como em Chen (2013), ou a porcentagem padronizada, como em Tetlock et al. (2008). Essa padronização pode ser necessária, se houver uma mudança brutal no estilo de escrita dos autores, quando ocorre uma mudança na distribuição das palavras no texto (TETLOCK et al., 2008). Outra medida é a contagem de palavras positivas menos a contagem de palavras negativas, dividida pela soma das contagens

de palavras positivas e negativas ou da contagem total (FEUERRIEGEL; WOLFF; NEUMANN, 2015).

Depois de construir a medida do sentimento, os pesquisadores aplicam em modelos econométricos e realizam o teste de hipótese, para analisar a interação entre o sentimento e as variáveis financeiras. Os principais modelos usados são regressão linear, vetoriais de regressão automática, logística e Probit, de volatilidade do tipo GARCH e de regressão quantílica.

A abordagem mais comum tem sido empregar o modelo de regressão linear em dados de séries temporais, que abrange o sentimento geral no nível do mercado e o desempenho do índice de ações. No lugar onde o sentimento transversal e as variáveis de desempenho no nível da empresa estão disponíveis, alguns pesquisadores usaram métodos de regressão de painel para controlar a heterogeneidade individual das empresas. Também é incluído variáveis de controle no modelo, pois assim é possível verificar se o sentimento tem previsibilidade, apesar do controle das variáveis da empresa e do mercado. Exemplos dessas variáveis são: fluxo de caixa, *book-to-market*, patrimônio líquido, alavancagem, índices do mercado, volume de negociações (KEARNEY; LIU, 2014).

Os pesquisadores utilizaram as metodologias de estudo de evento e causalidade reversa. A primeira busca analisar se o sentimento em divulgações corporativas ou notícias impactam os retornos normais e cumulativos das empresas, como fizeram Chen (2013) e Loughran e McDonald (2011). A segunda busca verificar se o desempenho da empresa ou do mercado pode impactar as medidas futuras do sentimento, como Garcia (2012) fez em seu estudo.

O modelo de autorregressão vetorial (VAR) é utilizado para capturar a evolução e as interdependências entre o desempenho no nível da empresa ou do mercado, as variáveis de controle e o sentimento. O VAR é um modelo de série temporal mais avançado do que os modelos de equação única, na medida em que todas as variáveis no VAR são determinadas endogenamente por seu próprio histórico e pelo histórico de todas as outras variáveis no modelo (KEARNEY; LIU, 2014).

As regressões logísticas são usadas para verificar se um sentimento textual ajuda a prever ou a identificar se um evento específico é provável de acontecer. Essas regressões não presumem uma relação linear entre as variáveis dependentes e independentes e a variável dependente deve ser dicotomia, ou seja, ter apenas duas categorias. Loughran e McDonald (2011) utilizaram a regressão logit para analisar se as empresas acusadas de fraude contábil e as empresas que relatam fracos controles internos usam uma linguagem diferente das outras empresas em seus 10-Ks.

Os modelos de volatilidade são empregados para testar se o sentimento é relacionado ao risco. A abordagem do tipo GARCH é a mais utilizada. No geral, as notícias estão inversamente relacionadas com a volatilidade. Um sentimento textual positivo está associado a níveis mais baixos de volatilidade (MARTY, VANSTONE, HAHN, 2020).

A regressão quantílica é usada para analisar o sentimento para cada percentil das empresas. Loughran e McDonald (2015) analisaram qual era o melhor indicador de sentimento para medir o tom e o otimismo ou pessimismo do texto financeiro, que impactam os preços das ações. Para tal feito, separou as empresas em três quartis.

2.4 Estudos anteriores sobre o sentimento textual em finanças

Para este tópico, realizou-se uma revisão sistemática da literatura, obtida mediante busca na base de dados Scopus, utilizando os termos, em inglês, de “Sentimento textual” e “finanças”. Até o momento da pesquisa, foram encontrados 57 artigos. Outros estudos também foram acrescentados para incrementar a revisão.

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 era 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. Foram escolhidos 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.

Mao, Counts e Bollen (2011) 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 de causalidade de Granger, notícias e consultas de pesquisas de opinião. Como resultado, foi verificado que pesquisas tradicionais não têm poder preditivo do mercado. No entanto, o Twitter e a frequência de termos financeiros na rede social, nos dois dias anteriores, foram considerados estatisticamente preditores do retorno diário das ações.

As opiniões dos indivíduos expostos no Twitter, entre 2009 e 2012, foram estudadas e analisadas, se elas poderiam prever seus ganhos e retornos. Os autores utilizaram o modelo bayenesiano como forma de mensurar os *tweets*, e como metodologia duas regressões 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 investidor 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 (BARTOV; FAUREL; MOHARAM, 2018).

Os indicadores de características de qualidade textual cognitiva — número de palavras, legibilidade, erros de ortografia e erros gramaticais — foram analisados em descrições de empréstimos, para saber se eles influenciavam no sucesso do financiamento. Os dados foram obtidos por meio de uma ferramenta de análise de texto online (*Readable*) e, depois, foram feitas a classificação manual do número de palavras, avaliação da legibilidade e correção de 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, por meio da qualidade textual, para assim, atrair os credores (PENGNATE; RIGGINS, 2020).

McGurk, Nowak e Hall (2020) e Seng e Yang (2017) também fizeram uma análise subjetiva e manual para a classificação textual, com objetivo de 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, onde a semântica é investigada de maneira automatizada nas notícias, relatórios ou cartas do CEO (YADAV et al., 2020; CHE; ZHU; LI, 2020).

Ito et al. (2017) apresentaram uma nova forma de desenvolver o indicador de sentimento textual, denominado GAP, 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). Uma análise baseada em características linguísticas foi objeto de pesquisa 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 advérbio. O sentimento textual teve efeito significativo sobre o movimento do mercado (MAO et al., 2014).

Li, Wu e Wang (2020) e Mangee (2018) analisaram informações financeiras e o impacto delas 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 e possui uma relação mais forte com o comportamento do preço das ações. Loughran e McDonald (2015) verificaram que o dicionário *Loughran-McDonald Financial* (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 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 da 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 é feita 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 (MAI et al., 2018; WANG, 2017). A análise do sentimento deve ser feita pela união dos métodos (DARENA et al., 2018).

Constatou-se que o sentimento agregado expresso em *tweets* financeiros transmite informações sobre a avaliação das empresas. Só foram utilizados os *tweets* com a *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 (RANCO et al., 2015). Com o mesmo algoritmo, foi verificado que os *tweets* possuem capacidade de previsão (FANG et al., 2017), e que as notícias financeiras publicadas recentemente também possuem capacidade de previsão do retorno dos preços (AB. RAHMAN, ABDUL-RAHMAN E MUTALIB, 2017).

O estudo de Pagolu et al. (2016) concluiu que as mudanças nos preços das ações de uma empresa possuem relação 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, consequentemente, os preços das ações aumentaram. Ito et al. (2017) 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.

Os estados emocionais do público foram analisados por meio de *feeds* do Twitter para prever tendências. Os *tweets* foram classificados como positivos e negativos com o buscador de opinião e pelo algoritmo *Google-Profile of Mood States*, que classifica o 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 (BOLLEN; MAO; ZENG, 2011).

Twitter de pessoas importantes para a sociedade também foram investigados. Foi verificado que os *tweets* do então Presidente dos Estados Unidos possuem influência no mercado acionário americano, aumentando assim o volume e a volatilidade dos preços, além de atrair maior atenção do investidor. No entanto, esse efeito se dissipa em poucos dias (GE; KUROV; WOLFE, 2018).

No Brasil, a análise de sentimento por meio das redes começou a ser abordada de forma lexical. Um dos primeiros estudos no Brasil analisou o sentimento nas opiniões de produtos escritos em português. O método que se mostrou mais eficiente foi o Sentilex (AVANÇO; NUNES, 2014). Posteriormente, os *tweets* foram analisados junto às notícias sobre 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-se 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).

O estudo de Souza, Lucena e Queiroz (2019) avaliou o sentimento do investidor no 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 nas línguas portuguesa e inglesa, 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 (B*ayes, 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 atual Presidente da República, 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, pelo pouco tempo de governo do Presidente e pelo conteúdo das postagens, que podem ser interpretadas mais voltadas para o meio político do que ao mercado (TOMMASETTI; MAIA; MACEDO, 2021).

Além dos estudos anteriores, para analisar o sentimento do investidor, via Twitter, sobre o 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. Dados do mercado de capitais e da mídia social foram coletados para entender melhor o comportamento dos investidores e o que os influenciam nas suas tomadas de decisão.

