

Como somos vistos? Análise da imagem organizacional pública utilizando ciência de dados¹

How are we seen? public organizational image analysis using data science

¿cómo somos vistos? análisis de imagen organizacional pública utilizando ciencia de los datos

Carolina Coelho da Silveira, Carla Bonato Marcolin e Carlos Henrique Rodrigues

<https://doi.org/10.36428/revistadacgu.v14i26.559>

Resumo: Entende-se por imagem organizacional as percepções públicas em torno de uma organização. Neste sentido, os veículos de imprensa têm grande influência, contudo existem desafios relacionados ao formato e volume dos dados destas fontes. Assim, o objetivo deste artigo é utilizar técnicas de Ciência de Dados para analisar a imagem organizacional de uma organização pública através da imprensa nacional, com foco em portais de notícias e jornais. Para validação, foi utilizado o caso da Agência Nacional de Energia Elétrica (ANEEL). As etapas metodológicas consistiram em definição, coleta, preparação e análise dos dados através de técnicas de Processamento de Linguagem Natural. Os principais resultados reforçam os indícios de alta relação entre a imagem da ANEEL e a imagem do Governo, e que existe uma disparidade entre a imagem passada pela manchete e pela notícia na íntegra. Ainda, observou-se a utilização de estratégias de comunicação (rotulagem, *agenda-setting*, *linkage* e *framing*) por parte dos veículos de imprensa. Para estudos futuros, sugere-se a utilização de outras fontes de dados e também a validação a partir de outros casos.

Palavras-chave: Imagem Organizacional Pública; Notícias; Ciência de Dados, NLP.

Abstract: Organizational image is understood as the public perceptions around an organization. In this sense, press vehicles have a great influence on the organizational image, but there are challenges when analyzing the organizational image through these sources. Therefore, the objective of this article is to use Data Science techniques to analyze the organizational image of a public organization through the national press, focusing on news portals and newspapers. For validation, the case of the Agência Nacional de Energia Elétrica (ANEEL) was used. The methodological steps consisted of defining, collecting, preparing and analyzing data, and Natural Language Processing techniques were used. The main results reinforce the evidence of a strong relationship between the image of ANEEL and the image of the Government, and that there is a disparity between the image conveyed by the headline and the news in full. Also, the use of communication strategies (labeling, *agenda-setting*, *linkage* and *framing*) by the press vehicles was observed. For future studies, it's suggested the use of other data sources and also the validation from other cases.

Key-words: Public Organizational Image; News; Data Science, NLP.

Resumén: La imagen organizacional se entiende como las percepciones del público en torno a una organización. En este sentido, los vehículos de prensa tienen una gran influencia en la imagen organizacional, pero existen desafíos al analizar la imagen organizacional a través de estas fuentes. Por ello, el objetivo de este artículo es utilizar

1. Artigo submetido em 18/07/2022 e aceito em 25/11/2022.

técnicas de Ciência de los Datos para analizar la imagen organizacional de una organización pública a través de la prensa nacional, centrándonos en portales de noticias y periódicos. Para la validación se utilizó el caso de la Agência Nacional de Energia Elétrica (ANEEL). Los pasos metodológicos consistieron en definir, recopilar, preparar y analizar datos, y se utilizaron técnicas de Procesamiento del Lenguaje Natural. Los principales resultados refuerzan la evidencia de una fuerte relación entre la imagen de ANEEL y la imagen del Gobierno, y que existe una disparidad entre la imagen transmitida por el titular y la noticia en su totalidad. Asimismo, se observó el uso de estrategias de comunicación (etiquetado, *agenda-setting*, *linkage* y *framing*) por parte de los vehículos de prensa. Para futuros estudios, se sugiere el uso de otras fuentes de datos y también la validación de otros casos.

Palavras-clave: Imagen Organizacional Publica; Noticias; Ciência de los Datos; NLP.

