

Análise de Impacto dos Posicionamentos de Presidentes do Brasil em Relação ao Mercado Acionário Brasileiro Utilizando Machine Learning

Pedro Santos Mendonça Neto¹, João Gabriel Gama Vila Nova

Ciência da Computação, Centro Universitário de Maceió, Brasil

¹opedrosantos16@gmail.com.

²joao.nova@unima.com.br

RESEARCH ARTICLE

Annals of Data Analysis

<https://periodicos.ufpe.br/revistas/ADA/>

Submitted on: 12.12.2025.

Accepted on: 01.08.2026.

Published on: 01.08.2026.

Copyright © 2026 by author(s).

This work is licensed under the Creative

Commons Attribution International

License CC BY-NC-ND 4.0

<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>

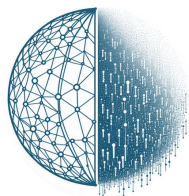


Resumo

Este estudo investigou a correlação entre os discursos dos presidentes Dilma Rousseff, Michel Temer e Jair Bolsonaro e o mercado de ações brasileiro, utilizando Inteligência Artificial para classificar as falas por tema e sentimento. A pesquisa concluiu que, embora os discursos tenham tido uma correlação estatística limitada e pontual com o preço dos ativos, eles demonstraram uma correlação relevante com o volume de negociações. Esse efeito foi particularmente notável durante o governo Bolsonaro, onde há indícios de que discursos com sentimento negativo estavam fortemente correlacionados a um aumento na atividade do mercado.

Keywords

Mercado Acionário Brasileiro; Discursos Presidenciais; Política e Finanças; Processamento de Linguagem Natural; Aprendizagem de Máquina.



1. Introdução

O mercado financeiro é uma das invenções mais fascinantes da nossa era, pois a aquisição de ações de uma empresa torna o investidor um sócio desta, implicando tanto na aceitação de riscos quanto na possibilidade de lucro. Além disso, outro fator que o torna impressionante e complexo é a ampla influência causada por fatores diversos, incluindo discursos de figuras relevantes, notícias divulgadas pela mídia, eventos epidêmicos, entre outros. (Shah et al., 2019). Como por exemplo, quando Donald Trump venceu a disputa presidencial dos EUA em 2017, contrariando todas as projeções estatísticas, bolsas de valores no mundo todo sofrem reações voláteis de valores (Wagner et al., 2017). Devido a estes fatores, a manipulação e análise das variáveis que regem o mercado financeiro é um tema de grande interesse e consequentemente objeto de pesquisa tanto por parte de acadêmicos quanto por profissionais atuantes na área.

A Inteligência Artificial (IA) e mais especificamente o Machine Learning (ML) são técnicas que se provaram eficazes quando usadas para resolver problemas relacionados ao setor financeiro, especialmente devido à grande quantidade de dados que precisa ser analisada nestes casos. (Lakhchini et al., 2022).

Além disso, os modelos que utilizam algoritmos de ML são capazes de proporcionar planejamento financeiro e assistência com estratégias de gerenciamento de risco. Como os trading systems que usam ML como apoio à decisão, eles podem ser treinados com o auxílio da análise de dados usando a econometria, para descobrir regras em espaços de alta dimensionalidade (Kniazieva, 2023). Porém, embora seja possível criar modelos somente com base nos dados de mercado, é crucial que consideremos também os dados não estruturados (Tatsat et al., 2020). Afinal, investidores são influenciados pelo sentimento e reagem muitas vezes aos acontecimentos ao seu redor de forma intuitiva (Baker and Wurgler, 2007).

A história do mercado de ações está repleta de eventos marcantes o suficiente para ganhar seus próprios nomes: The Great Crash (A Grande Depressão de 1929), o Tronics Boom do início dos anos 1960, a bolha Nifty Fifty do início dos anos 1970, o crash da Black Monday de outubro de 1987 e a bolha da Internet ou pontocom dos anos 1990 (Baker and Wurgler, 2007). Todos esses acontecimentos acarretaram em mudanças drásticas nos valores de ações, causados pelo chamado sentimento do investidor.

Compreender e analisar o sentimento de mercado é essencial para um bom entendimento do comportamento da bolsa de valores. A análise de sentimento envolve a interpretação de grandes volumes de dados não estruturados, como vídeos, fotos, arquivos de áudio, postagens em redes sociais, artigos, entre outros. Que age com o objetivo de definir se os dados passam uma mensagem positiva, negativa ou neutra (Sousa, 2019). Seguindo nesse contexto, Maligkris (2017), demonstrou que discursos de candidatos à presidência dos EUA influenciaram o mercado financeiro, afetando retornos, volatilidade e volume de negociações. A magnitude desse efeito depende do conteúdo dos discursos; por exemplo, discursos políticos informativos tendem a aumentar os retornos das ações e o volume de negócios, ao mesmo tempo que reduzem a volatilidade. De acordo com os estudos citados acima, é perceptível o impacto político no mercado acionário.

Dada a importância destes dados não estruturados e a grande quantidade de docu-



mentos textuais contida neles, compreende-se que o uso de técnicas de Natural Language Processing (NLP) pode oferecer insights sobre como fatores externos ao mercado influenciam a percepção dos investidores. Isso é viabilizado pela transformação de dados não estruturados em dados estruturados, permitindo análises mais precisas, o que possibilita a identificação de novas tendências e sinais reveladores. Um conjunto de ferramentas essenciais no âmbito do NLP para essa transformação são os algoritmos baseados em Deep Learning (DL), que têm apresentado avanços significativos nos últimos anos. Dentre esses algoritmos, destacam-se o GEMMA (Google DeepMind, 2025) e o LLAMA (Meta, 2025), ambos de código aberto e amplamente reconhecidos como LLMs (Large Language Models), devido à grande quantidade de parâmetros que os compõem, o que os torna modelos robustos e eficazes.

Diante do exposto, o objetivo deste projeto é analisar o impacto dos discursos e falas dos últimos presidentes brasileiros sobre o mercado acionário do país. Utilizando técnicas de NLP e com o uso de LLM, este estudo busca mensurar como as declarações presidenciais influenciam a volatilidade e o comportamento dos investidores. O objetivo, portanto, é analisar e mensurar a influência da comunicação presidencial nas movimentações do mercado, contribuindo para uma revisão aprofundada da relação entre o ambiente político e o financeiro no Brasil.

