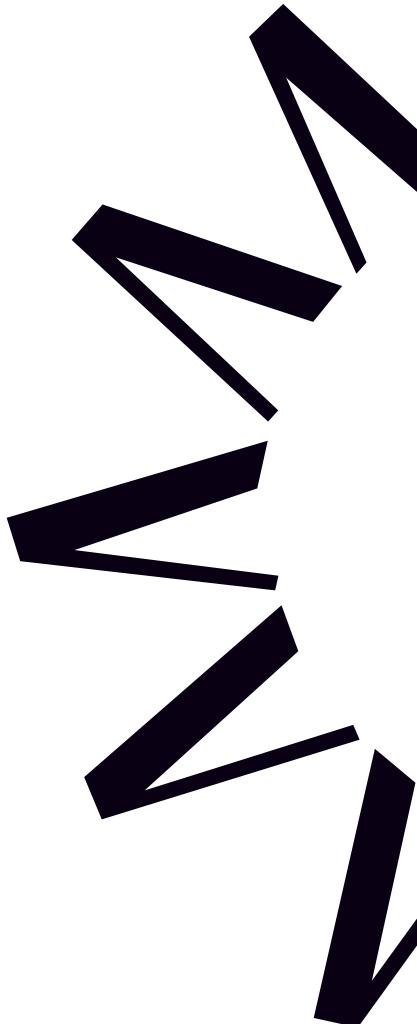

Atas do Banco Central mantêm mesmo padrão entre governos: Uma análise de texto com machine learning



Geraldo Leite

Cientista de dados, Doutorando em Ciências Cognitivas e Linguagem pela Universidade Autônoma de Barcelona

Wagner Vargas

Cientista de Dados, MBA Finanças Públicas, Mestre Gestão e Políticas Públicas

I. Introdução

O presente texto descreve alguns resultados de uma abordagem sistematizada para o pré-processamento de dados textuais, especificamente aplicada às atas do Banco Central do Brasil (Bacen), documento responsável por comunicar o mercado e a sociedade das decisões do Comitê de política Monetária sobre o patamar a ser adotado na SELIC, ou taxa básica de juros.

A justificativa para realizar a análise lexical está fundamentada no fato de que não apenas a decisão em si, mas também a forma que ela é comunicada pode impactar as decisões e as expectativas dos agentes de mercado, uma vez que as atas das reuniões do Bacen, por exemplo, são documentos ricos em informações sobre as decisões econômicas tomadas e os motivos por trás delas.

Com base nestes pressupostos, o artigo em tela, vale-se das técnicas de pré-processamento de texto e de Aprendizado de Máquina (Ramo da Inteligência Artificial) para buscar evidências de mudanças lexicais nos textos das atas entre abril de 1999 e maio de 2023 e se as mesmas são significativas a ponto de justificarem a hipótese ou alegação de que houve “endurecimento” no texto das atas no atual governo, primeiro período no qual a autoridade monetária desfruta de independência formal, com mandato fixo para presidente da instituição - cuja indicação ocorreu por um presidente da República de campo ideológico diferente do que pertence o atual presidente.



II. Metodologia

Em meio à era da informação digital em que texto, áudio, e imagem podem se transformar em dados estruturados, lidar e entender grandes conjuntos de dados, com variedade e velocidade, o tal do Big Data, torna-se tarefa cada vez mais presente na vida de quem atua com tecnologia. O Chat GPT, forjado neste contexto, foi treinado com mais de 400 bilhões de palavras com alguns tipos de aprendizado, entre eles, o por reforço, seguido de revisões humanas.

O setor bancário, onde a tomada de decisões informada é fundamental para a estabilidade econômica, a análise de grandes bases de dados assume um papel crucial e as técnicas de processamento de linguagem natural servem de subsídio na busca por objetividade na compreensão acerca de informações textuais sobre o tema.

Coleta dos dados

A coleta de dados das atas do Banco Central do Brasil, referentes ao período de 1999 a 2023 (até a ata 254), foi conduzida de forma automatizada, aproveitando bibliotecas de Python especializadas em web scraping, como BeautifulSoup e Scrapy.