1. INTRODUÇÃO

Uma organização é um sistema complexo de comunicações com um objetivo funcional e, durante sua existência e atuação, ela vai formando uma imagem junto aos seus diversos públicos em função de todo um conjunto de contatos realizados, e não somente pelas iniciativas planejadas de comunicação que ela toma para formar a imagem pretendida (Schuler, 2004). Essa imagem é chamada imagem organizacional, e pode ser descrita como uma compilação de opiniões e pontos de vista baseados em informações processadas de várias fontes ao longo do tempo, gerando uma imagem mental dos atributos da organização (Valerio & Pizzinatto, 2003; Zhang *et al.*, 2016).

Seguindo uma tendência já percebida na iniciativa privada, os órgãos públicos têm, cada vez mais, se preocupado com sua imagem organizacional. A imagem organizacional pública tem valor estratégico na medida em que interfere diretamente no relacionamento da organização com diferentes atores, bem como na sua legitimação e credibilidade perante a sociedade. No entanto, apesar de sua relevância, poucos estudos endereçam a imagem organizacional no contexto público, tanto nacional quanto internacionalmente (Da Silva & Gutierrez, 2013; De Sousa, Ardigo & Santos, 2021; Steppacher, 2012; Valerio & Pizzinatto, 2003).

Quando se trata da construção da imagem de uma organização, ressalta-se o papel dos veículos de imprensa. Estes veículos, ao publicarem uma notícia, atuam na associação e formação das crenças, ideias, sentimentos e impressões que uma pessoa ou grupo tem ou passa a ter sobre aquela organização (Otto *et al.*, 2022). Não obstante, a migração desses meios de comunicação para a internet aumenta o volume de informações em circulação, elevando seu poder de influência, uma vez que é possível acessar dados e in-

formações de qualquer lugar, a qualquer momento, expandindo a comunicação entre as pessoas e organizações (Bartoluzzio *et al.*, 2019).

Contudo, a popularização dos portais eletrônicos, o grande volume de informações geradas e o formato não-estruturado desses dados se apresentam como um desafio para as organizações que tentam se manter à par do que está sendo veiculado na mídia sobre elas. Existe uma dificuldade de resumo e análise através de esforço humano: não só altos custos seriam gerados, mas os resultados estariam expostos à subjetividade dos analistas (Marcolin *et al.*, 2021). Logo, são necessárias técnicas inovadoras para endereçar as complexidades que cercam a imagem de uma organização conforme veiculada pela imprensa. Nesse sentido, as técnicas de Ciência de Dados se apresentam como uma alternativa.

Considerando contexto, este artigo traz como pergunta de pesquisa: como aferir a imagem de uma organização pública através do que a imprensa veicula sobre ela? A fim de responder esta pergunta, o objetivo desse artigo é utilizar técnicas de Ciência de Dados para construir um instrumento capaz de analisar e oferecer insights sobre a imagem organizacional de uma instituição pública através da imprensa nacional, com foco em portais de notícias e jornais.

Para validação, foi utilizado o caso da Agência Nacional de Energia Elétrica - ANEEL. A ANEEL é uma agência reguladora, estabelecida como uma autarquia, vinculada ao Ministério das Minas e Energia, que tem como objetivo regular e fiscalizar a geração, transmissão e comercialização de energia elétrica, em conformidade com as políticas e diretrizes do governo federal. Apesar disso, a ANEEL tem autonomia decisória, financeira e gerencial, além de ter a competência técnica para normatizar questões do setor elétrico (Agência Nacional de Energia Elétrica, 2021).

Este trabalho está organizado da seguinte forma: na seção 2, apresenta-se o conceito de imagem organizacional e qual sua relevância no setor público, bem como discute-se o papel dos veículos de imprensa na formação dessa imagem. Na seção 3, apresenta-se a metodologia utilizada para se atingir o objetivo desta pesquisa. Os principais resultados são apresentados na seção 4. Na seção 5, são feitas as considerações finais, ressaltando as limitações e possibilidades de estudos futuros.