2. Trabalhos Relacionados

A compreensão da relação entre eventos políticos e o mercado financeiro tem sido objeto de estudo em diversas áreas, com a crescente disponibilidade de dados não estruturados impulsionando novas metodologias de análise. Esta seção revisa a literatura relevante em duas frentes principais: primeiro, o impacto da comunicação política na dinâmica do mercado de capitais e, segundo, o uso de NLP para analisar o sentimento de mercado. A revisão busca identificar as lacunas existentes na pesquisa que justificam a abordagem e os objetivos deste trabalho.

a. Impacto da Retórica Política no Mercado Financeiro

A influência de fatores políticos no mercado de ações é um tema bem documentado. Eventos de grande escala, como a vitória de Donald Trump na eleição presidencial dos EUA em 2017, demonstraram a capacidade de pronunciamentos políticos gerarem volatilidade e incerteza nos mercados globais (Wagner et al., 2017). Em um contexto mais granular, a análise do conteúdo dos discursos de figuras políticas relevantes, tem se mostrado uma boa ferramenta para compreender a forma como a informação política é absorvida pelo mercado.

A pesquisa de Maligkris (2017) representa um marco nesta área, ao demonstrar que discursos de candidatos à presidência dos EUA influenciaram retornos, volatilidade e volume de negociações no mercado acionário, resultando em um aumento médio de 26 pontos-base nos retornos excedentes do mercado. Os autores concluíram que a magnitude desse efeito dependia do conteúdo dos discursos, com informações de cunho econômico tendo um impacto particularmente relevante. Adicionalmente, o tom linguístico negativo dos discursos também teve um efeito perceptível, com um aumento de 1% no tom negativo



diminuindo os retornos excedentes em 1 ponto-base. Apesar da importância deste estudo, sua análise se concentra no cenário político e econômico norte-americano, deixando uma lacuna sobre como esses efeitos se manifestam em mercados emergentes, como o Brasil, que possuem características e dinâmicas políticas distintas.

Além disso, a literatura demonstra que o sentimento do investidor é um fator crucial nas movimentações de mercado (Baker and Wurgler, 2007), e que eventos negativos, em particular, podem ter um impacto desproporcional. Estudos como o de Baek et al. (2020) sobre a reação do mercado à pandemia de COVID-19 corroboram essa ideia, apontando que notícias com sentimento negativo tendem a gerar maior volatilidade do que notícias com sentimento positivo. Essa "assimetria de reação" sugere que os discursos presidenciais com tom negativo podem ter um efeito mais pronunciado no comportamento do mercado, um ponto que o presente trabalho busca investigar em profundidade no contexto brasileiro.

b. Aplicações de Machine Learning e NLP na Análise de Mercado

O avanço do ML tem transformado a análise financeira, oferecendo ferramentas para processar grandes volumes de dados de forma mais eficiente do que métodos tradicionais (Lakhchini et al., 2022). Um aspecto central dessa evolução é a capacidade de processar e extrair significado de dados não estruturados, como textos. Conforme destaca Tatsat et al. (2020), notícias de teor macroeconômico, político ou corporativo influenciam fortemente o mercado, e a crescente geração de dados textuais torna a análise assistida por NLP crucial para obter uma visão completa sobre o sentimento e a percepção dos investidores.

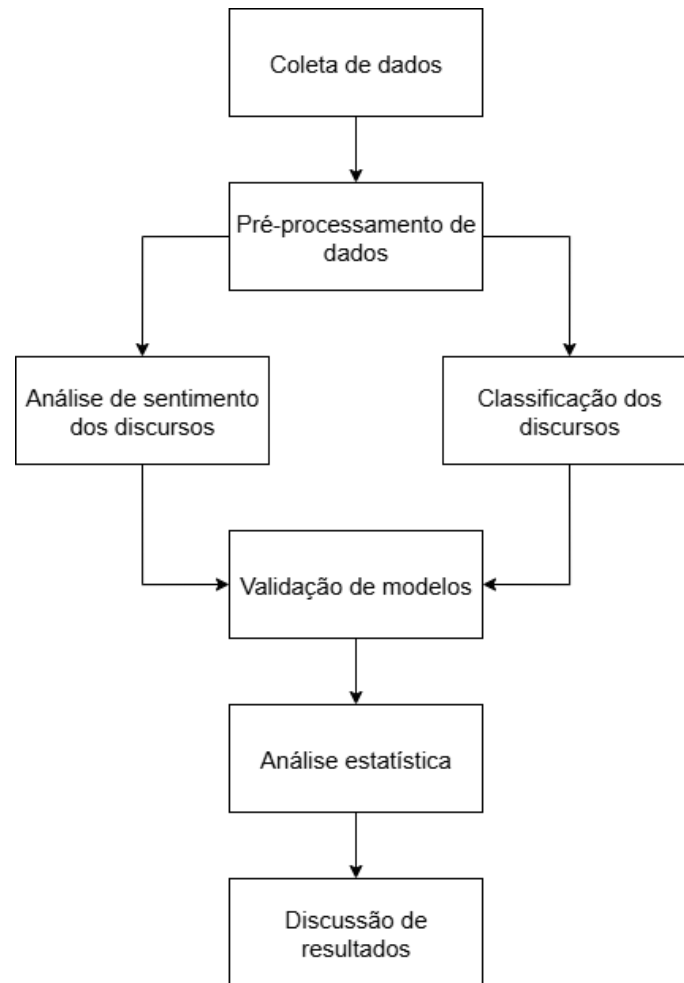
Neste contexto, o NLP emergiu como um campo fundamental. Uma de suas principais aplicações é a análise de sentimento, que, segundo Tatsat et al. (2020), analisa fontes como notícias financeiras e mídias sociais para prever tendências e capturar a percepção do mercado, replicando e aprimorando a capacidade humana de interpretar a atividade financeira. Trabalhos recentes têm se beneficiado do uso de modelos avançados de Deep Learning, como os LLMs, que demonstraram uma precisão excepcional em tarefas de classificação e compreensão textual. Essa alta performance é evidenciada em benchmarks como o MMLU, que avalia a capacidade de resolução de problemas em dezenas de assuntos, onde modelos de ponta consistentemente alcançam pontuações superiores a 80% (Naveed, 2025).