O objetivo deste procedimento foi transmutar as informações contidas nos documentos PDF em um formato textual apto para análise e interpretação. Para atingir essa meta, recorreu-se à técnica denominada Optical Character Recognition (OCR), com o uso das bibliotecas PyTesseract e PDF2Text para realizar a conversão de imagens em texto, viabilizando a extração do conteúdo das atas. O algoritmo foi concebido para identificar precisamente os links para os arquivos PDF e realizar

o download de cada documento, englobando todas as atas publicadas no período de 1999 a 2023.

Sumarização

Na sumarização do texto, foi utilizada a abordagem de geração de resumos com o modelo Transformer T5. Essa abordagem utiliza um modelo pré-treinado T5 para gerar resumos dos textos originais. Em seguida, o texto é codificado em um formato comprehensível pela máquina. A saída gerada pelo modelo é então decodificada de volta em texto legível para humanos, proporcionando um resumo conciso do texto original. Este método gera resumos mais coerentes e informativos em comparação com as técnicas tradicionais de sumarização de texto, pois utiliza aprendizado profundo para entender o contexto e a semântica do texto original.

Vetorização

Os textos sob análise passam por um processo de vetorização, onde palavras mais recorrentes e relevantes são selecionadas, criando um tipo de índice para representar cada documento.

Modelagem

Logo após, é aplicado um método estatístico chamado Latent Dirichlet Allocation (LDA), que tem como finalidade identificar quais são os tópicos principais presentes nesses textos. Esses tópicos são agrupamentos de palavras que costumam aparecer juntas e que, portanto, provavelmente estão relacionadas a um mesmo tema.

III. Análise de Tópicos de texto (Topic Modeling)

No caso específico desse estudo, o LDA tem como output inicial 21 tópicos distintos, tendo como base o intervalo de cerca um ano para mudanças de conjuntura, partindo do pressuposto de que há mudanças econômicas até mesmo dentro um mesmo governo. De modo que o modelo realiza, em seguida, o agrupamento de quais atas teriam a probabilidade de estarem mais relacionadas entre si, ainda que façam parte de períodos ou contextos diversos, diminuindo assim o número de tópicos de texto relacionados.

Ao final desse processo, obtemos uma matriz de tópicos para cada documento. Essa matriz oferece uma visão resumida dos principais temas abordados em cada texto, tornando mais fácil e rápido o entendimento sobre o que se trata cada documento.