O modelo geral do estudo analisou a influência do sentimento textual do Twitter no retorno da Petrobras, 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, têm-se as hipóteses de pesquisa:

H1: O sentimento textual se relaciona com o retorno;

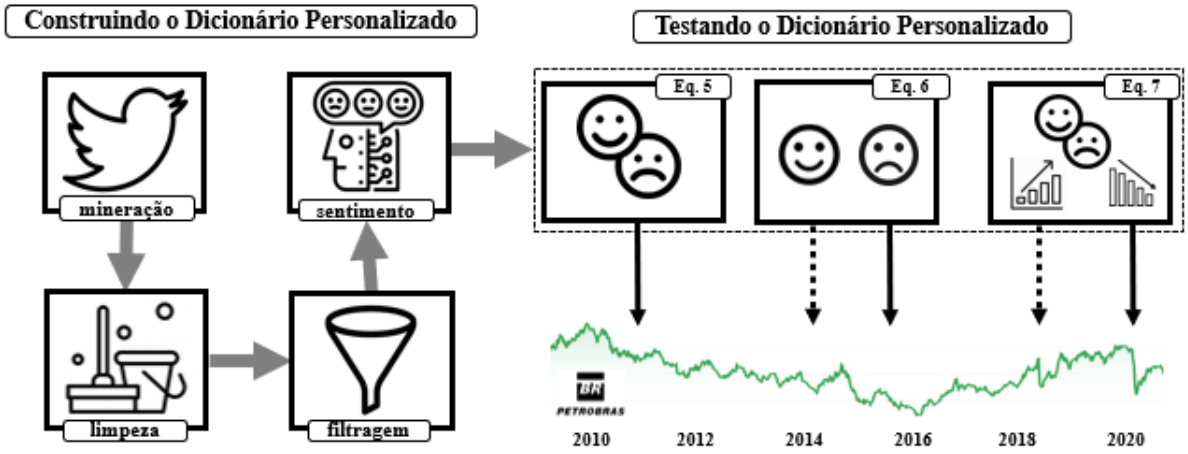
H1a: O sentimento textual positivo e sentimento textual negativo se relacionam com o retorno; e,

H1b: O sentimento textual impacta 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 a Petrobras e as mensagens expostas no Twitter entre 2010 e 2020. O histórico da série diária de preços da Petrobras foi coletado na *Bloomberg*. A metodologia foi realizada em duas fases: construção da variável variação do sentimento e os testes das hipóteses. A Figura 2 ilustra o fluxo de trabalho utilizado no estudo.

Figura 2. Etapas para o Processamento da Linguagem Natural usados no estudo



Fonte: Elaboração própria.

Para realizar a análise do sentimento exposto no Twitter, foi necessário coletar as postagens acerca da Petrobras em pré-processamento dos dados. Os *tweets* foram coletados por meio da biblioteca *GetOldTweets3*, no *Python*. Foram selecionadas as postagens que estão em língua portuguesa, que continham o nome da instituição (Petrobras), os *tickers* da ação (PETR3 e PETR4) e as respostas às postagens da empresa na sua conta oficial (@petrobras), entre 2010 e junho de 2020.

A escolha do banco de dados foi feita com o propósito de avaliar qual informação seria mais importante para o investidor: sobre a Petrobras, 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), em que coletaram postagens, em português ou inglês, que continham nomes das instituições participantes do índice IBOVESPA e/ou termos escolhidos pelos autores: IBOVESPA, IBOV, Brasil, Bolsa, Balcão e Bolsa de Valores Brasileira.

A construção do indicador de sentimento parte da mineração das mensagens postadas nas redes sociais. Como os textos são longos, foi necessário realizar a tokenização, que converte cada *tweet* em um vetor de palavras ("unigrama") (LI; WU; WANG, 2020). Em seguida, foi necessário limpar o banco de dados removendo *stopwords*, que são palavras que não possuem valor semântico. Essa mineração de dados foi construída com algoritmos em Python e, contendo esse tratamento de limpeza nos dados, a biblioteca usada foi a *tm – Text Mining*.

A fase subsequente foi descartar as palavras neutras. Estas aparecem todos os dias e na mesma frequência, independentemente da variação diária do preço das ações da empresa. O filtro foi palavras com diferença média e de desvio padrão entre dias de retorno positivo e negativo inferiores a 20%.

Posteriormente, as palavras filtradas foram inseridas no mapa perceptual. O dimensionamento multidimensional é um método para visualizar o grau de dissimilaridade entre os objetos. Com base nestes coeficientes de diferença, uma série de distâncias euclidianas é calculada. Os dados são geralmente plotados em dois espaços cartesianos multidimensionais (GOWER; LEGENDRE, 1986). Preliminarmente, é necessário calcular a semelhança, por meio do conjunto de objetos, para construir um mapa perceptivo. Suponha que as palavras "a" e "b" recebam valores semelhantes em comparação com outros objetos possíveis ("c", "d", "e"). Nesse caso, a escala multidimensional atribui uma distância euclidiana menor entre "a" e "b" do que entre os outros objetos.

A lista de palavras do estudo foi posicionada no mapa. Cada vetor do mapa é representado por uma palavra dessa lista. Os eixos representam os polos positivos e negativos, para que, quanto mais extrema a palavra for, mais positiva ou negativa ela será. Com isso, foi realizada a classificação de polaridade, de acordo com o eixo que o termo se encontra.

A abordagem de classificação utilizada por essa metodologia é o chamado "*bag of words*", que separa os termos do dicionário em positivos e negativos. Assim, as palavras recebem escores numéricos, de acordo com sua polaridade, sendo positivo (+1) e negativo (-1) (JIANG et al., 2019).

Essa lista de palavras positivas e negativas relacionadas à empresa consistiu no dicionário personalizado. O valor de um sentimento de *tweet* foi estimado com base no peso das palavras positivas e negativas. A medida é o total de palavras positivas menos negativas, como segue na Equação 1.

$$STO_t = \sum PP_t + \sum PN_t \tag{1}$$

Onde:

STO_t: Sentimento Total dos Tweets j;

PP_t: Quantidade de palavras positivas (Sua polaridade é +1);

PN_t: Quantidade de palavras negativas (Sua polaridade é -1).

Com o objetivo de realizar um teste de robustez da variável sentimento, foi construída a segunda variável de sentimento diário, sendo essa a média de sentimento dos *tweets* em determinado dia. Essa decisão teve o objetivo de investigar se haveria diferença nos resultados, caso a variável de sentimento fosse construída de diferentes formas. Foi um teste de robustez da classificação de polaridade dos termos do estudo. 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.

$$SME_t = \frac{\sum PP_t + \sum PN_t}{\sum PT_t} \tag{2}$$

Onde:

SME_t: Sentimento Médio do Tweet j;

PT_t: Quantidade de palavras totais.

A primeira diferença foi usada nas duas variáveis de sentimento (ΔSTO e ΔSME) para capturar a mudança no sentimento global em vez do status absoluto (positivo ou negativo). Por exemplo, se o índice passar de 0,90 para 0,75, continua positivo em termos absolutos, mas sua tendência — que é nossa variável de foco — é negativa.

Para calcular o retorno observado no período da Petrobras, será utilizada a seguinte função logarítmica que analisa o retorno sobre o preço da ação no dia presente e no dia anterior (FAMA, 1995):

$$Ret_t = ln\left(\frac{P_t}{P_{t-1}}\right) \tag{3}$$

Onde:

Ret_t: Retorno observado da Petrobras;

P_t: Preço da ação no dia t;

P_{t-1}: Preço da ação no dia *t-1;*

Sharpe (1963) apresentou um modelo chamado Modelo de Índice Único. Nesse modelo, os retornos dos ativos estão correlacionados com um índice único, que representa todos os retornos do mercado. A fim de analisar se o sentimento provocado pelo Twitter possui impacto no retorno das ações da Petrobras, será utilizada a seguinte equação de modelo de mercado desenvolvida por Sharpe (1963).

$$Ret_t = \propto + \beta R_{m,t} + \varepsilon_t \tag{4}$$

Onde:

$R_{i,t}$: Retorno observado do ativo;

$\propto$: Coeficiente linear da reta de regressão;

β: Beta da ação, ou a covariância entre os retornos da ação e do mercado;

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

ε_t: Termo de erro.

A partir do modelo de mercado desenvolvido por Sharpe (1963), foi construída uma equação para verificar se o retorno da Petrobras está 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 5 foi estimada para analisar se o sentimento textual se relaciona com o retorno do ativo.

$$Ret_t = \alpha + \beta_1 R_{m,t} + \beta_2 \Delta Sent_t + \varepsilon_t \tag{5}$$

Onde:

Ret_t: Retorno da Petrobras;

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

$\Delta Sent_t$: Variação do Sentimento Textual (STO ou SME);

É esperado que o sinal de β_2 seja positivo. A variação do sentimento impacta, de modo direto, o retorno. Se a variação do sentimento for positiva, é gerado um efeito positivo no retorno. No caso de variação negativa do sentimento, o retorno recebe um efeito negativo.

A fim de aprofundar o tema, a variável sentimento foi dividida em sentimento positivo e sentimento negativo. Optou-se, então, por esse caminho para analisar se existe alguma diferença entre o efeito do sentimento separadamente como está presente na Equação 6.

$$Ret_t = \alpha + \beta_1 R_{m,t} + \beta_2 \Delta PST_t + \beta_3 \Delta NST_t + \varepsilon_t \quad (6)$$

Onde:

ΔPST_t: Variação Positiva do Sentimento Textual (STO ou SME);

ΔNST_t: Variação Negativa do Sentimento Textual (STO ou SME);

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 em β_2 e β_3. Com isso, a variação positiva do sentimento aumenta os retornos e a variação negativa do sentimento diminui os retornos.