3. Metodologia

A metodologia deste trabalho foi estruturada em etapas sequenciais que garantiram a coleta, o processamento e a análise dos discursos presidenciais de forma eficiente, para que fosse possível uma análise mais detalhada da correlação desses dados com os do mercado acionário. Essas etapas foram organizadas em um processo integrado, como ilustrado na Figura 1, que apresenta o passo-a-passo da abordagem adotada.



Figura 1. Processos da Metodologia do Projeto, Brasil, 2025



a. Coleta de Dados

Com o objetivo de analisar padrões e implementar sistemas capazes de verificar possíveis correlações entre declarações feitas por figuras presidenciais e as variações do mercado financeiro, foi necessário realizar a coleta das transcrições desses discursos, disponibilizados pela Biblioteca Presidencial (Brasil, 2025).

Para a extração dos dados, utilizou-se o framework Selenium, desenvolvido em Python, a fim de realizar o web scraping das transcrições presidenciais. Essa abordagem foi necessária devido ao fato de que os discursos estão disponíveis apenas no HTML do site oficial da Biblioteca Presidencial, sem a opção de download direto. Os dados coletados abrangem o período de 01/01/2011, início do mandato de Dilma Rousseff, até 31/12/2022, término do mandato de Jair Bolsonaro. A quantidade de discursos para cada presidente pode ser verificada na Tabela 1.

Para a coleta dos dados de mercado, utilizou-se a biblioteca Python yFinance (Aroussi, 2025), uma ferramenta de código aberto que permite o acesso a informações públicas disponibilizadas pelo Yahoo Finance (Yahoo, 2025). Foram obtidos, para o mesmo período de análise, dados referentes aos valores de retorno e ao volume de nego-



Tabela 1. Quantidade de discursos presidenciais coletados por presidentes, Brasil, 2025

Presidente	Discursos Presidenciais
Dilma Rousseff	958
Michel Temer	408
Jair Bolsonaro	611

ciação do índice Ibovespa, bem como dos 25 ativos com maior percentual de participação do índice (B3, 2025). Essa seleção visa proporcionar uma perspectiva mais precisa sobre o possível impacto dos discursos presidenciais.

b. Pré-Processamento de Dados

A etapa de pré-processamento dos dados é fundamental para facilitar o manuseio e a manipulação do conteúdo, garantindo uma análise e compreensão mais clara dos resultados. Nesse sentido, as datas dos discursos foram extraídas diretamente do texto, convertidas para o formato datetime e armazenadas em uma nova coluna de um DataFrame do Pandas, permitindo a organização temporal dos dados para análises subsequentes. De forma análoga, os dados de mercado foram processados, e para a variável de retorno, utilizou-se o logaritmo, uma medida com propriedades estatísticas favoráveis, como simetria, e amplamente empregada na literatura financeira (Hudson and Gregoriou, 2015), calculado como:

$$\ln \frac{preco_t}{preco_{t-1}}$$

Já no caso do volume, foi empregada a variação percentual do volume negociado, definida como:

$$\frac{volume_t - volume_{t-1}}{volume_{t-1}}$$

Essa escolha se justifica pois o volume é uma variável de natureza distinta, uma quantidade física negociada, que pode inclusive assumir valor zero, inviabilizando o uso de logaritmos. Além de que é necessário manter a assimetria da variável, levando em consideração que só é possível cair no máximo 100%, porém é possível subir mais que isso.

c. Classificação dos Discursos

Levando em consideração o estudo feito por Maligkris (2017), no qual discursos políticos foram utilizados para comparar com o movimento do mercado financeiro, especificamente o dos Estados Unidos, uma das métricas adotadas para a análise foi a quantificação de informações econômicas. Nesse estudo, as políticas públicas foram separadas em dez categorias específicas, sendo elas: *Gastos do governo*, *Impostos*, *Política fiscal*, *Política mon-*



etária, Segurança nacional, Programas de assistência social, Saúde, Regulação, Dívida soberana/crises atuais e Política comercial."

Para realizar a quantificação no presente trabalho, todos os discursos foram separados em sentenças com três ou mais palavras, e cada sentença foi classificada por meio de um LLM, mais especificamente o GEMMA 3 (Google DeepMind, 2025) e o LLAMA

3.2 (Meta, 2025), com o intuito de verificar o que apresenta uma melhor acurácia. Foi utilizado um prompt que orientava o modelo a classificar cada sentença em uma das categorias de políticas públicas mencionadas ou, caso não se enquadrasse em nenhuma, a classificá-la como nula. Os resultados da classificação foram armazenados em um banco de dados SQLite para posterior análise. Após a catalogação de todas as sentenças de um discurso, realizou-se a contagem das categorias identificadas e o retorno dos resultados em forma de percentuais.

A escolha por utilizar um LLM foi motivada por sua robustez e precisão, características que permitem uma análise textual profunda e de alta qualidade. Esses modelos têm demonstrado um desempenho excepcional em tarefas de leitura e compreensão de texto, sendo capazes de classificar sentenças com alta confiabilidade (Naveed, 2025). Adicionalmente, a velocidade de processamento de um LLM permite analisar um volume de dados em uma escala que seria logisticamente inviável por meio de uma análise manual.

Essa abordagem não só permite uma análise mais granular dos discursos políticos, mas também oferece uma perspectiva quantitativa que facilita a comparação com indicadores econômicos, potencialmente revelando padrões e correlações relevantes.

d. Análise de Sentimento dos Discursos

Com base no estudo de Maligkris (2017), foi realizada uma análise de sentimento nos discursos transcritos de ex-presidentes do país, com o objetivo de identificar o tom linguístico predominante e utilizá-lo como parâmetro adicional no estudo de correlação com o mercado acionário. A análise de sentimento foi conduzida utilizando o modelo Distilbert base multilingual cased sentiments student (Yuan, 2025), disponibilizado na plataforma Hugging Face Transformers, garantindo uma abordagem moderna e eficiente para o processamento dos textos.