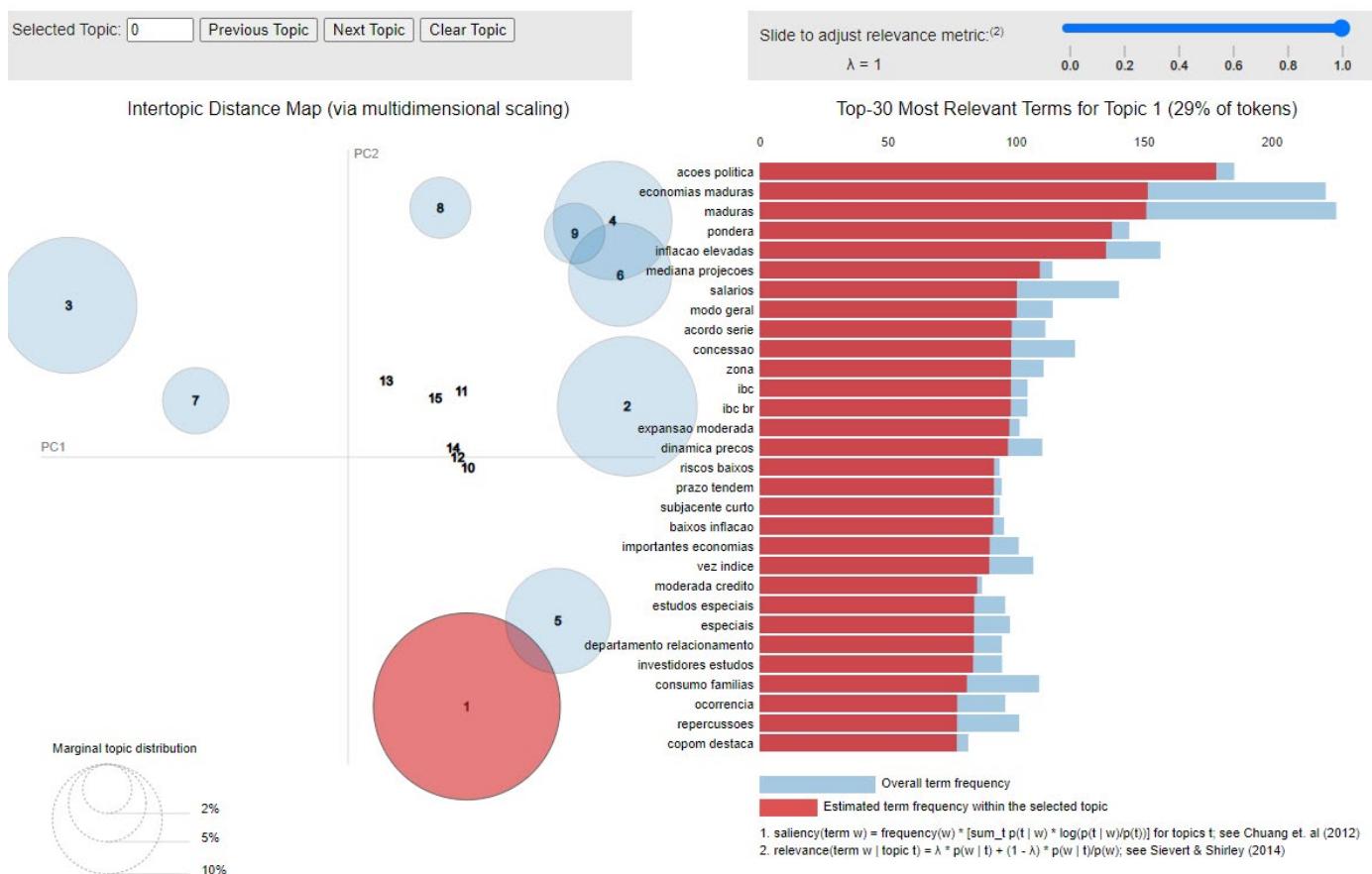
Tópicos associados inicialmente



Tópicos associados apenas com alta probabilidade

Neste caso, o modelo associa os mesmos 220 documentos (atas do Bacen) e agrupa aqueles temas que estariam mais próximos entre si. A quantidade de tópicos não busca uma exata divisão, o ponto focal é probabilidade de associação de ata do Banco Central a um determinado tópico detectado.

Com base nisso, realizamos a descrição dos principais tópicos por período de governo, considerando apenas as atas cuja probabilidade de pertencerem àquele determinado tema é essencialmente alta.



Tópico 3: Flexibilização Monetária e Desinflação

Neste tópico, são discutidas medidas de flexibilização monetária e os esforços para alcançar a desinflação. As reuniões destacadas parecem ter se concentrado em análises e propostas relacionadas a políticas monetárias que visam controlar a inflação e estimular o crescimento econômico. Isso sugere que o Comitê de Política Monetária (COPOM) considerou possíveis estratégias e concordou em adotar medidas de flexibilização para atingir suas metas.

Reunião	Probabilidade
205ª Reunião:	99,88%
204ª Reunião:	99,87%
203ª Reunião:	99,87%
201ª Reunião:	99,86%
200ª Reunião:	99,84%
202ª Reunião:	96,63%
206ª Reunião:	96,56%
207ª Reunião:	81,54%
208ª Reunião:	78,99%
209ª Reunião:	50,24%

Tópico 6: Reformas Econômicas e Discussões sobre Pesquisa Focus

Este tópico trata de reformas econômicas e está relacionado a discussões envolvendo a pesquisa Focus, pesquisa semanal que reúne projeções econômicas de várias instituições financeiras. As reuniões destacadas indicam que houve discussões abrangentes sobre reformas e atualizações na economia, com foco nas medidas necessárias para enfrentar os desafios decorrentes da pandemia e garantir a estabilidade e o crescimento sustentável.