2. CONTEXTUALIZAÇÃO

Nesta seção é feita uma contextualização sobre o tema imagem organizacional e a influência da imprensa na sua formação. Esta seção inicia com uma caracterização da imagem organizacional (seção 2.1) e porquê é importante para o ente público gerenciá-la. Na seção 2.2 discute-se o papel da imprensa na formação dessa imagem, ressaltando as principais estratégias utilizadas por esses veículos.

2.1. A caracterização da imagem organizacional e a imagem organizacional pública

São três os aspectos que caracterizam a imagem organizacional. Primeiro, a imagem organizacional representa limitações da realidade na mente do público - ou seja, as pessoas formam uma ideia de uma determinada organização a partir do que consideram real sobre ela. Em segundo lugar, a formação de uma imagem organizacional para um determinado público decorre das mensagens recebidas e decodificadas de forma acumulada ao longo do tempo processadas a partir de traços psicológicos de cada pessoa. Por fim, a imagem organizacional representa as percepções individuais da realidade das pessoas - diferentes públicos podem ter diferentes impressões de uma única organização com base em suas origens econômicas, sociais e pessoais (Barich & Kotler, 1991; Bromley, 1993; Dowling, 1986).

Sendo o conceito de imagem organizacional ligado à área de marketing, Cezar (2019) destaca que se apoiar em suas teorias, conceitos e práticas enquanto instrumento para a melhoria de desempenho das organizações públicas é pertinente e necessário, porém ressalta que a maioria foi desenvolvido sob uma ótica mercadológica, e as organizações públicas não se orientam por uma lógica de lucratividade ou venda

como acontece no setor privado. Logo, é necessário delinear os pressupostos de marketing às lógicas das organizações públicas, considerando a forma de atuação dos órgãos, as leis a que estão expostos, as formas de governança e os serviços que oferecem, entre outros.

Mesmo com tais considerações, é fundamental para o ente público conhecer e gerenciar sua imagem organizacional, na medida em que ela interfere diretamente no relacionamento da organização com os seus diversos públicos (Valerio & Pizzinatto, 2003) e na legitimação em relação ao seu ambiente técnico-organizacional (Steppacher, 2012). Além disso, o alinhamento entre a imagem projetada pela organização e a percepção de seus públicos é de fundamental importância, uma vez que, quando este alinhamento não existe ou é deficiente, a credibilidade do discurso organizacional fica afetada (Da Silva & Gutierrez, 2013).

Entende-se, portanto, que a imagem real de uma organização não é necessariamente aquela cujos seus esforços se propõem a divulgar, sendo construída de forma diversa e quase que individual na mente do público. E, considerando que as organizações públicas devem se preocupar com diferentes questões ao exercer sua função de servir aos cidadãos, o estudo de imagem organizacional no setor pode trazer benefícios.

2.2. O papel da imprensa na formação da imagem organizacional

Em comunicação, entende-se por mídia os canais ou ferramentas usadas para armazenamento e transmissão de informações ou dados. Neste universo, existe a imprensa, que é um conjunto de veículos de comunicação que exercem o Jornalismo e outras funções de comunicação informativa (Martins, 2014).

O papel da imprensa, em sua essência, é produzir e transmitir informação, conteúdo de relevância e de interesse público; nesse contexto, a imprensa é vista como distribuidora social de mensagens, consideradas em sua maioria fidedignas pela população (Martins, 2014; Otto *et al.*, 2022). No entanto, a imprensa não apenas transmite um conjunto de estruturas interpretativas, mas também sistemas de relevância, reforçando valores, interesses, ênfases e convicções sobre o que é verdadeiro ou falso, o que vale a pena conhecer ou o que é negligenciável (Hjarvard, 2012).