Por fim, foi investigado se o sentimento amplifica ou diminui os retornos (positivos ou negativos) do mercado. Conforme pode ser visualizado na Equação 7.

$$Ret_t = \alpha + \beta_1 R_{m,t} + \beta_2 \Delta Sent_t + \beta_3 RP_{m,t} \Delta Sent_t + \beta_4 RN_{m,t} \Delta Sent_t + \varepsilon_t \quad (7)$$

Onde:

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

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

O objetivo foi determinar se, quando o mercado está pessimista, o sentimento positivo pode moderar, suavizando o impacto no retorno da empresa, e o negativo, em exacerbar o efeito sobre os retornos firmes. A expectativa oposta é quando o mercado está otimista, analisar se o sentimento positivo amplifica esse efeito ou se o sentimento negativo modera. Portanto, o sinal esperado das variáveis β_3 é positivo e β_4 é negativo. Assim, a variação do sentimento possui algum efeito no mercado.

O modelo de regressão linear para séries temporais com a estimativa por Mínimos Quadrados Ordinários (MQO) foi realizado para todas as regressões do estudo. Executou-se os seguintes testes para identificar possíveis problemas de distribuição residual: teste jarque-bera para normalidade, teste breusch-pagan para heteroscedasticidade, teste de Fator de Inflação da Variância (VIF) para multicolinearidade e teste Ljung-Box para autocorrelação. Quando um dos testes acima rejeitou a hipótese de distribuição residual, foram aplicados erros padrões consistentes de correção automática (HAC).

ANÁLISE DE RESULTADOS

4.1 Análise das Amostras

Para investigar qual seria a amostra mais adequada para o estudo, optou-se por verificar se o impacto da variação do sentimento no retorno ocorreu com postagens coletadas diretamente do perfil da empresa no Twitter (@Petrobras), com alguma citação do termo "petrobras", ou com a citação dos *tickers* da ação (PETR3 e/ou PETR4). O objetivo foi avaliar qual coleta da amostra seria mais fácil e menos custosa, pois poderiam ser avaliados menos *tweets* para obter o mesmo resultado. Por isso, foi realizada a mesma metodologia, porém, com amostras diferentes.

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.

No caso da amostra dos *tweets* que tiveram alguma citação dos *tickers* da ação, não apresentaram sentimento. As palavras apareceram, tanto em dias positivos como negativos, e não apresentaram desvios padrões. Para demonstrar o que ocorreu, parte dos *tweets* foi exposto na Tabela 1.

Tabela 1: Parte da amostra dos *tweets* que continha "PETR3" e/ou "PETR4"

Dias	***Tweets***
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 http://t.co/4sp7US4u
13/10/2015	$PETR3 - Relatório sobre os resultados operacionais e financeiros da Petrobras (PETR3, PETR4, PBR, PBR.A, XPBR http://br.advfn.com/noticias/PAPERBR/2015/artigo/68864874?adw=1126416
23/08/2016	$PETR3 - Bovespa: Índice Ibovespa fecha em alta de +0,41% nesta terça-feira, 23 de agosto de 2016 http://br.advfn.com/noticias/PAPERBR/2016/artigo/72274612?adw=1126416
30/12/2019	$PETR3 - PETROBRAS (PETR-N2) - Regimento Comite Auditoria Estatutario - 18/12/19 http://br.advfn.com/noticias/BOV/2019/artigo/81437435?adw=1126416

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

Na Tabela 1, 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 ressaltar que *retweets* foram retirados das amostras, mas mensagens copiadas não foram critérios 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 às outras: 3.826.463 *tweets*, enquanto as outras foram de 139.530 e 210.881. Também apresentou um conteúdo mais diversificado, com mais opiniões e destaques diários. Não há como comprovar que todos os *tweets* que continham o termo "petrobras" foram escritos por investidores. A relação entre sentimento do investidor e sentimento textual da mídia social é complexa. O último inclui o primeiro. Contudo, é importante ressaltar que as mídias sociais demonstram as condições dentro das empresas e o mercado na visão da sociedade, o que inclui os investidores e influenciam os mesmos.

Por isso, essa foi a amostra escolhida para a análise dos resultados do estudo. A Tabela 2 apresenta dados sobre a amostra que tinham *tweets* coletados a partir da palavra-chave "Petrobras", do ano de 2010 até junho de 2020.

Tabela 2: Quantidade de *tweets*

Ano	Nº de Obs.
2010	147.951
2011	170.844
2012	213.645
2013	216.071
2014	833.109
2015	792.601
2016	350.400
2017	185.125
2018	638.913
2019	207.842
2020*	69.962
2010-2020	3.826.463

Nota: *dados até junho de 2020.

Fonte: Elaboração própria com base nos *Tweets* sobre a Petrobras nos anos de 2010 a 2020.

O ano de 2014 foi um ano de eleição presidencial no Brasil, e aconteceu uma alta das citações da empresa, principalmente nos meses de outubro e novembro. O mesmo ocorreu em 2018. Nas postagens, o nome do Presidente do país da época e dos principais candidatos à presidência, foram muito comentados. Fato que demonstrou, assim, a importância da escolha do presidente para realizar previsões sobre o futuro da estatal.

A data de 26 de outubro de 2014 foi marcada por ser um dia atípico, com total de 4.016 observações. Nesse dia, o tema principal era a possível privatização da Petrobras. Na corrida eleitoral, surgiram mensagens no Twitter que afirmavam que o ex-candidato Serra tinha a pretensão de privatizar a estatal, caso ele vencesse a eleição presidencial. Nesse mesmo dia, aconteceu um debate com candidatos na televisão, e foi o momento com mais citações da Petrobras na rede social durante o ano.

De todo o período, o dia 06 de fevereiro de 2015 registrou mais *tweets*, durante a mudança de presidente da estatal, com a nomeação de Aldemir Bendine. Os anos que tiveram maiores manifestações dos usuários sobre a Petrobras foram 2014 e 2015, respectivamente. No ano de 2014, a Petrobras foi citada pela primeira vez na operação Lava Jato da Polícia Federal. No ano seguinte, os escândalos de corrupção envolvendo o nome da empresa (OTOBONI; FREIRE, 2020) continuaram. Possivelmente, essas foram as causas do aumento nesses anos. Os usuários do Twitter comentaram o fato, como pode ser visto nos exemplos de *tweets* expostos no anexo A. Nos anos seguintes desse período, o número de postagens diminuiu.

4.2 Construção do sentimento

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

No anexo B, são apresentadas as frequências médias e desvios padrões da frequência de cada termo, tanto em dias positivos quanto em negativos. Nomes de políticos foram retirados, pois são autoridades que têm mandato temporário e não agregaria na lista de palavras. Com isso, palavras com diferença média e de desvio padrão, entre dias de retorno positivo e negativo, inferiores a 20%, foram excluídas.

Depois da filtragem, a frequência dos termos em dias positivos e negativos foi incluída no mapeamento perceptual. O mapa perceptivo demonstrou o posicionamento relativo de todas as palavras. Os termos do estudo e sua polaridade estão presentes no anexo C. O objetivo foi analisar como as palavras se comportam próximas do retorno positivo e do retorno negativo.

Com a classificação da polaridade dos termos concluída, foi realizada a construção das variáveis do sentimento, total e médio. Vale ressaltar que nessas foram avaliadas as variações e não os valores absolutos diários. Os gráficos da Figura 3 apresentam as oscilações que ocorreram nas variáveis em análise, de 2010 até 2020.

