

---

## **Análise de sentimentos e impacto de postagens do Twitter na bolsa de valores brasileira**

Fabio Fogliarini Brolesi<sup>1\*</sup>; André Correia Bueno<sup>2</sup>

<sup>1</sup> Especialista em Didática e Metodologia do Ensino Superior e Mineração de Dados Complexos. Campinas, São Paulo, Brasil – brolesi@gmail.com

<sup>2</sup> Instituto de Economia, Universidade Estadual de Campinas. Doutorando em Teoria Econômica pela UNICAMP. Rua Santos, 325 – Barra Funda; 11410-330, Guarujá, São Paulo, Brasil

\* brolesi@gmail.com

## **Análise de sentimentos e impacto de postagens do Twitter na bolsa de valores brasileira**

### **Resumo**

O crescente número de pessoas na bolsa de valores mostrou que ela possui relevância no contexto econômico brasileiro e que é uma opção para o ganho de capital também para pequenos e médios investidores. Conseguir monitorar o mercado para aumentar os ganhos é algo que também é desejável. O uso de redes sociais para este monitoramento possui relevância para realizar as análises. Dado este contexto, o objetivo geral do trabalho foi entender quais as relações que existem entre a análise de sentimentos de textos do Twitter com os valores de fechamento do mercado de ações específicas da bolsa de valores brasileira ao longo do tempo. A metodologia foi um estudo de caso, com a análise temporal de textos na rede social bem como dos resultados de fechamento de papéis específicos da bolsa brasileira. Para o presente trabalho não foi identificada relação significativa entre os sentimentos da rede social com o fechamento do mercado de ações.

**Palavras-chave:** B3; mineração de textos; dados não estruturados; rede social; investimentos.

### **Introdução**

O crescimento de pessoas físicas na bolsa de valores brasileira tem se mostrado bastante proeminente. Ainda que em período pandêmico, a B3, bolsa de valores brasileira, teve um aumento de 30% de CPFs comparando-se o terceiro trimestre de 2021 em relação ao mesmo período de 2020 conforme B3 (2022), totalizando mais de 3,3 milhões de pessoas. Ainda assim, o país não vigora nem mesmo entre os 10 países de maior participação no valor total do mercado acionário mundial CREDIT SUISSE (2021).

Os retornos do mercado de ações chamaram a atenção do público em geral, e trazem novas perguntas às pessoas que estão investindo neste mercado, como por exemplo: é possível, em algum grau, predizer preços, volatilidade ou resultados?

Yoshinaga e Junior (2012) afirma que existem eventos financeiros nos quais os investidores, movidos por algum sentimento, podem fazer uma aposta sobre o futuro de determinado papel, não necessariamente levando em conta o valor corrente. Neste sentido, os eventos financeiros talvez não possam mais ser completamente explicados pela teoria tradicional de finanças.

Uma avaliação do sentimento geral do mercado sobre um determinado tópico pode ser pertinente para esse tipo de análise. Origens de sentimento do público em geral, como textos em redes sociais nas quais os usuários se expressam livremente podem ser relevantes para análises. Conforme Souza et al. (2019), o sentimento que é colocado nos textos publicados em redes sociais tem potencial de explicar os movimentos que ocorrem no mercado de ações brasileiro, com relação a preços e volume de negociações. Também Peres et al. (2019) fala da importância da geração de valor a partir de dados de engajamento do

usuário, especificamente no mercado financeiro. Portanto, faz-se necessário identificar os desdobramentos das mídias sociais nas flutuações e/ou movimentos do mercado financeiro do Brasil.

No presente trabalho, utiliza-se a plataforma do Twitter<sup>1</sup> para avaliar se o sentimento dos conteúdos produzidos pelos usuários brasileiros num contexto de mercado de ações também do Brasil tem relação com o resultado das variações de bolsa e ações ao longo dos dias. Essa escolha se dá uma vez que do ponto de vista de acesso a redes sociais e produção de conteúdo, realizando um recorte para o público brasileiro, SLIDESHARE (2021) mostra que o país está entre os 5 maiores países em relação ao número de usuários nesta rede social.

Cocilova (2014) afirma que o Twitter tornou-se uma importante fonte de informação a partir de textos gerados pelo usuário e sua possibilidade de compartilhamento por vários outros usuários, além da possibilidade de análise da relevância dos textos. Isso fez com que pesquisas de diferentes esferas do conhecimento fossem executadas, como o estudos de propagação de epidemias, previsão de resultados eleitorais ou resultados de partidas futebol. No setor de finanças, estudos têm sido feitos para descobrir se existe uma correlação entre os textos e tendências do mercado de ações, e se os tweets afetam (ou preveem) os mercados financeiros.

Também Souza et al (2015) vê relevância em dados do Twitter: além de ser um canal abrangente e democrático de informações em tempo real, ele inclui notícias das grandes mídias, mas também pequenos eventos, que podem vir a fornecer informações sobre o mercado antes mesmo dos principais veículos de comunicação.

O embasamento do uso da ferramenta se dá por referências do trabalho de Bollen et al. (2011) e Guo e Li (2019). No caso de Bollen et al. (2011), o trabalho foi um dos primeiros a demonstrar que a análise de sentimentos no Twitter pode ser útil na análise de mercados. Na ocasião, a precisão de previsão de precificação de ações ultrapassou 87%, o que mostra a vasta possibilidade de análise do campo de processamento de linguagem natural associado ao mercado financeiro. Guo e Li (2019) também utilizaram o Twitter para análise de sentimentos e mercado, utilizando o que foi definido como “Twitter Sentiment Score”, uma medida de acurácia bastante alta (cerca de 66%) para prever a tendência futura do mercado (se de ganhos ou de perdas).

Portanto, o objetivo do trabalho é avaliar a existência de padrões entre sentimentos (positivo, neutro ou negativo) de *posts* de rede social em comparação com o fechamento do mercado brasileiro de ações (B3), num período de 5 meses (nov. 2021 a mar. 2022).

---

<sup>1</sup> <https://twitter.com/>

## Material e Métodos

O presente trabalho trata-se de um estudo de caso que utiliza a metodologia presente na Figura 1. Inicialmente é realizado dois tipos de extração: um referente aos textos da rede social Twitter e outro para recuperação dos valores dos papéis listados na B3 referente aos seguintes papéis: PETR4, VALE3, CASH3, BBSA3, MGLU3 e o próprio índice BOVESPA.