Os discursos foram divididos em blocos (chunks) de até 512 tokens, respeitando a limitação do modelo. Cada chunk foi analisado individualmente, gerando resultados que indicam as probabilidades de o texto ser classificado como "positivo", "negativo" ou "neutro". Após a análise de todos os chunks de um discurso, os resultados foram agregados pelo cálculo da média dos valores de sentimento. Essas abordagens permitem uma visão consolidada do tom geral de cada discurso, minimizando a influência de possíveis outliers em chunks específicos.

e. Validação de Modelos

A validação dos modelos foi executada para assegurar a confiabilidade e a precisão dos resultados. Nesta análise, avaliou-se a capacidade do modelo em explicar o fenômeno



estudado e a consistência de suas previsões, utilizando métricas de acurácia e F1-score.

i. Classificador de Discursos

Com o intuito de verificar a precisão do classificador de discursos empregado, foram conduzidos testes estatísticos utilizando uma base de dados pré-classificada e balanceada. Foi selecionada uma amostra de 100 textos, distribuídos equitativamente, de modo que cada porção contivesse textos com foco em uma das dez categorias de políticas públicas escolhidas.

Dessa forma, tornou-se viável comparar os resultados obtidos com o modelo utilizado neste projeto. Após processar esses textos no classificador, a categoria prevista para cada texto foi comparada com a categoria real, permitindo a verificação das métricas de desempenho. O modelo LLAMA 3.2, com cerca de 3 bilhões de parâmetros, atingiu 45% de acurácia e um F1-score de 0,2322. Já o modelo GEMMA 3, com aproximadamente 4 bilhões de parâmetros, apresentou um desempenho superior, alcançando 73% de acurácia e um F1-score de 0,6780, sendo assim, o modelo escolhido para o prosseguimento do trabalho. Modelos mais robustos poderiam resultar em acurácias ainda maiores, porém devido à limitações computacionais, este estudo se limitou a modelos de até 4B parâmetros.

ii. Análise de Sentimento dos Discursos

Para a validação do modelo de sentimento textual, empregou-se uma base de 100 textos pré-classificados e balanceados. Neste contexto, a análise concentrou-se no sentimento negativo, visto que esta é a variável a ser utilizada na regressão linear múltipla. Os textos que apresentaram mais de 50% de conteúdo negativo foram classificados como negativos, possibilitando assim a comparação com seu sentimento verdadeiro. O modelo Distilbert base multilingual cased sentiments student demonstrou alta eficácia, atingindo 97% de acurácia e um F1-score de 0,9690.

f. Regressão Linear Múltipla

Com o objetivo de realizar uma análise direcionada entre os discursos presidenciais e o comportamento do mercado financeiro, foi desenvolvido um sistema que executa, para cada presidente, no período de mandato, e para cada um dos 25 ativos selecionados, bem como para o índice Ibovespa, uma regressão linear múltipla.

Nessa modelagem, a variável dependente é representada pelo retorno ou pelo volume do ativo, enquanto as variáveis independentes correspondem às categorias temáticas e ao medidor de sentimento extraídos dos discursos presidenciais, no caso o nível do sentimento negativo, dessa forma assim, possibilitando capturar os efeitos positivos ou neutro, e evitando a multicolinearidade na regressão.

$$dependente_t = \sum_{i=1}^{10} \beta_i \cdot categoria_{i,t} + \beta_{11} \cdot negatividade_t + \varepsilon_t$$



Em que:

- $dependente_t$ representa a variável dependente da regressão (pode ser o retorno ou volume) no tempo t ;
- $categoria_{i,t}$ são variáveis com valores percentuais, para a presença de cada uma das dez categorias temáticas nos discursos presidenciais, a saber: gastos do governo, impostos, política fiscal, política monetária, segurança nacional, programas de assistência social, saúde, regulação, política comercial e dívida soberana/crises atuais;
- $negatividade_t$ é o valor do sentimento negativo no discurso;
- ε_t é o erro aleatório

Dessa forma, é possível observar se o conteúdo dos discursos tem algum possível efeito no preço ou na quantidade de ações negociadas, além disso, verificar as empresas mais afetadas e os temas mais relevantes.

4. Resultados

Foi conduzida uma análise de regressão linear com o objetivo de avaliar o impacto dos temas abordados em discursos presidenciais sobre os retornos de 25 ativos do mercado financeiro brasileiro, bem como do índice IBOVESPA. Os modelos foram estimados separadamente para os governos de Dilma Rousseff, Michel Temer e Jair Bolsonaro, além de um modelo consolidado, que agrega os dados de todos os períodos em uma única análise. Para a interpretação dos resultados, adotou-se um nível de significância de 5% (p-value

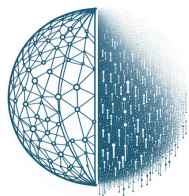
$\leq 0,05$), ou seja, apenas coeficientes estatisticamente significativos a 95% de confiança foram considerados relevantes na análise.

a. Diagnóstico e Robustez dos Modelos

Antes de analisar os coeficientes das regressões, foi realizada uma bateria de testes de diagnóstico para assegurar a validade estatística dos modelos estimados. A Tabela 2 apresenta o resumo dos testes de Multicolinearidade (VIF), Normalidade dos resíduos (Shapiro-Wilk) e Homocedasticidade (Breusch-Pagan), segmentados por presidente e variável dependente.

A análise dos diagnósticos revela três pontos fundamentais para a interpretação dos resultados:

- **Multicolinearidade Controlada:** O *Variance Inflation Factor* (VIF) manteve-se consistentemente baixo (máximo de 3,45), muito aquém do limite crítico de 5, indicando que as categorias temáticas extraídas via NLP são independentes entre si.
- **Normalidade e Robustez:** O teste de Shapiro-Wilk rejeitou a normalidade dos resíduos em todos os modelos, um comportamento esperado em séries financeiras de alta frequência ("caudas gordas"). Para mitigar isso, baseou-se no Teorema do Limite Central dada a robustez da amostra ($N > 100$).

**Tabela 2. Resumo dos Diagnósticos Estatísticos dos Modelos, Brasil, 2025**

Presidente	Target	VIF Máx.	% Homocedásticos	R ² Médio
Bolsonaro	Retorno	1,97	100,0%	0,40%
Bolsonaro	Volume	1,97	84,0%	0,78%
Dilma	Retorno	3,10	33,3%	1,20%
Dilma	Volume	3,10	100,0%	-0,27%
Temer	Retorno	3,45	40,0%	1,43%
Temer	Volume	3,41	88,0%	1,30%

Nota: % Homocedásticos refere-se à taxa de aprovação no teste de Breusch-Pagan ($p > 0.05$). O teste de Shapiro-Wilk rejeitou a normalidade em 100% dos casos.