Figura 3: Oscilações das variáveis ao longo do tempo

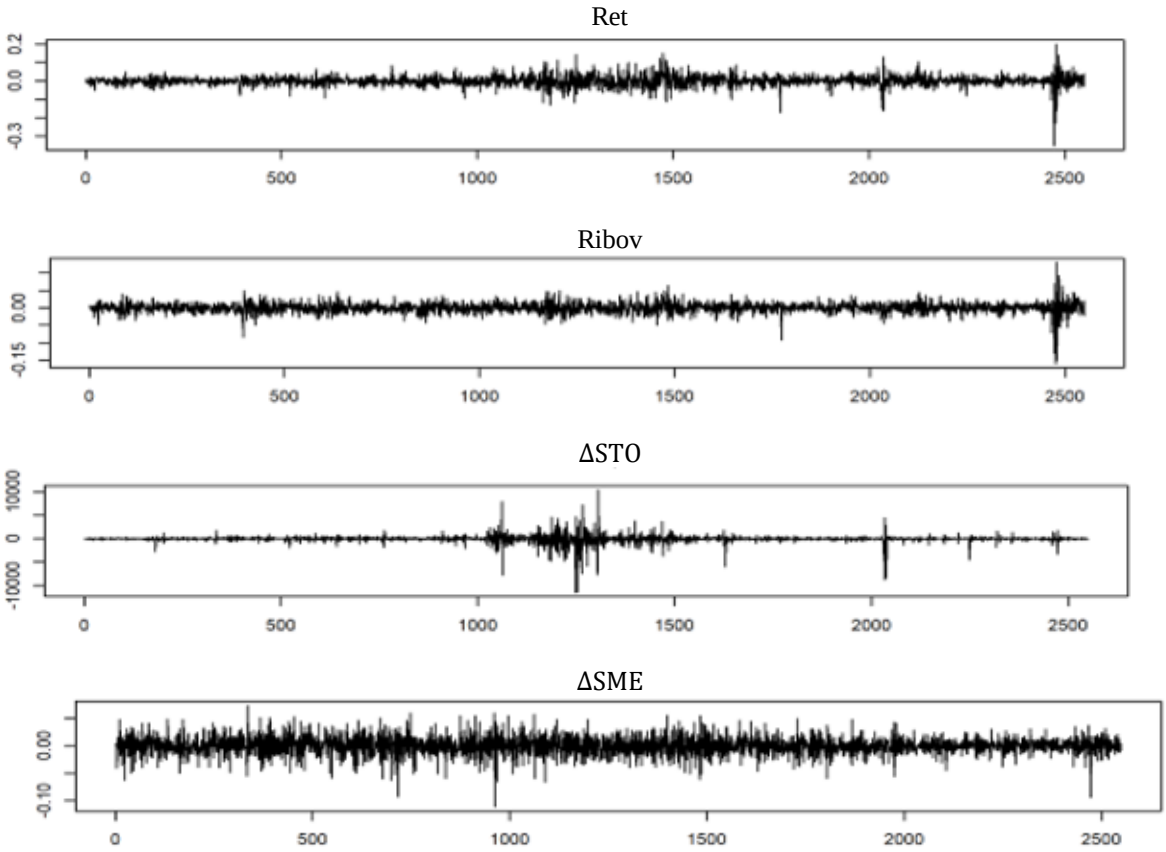


Figura 3: Oscilações das variáveis ao longo do tempo
Fonte: Elaborado pelos autores a partir de dados da pesquisa.

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
ΔSTO	-55.401	891.69	-11491	10366
ΔSME	-0.00033	0.01774	-0.110922	0.07260
Ret	-0.00021	0.03022	-0.35236	0.20067
Ribov	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 Petrobras 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.

4.3 Análise Das Regressões

A Tabela 4 apresenta os resultados do teste Kwiatkowski-Phillips-Schmidt-Shin (KPSS) que verificou a estacionariedade da série.

Tabela 4: Teste de estacionariedade das séries

Teste KPSS	**Ret**	**Ribov**	ΔSTO	ΔSME
P-value	0.1	0.1	0.1	0.1

Fonte: Elaborado pelos autores a partir de dados da pesquisa. *p < 0.10; ** 0.01 < p < 0.05; *** p < 0.01.

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. Os resultados obtidos já eram esperados, visto que não foi analisado se o sentimento impacta o retorno, e sim, se a variação do sentimento impacta o retorno. Foram realizados testes de autocorrelação e autocorrelação parcial, e nenhuma das variáveis apresentaram tendência ou sazonalidade.

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 5.

Tabela 5: Correlação Das Variáveis Independentes Com A Dependente

Ret	**Ribov**	ΔSTO	ΔSME
Teste de Pearson	0.78	0.21	0.14
	(0)***	(0)***	(0)***

Fonte: Elaborado pelos autores a partir de dados da pesquisa. Correlação: Pearson. *p < 0.10; ** 0.01 < p < 0.05; *** p < 0.01

O retorno do mercado apresentou alta correlação com o retorno da Petrobras (78%). Todos os dois indicadores de variação de sentimento tiveram correlação e significância com o retorno da estatal (21% e 14%). O resultado demonstra que existe uma associação 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 variação de sentimento, foi realizado o cálculo de duas variáveis com uma equação distinta da outra. Com isso, foi possível a comparação dos modelos e a verificação se o modelo possuía robustez. Os resultados do primeiro indicador, sentimento total, estão demonstrados na Tabela 6.

Tabela 6: Impacto da variação do sentimento *total* no retorno da Petrobras (2,550 Observações)

Regressores	**Sinal**	**Equação 5**	**Equação 6**	**Equação 7**

$Ribov$	+	1.4466e+00 *** (< 2.2e-16)	1.4435e+00*** (< 2.2e-16)	1.3961e+00*** (< 2.2e-16)
$\Delta Sent$	+	4.7366e-06 *** (1.642e-06)	-	3.3927e-06* (0.01156)
$\Delta\ Pos\ Sent$	+	-	3.4063e-07 (0.71833)	-
$\Delta\ Neg\ Sent$	+	-	7.1590e-06*** (3.554e-08)	-
$RPibov * \Delta\ Sent$	+	-	-	-1.6263e-04 (0.4407)
$RNibov * \Delta\ Sent$	-	-	-	-2.9479e-04*** (3.78e-06)
α	na	-1.2957e-04 (0.7122)	1.1165e-03 (0.00592)	1.5154e-04 (0.66403)
Adjusted R^2		0.6334	0.6407	0.6507
Normalidade p-value		< 2.2e-16	< 2.2e-16	< 2.2e-16
Heteroscedasticidade p-value		< 2.2e-16	< 2.2e-16	5.531e-12
Multicolinearidade VIF		Ribov:1.007288 Sent: 1.007288	Ribov: 1.07650 $\Delta\ Pos\ Sent$:1.011903 $\Delta\ Neg\ Sent$:1.017039	Ribov:1.074142 Sent:2.166817 $RPibov *$ $\Delta\ Sent$:1.604941 $RNibov *$ $\Delta\ Sent$:1.659181
Autocorrelação p-value		0.9429	0.7624	0.6099

Nota: Normalidade: teste Jarque-Bera; Heteroscedasticidade: teste Breusch-Pagan. Teste Ljung-Box para autocorrelação (10 lags). Método de estimação: MQO com Autocorrelation-Consistent Standard Errors (HAC). *p < 0.10; ** 0.01 < p < 0.05; *** p < 0.01.

Os resultados demonstraram estar de acordo com a teoria. O retorno do mercado foi significativo e positivo, da mesma forma que foi explicitada no modelo de Sharpe (1963). O indicador de variação de sentimento textual teve significância e foi positiva 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 relaciona com o retorno.

Na segunda regressão, disposta na Tabela 6, o objetivo foi analisar a segunda hipótese de pesquisa. Buscou-se verificar se (Hipótese 2) a variação positiva e a variação negativa dos sentimentos, separadamente, possuem impacto no retorno, e qual seria a direção desses relacionamentos. Nesse caso, somente a variação negativa dos sentimentos possuiu impacto negativo no retorno. Esse resultado vai ao encontro à 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 negativa do sentimento textual se relacionou com o retorno.

Na linha dos resultados da segunda regressão, a variação do sentimento textual total na Equação 7 teve efeito somente com o mercado negativo. O retorno de mercado foi significativo e positivo, da mesma forma que ocorreu nas equações passadas. A variação do sentimento obteve significância e com efeito negativo em relação ao retorno. Esse achado demonstra que no cenário de o mercado estar em queda, 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 suavizado. 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: a variação do sentimento textual impactou o movimento de mercado em relação ao retorno, porém, somente quando o mercado está negativo.

Os resultados também mostram que a proporção da variância no dependente explicada pelas variáveis independentes, representadas por R^2, aumenta da Equação 5, onde a variação do sentimento total é o regressor observado, para a Equação 6, onde o regressor acima é dividido em variações positivas e negativas do sentimento, e finalmente para a Equação 7, onde a mesma variável interage com dias de mercado favoráveis e desfavoráveis. Esses resultados representam uma confirmação adicional dos comportamentos assimétricos dos investidores nos cenários de ganhos ou perdas, com este último gerando mais impacto do que o primeiro, porque as pessoas são avessas ao risco. É um comportamento natural defender seus bens (JIANG et al., 2019).

Com o objetivo de realizar o teste de robustez do indicador variação do sentimento, os mesmos modelos foram calculados para a variável variação do sentimento médio. Os resultados encontram-se na Tabela 7.

Tabela 7: Impacto da variação do sentimento *médio* no retorno da Petrobras (2,550 Observações)

Regressores	Sinal	Equação 5	Equação 6	Equação 7
$Ribov$	+	1.457e+00 *** (< 2e-16)	1.453e+00 *** (< 2e-16)	1.432e+00*** (< 2.e-16)
$\Delta\ Sent$	+	1.579e-16 * (0.04854)	-	4.759e-17 (0.2760)
$\Delta\ Pos\ Sent$	+	-	-3.436e-17 (0.31805)	-
$\Delta\ Neg\ Sent$	+	-	2.036e-16 * (0.02601)	-

$RPibov * \Delta\ Sent$	+	-	-	-8.054e-15 (0.2569)
$RNibov * \Delta\ Sent$	-	-	-	-4.911e-15*** (<2e-16)
α	na	-2.823e-04 (0.43293)	-1.843e-04 (0.61690)	-2.575e-04 (0.4719)
Adjusted R^2		0.6185	0.6199	0.6268
Normalidade p-value		< 2.2e-16	< 2.2e-16	< 2.2e-16
Heteroscedasticidade p-value		4.12e-10	1.251e-10	0.8977
Multicolinearidade VIF		Ribov:1.00953; Sent: 1.0095	Ribov: 1.011323 $\Delta\ Pos\ Sent$:1.0001 $\Delta\ Neg\ Sent$:1.011316	Ribov:1.029258 Sent:1.738738 $RPibov * \Delta\ Sent$:1.297859 $RNibov * \Delta\ Sent$:1.465194
Autocorrelação p-value		0.928	0.9002	0.6641

Normalidade: teste Jarque-Bera; Heteroscedasticidade: teste Breusch-Pagan. Teste Ljung-Box para autocorrelação (10 lags). Método de estimação: MQO com Autocorrelation-Consistent Standard Errors (HAC). *p < 0.10; ** 0.01 < p < 0.05; *** p < 0.01.