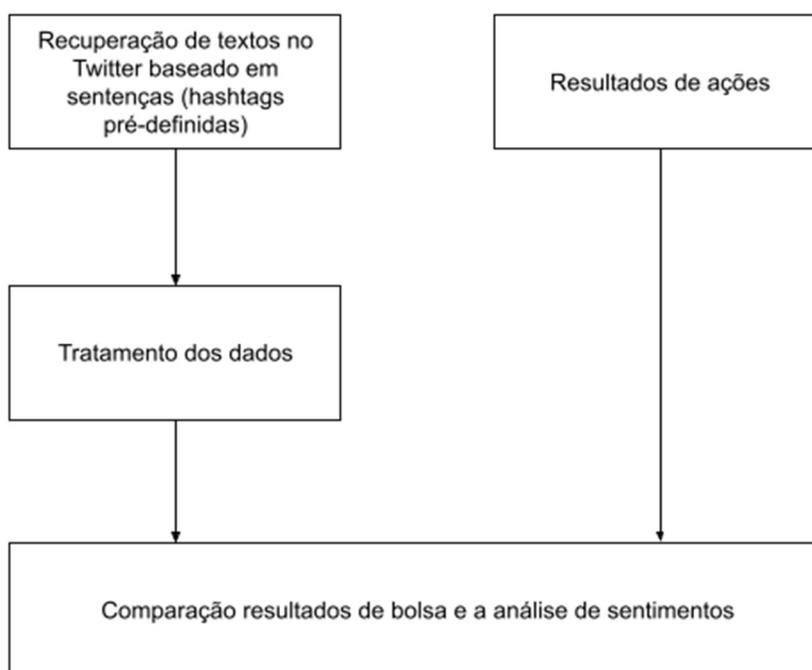


Figura 1. Visão esquemática da metodologia utilizada

Fonte: dados originais da pesquisa

A decisão destes papéis veio pelo fato de que a partir de todos os papéis listados no índice Brasil 100 (IBrX100 B3)<sup>2</sup>, estes cinco foram identificados como mais utilizados nos textos da rede social Twitter.

Do ponto de vista da extração de texto executada a partir da rede social Twitter, esta foi feita por meio de uma interface de programação de aplicativo (“*Application Programming Interface [API]*”), onde foram recuperadas informações baseadas em “*hashtags*” e “*cashtags*” entre nov. 2021 e jan. 2022. Todas as “*hashtags*” referiram-se a papéis que compunham o índice Brasil 100, a saber: PETR4, VALE3, CASH3, BBSA3, MGLU3, além do índice BOVESPA.

A Tabela 1 mostra parcialmente o esquema dos dados retornados pela API do Twitter com relação aos atributos mais significativos.

<sup>2</sup>[https://www.b3.com.br/pt\\_br/market-data-e-indices/indices/indices-amplos/indice-brasil-100-ibrx-100.htm](https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/indice-brasil-100-ibrx-100.htm)

Tabela 1: Esquema de dados parcial mostrando principais atributos do retorno da API do Twitter

Atributo	Descrição
id	Identificador único de entrada de texto na rede social
text	O texto criado pelo autor
created_at	Data e hora da criação do texto
hashtag	Combinação de palavras-chave ou frases precedidas pelo símbolo #
cashtag	Identificador do código de uma empresa precedido pelo sinal \$

Fonte: Dados originais da pesquisa.

Com relação à extração feita para os valores dos índices, utilizou-se uma biblioteca da linguagem de programação *Python* de nome “*yfinance*” que capturou dados de abertura, fechamento, máximo e mínimo das ações a partir do site Yahoo! Finance<sup>3</sup>.

Para a abordagem léxica, utilizou-se a ferramenta *OpLexicon* que é um conjunto de léxicos de sentimento em língua portuguesa. Conforme Souza (2011), esta ferramenta consiste na aplicação de técnicas já presentes na literatura: uma delas a *Thesaurus-based*, outra sendo *Turney’s Corpus-based*, e uma última como variação de *Mihalcea*, executando a automação afim de marcar a palavra como positiva, neutra ou negativa.

Inicialmente utilizou-se uma lista de palavras que foram removidas antes da análise, por não terem valor dentro do contexto que foi analisado. Esta lista é chamada de lista de “*stopwords*”. Utilizou-se uma lista da biblioteca *Python* “*nltk*” e removeu-se a palavra “não” da lista de “*stopwords*” porque ela tem valor para o contexto analisado.

Em seguida, converteu-se todo o texto para minúsculo, para equalizar todos os textos afim de palavras que possuísem inicialmente maiúsculas ou minúsculas, ao final não tivessem distinção.

Na próxima etapa foram removidos endereços de “*sites*” (em inglês, “*Uniform Resource Locator*” ou “*URL*”). A seguir, foi feita a conversão de números para texto do número por extenso. O resultado foi processado novamente para remoção de caracteres de pontuação e em seguida removeu-se o conjunto de “*stopwords*”. Por fim, removeu-se “*emojis*”, que são caracteres pictóricos presentes no conjunto de caracteres possíveis de serem gravados.

Após a etapa de preparação dos dados, realizou-se a análise de sentimentos dos textos do Twitter a partir da metodologia de Peres et al. (2019) do ponto de vista de análise da polaridade, seguindo a eq. (1):

<sup>3</sup> <https://finance.yahoo.com/>

$$S = \sum_{k=1}^N t(i,j) \quad (1)$$

onde  $t(i,j)$  é a  $i$ -ésima palavra (ou termo) e  $j$  é sua polaridade de acordo com a disponibilidade no dicionário léxico definido. Uma vez o somatório feito para cada *tweet*, estes são classificados conforme a Tabela 2.

Tabela 2: Classificação de sentimento a partir da eq. (1)

Valor Somatório	Classe
$(-\infty, -2]$	Muito negativo
$(-2, -0,5)$	Negativo
$[-0,5, 0,5]$	Neutro
$(0,5, 2)$	Positivo
$[2, +\infty)$	Muito positivo

Fonte: Peres et al (2019), adaptado.

Com os textos dos *tweets* sanitizados e a polaridade definida, executou-se uma agregação por dia considerando-se a soma e a média dos sentimentos dos textos do Twitter afim de obter um único valor numérico para que este fosse comparado com o valor de fechamento dos papéis comparando-se com fechamento de dias anteriores.

Com relação ao fechamento, fez-se a análise de variações de preços de  $d-1$  a  $d-10$ , onde  $d$  é um dia em particular, considerando-se os limites de corte de datas estabelecidos, e correlações de Pearson (correlação linear) e Spearman (monotônica), mas também utiliza-se testes de significância estatística não paramétricas para avaliar a correlação entre o fechamento e o sentimento para os papéis, dos quais teste U de Mann-Whitney, teste Wilcoxon Signed-Rank e teste H de Kruskal-Wallis.

O teste U de Mann-Whitney é um teste de significância estatística para determinar se as amostras foram extraídas de uma população com a mesma distribuição. Ao avaliar os papéis com relação a comentários e o fechamento do mercado, os valores do resultado dos testes foram no sentido de rejeitar  $H_0$ .

Segundo Corder et al, 2009, a eq (2) é utilizada para determinar uma estatística de teste de Mann-Whitney para cada uma das amostras. A menor das duas estatísticas U é o valor obtido para o resultado final.

$$U_i = n_1 n_2 + \frac{n_i(n_i + 1)}{2} - \sum R_i \quad (2)$$

Onde  $U_i$  é a estatística de teste para a amostra de interesse,  $n_i$ , é o número de valores da amostra de interesse,  $n_1$  é o número de valores da primeira amostra,  $n_2$  é o número de valores da segunda amostra, e  $\sum R_i$  é a soma das classificações da amostra de interesse.

O teste Wilcoxon Signed-Rank pressupõe que os dados possam ser emparelhados, ou seja, podem de alguma forma representar medições da mesma técnica e pelo fato de não serem independentes, não seria possível utilizar o teste U de Mann-Whitney. A eq (3) mostra como é computada esta estatística

$$T = \min(\sum R_+, \sum R_-) \quad (3)$$

Onde  $\sum R_+$  é a soma das classificações com diferenças positivas e  $\sum R_-$  é a soma das classificações com diferenças negativas.