Reunião	Probabilidade
252ª Reunião - 31 janeiro - 1 fevereiro, 2023:	99,92%
224ª Reunião:	99,92%
253ª Reunião - 21-22 março, 2023:	99,92%
225ª Reunião:	99,91%
251ª Reunião - 6-7 dezembro, 2022:	99,91%
223ª Reunião:	99,91%
222ª Reunião:	99,91%
221ª Reunião:	99,91%
226ª Reunião:	99,91%
254ª Reunião - 2-3 maio, 2023:	99,91%

Tópico 7: Notas de Reuniões Anteriores e Preços Monitorados

Esse tópico está relacionado a notas e análises de reuniões anteriores, abordando principalmente a evolução dos preços monitorados pela instituição. Essas reuniões podem ter se concentrado na avaliação de indicadores de preços, como inflação e índices de preço, e em como esses indicadores afetam as decisões e as perspectivas do COPOM para a política monetária. As notas das reuniões anteriores podem ter servido como base para discussões subsequentes.

Reunião	Probabilidade
134ª Reunião:	99,91%
133ª Reunião:	99,91%
132ª Reunião:	99,91%
130ª Reunião:	99,91%
108ª Reunião:	99,91%
107ª Reunião:	99,91%
129ª Reunião:	99,90%
127ª Reunião:	99,90%
106ª Reunião:	99,90%
105ª Reunião:	99,90%

Tópico 8: Critério, Condições Financeiras e Economias Maduras

Neste tópico, são abordados critérios, como condições financeiras, que podem afetar as economias maduras. As reuniões destacadas parecem ter enfatizado a importância de avaliar e monitorar de perto as condições financeiras e a confiança dos consumidores nas economias maduras. Isso indica uma abordagem prospectiva na tomada de decisões, com foco na análise de riscos e incertezas associados às economias maduras.

Reunião	Probabilidade
144ª Reunião:	99,94%
145ª Reunião:	99,94%
142ª Reunião:	99,94%
143ª Reunião:	99,94%
146ª Reunião:	99,94%
148ª Reunião:	99,94%
147ª Reunião:	99,94%
149ª Reunião:	99,94%
150ª Reunião:	99,94%
151ª Reunião:	99,93%

Tópico 10: Fundos, Desvalorização Cambial e Mercado Cambial

Este tópico trata de fundos, desvalorização cambial e do mercado cambial em geral. As reuniões destacadas sugerem que houve discussões sobre a liquidez de fundos, desvalorização da moeda nacional e as dinâmicas do mercado cambial. Isso pode refletir o monitoramento das flutuações cambiais e seu impacto na economia, bem como medidas para lidar com possíveis desequilíbrios.

Reunião	Probabilidade
21ª Reunião:	99,92%
30ª Reunião:	99,91%
24ª Reunião:	99,91%
23ª Reunião:	99,91%
26ª Reunião:	99,91%
29ª Reunião:	99,91%
27ª Reunião:	99,91%
25ª Reunião:	99,91%
33ª Reunião:	99,91%
31ª Reunião:	99,91%

Tópico 12: Leilões do Tesouro e Vencimento de Títulos Públicos

Neste tópico, são discutidos leilões do Tesouro Nacional e o vencimento de títulos públicos. As reuniões destacadas parecem ter se concentrado em questões relacionadas a leilões de títulos, estratégias de captação de recursos e gerenciamento da dívida pública. Isso sugere uma abordagem voltada para a eficiência e a eficácia das operações do Tesouro e a importância de equilibrar as necessidades de financiamento do governo.

Reunião	Probabilidade
77ª Reunião:	99,90%
78ª Reunião:	99,90%
60ª Reunião:	99,90%
74ª Reunião:	99,90%
69ª Reunião:	99,90%
64ª Reunião:	99,90%
71ª Reunião:	99,90%
63ª Reunião:	99,90%
79ª Reunião:	99,90%
62ª Reunião:	99,90%

Tópico 13: Ações Políticas e Inflação Elevada

Esse tópico aborda ações políticas para combater a inflação elevada, com ênfase em análises relacionadas às projeções de inflação e a série de medidas adotadas. As reuniões destacadas indicam uma preocupação particular com o controle da inflação e o impacto de medidas econômicas nas projeções de médio e longo prazo. As discussões também podem ter abordado o papel dos salários e o impacto geral dessas medidas na economia.