O teste de robustez com a variável de variação do sentimento médio apresentou, em geral, os mesmos resultados que o indicador de variação do sentimento total. A Tabela 7 demonstra resultados com as mesmas significâncias do que a Tabela 6, no geral. A diferença está na Equação 7 das duas tabelas: a variação do sentimento só foi significativa com a construção do sentimento total. Com isso, conclui-se que a construção do dicionário personalizado apresentou significância, independente da forma que a variável sentimento textual foi gerada.

Com esses achados, foi possível concluir que criar um dicionário personalizado explicou ganhos ou perdas no valor da empresa. A contribuição do estudo foi personalizar o dicionário para organizações individuais, enquanto os dicionários existentes são genéricos e, portanto, com menor qualidade. Dessa forma, o indicador de sentimento particular pode mitigar ou melhorar o movimento de mercado dessa empresa. A desconfiança do mercado sobre alguns aspectos pode diminuir o valor das empresas. Com a construção desses mapas, pode surgir uma poderosa ferramenta para gestão de marketing e análise do sentimento sobre a organização.

O estudo se mostrou relevante, pois as palavras são componente importante dos sistemas de linguagem natural. A análise de sentimentos oferece às organizações uma maneira rápida de monitorar os sentimentos do público em relação à marca, aos negócios, diretores, etc. Isso auxiliará não apenas para atrair novos investidores ou clientes, mas também a manter os atuais. Com esse método, foi possível identificar padrões de comportamento nas mídias sociais ligados às oscilações do mercado, contribuindo assim para a gestão de riscos e otimização das tomadas de decisão no mercado de ações. Os resultados apresentados contribuíram para a teoria, com uma nova forma de construir uma lista de palavras customizada para as empresas. Quanto à

implicação prática, o Dicionário Personalizado de Mídia Social pode ser usado para prever o retorno de ativos futuros.

4. CONCLUSÃO

O objetivo deste trabalho foi analisar a relação entre o sentimento exposto no Twitter e o retorno das ações. Para esse fim, os *tweets* sobre a Petrobras foram analisados, a fim de verificar se o sentimento textual dessas mensagens tem relação com o retorno da empresa.

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 vista de forma positiva ou negativa na internet.

Todos os indicadores de variação do sentimento, total e médio, 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).

A variação negativa do sentimento, total e médio, possuía maior impacto nos indivíduos do que a variação positiva do sentimento. As pessoas possuem comportamento avesso ao risco, sentem mais pela perda do que pelo ganho, mesmo que seja na mesma magnitude. Os mesmos achados foram encontrados no trabalho de Silva (2017).

Somente a variação do sentimento negativo, total e médio, mostrou resultados negativos e significativos no movimento de mercado. Desse modo, a variação do sentimento negativo potencializa o movimento negativo do mercado e ameniza o movimento positivo do mercado. Esse resultado corrobora com 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 no Twitter e os *tweets* que continha algum dos *tickers* das ações (PETR3 e 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 apresentou mais sentimento, diversidade de palavras e quantidade de *tweets* foi a primeira amostra, a que continha apenas a palavra “petrobras” como

filtro de seleção. Essa amostra possui uma lista de termos que apresentam desvios relacionados aos retornos diários. Por isso, o presente estudo optou por aprofundar-se na lista de termos formada por essa amostra.

A pesquisa apresentou as limitações de ter coletado somente os dados da mídia social Twitter e da escolha da organização, por não ser um não ser um nome homônimo, o que auxiliou na obtenção das informações.

Esses resultados são relevantes para os gestores e agentes de políticas públicas, uma vez que é importante entender a opinião pública sobre a empresa Petrobras, para realizar uma tomada de decisão, por exemplo. Assim, o estudo presente 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, analisar como o movimento da mídia social Twitter e/ou outras se relacionam com o movimento de mercado e a inclusão de uma ponderação dos termos.

REFERÊNCIAS

AB. RAHMAN, A. S.; ABDUL-RAHMAN, S.; MUTALIB, S. Mining textual terms for stock market prediction analysis using financial news. **Communications in Computer and Information Science**, v. 788. 2017

AHMADI, Z.; MARTENS, P.; KOCH, C.; GOTTRON, T.; KRAMER, S. **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, DSAA 2018. **Anais**...2019

ALKUBAISI, G. A. A. J.; KAMARUDDIN, S. S.; HUSNI, H. Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naive Bayes Classifiers. **Computer and Information Science**, v. 11, n. 1, 2018.

AVANÇO, L. V.; NUNES, M. D. G. V. Lexicon-based sentiment analysis for reviews of products in Brazilian Portuguese. **Proceedings - 2014 Brazilian Conference on Intelligent Systems, BRACIS 2014**, p. 277–281, 2014.

BAKER, M.; WURGLER, J. Investor sentiment and the cross-section of stock returns. **Journal of Finance**, v. 61, n. 4, p. 1645–1680, 2006.

BANNIER, C.; PAULS, T.; WALTER, A. Content analysis of business communication: introducing a German dictionary. **Journal of Business Economics**, v. 89, n. 1, p. 79–123, 8 fev. 2019.

BARBER, B. M.; ODEAN, T. Boys will be Boys: Gender, Overconfidence, and Common Stock Investment. **The Quartely Journal of Economics**, v. 116, 2001.

BARBERIS, N.; SHLEIFER, A.; VISHNY, R. A model of investor sentiment. **Journal of Financial Economics**, v. 49, n. 3, p. 307–343, 1998.

BARTOV, E.; FAUREL, L.; MOHANRAM, P. S. Can twitter help predict firm-level earnings and stock returns? **The Accounting Review**, v. 93, n. 3, p. 25–57, 2018.

BAUER, M. W.; GASKELL, G. **Pesquisa qualitativa com texto, imagem e som: um manual prático**. 2º ed. Petrópolis: Vozes, 2003.

BEATTIE, V. Accounting narratives and the narrative turn in accounting research: Issues, theory, methodology, methods and a research framework. **British Accounting Review**, v. 46, n. 2, p. 111–134, 2014.

BERMINGHAM, A.; SMEATON, A. Classifying sentiment in microblogs: Is brevity an advantage? **International Conference on Information and Knowledge Management, Proceedings**, 2010.

BIAŁKOWSKI, J.; BOHL, M. T.; KAUFMANN, P.; WISNIEWSKI, T. P. Do mutual fund managers exploit the Ramadan anomaly? Evidence from Turkey. **Emerging Markets Review**, v. 15, p. 211–232, 2013.

BOLLEN, J.; MAO, H. Twitter mood as a stock market predictor. **Computer**, 2011.

BOLLEN, J.; MAO, H.; ZENG, X. Twitter mood predicts the stock market. **Journal of Computational Science**, v. 2, n. 1, p. 1–8, 2011.

BUKOVINA, J. Social media big data and capital markets-An overview. **Journal of Behavioral and Experimental Finance**, v. 11, p. 18–26, 2016.

Buzz Rankings de 2019: Brasil | YouGov - BrandIndex. Disponível em: <https://www.brandindex.com/ranking/brazil/2019-buzz>. Acesso em: 13 jun. 2020.

CAROSIA, A. E. O.; COELHO, G. P.; SILVA, A. E. A. Analyzing the Brazilian Financial Market through Portuguese Sentiment Analysis in Social Media. **Applied Artificial Intelligence**, v. 34, n. 1, p. 1–19, 2019.

CHEN, M. K. The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets. **American Economic Review**, v. 103, n. 2, p. 690–731, 2013.

CLEMENT, J. **Countries with most Twitter users 2020**. Disponível em:

<https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>. Acesso em: 15 mar. 2020.

DAŘENA, F.; PETROVSKÝ, J.; PŘICHYSTAL, J.; ŽIŽKA, J. Machine learning-based analysis of the association between online texts and stock price movements. **Inteligencia Artificial**, v. 21, n. 61, p. 95–110, 2018.

DE LONG, J. B. ; SHLEIFER, A. ; SUMMERS, L. H. ; WALDMANN, R. J. Noise Trader Risk in Financial Markets. **Journal of Political Economy**, v. 98, n. 4, p. 703–738, 1990.

DOMINGUES, R. **Previsibilidade de retornos diários no mercado de ações a partir de indicadores de sentimento do investidor**. Dissertação (Mestrado Profissional em Economia), Ibmec São Paulo, 2008.

DOW, S. C. Cognition, market sentiment and financial instability. **Cambridge Journal of Economics**, v. 35, n. 2, p. 233–249, 2011.

FAMA, E. F. Random Walks in Stock Market Prices. **Financial Analysts Journal**, v. 51, n. 1, p. 75–80, 1995.