- **Heterocedasticidade e Correlação:** Observou-se uma distinção clara na estrutura de risco. Os modelos de retorno dos governos Dilma e Temer apresentaram alta heterocedasticidade (apenas 33,3% e 40% de estabilidade, respectivamente), indicando que discursos nesses períodos geravam volatilidade instável. Em contraste, o período Bolsonaro mostrou-se estatisticamente mais estável (100% de homocedasticidade no retorno). Devido a essas discrepâncias, todas as regressões foram estimadas utilizando matrizes de covariância robustas (HAC/HC3), garantindo a validade dos p-valores apresentados a seguir.

b. Resultados Significativos com Retorno Sendo a Variável Dependente

De um total de 78 combinações de presidente-ativo, 6 modelos apresentaram significância estatística global (F p-value $< 0,05$). A análise consolidada, agrupando todos os períodos, não revelou nenhum modelo com significância estatística sob este critério.

Um ponto central observado em todos os modelos significativos foi o baixo poder explicativo, com valores de R^2 Ajustado consistentemente abaixo de 1%, indicando que as variáveis de discurso explicam uma porção mínima da variabilidade dos retornos dos ativos.

i. Análise por Período Presidencial

Para a adequada interpretação dos resultados estatisticamente significativos, considerando o retorno como variável dependente, é necessário observar cada período presidencial de forma individual. Essa abordagem permite identificar quais presidentes exerceram maior influência sobre o mercado, quais ativos foram mais impactados e quais categorias de discurso apresentaram maior correlação com as variações de retorno. Essas informações estão detalhadas na Tabela 2.

1. Governo Bolsonaro

Dos 6 modelos estatisticamente significante, o período do governo Bolsonaro contém 4, sendo eles:



- a. ITUB4 (F p-value = 0,005)
- b. ITSA4 (F p-value = 0,007)
- c. BBAS3 (F p-value = 0,014)
- d. BBDC4 (F p-value = 0,036)

Todos os quatro ativos pertencentes ao setor bancário. Em cada regressão, olhando para a Tabela 2, é perceptível que a variável “Saúde” apresenta um coeficiente negativo e significativo, causando um impacto contrário nos retornos de todos esses ativos. Da mesma maneira foi a variável “Dívida soberana e crises atuais”.

2. *Governo Temer*

Durante o governo Temer, apenas a regressão para o ativo ABEV3 se mostrou significativa (p-value = 0,025). A análise dos coeficientes individuais revelou que menções a "Programas de assistência social" estavam associadas a um impacto positivo e significativo nos retornos do ativo.

3. *Governo Dilma*

Para o período do governo Dilma, somente um modelo se mostrou significativo também, para o ativo SBSP3 (p-value = 0,041). Os resultados indicam que os temas "Segurança nacional" (p-value = 0.018) e "Programas de assistência social" (p-value = 0.040) tiveram um impacto negativo nos retornos da companhia de saneamento. Adicionalmente, a variável de controle para o sentimento, demonstrou uma correlação positiva e altamente significativa com os retornos (p-value < 0.001). Na Tabela 2, é possível visualizar os coeficientes dessas variáveis.

Tabela 3. Variáveis estatisticamente significativas por Presidente, Ativo e Tema do Discurso

Presidente	Ativo	Tema do Discurso	Coeficiente	P-value
Bolsonaro	ITUB4	Saúde	-0,0005	0,005
Bolsonaro	ITUB4	Dívida Soberana e Crises Atuais	-0,0017	0,023
Bolsonaro	ITSA4	Saúde	-0,0005	0,003
Bolsonaro	ITSA4	Dívida Soberana e Crises Atuais	-0,0018	0,040
Bolsonaro	BBAS3	Saúde	-0,0007	0,008
Bolsonaro	BBAS3	Dívida Soberana e Crises Atuais	-0,0025	0,008
Bolsonaro	BBDC4	Saúde	-0,0005	0,033
Bolsonaro	BBDC4	Dívida Soberana e Crises Atuais	-0,0017	0,032
Temer	ABEV3	Programas de Assistência Social	0,0001	0,046
Dilma	SBSP3	Segurança Nacional	-0,0001	0,018
Dilma	SBSP3	Programas de Assistência Social	-0,0001	0,040

Em suma, os resultados indicam que o conteúdo dos discursos presidenciais possui um efeito estatisticamente detectável, porém limitado, sobre os retornos de ativos



específicos, com padrões que variam conforme o governo e o setor do ativo.

c. Resultados Significativos com Volume Sendo a Variável Dependente

Diferente dos resultados dos casos analisando o retorno, a principal conclusão aqui é que os discursos presidenciais, especialmente durante o governo Bolsonaro, tiveram um impacto significativo sobre o volume de negociações. Porém ainda mantendo um R^2 Ajustado médio baixo.

i. Análise por Período Presidencial

Para compreender os resultados significativos relacionados ao volume de negociação, é essencial analisar separadamente cada mandato presidencial. Essa análise possibilita identificar quais períodos apresentaram maior sensibilidade nas movimentações de mercado, além de destacar os ativos mais impactados e as categorias discursivas com maior influência sobre o volume. Os detalhes dessa análise estão apresentados na Tabela 3.

1. Governo Bolsonaro

Dos 26 modelos estimados, 25 apresentaram significância estatística, o que indica que, para a quase totalidade dos ativos analisados, houve correlação entre o conteúdo dos discursos do presidente Bolsonaro e o volume de negociação no mercado financeiro.

As principais variáveis independentes desses casos foram:

- a. Política Monetária: Causou um impacto significativo e negativo nos ativos RAIL3, BRFS3 e SUZB3.
- b. Gastos do Governo: Também apresentou impacto significativo e negativo, mas nos ativos LREN3, BBSE3 e RENT3.
- c. Sentimento Negativo: Esta é a variável mais influente. Ela possui um coeficiente positivo e significativo em pelo menos 11 dos modelos analisados, é possível ver mais afundo na Figura 2, onde também é apresentado o número do coeficiente de cada caso.