Reunião	Probabilidade
193 ^a Reunião:	99,95%
191 ^a Reunião:	99,95%
195 ^a Reunião:	99,95%
192 ^a Reunião:	99,95%
196 ^a Reunião:	99,95%
194 ^a Reunião:	99,95%
190 ^a Reunião:	99,95%
169 ^a Reunião:	99,95%
187 ^a Reunião:	99,95%
171 ^a Reunião:	99,95%

Tópico 14: Agregados, Manufaturados e Políticas Direcionadas

Nesse tópico, são discutidos aspectos relacionados aos agregados econômicos, como produção industrial e manufaturados, e as diretrizes de políticas para esses setores. As reuniões destacadas podem ter se concentrado em estratégias para impulsionar a produção e os investimentos em setores específicos, bem como em questões relacionadas a preços relativos e resgates de títulos do Tesouro Nacional.

Reunião	Probabilidade
51 ^a Reunião:	99,90%
52 ^a Reunião:	99,90%
50 ^a Reunião:	97,35%
53 ^a Reunião:	87,61%
54 ^a Reunião:	84,22%
48 ^a Reunião:	82,34%
55 ^a Reunião:	60,94%
47 ^a Reunião:	54,94%
49 ^a Reunião:	45,10%
56 ^a Reunião:	39,96%

Tópico 15: Tesouro Nacional e Oferta Agregada

Este tópico trata do papel do Tesouro Nacional e das diretrizes de política relacionadas à oferta agregada. As reuniões destacadas sugerem que houve discussões sobre as metas de taxas definidas pelo COPOM e como o Tesouro Nacional pode contribuir para o equilíbrio entre oferta e demanda agregadas na economia. Essas discussões podem envolver análises de indicadores econômicos e o papel da política fiscal na estabilização econômica.

Reunião	Probabilidade
46 ^a Reunião:	99,90%
45 ^a Reunião:	99,88%
44 ^a Reunião:	99,87%
43 ^a Reunião:	99,87%
41 ^a Reunião:	99,87%
38 ^a Reunião:	99,86%
40 ^a Reunião:	99,85%
39 ^a Reunião (Extraordinária):	99,84%
42 ^a Reunião:	99,83%
37 ^a Reunião:	55,02%

IV. Conclusão

O presente estudo utilizou técnicas de processamento de linguagem natural e aprendizado de máquina para analisar as atas do Comitê de Política Monetária do Banco Central do Brasil, com o objetivo de identificar possíveis mudanças lexicais ao longo do tempo e verificar se houve um endurecimento no texto das atas durante o governo atual, cuja indicação ocorreu por um presidente de campo ideológico diferente.

Através da aplicação de métodos como OCR, sumarização, vetorização e modelagem com o algoritmo LDA, foram identificados os tópicos principais, presentes nos textos das atas ao longo do período analisado. Com base nessas análises,

podemos afirmar que não foi encontrado um endurecimento significativo no texto das atas durante o governo atual em comparação com governos anteriores.

Ao examinar os principais tópicos associados à alta probabilidade nas atas, verificou-se que os temas discutidos englobam questões como flexibilização monetária e desinflação, reformas econômicas, notas de reuniões anteriores e preços monitorados, critérios e condições financeiras, fundos e mercado cambial, leilões do Tesouro Nacional e vencimento de títulos públicos, ações políticas e inflação elevada, agregados econômicos e políticas direcionadas, e o papel do Tesouro Nacional e oferta agregada.

Esses resultados indicam que as discussões e decisões do Comitê de Política Monetária ao longo do governo atual não apresentaram um endurecimento significativo em relação aos governos anteriores. Portanto, a hipótese de que as atas do Banco Central teriam endurecido no atual governo, devido à diferença ideológica, não foi confirmada pelos resultados deste estudo.

É importante ressaltar que a análise realizada neste estudo se baseou exclusivamente nas características textuais das atas do Bacen. Outros fatores econômicos e políticos podem influenciar as decisões e a comunicação do Comitê de Política Monetária, e esses aspectos não foram considerados neste estudo.

Em resumo, a abordagem sistemática utilizada neste estudo, combinando técnicas de processamento de linguagem natural e aprendizado de máquina, permitiu uma análise detalhada das atas do Banco Central do Brasil. Os resultados indicam que as mudanças lexicais não sustentam a alegação de endurecimento no texto das atas durante o governo atual, fornecendo uma visão mais objetiva sobre as decisões e as comunicações do Comitê de Política Monetária ao longo do tempo.