FANG, B. ; LIANG, S. ; ZOU, Q. ; HUANG, W. **Research on sentiment analysis of financial texts**. Proceedings of 2017 IEEE 7th International Conference on Electronics Information and Emergency Communication, ICEIEC 2017. **Anais**...2017

FEUERRIEGEL, S.; WOLFF, G.; NEUMANN, D. **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. **Anais**...2015

GANDHI, P.; LOUGHRAN, T.; MCDONALD, B. Using Annual Report Sentiment as a Proxy for Financial Distress in U.S. Banks. **Journal of Behavioral Finance**, v. 20, n. 4, p. 424–436, 2 out. 2019.

GARCIA, D. Sentiment During Recessions. **SSRN Electronic Journal**, n. 2010, 2012.

GE, Q.; KUROV, A.; WOLFE, M. H. Stock Market Reactions to Presidential Statements: Evidence from Company-Specific Tweets. **Economics Faculty Scholarship**, p. 29, 2018.

GENESOVE, D.; MAYER, C. Loss Aversion and Seller Behavior: Evidence from the Housing Market. **The Quarterly Journal of Economics,** Oxford University Press, v. 116, n. 4, 2001.

GO, A.; BHAYANI, R.; HUANG, L. Twitter Sentiment Classification using Distant Supervision. **Processing**, v. 150, p. 1–6, 2009.

GODSAY, M. The Process of Sentiment Analysis: A Study. **International Journal of Computer Applications**, v. 126, n. 7, p. 26–30, 2015.

GOWER, J.C; LEGENDRE, P. Metric and Euclidean properties of dissimilarity coefficients. **Journal of Classification**, v. 3, p. 5-18, 1986.

GRAHAM, J. R.; HARVEY, C. R.; HUANG, H. Investor competence, trading frequency, and home bias. **Management Science**, v. 55, n. 7, p. 1094–1106, 2009.

GUPTA, A.; SIMAAN, M.; ZAKI, M. J. **Investigating bank failures using text mining**. 2016 IEEE Symposium Series on Computational Intelligence, SSCI 2016. **Anais**...2017

HÁJEK, P. Combining bag-of-words and sentiment features of annual reports to predict abnormal stock returns. **Neural Computing and Applications**, v. 29, n. 7, p. 343–358, 1 abr. 2018.

HALFELD, M.; TORRES, F. DE F. L. Finanças Comportamentais: aplicações no contexto brasileiro. **RAE - Revista de Administração de Empresas**, p. 64–71, 2001.

HIRSHLEIFER, D. A.; SHUMWAY, T. Good Day Sunshine: Stock Returns and the Weather. **The Journal of Finance**, 2005.

HOULIHAN, P.; CREAMER, G. G. Can Sentiment Analysis and Options Volume Anticipate Future Returns? **Computational Economics**, v. 50, n. 4, p. 669–685, 2017.

HUANG, S.; HUANG, Y.; LIN, T. C. Attention allocation and return co-movement: Evidence from repeated natural experiments. **Journal of Financial Economics**, v. 132, n. 2, p. 369–383, 2019.

ITO, T.; SAKAJI, H.; IZUMI, K.; TSUBOUCHI, K.; YAMASHITA, T. **Development of an interpretable neural network model for creation of polarity concept dictionaries**. IEEE International Conference on Data Mining Workshops, ICDMW. **Anais**...2017

ITO, T.; TSUBOUCHI, K.; SAKAJI, H.; YAMASHITA, T.; IZUMI, K. **Word-level Sentiment Visualizer for Financial Documents**. CIFEr 2019 - IEEE Conference on Computational Intelligence for Financial Engineering and Economics, p. 1-7, **Anais**...2019

IYENGAR, S. S.; LEPPER, M. R. When choice is demotivating: Can one desire too much of a good thing? **Journal of Personality and Social Psychology**, v. 79, n. 6, p. 995–1006, 2000.

JEGADEESH, N.; TITMAN, S. Returns to Buying Winners and Selling Losers : Implications for Stock Market Efficiency. **The Journal of Finance**, v. 48, n. 1, p. 65–91, 1993.

JIANG, F.; LEE, J.; MARTIN, X.; ZHOU, G.. Manager Sentiment and Stock Returns. **Journal of Financial Economics**, v. 132, n. 1, pp. 126-149, 2019.

JIN, L.; KOTHARI, S. P. Effect of personal taxes on managers' decisions to sell their stock. **Journal of Accounting and Economics**, v. 46, n. 1, p. 23–46, 2008.

KAHNEMAN, D.; TVERSKY, A. Prospect Theory: An Analysis of Decision under Risk. **Jornal of the Econometrica society**, v. 47, n. 2, p. 263–292, 1979.

KAPLANSKI, G.; LEVY, H. The holiday and Yom Kippur War sentiment effects: The Tel Aviv Stock Exchange (TASE). **Quantitative Finance**, v. 12, n. 8, p. 1283–1298, 2012.

KAWAKATSU, H.; MOREY, M. R. An Empirical Examination of Financial Liberalization and the Efficiency of Emerging Market Stock Prices. **The Journal of Financial Research**, v. 22, n. 4, p. 385–411, 1999.

KEARNEY, C.; LIU, S. Textual sentiment in finance: A survey of methods and models. **International Review of Financial Analysis**, v. 33, p. 171–185, 2014.

KO, K. J.; HUANG, Z. J. Arrogance can be a virtue: Overconfidence, information acquisition, and market efficiency. **Journal of Financial Economics**, v. 84, n. 2, p. 529–560, 2007.

LAKONISHOK, J.; SHLEIFER, A.; VISHNY, R. W. The Impact of Institutional Trading on Stock Prices. **Journal of Finance Economics**, p. 23–43, 1992.

LI, G.; LI, J.; LIU, M.; ZHU, X. Evolutionary Mechanism of Risk Factor Disclosure in American Financial Corporation Annual Report. **Communications in Computer and Information Science**, v. 1179, 2020.

LI, Q.; WANG, J.; WANG, F.; LI, P.; LIU, L.; CHEN, Y. The role of social sentiment in stock markets: a view from joint effects of multiple information sources. **Multimedia Tools and Applications**, v. 76, n. 10, p. 12315–12345, 2017.

LI, X.; WU, P.; WANG, W. Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. **Information Processing and Management**, v. 57, n. 5, 2020.

LOUGHRAN, T.; MCDONALD, B. Internet appendix for "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks". **The Journal of Finance**, v. 66, n. 1, p. 35–65, 2011.

LOUGHRAN, T.; MCDONALD, B. The Use of Word Lists in Textual Analysis. **Journal of Behavioral Finance**, v. 16, n. 1, p. 1–11, 2015.

MAI, F.; SHAN, Z.; BAI, Q.; WANG, X. S.; CHIANG, R. H. L.. How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis. **Journal of Management Information Systems**, v. 35, n. 1, p. 19–52, 2018.

MANGEE, N. Stock Returns and the Tone of Marketplace Information: Does Context Matter? **Journal of Behavioral Finance**, v. 19, n. 4, p. 396–406, 2 out. 2018.

MAO, H.; WANG, K.; MA, R.; GAO, Y.; LI, Y.; CHEN, K.; XIE, D.; ZHU, W.; WANG, T.; WANG, H. **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. **Anais**...2014

MAO, H.; COUNTS, S.; BOLLEN, J. Predicting Financial Markets: Comparing Survey, News, Twitter and Search Engine Data. **Economics, Computer Science, Physics - ArXiv**, p. 1–10, 2011.

Marcos Históricos da Legislação do Petróleo. 2020. Disponível em: <https://www.presalpetroleo.gov.br/ppsa/a-pre-sal-petroleo/marcos-historicos>. Acesso em: 27 ago. 2020.

MARTINS, R.; PEREIRA, A.; BENEVENUTO, F. An approach to sentiment analysis of web applications in Portuguese. **WebMedia 2015 - Proceedings of the 21st Brazilian Symposium on Multimedia and the Web**, p. 105–112, 2015.

MARTY, T.; VANSTONE, B.; HAHN, T. News media analytics in finance: a survey. **Accounting & Finance**, v. 60, p. 1385-1434. 2020

MCGURK, Z.; NOWAK, A.; HALL, J. C. Stock returns and investor sentiment: textual analysis and social media. **Journal of Economics and Finance**, v. 44, n. 3, p. 458–485, 1 jul. 2020.

NARDO, M.; PETRACCO-GIUDICI, M.; NALTSIDIS, M. Walking down wall street with a tablet: A survey of stock market predictions using the web. **Journal of Economic Surveys**, p. 1–14, 2015.

NAYLOR, R. W.; LAMBERTON, C. P.; WEST, P. M. Beyond the "like" button: The impact of mere virtual presence on brand evaluations and purchase intentions in social media settings. **Journal of Marketing**, v. 76, n. 6, p. 105–120, 2012.

NOFSINGER, J. **The Psychology of Investing**. 4. ed. New Jersey: Prentice Hall, 2010.

OLIVEIRA, N.; CORTEZ, P.; AREAL, N. Some experiments on modeling stock market behavior using investor sentiment analysis and posting volume from twitter. **ACM International Conference Proceeding Series**, 2013.