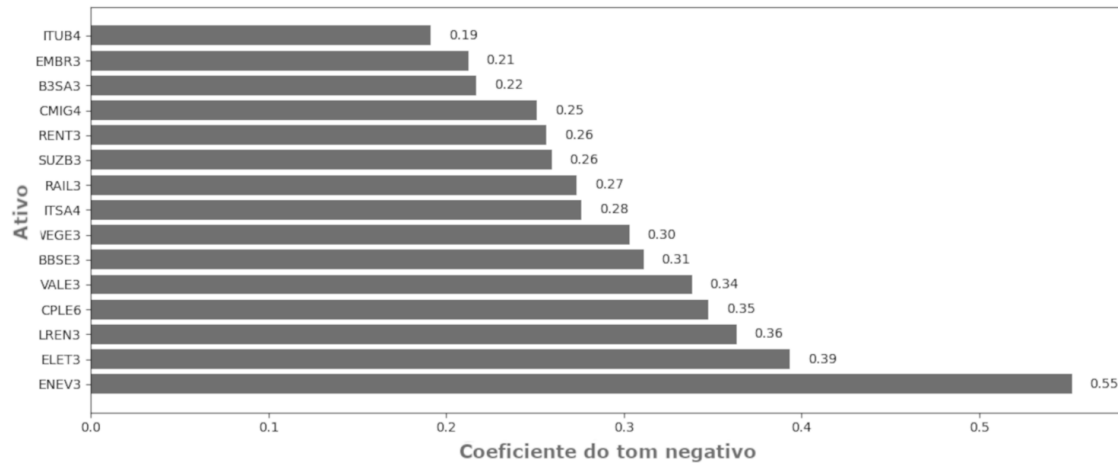
É possível analisar então, que discursos com tom mais negativo estavam associados a um aumento no volume de negociações, se mostrando mais impactante principalmente com o ativo ENEV3.

2. Governo Temer

O modelo com o ativo ENEV3 foi o único a apresentar significância, com a variável de Tom Negativo exibindo um coeficiente positivo e significativo ($p\text{-value} = 0,021$). Isso indica que discursos negativos do presidente Temer estão correlacionados a um aumento no volume de transações para o ativo ENEV3.



Figura 2. Impacto do Tom Negativo no Volume de Negociação (Gov. Bolsonaro), Brasil, 2025



3. Governo Dilma

Apenas o índice BVSP teve significância no modelo, com o tema "impostos" impactando negativamente o volume do Ibovespa ($p\text{-value} = 0,006$), sugerindo que discussões sobre impostos estão ligadas a menor atividade no mercado.

4. Análise Consolidada

No caso da análise consolidada que analisa todos os períodos juntos, três modelos se mostraram estatisticamente significativos, sendo eles: RAIL3, SUZB3 E BVSP. No modelo SUZB3, o tema "segurança nacional" teve impacto negativo no volume, enquanto a variável de Tom Negativo, novamente mostrou um impacto positivo ($p=0.030$).

Tabela 4. Variáveis estatisticamente significativas nos casos de Volume, Brasil, 2025

Presidente	Ativo	Tema do Discurso	Coefficiente	P-valor
Presidente	Ativo	Tema do Discurso	Coefficiente	P-valor
Bolsonaro	RAIL3	Política Monetária	-0,0476	0,014
Bolsonaro	RAIL3	Regulação	0,0417	0,025
Bolsonaro	RAIL3	Sentimento Negativo	0,2734	0,015
Bolsonaro	SBSP3	Segurança Nacional	0,006	0,006
Bolsonaro	LREN3	Gastos do Governo	-0,0172	0,002
Bolsonaro	LREN3	Sentimento Negativo	0,3633	0
Bolsonaro	WEGE3	Sentimento Negativo	0,3031	0
Bolsonaro	TOTS3	Saúde	-0,0142	0,049
Bolsonaro	TOTS3	Dívida Soberana e Crises Atuais	-0,0452	0,027

Continua na próxima página

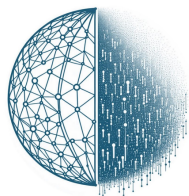


Tabela 4 – Continuação

Presidente	Ativo	Tema do Discurso	Coefficiente	P-valor
Bolsonaro	BBDC4	Dívida Soberana e Crises Atuais	0,0354	0,049
Bolsonaro	GGBR4	Dívida Soberana e Crises Atuais	0,0429	0,034
Bolsonaro	BBSE3	Gastos do Governo	-0,0122	0,027
Bolsonaro	BBSE3	Sentimento Negativo	0,3111	0,002
Bolsonaro	RENT3	Gastos do Governo	-0,0145	0,043
Bolsonaro	RENT3	Sentimento Negativo	0,2566	0,021
Bolsonaro	B3SA3	Sentimento Negativo	0,217	0,01
Bolsonaro	ITSA4	Regulação	-0,0165	0,041
Bolsonaro	ITSA4	Política Comercial	0,0081	0,003
Bolsonaro	ITSA4	Sentimento Negativo	0,2762	0,016
Bolsonaro	CMIG4	Sentimento Negativo	0,2513	0
Bolsonaro	BRFS3	Política Monetária	-0,0392	0,001
Bolsonaro	ELET3	Sentimento Negativo	0,3933	0,028
Bolsonaro	SUZH3	Política Monetária	-0,0301	0,034
Bolsonaro	SUZH3	Sentimento Negativo	0,2594	0,006
Bolsonaro	ENEV3	Sentimento Negativo	0,5521	0,024
Bolsonaro	ITUB4	Sentimento Negativo	0,1916	0,039
Bolsonaro	VALE3	Sentimento Negativo	0,3383	0,005
Bolsonaro	CPL6	Sentimento Negativo	0,3476	0,011
Bolsonaro	EMBR3	Sentimento Negativo	0,2124	0,049
Dilma	BVSP	Impostos	-0,0198	0,006
Temer	ENEV3	Sentimento Negativo	8,9956	0,021
Todos	SUZH3	Segurança Nacional	-0,006	0,04
Todos	SUZH3	Sentimento Negativo	0,7039	0,03