Já o teste H de Kruskal-Wallis, diferente dos dois anteriores, não faz a comparação aos pares, mas avalia diferentes distribuições. Ele é uma versão não paramétrica do teste ANOVA, ou análise de variância. Este teste pode ser usado para determinar se as amostras independentes têm uma distribuição diferente. Pode ser pensado como a generalização do teste U de Mann-Whitney.

Para este teste, a eq (4) define o cálculo como:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1) \quad (4)$$

Onde N é o número de valores de todas as amostras combinadas,  $R_i$  é a soma das classificações de uma amostra em particular, e  $n_i$  é o número de valores soma da classificação da amostra correspondente.

Os graus de liberdade, (df, ou “degrees of freedom”), para o teste H de Kruskal-Wallis são determinados usando a eq (5).

$$df = k - 1 \quad (5)$$

Onde df é o grau de liberdade e k é o número de grupos.

Para a avaliação de normalidade, tanto da média de polaridade quanto dos valores de fechamento, utilizou-se o teste de normalidade de Shapiro-Wilk, conforme [NIST] 2012 na eq. (6):

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (6)$$

Onde  $x_{(i)}$  é a i-ésima estatística de ordem, ou seja, o i-ésimo menor número da amostra,  $x_1, \dots, x_n$  vêm de uma população normalmente distribuída.

Para este caso, tem-se o seguinte:

- Hipotese nula: os dados seguem distribuição normal
- Hipotese alternativa: os dados não seguem distribuição normal

## Resultados e Discussão

Ao longo da análise observou-se que os papéis com mais textos vinculados eram os seguintes: PETR4 (com 16.291 textos), MGLU3 (com 8.474 menções), CASH3 (com 7.137 referências), VALE3 (com 7.884 citações), BBAS3 (com 5.664 tweets). Do total de textos (111.080 postagens na rede social) que foram coletados entre 31 out. 2021 e 01 abr. 2022.

A Figura 2 mostra a distribuição de ocorrências de cada um dos papéis e do índice por dia, considerando o parâmetro “created\_date”. Nota-se que há um pico para MGLU3 no dia 14 dez. 2021 e dois valores bastante discrepantes no resultado da BOVESPA nos dias 29 e 30 de mar. 2022.

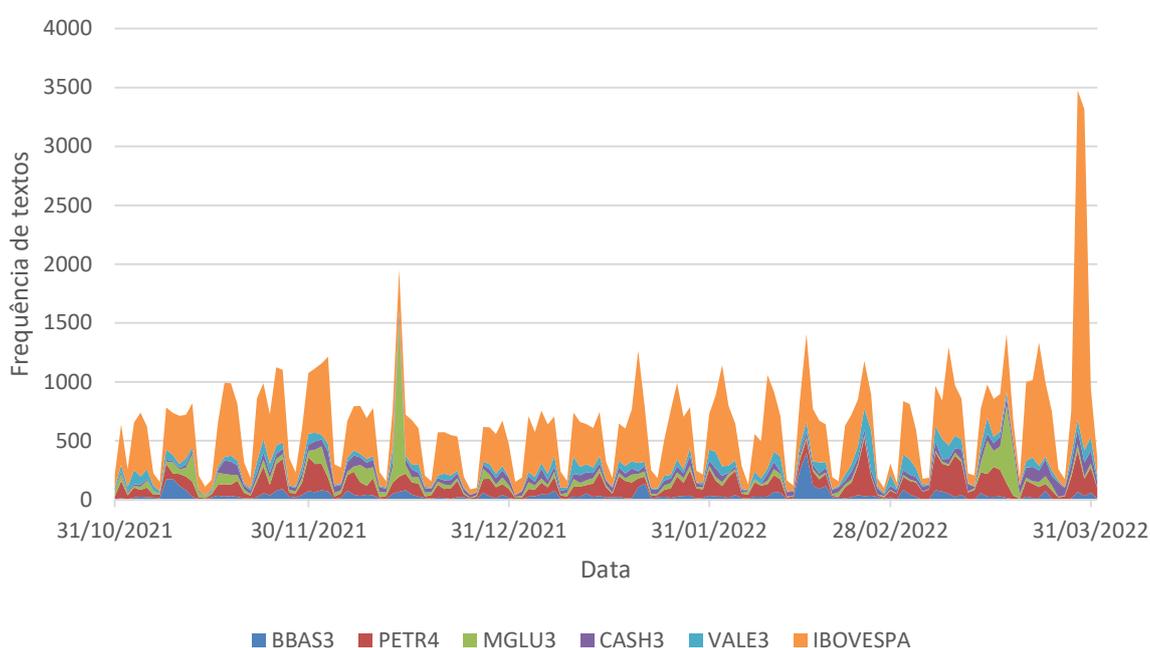


Figura 2. Ocorrência de tweets por dia considerando o código de cada um dos papéis de interesse

Fonte: Resultados originais da pesquisa

Do ponto de vista de frequência mensal de textos referenciando determinado papel ou índice, a Tabela 3 mostra a quantidade de textos que contém cada um dos papéis ou índice por mês que foram recuperados a partir da API do Twitter para o presente trabalho. A mesma Tabela 3 está materializada na forma de gráfico na Figura 3.

Tabela 3: Frequência de textos por mês para cada um dos papéis ou índice

Mês	BBAS3	CASH3	IBOVESPA	MGLU3	PETR4	VALE3
out. 2021	3	9	111	52	24	15
nov. 2021	1.270	984	10.411	1.485	2.924	1.482
dez. 2021	1.024	1.606	9.226	3.021	2.922	1.003
jan. 2022	944	1.335	10.017	987	2.795	1.373
fev. 2022	1.525	995	9.344	605	3.020	1.943
mar. 2022	890	2.155	15.963	2.323	4.514	2.029
abr. 2022	8	53	165	1	92	39