Dada a força das suas posições e influência sobre a sociedade, Maia (2009) ressalta que a imprensa não é composta de meros veículos neutros, mas de organizações que controlam o acesso dos atores sociais aos

seus canais e regulam os fluxos de comunicação. Os profissionais da imprensa selecionam e editam eventos e discursos, enquadrando significados partir da própria lógica e de seus *modus operandi*. Consequentemente, quando a imprensa transmite uma notícia, ela começa a construir uma realidade social – no âmbito deste estudo, essa realidade social pode ser entendida como a imagem organizacional.

Nesse processo, a imprensa faz uso de algumas estratégias que impactam a imagem organizacional. A rotulagem, por exemplo, associa a organização a uma ação, fato ou tema (Otto *et al.*, 2022). Já na *agenda-setting*, há uma atenção e repetição da mídia sobre uma questão específica, concentrando a atenção do público em uma questão particular (Martins, 2014).

Ainda, há o *framing*, usado pela imprensa quando se deseja modelar, destacar ou enfatizar algo, dando enfoque utilizando certas expressões, palavras e adjetivos, permitindo que o receptor compreenda de forma recortada uma informação (Martins, 2014). Outra estratégia é o *linkage*, processo que acontece quando a imprensa vincula um acontecimento a outro e, consequentemente, um conjunto de tópicos torna-se associado a uma organização, fenômeno ou evento (Otto *et al.*, 2022).

Portanto, a mídia tem se constituído como fonte de definição e interpretação da realidade, e adquire grande importância enquanto meio pelo qual é possível oferecer uma visão particular das organizações. Como os órgãos públicos não são entidades isoladas dessa realidade, eles não podem ignorar a necessidade de conhecer e gerenciar a sua imagem organizacional, conforme veiculada pela imprensa (Da Fonseca, 2020).

3. PROCESSO METODOLÓGICO

3.1. Ciência de Dados e Processamento de Linguagem Natural

Entende-se *big data* como um conjunto de dados identificado por quatro características principais: volume, devido à grande quantidade de dados; velocidade, devido à rapidez com que os dados são gerados ou atualizados; variedade, devido aos dados estrutu-

rados e não-estruturados, provenientes de diversas fontes; e veracidade, devido às estruturas complexas dos grandes conjuntos de dados que os tornam ambíguos, imprecisos e inconsistentes (Tabesh, Mousavidin & Hasani, 2019).

Considerando tais aspectos, é necessária a utilização de ferramentas inovadoras para a coleta, armazenamento e análise desses dados (Tabesh, Mousavidin & Hasani, 2019). Logo, um dos objetivos da Ciência de Dados é considerar diferentes tipos de dados de forma conjunta, a fim de facilitar sua análise e apresentá-los como suporte na tomada de decisão organizacional (Chuprina, Postanogov & Kostareva, 2017).