5 Discussão

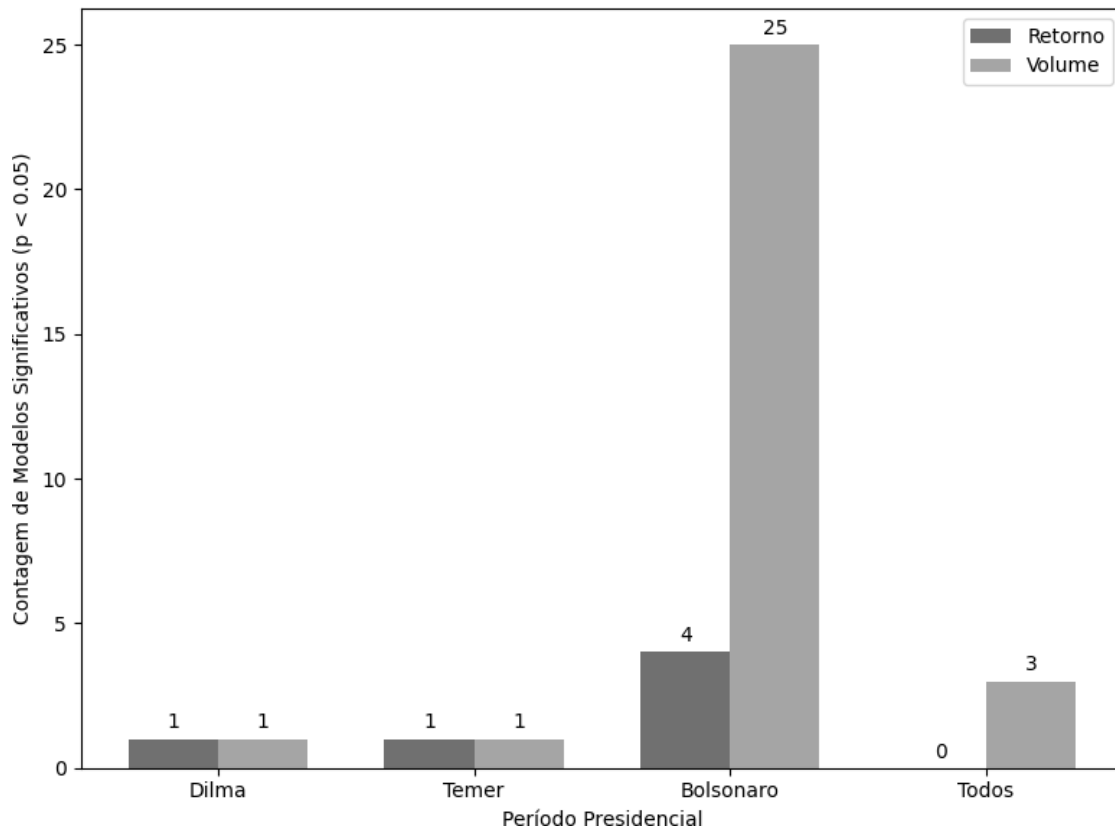
Apesar da alta significância estatística observada nos modelos, indicada pelos baixos valores de p , que demonstram que há uma relação estatisticamente confiável entre as variáveis independentes e o volume, o R^2 Ajustado permanece baixo, geralmente entre 1% e 4%. Isso indica que, embora os efeitos sejam estatisticamente significativos, a capacidade explicativa dos modelos é limitada. Em outras palavras, as variáveis relacionadas ao conteúdo dos discursos explicam apenas uma pequena fração da variação no volume de negociação, sugerindo que outros fatores não incluídos nos modelos exercem influência mais relevante. Esse resultado, no entanto, é esperado, uma vez que o volume de negociação tende a ser fortemente impactado por variáveis macroeconômicas que vão além do conteúdo dos discursos presidenciais.

Porém, um achado interessante pode ser verificado na comparação entre Retorno e Volume. Os resultados indicam que os discursos presidenciais não geraram informação suficiente para formar um consenso de mercado que movesse os preços. Contudo, eles criaram ruído e incerteza suficientes que podem impulsionar a negociação, o que, por sua vez, aumentou o volume. Enquanto o impacto dos discursos presidenciais nos retornos dos ativos foi limitado e pontual, o efeito sobre o volume de transações foi mais significativo,



particularmente no governo Bolsonaro, como ilustrado na Figura 3.

Figura 3. Modelos Significativos por Presidente e Tipo de Variável dependente, Brasil, 2025



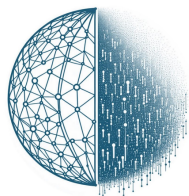
Este padrão sugere que os discursos funcionam menos como uma fonte de informação para direcionar preços e mais como um sinal de atenção e incerteza.

5.1 Contextualização do Período Bolsonaro

Um possível fator explicativo para esse resultado é o período do governo Bolsonaro, em que o Brasil foi marcado por um ambiente de elevada instabilidade, caracterizado pela crise sanitária e econômica provocada pela COVID-19. Esse cenário complexo e volátil contribui para um clima generalizado de apreensão nos mercados financeiros, que podem ter operado em estado de “alerta máximo”. Nesse contexto, qualquer pronunciamento do chefe de Estado teria potencial para gerar reações intensificadas.

5.2 Sentimento Negativo

No presente estudo, observou-se uma correlação entre o Sentimento Negativo presente nos discursos e a intensificação da atividade no mercado acionário. No entanto, discursos com tom mais negativo não implicam, necessariamente, um movimento coordenado de venda que leve à queda nos preços dos ativos. Em vez disso, esses pronunciamentos



parecem estar correlacionados a um aumento no volume de negociações, possivelmente em função da diversidade de interpretações por parte dos investidores e do consequente reposicionamento de carteiras diante de um ambiente percebido como mais incerto ou volátil. Resultado semelhante foi identificado por Baek et al. (2020), ao analisar a volatilidade do mercado norte-americano diante de notícias relacionadas à pandemia da COVID- 19: tanto informações positivas quanto negativas geram impacto, mas as negativas apresentaram efeito mais pronunciado, evidenciando a presença de um viés de negatividade (negativity bias).

5.3 *Heterogeneidade de Expectativas e Especulação*

O aumento do volume na ausência de variação de preços é um indicativo clássico de *Divergência de Opiniões* (Divergence of Opinion). Quando um presidente realiza um pronunciamento polêmico ou ambíguo, diferentes agentes de mercado interpretam a fala de maneiras opostas.