Fonte: Resultados originais da pesquisa

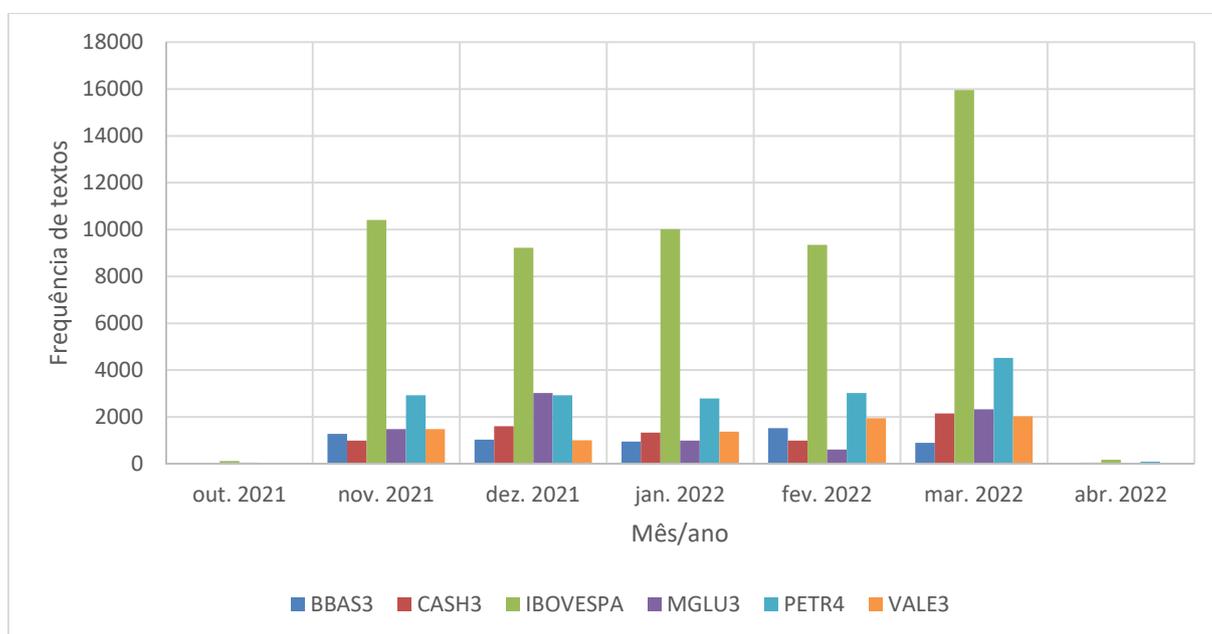


Figura 3. Frequência de textos por mês para cada um dos papéis ou índice

Fonte: Resultados originais da pesquisa

Sobre estes textos dos 5 papéis mais comentados, observou-se uma alta frequência de valores de sentimento neutro, que correspondeu a 60,00% do total de textos para os papéis citados. Isso significa que as palavras dos textos somadas resultaram em 0 embora existisse alguma polaridade, ou que não houve correspondência de palavras no dicionário léxico utilizado. A Tabela 4 mostra a distribuição de polaridade entre os papéis e o índice selecionados para o presente trabalho. Adicionalmente, a Figura 4 materializa a Tabela 4 em forma de gráfico.

Tabela 4: Frequência de polaridade para cada um dos papéis ou índice

Polaridade	BBAS3	CASH3	IBOVESPA	MGLU3	PETR4	VALE3
-7	0	0	1	0	0	0
-6	0	0	2	0	1	0
-5	0	0	2	0	2	0
-4	2	5	35	5	18	11
-3	11	6	166	24	63	19
-2	71	31	947	297	348	112
-1	715	218	5975	924	1812	601
0	3615	2273	34775	4395	9885	5518
1	889	3541	9515	2479	3058	1178
2	229	899	2937	241	767	311
3	88	143	693	80	253	90
4	34	16	126	21	67	31
5	10	3	51	5	11	11
6	0	0	8	2	4	2
7	0	0	2	1	2	0
8	0	1	1	0	0	0
9	0	0	0	0	0	0
10	0	1	0	0	0	0
11	0	0	0	0	0	0
12	0	0	0	0	0	0
13	0	0	1	0	0	0

Fonte: Resultados originais da pesquisa

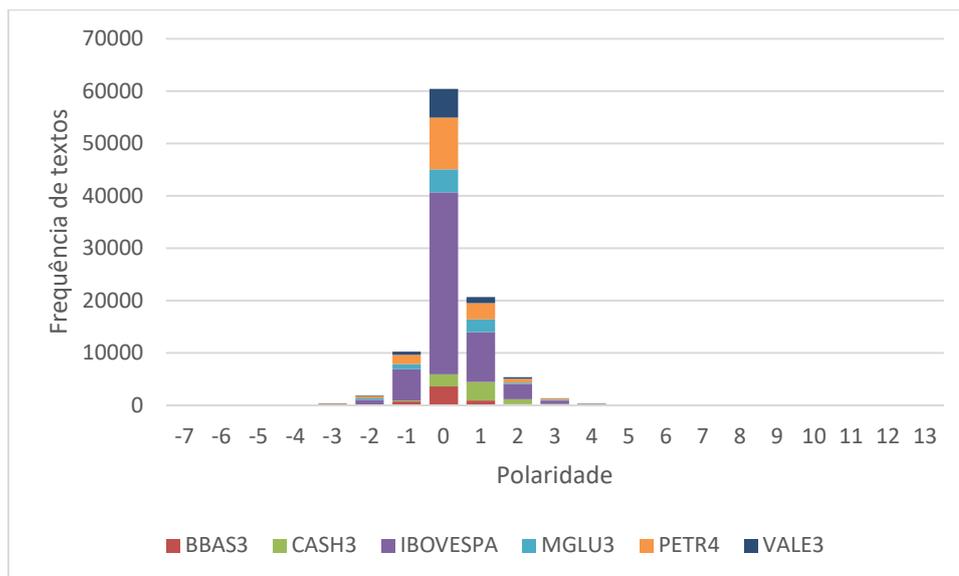


Figura 4. Frequência de polaridade para cada um dos papéis ou índice  
Fonte: Resultados originais da pesquisa

Do ponto de vista metodológico, considerando a Tabela 2 conforme Peres et al (2019), a Tabela 4 seria convertida na Tabela 5. A Figura 5 ilustra a Tabela 5 no formato gráfico, para fins comparativos.

Tabela 5: Frequência de categoria de sentimento para cada um dos papéis ou índice

Categoria	BBAS3	CASH3	IBOVESPA	MGLU3	PETR4	VALE3
Muito negativo	84	42	1153	326	432	142
Negativo	715	218	5975	924	1812	601
Neutro	3615	2273	34775	4395	9885	5518
Positivo	889	3541	9515	2479	3058	1178
Muito positivo	361	1063	3819	350	1104	445

Fonte: Resultados originais da pesquisa

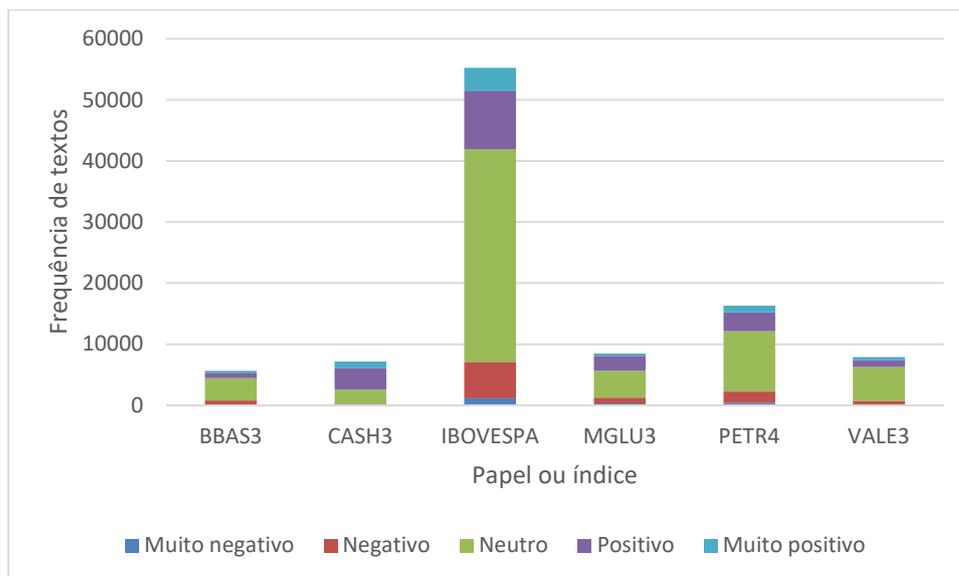


Figura 5. Frequência de polaridade de textos cumulativa para cada um dos papéis ou índice  
Fonte: Resultados originais da pesquisa

Considerando-se a Tabela 5, tem-se em termos de distribuição em classes, o resultado da Figura 6, com as frequências somadas.