OTOBONI, J.; FREIRE, D. **Seis anos da Lava Jato: relembre todas as fases da operação**. 2020. Disponível em: <https://www.cnnbrasil.com.br/politica/2020/03/07/seis-anos-da-lava-jato-relembre-todas-as-fases-da-operacao>. Acesso em 11 jan. 2021.

PENGNATE, S. F.; RIGGINS, F. J. The role of emotion in P2P microfinance funding: A sentiment analysis approach. **International Journal of Information Management**, v. 54, 2020.

PINTO, A. E. DE S. **Pelo 4º ano seguido, Petrobras é a Marca que Representa o Brasil**. Disponível em: <https://top-of-mind.folha.uol.com.br/2019/10/pelo-4o-ano-seguido-petrobras-e-a-marca-que-representa-o-brasil.shtml>. Acesso em: 11 out. 2020.

RANCO, G.; ALEKSOVSKI, D; CALDARELLI, G.; GRČAR, M.; MOZETIČ, I. The effects of twitter sentiment on stock price returns. **PLoS ONE**, v. 10, n. 9, p. 1–21, 2015.

RAO, T.; SRIVASTAVA, S. Analyzing Stock Market Movements Using Twitter Sentiment Analysis. **Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining**, p. 119–123, 2012.

SANTOS, H. S.; LAENDER, A.; PEREIRA, A. C. Uma Visão do Mercado Brasileiro de Ações a partir de Dados do Twitter. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING (BRASNAM), 4. , 2015, Recife. **Anais** [...]. Porto Alegre: Sociedade Brasileira de Computação, 2015 . p. . ISSN 2595-6094. v. 10, n. 3, p. 369–394, 2015.

SARIGIANNIDIS, A.; KARYPIDIS, P. A.; SARIGIANNIDIS, P.; PRAGIDIS, I. C. A Novel Lexicon-Based Approach in Determining Sentiment in Financial Data Using Learning Automata. **Lecture Notes in Computer Science**, v. 10750, 2018.

SENG, J.-L.; YANG, H.-F. The association between stock price volatility and financial news – a sentiment analysis approach. **Kybernetes**, v. 46, n. 8, p. 1341–1365, 2017.

SHARPE, W. F. A Simplified Model For Portfolio Analysis. **Magazine Science**, v. 9, n. 2, p. 277–293, 1963.

SILVA, M. D. DE O. P. DA. **O Efeito do Sentimento das Notícias sobre o Comportamento**

dos Preços no Mercado Acionário Brasileiro. Universidade de Brasília, 2017.

SIMON, H. A. **Administrative Behavior**. 2º ed. New York: The Free Press, 1947.

SOUZA, D. M. S. DE; LUCENA, W. G. L.; QUEIROZ, D. B. DE. O Efeito do Sentimento do Investidor Expresso via Twitter sobre o Comportamento do Mercado Acionário Brasileiro Durante o Período Eleitoral. **XIX Congresso USP**, p. 1–19, 2019.

SPRENGER, T. O. et al. News or noise? Using twitter to identify and understand company-specific news flow. **Journal of Business Finance and Accounting**, v. 41, n. 7–8, p. 791–830, 2014.

TARDE, G. **Psychologie Économique**. Paris: Bibliothèque de philosophie contemporaine, 1902.

TETLOCK, P. C.; SAAR-TSECHANSKY, M.; MACSKASSY, S. More Than Words: Quantifying Language to Measure Firms' Fundamentals. **Journal of Finance**, v. 63, n. 3, p. 1437–1467, 2008.

TOMMASETTI, R.; MAIA, V. M.; MACEDO, M. Á. DA S. Twitter as 'bully pulpit': Brazilian banking sector case. **Applied Economics Letters**, v. 28, n. 01, p. 61–64, 2021.

TVERSKY, A.; KAHNEMAN, D. Judgment under uncertainty: Heuristics and biases. **Science**, v. 185, n. 4157, p. 1124–1131, 27 set. 1974.

UHR, P.; ZENKERT, J.; FATHI, M. **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. **Anais**...2014

VIANEZ, J. P.; MARTÍNEZ, R. G.; ROMÁN, C. P. A bibliometric analysis of behavioural finance with mapping analysis tools. **European Research on Management and Business Economics**, p. 1–7, 2020.

WANG, Y. Stock market forecasting with financial micro-blog based on sentiment and time series analysis. **Journal of Shanghai Jiaotong University (Science)**, v. 22, n. 2, p. 173–179, 2017.

YADAV, A.; JHA, C. K.; SHARAN, A.; VAISH, V. **Sentiment analysis of financial news using unsupervised approach**. Procedia Computer Science. **Anais**...2020

YOSHINAGA, C. E. **A Relação Entre Índice de Sentimento de Mercado e as Taxas de Retorno das Ações: Uma Análise com Dados em Painel**. Tese (Doutorado em Administração) - Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, 2009.

ANEXO A – EXEMPLOS DE TWEETS PRESENTES EM PERÍODOS DE PICOS DE NÚMERO DE OBSERVAÇÕES

Data	*Likes*	*Retweets*	Texto
13/03/2014	170	46	"E o poço não tem fim ...Escândalo Bilionário na Petrobras é Maior que o Mensalão - http://migre.me/iixMM"
11/04/2014	1091	9	"A #DilmaTemMedoDaCPI da Petrobras por saber do que participou lá dentro com Palocci, Dirceu.. http://migre.me/iHvdH"
05/09/2014	17	10	"As Denúncias do envolvimento do Presidente do Congresso e de Senadores em esquemas de saque da Petrobras são graves. http://m.estadao.com.br/noticias/politica,renan-calheiros-e-citado-em-depoimentos-de-delator-da-operacao-lava-jato,1555616,0.htm"
21/09/2014	10	4	"Caso Petrobras faz Mensalão parecer coisa miúda. Que tipo de brasileiro vota em Dilma Rousseff e seu bando? http://ln.is/glo.bo/cbfG8"
15/12/2014	137	49	"Acabou de sair no New York Times, o caso Petrobras é o maior caso de corrupção em um país democrático na história do mundo moderno."
29/01/2015	123	57	"Em 2 dias, o valor da Petrobras caiu quase 17 bilhões.Dilma mandou na empresa durante os 12 anos do PT s/saber o q acontecia ali?"
02/06/2015	17	13	"Petrobras morre de rir do Moro e da Urubóloga | Conversa Afiada http://t.co/ZGoEhxkh4L http://t.co/W4WNLBFPA4"
02/10/2015	23	19	"Defender a Petrobras é defender o Brasil http://t.co/c1FjvrO7SQ #dia3Brasilprafrente"
13/11/2015	89	44	"NA SUÍÇA: Dono da Itaipava repassou dinheiro desviada da Petrobras para a #BanDilma"
30/11/2015	11	5	"Ex-gerente da Petrobras diz que pagou R$ 14 mi de propina ao Senador pernambucano Humberto 'Sanguessuga' Costa https://t.co/AVvAqxL6Za"

Fonte: Twitter (2010).

ANEXO B –FREQUÊNCIA, MÉDIA E DESVIO PADRÃO DAS PALAVRAS MAIS CITADAS NOS TWEETS

Palavras	Frequência pos/neg	Média pos/neg	Desvio padrão pos/neg
blog	0,995617	0,984609	0,897922
petróleo	1,000883	0,996249	0,956964
valor	1,010427	0,79955	0,6348
dia	1,00081	0,889943	0,691502
milhões	1,005931	0,977147	0,934471
produção	0,993138	1,038823	0,995072
recorde	1,037519	1,210028	1,280291
economia	0,999831	0,918978	0,737733
eua	1,006425	0,952294	0,922616
negócios	0,992387	0,84566	0,713622
refinaria	1,018241	1,061051	1,148611
posto	0,995975	0,870321	0,752988
bilhões	0,997871	0,886442	0,714705
maior	1,004227	1,002604	0,878476
fecha	0,985436	0,806124	0,732648
anos	1,012797	0,9005	0,810315
quer	1,003401	0,915805	0,741092
brasil	1,00081	0,926489	0,698729
ano	1,000036	0,94125	0,887071
hoje	0,992443	0,893276	0,802525
distribuidora	0,96033	0,849587	0,482647
programa	0,996247	1,069026	1,493535
campanha	0,988516	0,896726	1,359739
gasolina	1,016386	0,906037	1,064625
empresas	0,994756	0,929267	0,752314
rj	0,998228	0,980115	0,940467
desde	1,005964	0,707524	0,567426
bem	1,018613	1,161365	1,055116
deve	0,99545	0,947642	0,837046
novo	0,991408	0,786396	0,373054
óleo	1,000071	1,037443	1,055625
agora	1,000542	0,901737	1,035657
não	1,00081	0,838828	0,688936
uol	1,000764	0,970878	0,896518
concurso	0,954853	0,823849	0,768485
plataforma	0,995585	1,442885	4,369253
contra	1,001102	0,921214	0,848444
todo	0,987361	0,809203	0,538999
mundo	1,004564	0,855534	0,710085
parte	1,032697	1,021009	1,306433