Se um grupo de investidores interpreta um discurso negativo como um sinal de venda (pessimismo) e outro grupo o interpreta como uma oportunidade de compra (especulação de que o mercado exagerou), ocorre uma intensa "troca de mãos" dos ativos. Esse fenômeno infla o volume de transações, mas, como as pressões de compra e venda se anulam, o preço final do ativo permanece relativamente estável. O forte impacto observado no governo Bolsonaro, correlacionado ao sentimento negativo, reforça a hipótese de que a polarização política se traduziu em polarização de expectativas no mercado, fomentando comportamentos especulativos.

5.4 *Correlação e Causalidade*

É importante destacar que os resultados deste estudo indicam uma correlação significativa entre os discursos presidenciais e a variação na atividade do mercado, porém não permitem inferir uma relação de causalidade direta. Ou seja, embora exista associação estatística, não se pode afirmar que o discurso negativo tenha sido a causa primária das movimentações observadas. É plausível que, em dias marcados por grande expectativa em torno de um pronunciamento presidencial, outros fatores já estivessem contribuindo para uma mudança no mercado, refletindo um ambiente de incerteza.

6 Conclusão

A conclusão central deste estudo é que os discursos presidenciais no Brasil impactam significativamente o volume de negociações, mas exercem efeito limitado sobre os preços dos ativos. A retórica política, especialmente em períodos de instabilidade como no governo Bolsonaro, atua gerando ruído e incerteza. Notavelmente, o sentimento negativo não precipitou quedas de preços, mas sim um aumento na atividade de mercado, sugerindo um reposicionamento estratégico de carteiras por parte dos investidores diante da divergência de expectativas.

Contudo, é necessário destacar as limitações e desafios enfrentados. A restrição de recursos computacionais limitou a aplicação a modelos de linguagem de até 4 bilhões de parâmetros, o que pode ter restringido a capacidade de capturar nuances mais sutis e complexas do discurso político brasileiro. Além disso, os baixos coeficientes de determinação (R^2) nos modelos de regressão reforçam a dificuldade de isolar o efeito da



comunicação em um mercado regido majoritariamente por variáveis macroeconômicas e fluxos globais de capital.

Para trabalhos futuros, recomenda-se a utilização de LLMs de maior escala e capacidade de processamento para refinar a classificação temática e de sentimento. Sugere-se também a exploração de dados de alta frequência (intraday) para capturar a reação imediata do mercado nos minutos seguintes aos pronunciamentos, bem como a inclusão de variáveis macroeconômicas de controle mais robustas para fortalecer o poder explicativo dos modelos estatísticos.

Referências

1. Ran Aroussi. yfinance, 2025. Versão 0.2.65. Disponível em: <https://github.com/ranaroussi/yfinance>. Acesso em: 21 jul. 2025.
2. B3. Índice ibovespa - ibovespa - composição da carteira, 2025. Disponível em: https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/indice-ibovespa-ibovespa-composicao-da-carteira.htm. Acesso em: 21 jul. 2025.
3. S. Baek, S. K. Mohanty, and M. Glambosky. Covid-19 and stock market volatility: An industry level analysis. *Finance Research Letters*, 37:101748, nov 2020.
4. M. Baker and J. Wurgler. Investor sentiment in the stock market. Technical report, National Bureau of Economic Research, Cambridge, MA, jun 2007. Disponível em: <https://doi.org/10.3386/w13189>. Acesso em: 28 maio 2024.
5. Brasil. Biblioteca presidência da república, 2025. Disponível em: <https://www.biblioteca.presidencia.gov.br/>. Acesso em: 21 jul. 2025.



6. Google DeepMind. Gemma, 2025. Disponível em: <https://deepmind.google/models/gemma/?hl=pt-br>. Acesso em: 21 jul. 2025.
7. R. S. Hudson and A. Gregoriou. Calculating and comparing security returns is harder than you think: A comparison between logarithmic and simple returns. *International Review of Financial Analysis*, 38:151–162, mar 2015.
8. Yuliia Kniazieva. Machine learning in the finance industry: is it worth taking the risk?, 2023. Disponível em: <https://labelyourdata.com/articles/machine-learning-in-finance>. Acesso em: 30 maio 2024.
9. Wassima Lakhchini, Rachid Wahabi, and Mounime El Kabbouri. Artificial intelligence & machine learning in finance: a literature review. *International Journal of Accounting, Finance, Auditing, Management and Economics*, 2022.
10. A. Maligkris. Presidential candidates, political speeches and stock market re- turns, 2017. Disponível em: <https://ssrn.com/abstract=2980475>. DOI: 10.2139/ssrn.2980475.
11. Meta.Llama 3, 2025. Disponível em: <https://www.llama.com/models/llama-3/>. Acesso em: 21 jul. 2025.
12. H. et al. Naveed. A comprehensive overview of large language models, jun 2025.
13. D. Shah, H. Isah, and F. Zulkernine. Stock market analysis: A review and taxonomy of prediction techniques. *International Journal of Financial Studies*, 7(2):26, maio 2019.
14. M. G. et al. Sousa. Bert for stock market sentiment analysis. In *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, nov 2019. Disponível em: <https://doi.org/10.1109/ictai.2019.00231>. Acesso em: 30 maio 2024.
15. Hariom Tatsat, Sahil Puri, and Brad Lookabaugh. *Machine learning and data science blueprints for finance*. O'Reilly Media, 2020.
16. A. Wagner, R. Zeckhauser, and A. Ziegler. Company stock reactions to the 2016 election shock: Trump, taxes and trade. Technical report, National Bureau of Economic Research, Cambridge, MA, fev 2017. Disponível em: <https://doi.org/10.3386/w23152>. Acesso em: 28 jul. 2025.
17. Yahoo. Yahoo finanças, 2025. Disponível em: <https://finance.yahoo.com/>. Acesso em: 21 jul. 2025.
18. L. X. Yuan. distilbert base multilingual cased sentiments student · hugging face, 2025. Disponível em: <https://huggingface.co/lxyuan/distilbert-base-multilingual-cased-sentiments-student>. Acesso em: 21 jul. 2025.