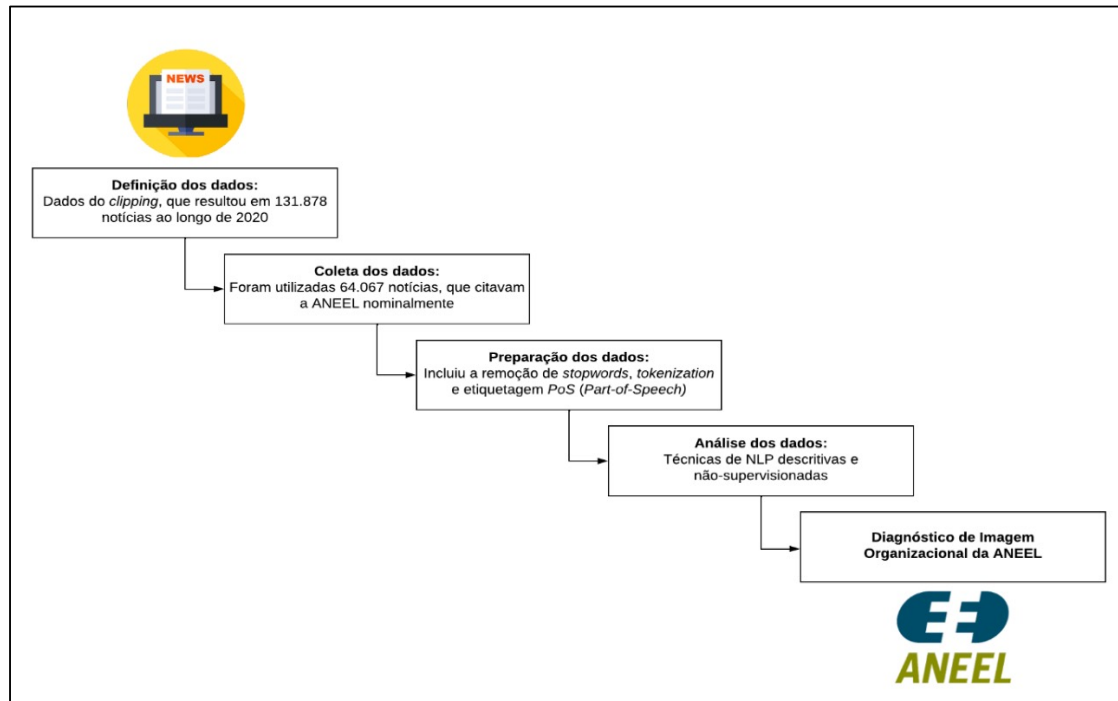
Para atingir os objetivos deste trabalho, foi realizada uma coleta de dados junto à ANEEL. A Agência contrata um serviço de *clipping*, que coleta notícias da imprensa especializada e não-especializada sobre a organização. As características e o volume dos dados do *clipping* de notícias aproximam-se do que Tabesh, Mousavidin & Hasani (2019) denominaram *big data*. Logo, esses conjuntos de dados são tão volumosos que técnicas metodológicas tradicionais de análise podem não ser adequadas para lidar com o volume e a multidimensionalidade.

3.2. As etapas metodológicas para a construção do instrumento

A partir da escolha das técnicas de análise mais recomendadas para o *clipping* de notícias, esta seção busca descrever as etapas metodológicas utilizadas para a construção do instrumento, cuja preocupação central foi permitir um processo de coleta, preparação, resumo, visualização e análise de dados (Fan & Gordon, 2014) produzidos pela imprensa nacional para extrair padrões e informações úteis acerca da imagem organizacional da ANEEL. Na prática, o instrumento criado para análise da imagem organizacional de notícias possui quatro etapas distintas: definição dos dados, coleta dos dados, preparação dos dados e análise dos dados (Stieglitz *et al.*, 2018). Os códigos na íntegra estão disponíveis na plataforma GitHub² dos autores.

2. Link para acesso: <https://github.com/carlamarcolin/topicmodelling/blob/master/codigos-aneel-imprensa.zip>.

FIGURA 1: ETAPAS METODOLÓGICAS PARA CRIAÇÃO DO INSTRUMENTO



Considerando a definição dos dados, o *clipping* de notícias contratado pela ANEEL mostrou-se a melhor opção, por possuir quantidade, qualidade e custos de obtenção mais adequados.

Para coleta, os dados utilizados foram as notícias do *clipping* para todo o ano de 2020. No total foram coletadas notícias de 3.232 veículos de imprensa diferentes, o que totalizou 131.878 notícias ao longo de 2020. Como o objetivo deste trabalho de pesquisa é desenvolver um instrumento diagnóstico para analisar a imagem organizacional a partir das publicações da imprensa nacional, foram filtradas apenas as notícias que citaram nominalmente a Agência ao longo de 2020 (cerca de 64.067 notícias, equivalente a 49%).