#### FREQUÊNCIAS DE CLASSES DE SENTIMENTOS

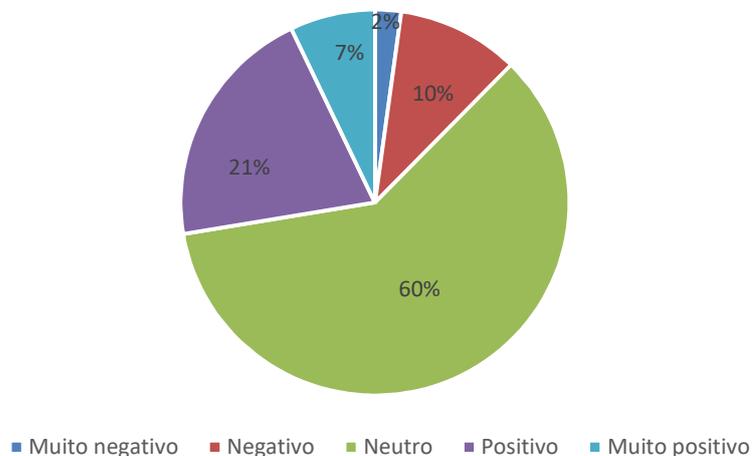


Figura 6. Soma dos fatores para todos os papéis e índice considerados.  
Fonte: Resultados originais da pesquisa

Abaixo, mostra-se um comparativo das médias de sentimentos por dia e o fechamento do mercado para os papéis PETR4 (Figura 7), MGLU3 (Figura 8), CASH3 (Figura 9), VALE3 (Figura 10), BBAS3 (Figura 11) e do próprio fechamento da bolsa e sentimentos vindos de textos que mencionavam a BOVESPA por dia (Figura 12).

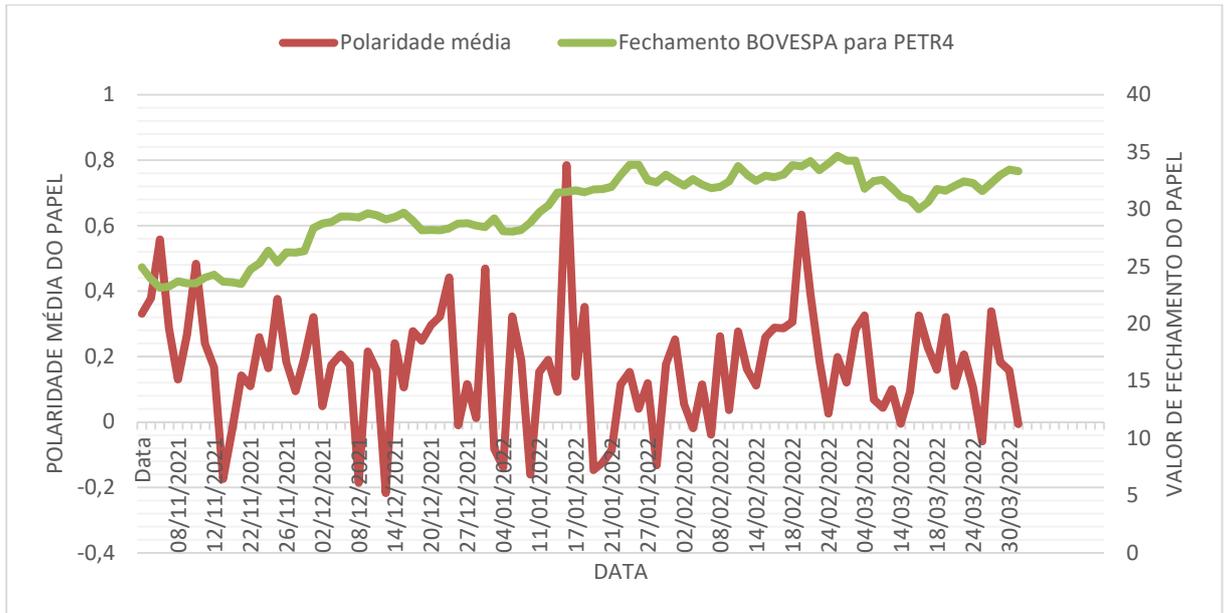


Figura 7. Gráfico de polaridade média por dia para o papel PETR4 e valor de fechamento por dia do papel PETR4

Fonte: Resultados originais da pesquisa

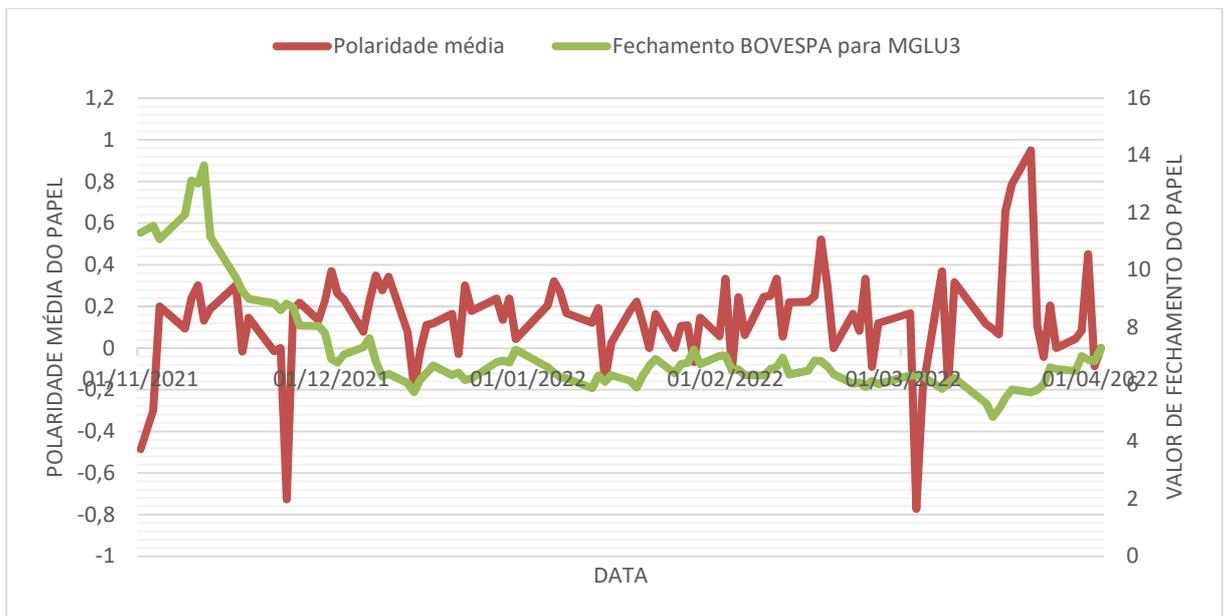


Figura 8. Gráfico de polaridade média por dia para o papel MGLU3 e valor de fechamento por dia do papel MGLU3