caixa	0,995262	0,988348	1,038219
globo	1,003752	1,022502	0,987218
presidente	1,005977	0,864894	0,564046
veja	0,999802	0,945368	1,06011
povo	1,012623	0,792385	0,491295
tudo	0,994982	0,901726	0,959464
bi	0,999227	0,953827	0,878305
ação	1,00081	0,93258	0,803843
quem	1,003317	0,885263	0,869676
nacional	1,010331	0,941414	0,878067
pode	0,997815	0,941943	0,933936
combustível	0,980474	0,773316	0,67462
mil	0,998522	0,949028	0,876297
us	1,00081	0,934774	0,855943
anuncia	0,99497	0,994639	0,98708
présal	1,093835	0,812354	0,438377
menos	0,968431	0,76226	0,876258
diz	0,997073	0,9818	0,985244
jornal	1,006144	0,943171	0,882701
empresa	1,000846	0,869145	0,69009
privatizar	1,03013	0,974379	1,266545
janeiro	1,005303	0,938882	0,736895
rio	1,00081	0,944297	0,871237
tv	0,989616	0,933164	0,926628
vídeo	1,004649	1,099286	1,39964
youtube	1,003496	1,033676	1,243204
bovespa	0,994658	0,865823	0,756758
brasileiro	1,016777	0,882883	0,716578
mercado	1,005646	0,84816	0,629194
ativos	1,003941	0,995503	1,209989
aumento	1,006859	0,831713	0,86758
gás	1,010657	1,047257	0,913225
preço	0,987862	0,855235	0,764355
preços	0,991949	0,750319	0,734782
vagas	0,950635	0,805026	0,744714
vai	1,000071	0,860505	0,732128
bom	0,993963	0,842217	0,648951
reuters	1,016472	0,90766	0,822475
conta	0,997267	0,871124	0,808626
política	1,009537	0,75255	0,528763
nesta	1,016771	0,878018	0,832337
cai	0,999011	0,668718	0,657585
estadão	0,996191	0,841632	0,752474
bolsa	1,012147	0,803702	0,60905
fim	0,998664	1,124795	1,496895

folha	0,998937	0,952908	0,976682
governo	0,993775	0,844379	0,733735
país	1,006271	0,788774	0,564768
afirma	0,99241	1,003386	1,201789
ações	1,000036	0,834337	0,756638
quase	0,981245	0,878971	0,949776
bacia	0,983359	1,181267	1,119539
caso	1,017556	0,900327	0,847927
diretor	1,013395	0,974897	1,136519
com	1	0,954217	0,823348
fazer	1,015922	0,923042	1,039252
todos	1,001273	0,832358	0,586534
crise	0,961654	0,898747	0,860356
ainda	0,99495	0,868232	0,843535
segundo	1,009928	1,010105	1,097483
nada	0,998402	0,811256	0,745943
estatal	0,993564	0,94877	0,832376
agência	0,998329	0,827081	0,489228
combustíveis	0,980252	0,737974	0,651441
notícias	1,002238	1,009325	1,06685
volta	1,031226	0,951365	1,2514
vazamento	1,025215	1,013401	1,018581
venda	0,999451	1,01128	1,013623
queda	0,908621	0,442476	0,485644
alta	1,011158	1,434223	1,207334
contrato	1,015274	0,975863	0,946857
diretoria	1,023105	0,73349	0,74694
grande	1,005712	0,871714	0,732706
projeto	0,999913	1,111177	2,181021
mi	1,00081	0,924466	0,76299
após	1,000131	0,766313	0,621374
expresidente	0,614518	0,664281	0,739736
contratos	1,025102	0,972784	1,047171
disse	0,993637	1,008243	0,936756
ibovespa	1,007826	0,8952	0,855718
sobe	1,087084	1,907901	1,795526
vale	1,00532	0,853587	0,828862
dinheiro	0,984228	0,896945	0,848655
operação	1,041783	1,073232	0,978749
pede	1,039295	0,650006	0,393088
público	1,011101	0,813203	0,921244
recebeu	0,981752	0,744522	0,713844
dias	1,020542	0,854546	0,702443
semana	0,986838	0,917212	0,903105
processo	0,98512	0,880162	0,598427

compra	1,00167	1,092028	1,386658
confirma	0,961209	1,058106	1,085033
oposição	1,01288	1,056116	1,235577
abre	1,007456	0,776749	0,648072
lucro	0,985789	0,794724	0,86436
trimestre	0,986449	0,809174	0,800652
defende	0,990916	0,786747	0,722821
pt	1,001176	0,938	0,986225
políticos	1,034988	1,250967	2,564744
mpf	0,998276	0,863902	0,871116
funcionários	0,978622	1,150788	1,435448
br	1	0,939882	0,82132
partir	0,991534	1,031103	1,014933
acordo	0,994042	1,060472	1,264441
decisão	0,983624	1,111568	1,227613
ministro	0,984502	0,837553	0,944732
pagar	1,00107	0,843489	0,678278
estadao	0,992141	0,965299	0,95747
dívida	1,017225	0,880082	0,848273
fala	0,98957	0,911964	0,911052
mais	1,00081	0,911916	0,878066
aumenta	0,973528	0,958566	1,343845
leia	0,984834	1,111678	2,249061
descoberta	1,072751	1,315067	1,328381
dólar	0,995486	0,884145	0,726352
vender	0,981329	1,024529	1,454755
bndes	1,009586	1,128217	1,151248
nova	1,00387	0,911825	0,933512
psdb	0,978251	0,945105	1,211643
sabia	1,009023	1,288843	2,029112
gestão	0,974124	0,851321	0,818375
conselho	0,976154	0,974003	0,947378
esquema	1,019644	0,91564	0,9757
banco	1,023706	0,883868	0,59057
justiça	1,015661	0,762021	0,596078
privatização	1,025642	0,903357	0,948225
federal	0,982655	0,87241	0,765928
investimentos	1,006812	0,900541	1,005963
petroleiros	1,007126	0,741943	0,732922
balanço	1,030452	0,815939	0,67064
bilhão	0,998863	1,080135	1,00502
nega	0,968776	0,926217	1,097809
diesel	1,013034	0,975292	1,050903
plano	1,012388	0,975161	1,284663
cpi	1,024409	1,211602	1,303123

refinarias	1,019973	0,956646	0,959344
nota	0,933042	0,696327	0,325554
desvios	1,037429	0,884285	0,708038
preso	1,040563	1,20677	1,034944
greve	0,926851	0,604385	0,643453
bancos	1,016904	0,894455	0,855563
redução	1,015375	0,842853	1,058495
comissão	0,946778	1,096282	1,379681
defesa	1,001078	0,751103	0,541262
aprova	1,053391	1,187759	1,286576
câmara	0,954347	1,275987	1,433753
denúncias	0,953319	0,657261	0,547409
corrupção	1,011493	0,817644	0,693999
prejuízo	0,988284	0,758163	0,685396
senado	0,998089	1,210321	1,355047
investigar	1,012261	0,787442	1,033853
reajuste	0,974599	1,08355	1,260134
tcu	1,099385	0,870152	1,064119
stf	1,019195	1,052547	1,490633
pmdb	0,991207	1,056009	2,019405
reduz	0,978702	0,808368	0,863696
pf	0,978596	0,906529	0,718448
exdiretor	1,170998	1,378009	1,188081
lava	1,035957	1,09153	0,93363
pasadena	1,079261	1,15812	0,998583
escândalo	1,012543	0,831048	0,780011
investigação	0,997955	0,682266	0,620919
depoimento	1,042182	0,968884	0,934481
propina	1,053853	0,850529	0,825945
exgerente	0,57355	0,782113	1,120198
jato	1,047084	1,082808	0,914635
delação	0,951137	0,923928	1,02
delator	0,960934	1,002667	0,992718

Fonte: Elaboração própria com base nos *Tweets* sobre a Petrobras nos anos de 2010 a 2020.

ANEXO C – LISTA DE TERMOS POSITIVOS E NEGATIVOS DO ESTUDO

POLARIDADE DAS PALAVRAS

Positivas:

"recorde", "plataforma", "alta", "sobe", "políticos", "descoberta", "sabia" , "cpi", "preso", "câmara", "senado", "exdiretor", "expresidente", "exgerente", "programa", "campanha", "parte", "privatizar", "vídeo", "ativos", "fim", "afirma", "volta", "projeto", "compra", "oposição", "funcionários", "acordo", "decisão", "aumenta", "leia", "vender", "plano", "comissão", "aprova", "reajuste", "stf", "economia", "quer", "brasil", "empresas", "présal", "janeiro", "grande", "mi", "processo", "dólar", "desvios", "pf".

Negativas:

"valor" , "desde", "novo", "povo", "combustível", "menos", "preços", "política", "cai", "país", "combustíveis", "queda", "diretoria", "após", "pede", “recebeu", "abre", "lucro", "defende", "justiça", "petroleiros", "nota", "greve" , "defesa", "denúncias", "prejuízo", "investigar", "investigação", "dia", "negócios", “posto", "bilhões", "fecha", "distribuidora", "não", "concurso", "todo", "mundo", "presidente", "empresa", "bovespa", "brasileiro", "mercado", "preço", "vagas", "vai" , "bom", "bolsa", "governo", "ações", "todos", "nada", "agência", "dias", "pagar", "banco", "federal", "balanço", "corrupção", "escândalo".