Na fase de preparação dos dados, utilizou-se a linguagem computacional Python, por ser a linguagem computacional mais utilizada para análise de dados em massa com características *big data*, que é o caso do *clipping* (Cielen & Meysman, 2016). Esta fase seguiu as melhores práticas indicadas em literatura (Altszyler *et al.*, 2017; Manning, Rhagavan & Schutze, 2010) considerando: (i) remoção de *stopwords*; (ii) *tokenization* ou tokenização e *n-grams*; e (iii) etiquetagem PoS (*Part-of-Speech*) ou “tagueamento”.

A remoção de *stopwords* consiste em remover palavras de alta frequência, mas baixo valor semântico, ou seja, expressões como “de”, “para”, “que”, entre

outras. A tokenização visa quebrar o texto em unidades mínimas, que podem ser palavras simples ou até mesmo compostas, bem como outras expressões como siglas. A tokenização auxilia a melhor performance computacional, permitindo a análise *n-grams*, que são uma sequência de *n*-caracteres. Por fim, a etiquetagem PoS consiste na identificação das classes gramaticais de cada palavra, que recebe uma etiqueta com sua respectiva classe, como verbo, substantivo e adjetivo. Também se realizou um tratamento no texto para a substituição de caracteres codificados incorretamente pelos caracteres corretos. Isso foi necessário pois o arquivo de texto das notícias não possuía padrão de codificação.

Por fim, na fase de análise dos dados, foram utilizadas técnicas de Processamento de Linguagem Natural (NLP) descritivas e não-supervisionadas. NLP é um campo dentro da Ciência de Dados formado por um subconjunto de técnicas computacionais baseadas em inteligência artificial, voltado para a compreensão da linguagem humana, sendo um método para extrair conhecimento a partir de um grande volume de textos e convertê-los de linguagem natural (falada ou escrita) em resultados utilizáveis (Aggarwal & Zhai, 2012; Maulud *et al.*, 2021). As técnicas descritivas tiveram como objetivo compreender o conjunto de dados e obter uma análise inicial da imagem organizacional.

Contudo, sendo o conceito de imagem organizacional complexo e multifacetado, optou-se por aprofundar essa análise através de uma técnica não-supervisionada, a modelagem de tópicos (*topic modeling*).

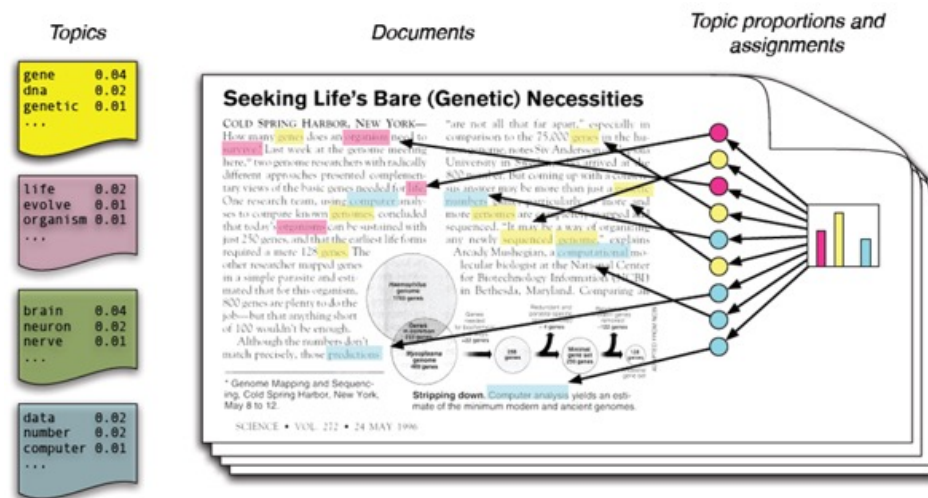
A modelagem de tópicos é adequada à tarefa e ao potencial de avanço para a pesquisa, fornecendo um caminho complementar aos métodos empíricos tradicionais (Schmiedel, Muller & Brocke, 2019; Langley & Klag, 2019). O objetivo desta técnica é uma representação multidimensional dos tópicos presentes em um conjunto de documentos, neste caso, notícias. Após o processo, cada documento é atribuído a cada tópico com diferentes pesos, especificando o grau de associação, revelando a estrutura latente do tópico formada por seu conjunto de palavras (Crain *et al.*, 2012). Embora se possa utilizar a modelagem de tópicos como técnica de agrupamento (Landauer, 2007), para este trabalho o objetivo é o processo de descoberta de tópicos, que representam as diferentes dimensões da imagem organizacional do ente público.