Fonte: Resultados originais da pesquisa

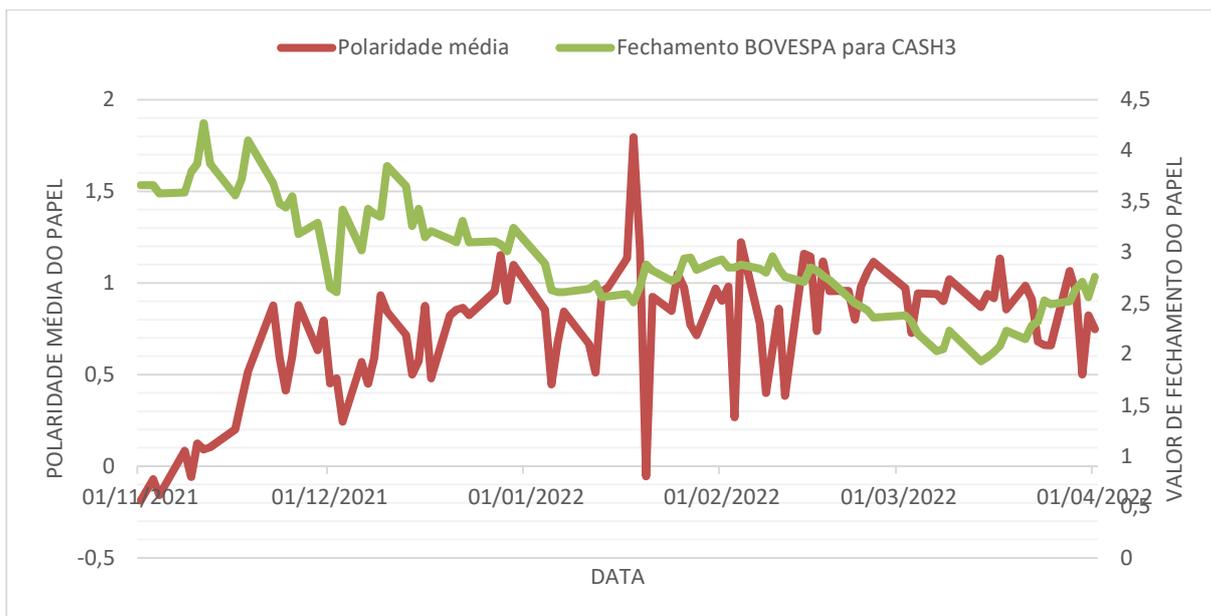


Figura 9. Gráfico de polaridade média por dia para o papel CASH3 e valor de fechamento por dia do papel CASH3

Fonte: Resultados originais da pesquisa

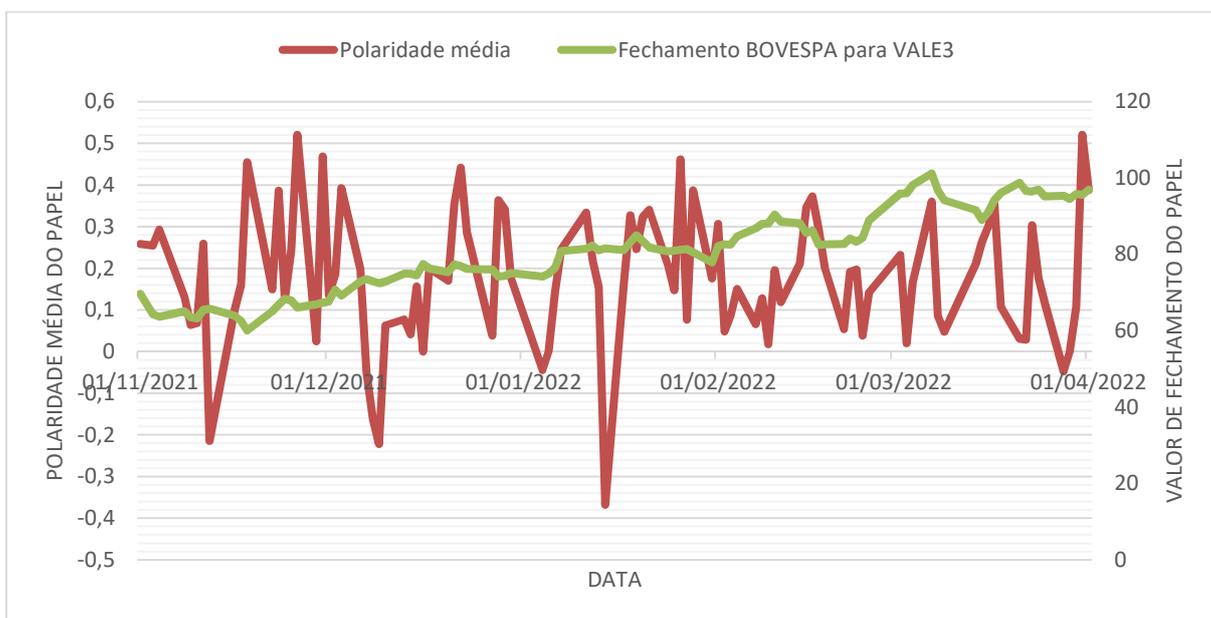


Figura 10. Gráfico de polaridade média por dia para o papel VALE3 e valor de fechamento por dia do papel VALE3

Fonte: Resultados originais da pesquisa

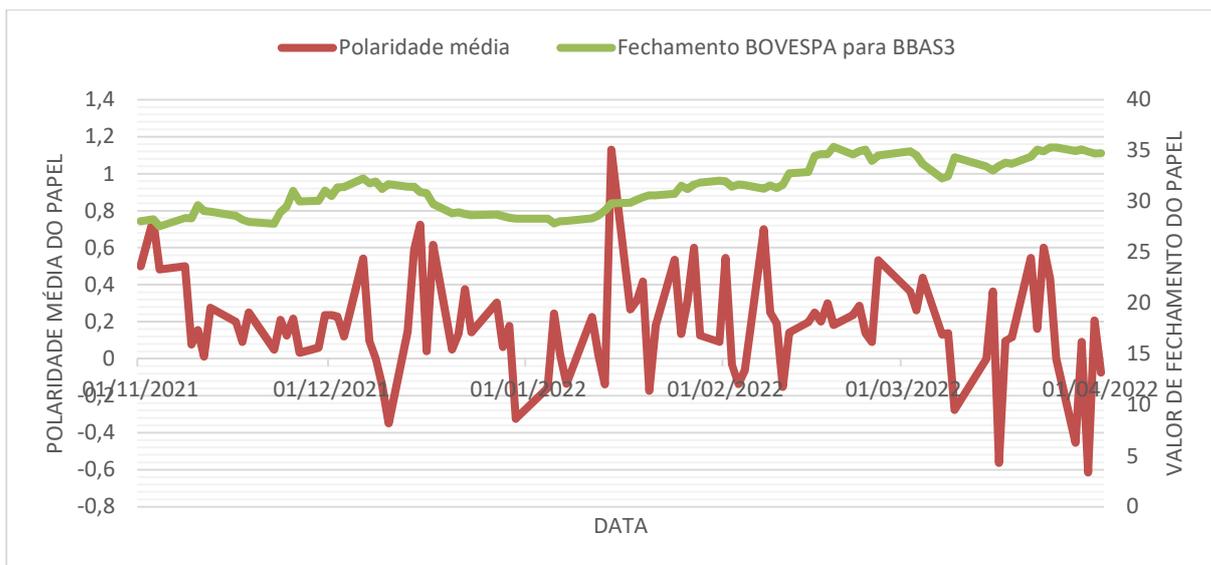


Figura 11. Gráfico de polaridade média por dia para o papel BBAS3 e valor de fechamento por dia do papel BBAS3

Fonte: Resultados originais da pesquisa

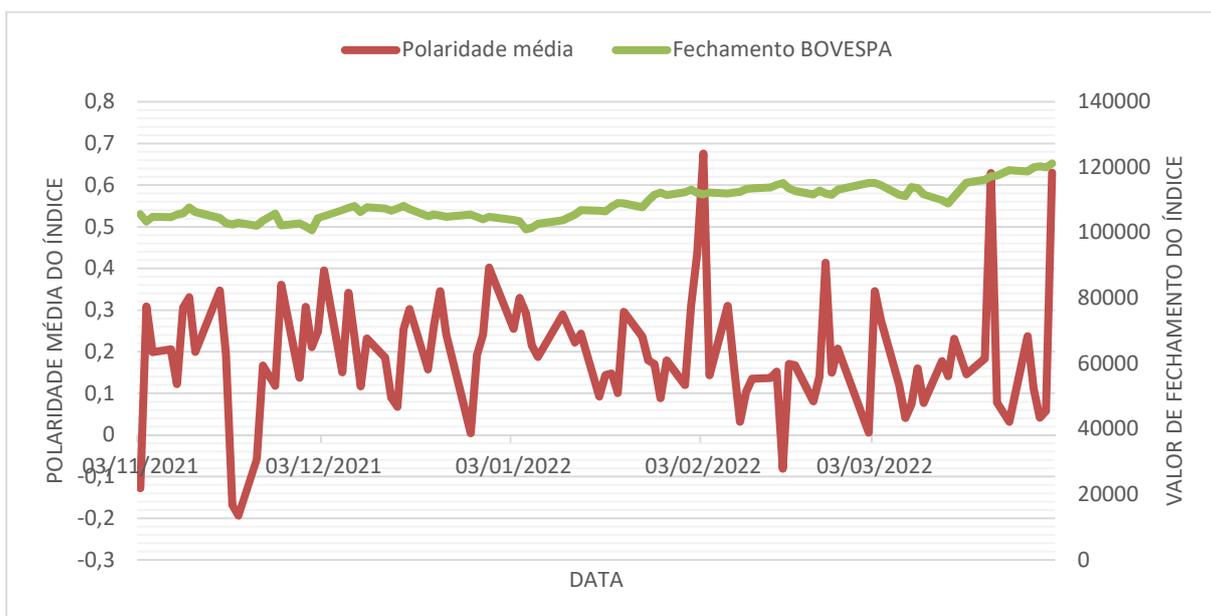


Figura 12. Gráfico de polaridade média por dia para o índice BOVESPA e valor de fechamento por dia do índice BOVESPA

Fonte: Resultados originais da pesquisa

Sobre a normalidade dos dados, utilizou-se o teste de Shapiro-Wilk para normalidade. A Tabela 5 mostra os resultados obtidos considerando o p-valor de 0,05:

Tabela 5: Avaliação de normalidade dos dados

Código	Tipo	Valor	Detalhe
PETR4	Polaridade	0,0561	A hipótese nula não pode ser rejeitada
PETR4	Fechamento	0,0	A hipótese nula pode ser rejeitada
MGLU3	Polaridade	0,0	A hipótese nula pode ser rejeitada
MGLU3	Fechamento	0,0	A hipótese nula pode ser rejeitada
CASH3	Polaridade	0,0127	A hipótese nula pode ser rejeitada
CASH3	Fechamento	0,0881	A hipótese nula não pode ser rejeitada
VALE3	Polaridade	0,053	A hipótese nula não pode ser rejeitada
VALE3	Fechamento	0,0042	A hipótese nula pode ser rejeitada
BBAS3	Polaridade	0,3052	A hipótese nula pode ser rejeitada
BBAS3	Fechamento	0,0	A hipótese nula pode ser rejeitada
BOVESPA	Polaridade	0,0	A hipótese nula pode ser rejeitada
BOVESPA	Fechamento	0,0013	A hipótese nula pode ser rejeitada

Fonte: Resultados originais da pesquisa

Quanto aos resultados de correlações e testes estatísticos, tem-se como resultado a Tabela 6:

Tabela 6: Papéis e índice e correlações e testes estatísticos com relação a fechamento do papel e polaridade média diária

	PETR4	MGLU3	CASH3	VALE3	BBAS3	BOVESPA
Correlação de Pearson	-0.1584	-0.1086	-0.5621	-0,0396	-0.0313	0.0652
Correlação de Spearman	-0.1497	-0.1267	-0.5100	-0,0264	-0.0418	-0.1554
Teste U de Mann-Whitney	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Wilcoxon Signed-Rank	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
teste H de Kruskal-Wallis	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Fonte: Resultados originais da pesquisa

Notou-se relações fracas entre o agrupamento dos sentimentos por dia, utilizando-se medidas estatísticas de centralidade. As identificações não foram suficientemente robustas para, a partir delas, obter-se conclusões mais evidentes acerca da análise de sentimento e do fechamento do mercado em determinado período. Neste sentido, pode-se afirmar que utilizar apenas um número – o resultado de cálculo de sentimento das palavras – a partir dos insumos utilizados – textos do Twitter e léxico de palavras abrangente –, não é possível estabelecer uma correlação seja a partir de medidas estatísticas paramétricas, seja a partir de medidas estatísticas não paramétricas.

Observou-se um conjunto de elementos que poderiam interferir na análise, a saber: o Brasil possui poucos investidores comparado a outros países como os Estados Unidos, por exemplo. Bollen et al. (2011) trabalhou com um volume significativamente maior (9.853.498) de textos de Twitter em cerca do dobro do tempo do trabalho corrente considerando o mercado norte-americano; nem todos os investidores brasileiros utilizaram e/ou escreveram

textos no Twitter a respeito de papéis comercializados na bolsa de valores; os textos do Twitter eram muito sucintos para trazerem toda a informação necessária, eventualmente sendo necessário um contexto maior; o período de análise pode ter sido muito pequeno; identificou-se volumes consideráveis de textos com polaridade neutra, o que indica que talvez possa ser utilizado um léxico específico para análise de sentimentos num escopo definido para mercado financeiro.

Outros aspectos importantes a serem considerados são os seguintes: um único número (a polaridade, no caso) é uma simplificação (ou redução de dimensionalidade) grande para o contexto e cenários de análise. Neste caso, uma possibilidade de enriquecimento poderia ser adentrar no contexto de análise de emoções como as de Ekman. Conforme Ekman (1976), as expressões faciais identificadas são nojo, raiva, medo, tristeza, alegria, surpresa e desprezo. Essa lista posteriormente foi expandida, adicionando emoções positivas e negativas a saber: diversão, desprezo, excitação, culpa, mentira, orgulho, alívio, satisfação, prazer e vergonha. Neste caso, a proposta seria identificar um léxico para cada tipo de emoção, e com isso realizar a análise (conforme eq. (1)) afim de gerar um número para cada tweet considerando cada uma das emoções e, a partir da composição destes números, entendendo que a dimensionalidade da análise se amplia, pode ser possível identificar resultados mais robustos.

Outras análises possíveis considerando os dados seriam a comparação com o volume negociado em bolsa, ou a variação entre o fechamento de um dia versus a abertura do dia seguinte. Também pode ser possível a análise de variação de valores (abertura, fechamento ou volume), proporcionalmente a dias anteriores, captando a porcentagem e fazendo uma análise a partir deste referencial.

## **Conclusão Final**

Para a análise proposta, não se observou relações significativas a partir da mineração de textos e análise de sentimentos baseado no léxico do OpLexicon comparadas ao fechamento da bolsa de valores brasileira, nem com testes paramétricos, tampouco com testes estatísticos não paramétricos. O estudo levou em conta o ferramental léxico já existente para português, porém de escopo abrangente. Para futuras pesquisas, sugerisse a criação de um léxico de sentimentos com escopo definido para o mercado financeiro brasileiro, bem como análise de outras fontes para avaliação de relações com fechamento de bolsa de valores, como por exemplo textos e artigos que possuem uma extensão maior que potencialmente pode trazer um contexto maior afim de que os resultados de sentimentos tenham mais relevância ao comparar com o fechamento do mercado. Também sugerisse que

se possa buscar outros léxicos que deem mais corpo à análise de sentimentos, como análise de emoções, para que possa ser aumentada a dimensionalidade e assim, a partir da estruturação dentro de um espaço vetorial maior, se possa realizar uma análise e avaliação mais robusta para tentar identificar as associações de forma mais precisa. Outra possibilidade é o uso de n-gramas para a partir destes tentar identificar mais e melhor o contexto do que está no discurso.

## Referências

- Bollen, J., Mao, H., Zeng, X-J.. 2011. Twitter mood predicts the stock market. In: Journal of computational Science. Disponível em <<https://arxiv.org/pdf/1010.3003>>. Acesso em: 10 nov. 2021.
- Pessoas físicas - Uma análise da evolução dos investidores na B3 [https://www.b3.com.br/data/files/35/17/D6/0D/B37EC710BD0885C7AC094EA8/Book\\_PF\\_Se t2021.pdf](https://www.b3.com.br/data/files/35/17/D6/0D/B37EC710BD0885C7AC094EA8/Book_PF_Se t2021.pdf) Acessado em 12 jan. 2022.
- Guo, X., Li, J.. 2019. A Novel Twitter Sentiment Analysis Model with Baseline Correlation for Financial Market Prediction with Improved Efficiency. In: Conference on Social Networks Analysis, Management and Security (SNAMS) Disponível em <<https://arxiv.org/ftp/arxiv/papers/2003/2003.08137.pdf>>. Acesso em: 10 nov. 2021.
- Souza, D. M. S., Lucena, W. G. L., Queiroz, D. B.. 2019. O Efeito do Sentimento do Investidor Expresso via Twitter sobre o Comportamento do Mercado Acionário Brasileiro Durante o Período Eleitoral. In: XIX USP International Conference in Accounting Disponível em <<https://congressosp.fipecafi.org/anais/19Usplnternational/ArtigosDownload/1621.pdf>>. Acesso em: 10 out. 2021.
- Credit Suisse, [CREDIT SUISSE] 2021. Credit Suisse Global Investments Returns Yearbook 2021 Summary Edition. Disponível em: <<https://www.credit-suisse.com/media/assets/corporate/docs/about-us/research/publications/credit-suisse-global-investment-returns-yearbook-2021-summary-edition.pdf>>. Acesso em: 15 jan. 2021
- Slide Share [SLIDE SHARE] 2021. Digital 2021 October Global Statshot Report v03. Disponível em <<https://www.slideshare.net/DataReportal/digital-2021-october-global-statshot-report-v03>>. Acesso em: 15 jan. 2021
- Peres, V., Vieira, R., & Bordini, R. (2019). Análises de Sentimentos: abordagem lexical de classificação de opinião no contexto mercado financeiro brasileiro. Disponível em <[http://www.comp.ita.br/labsca/waiaf/papers/VitorPeres\\_paper\\_6.pdf](http://www.comp.ita.br/labsca/waiaf/papers/VitorPeres_paper_6.pdf)> Acesso em: 16 jan. 2022
- Yoshinaga, C. E., Junior, F. H. F. C. The Relationship between Market Sentiment Index and Stock Rates of Return: a Panel Data Analysis. Disponível em <<https://www.scielo.br/j/bar/a/mYcPXrTgXGFWDRrybdQZrSJ/?lang=en&format=pdf>> Acesso em: 09 mai. 2022
- Souza, M.; Vieira, R.; Buseti, D.; Chishman, R. e Alves, I. M. Construction of a Portuguese Opinion Lexicon from multiple resources. 8th Brazilian Symposium in Information and Human Language Technology, 2011. Disponível em <<https://www.inf.pucrs.br/linatural/wordpress/wp-content/uploads/2017/08/STIL2011.pdf>> Acesso em: 16 jan. 2022
- Cocilova, A. Twitter data analysis for financial markets, Disponível em <[https://amslaurea.unibo.it/8126/1/cocilova\\_alessandro\\_tesi.pdf](https://amslaurea.unibo.it/8126/1/cocilova_alessandro_tesi.pdf)> Acesso em 09 mai. 2022

Corder, G. W, Foreman, D. I, Nonparametric Statistics for Non-Statisticians: A Step-by-Step 2009

Ekman, P.; Friensen, W. V. The facial action coding system. Palo Alto, CA: Consulting Psychologists Press, 1976.

National Institute of Standards and Technology [NIST]. 2012. Anderson-Darling and Shapiro-Wilk tests Engineering Statistics Handbook. Disponível em <<https://www.itl.nist.gov/div898/handbook/prc/section2/prc213.htm>> . Acesso em 22 jul. 2022

## Apêndice

Tabela 7: Ativos que compõe o índice IBRX 100

Código	Código	Código	Código	Código
ALPA4	CESP6	GGBR4	LAME3	RDOR3
ABEV3	CIEL3	GOAU4	LAME4	RAIL3
AMER3	COGN3	GETT11	LREN3	SBSP3
ASAI3	CPLE6	GOLL4	MGLU3	SAPR11
AZUL4	CSAN3	NTCO3	MRFG3	SANB11
B3SA3	CPFE3	SOMA3	CASH3	CSNA3
BIDI4	CVCB3	HAPV3	BEEF3	SULA11
BIDI11	CYRE3	HYPE3	MOVI3	SUZB3
BPAN4	DXCO3	IGTA3	MRVE3	TAAE11
BBSE3	ECOR3	GNDI3	MULT3	TASA4
BRML3	ELET3	IRBR3	PCAR3	VIVT3
BBDC3	ELET6	ITSA4	PETR3	TIMS3
BBDC4	EMBR3	ITUB4	PETR4	TOTS3
BRAP4	ENBR3	JBSS3	PRIO3	UGPA3
BBAS3	ENGI11	JHSF3	PETZ3	USIM5
BRKM5	ENEV3	KLBN11	PSSA3	VALE3
BRFS3	EGIE3	LIGT3	POSI3	VIIA3
BPAC11	EQTL3	RENT3	QUAL3	VBBR3
CRFB3	EZTC3	LCAM3	RADL3	WEGE3
CCRO3	FLRY3	LWSA3	RAPT4	YDUQ3
CMIG4				

Fonte: Dados originais da pesquisa