Das diferentes abordagens para modelagem de tópicos, foi escolhida a abordagem probabilística, que impõe um modelo probabilístico gerando a partir da estrutura semântica de um conjunto de documentos. Assumindo que os documentos são gerados de acordo com um determinado modelo probabilístico, descobrir sua estrutura semântica equivale a estimar os parâmetros do modelo, alguns deles representando um conjunto de tópicos (Stevens *et al.*, 2012). De certa forma,

como sugerido por Ashton *et al.* (2014), a abordagem probabilística, na modelagem de tópicos, expandiu os modelos de fatoração de matrizes incorporando um contexto de incerteza, permitindo a implementação de processos generativos. Neste trabalho, foi aplicado o LDA (*Latent Dirichlet allocation*), um modelo tradicional que assume que os documentos são misturas probabilísticas de tópicos latentes (Figura 2) e que os tópicos são distribuições de probabilidade sobre palavras (Blei, 2012). Além de ser um dos modelos mais comuns em uso em trabalhos de ciências sociais aplicadas (Eickhoff & Neuss, 2017), o LDA também tem a robustez necessária para lidar com dados qualitativos do discurso, resultando em um conjunto de tópicos mais coerente que representa a verdade, mais próximo do julgamento humano (Nikolenko *et al.*, 2017).

O modelo LDA é uma abordagem não supervisionada que requer três parâmetros: os hiperparâmetros de Dirichlet (α e β) e o número de tópicos (κ). A escolha de κ tem implicações significativas para os resultados gerados pelo modelo, uma vez que um número relativamente grande atribuído ao parâmetro pode dispersar a alocação de texto, e um número relativamente pequeno pode comprimir demais a alocação de texto, dificultando uma análise sólida. A escolha apropriada de κ é uma questão de pesquisa aberta (Lima Jr., 2020). Para este trabalho, foi utilizada uma abordagem iterativa, verificando a estabilidade do modelo considerando entre 3 e 10 tópicos, escolhendo $\kappa = 5$.

FIGURA 2: LATENT DIRICHLET ALLOCATION (LDA)



Fonte: Blei (2012).

4. RESULTADOS

Neste tópico são apresentados os resultados da análise das notícias sobre a ANEEL produzidas pela imprensa nacional no ano de 2020. Esta seção inicia com uma apresentação dos dados (seção 4.1), depois apresenta a aplicação de técnicas de NLP descritivas (seção 4.2) e finaliza com uma análise não-supervisionada a partir da modelagem de tópicos (seção 4.3).

4.1. Apresentação dos dados

Por se tratar de uma análise diagnóstica da imagem organizacional, primeiramente, identificou-se os principais veículos de imprensa que publicaram sobre a ANEEL no ano de 2020 (Figura 3). Em seguida, evidenciou-se as principais palavras associadas à organização nas notícias publicadas (Figura 4). Ambos resultados são apresentados na forma de nuvem de palavras e, portanto, convém ressaltar que quanto maior o tamanho em que uma palavra aparece na nuvem, maior a sua frequência e relevância para a análise em questão.

FIGURA 3: PRINCIPAIS VEÍCULOS DE IMPRENSA QUE PUBLICARAM SOBRE A ANEEL EM 2020



Percebe-se que os canais nacionais que mais produzem notícias citando o nome da ANEEL são aqueles que não possuem uma linha editorial especializada no setor de energia elétrica, tais como: Revista Isto É Online, Portal Terra, Valor Econômico, Portal UOL e G1. Dos veículos que têm uma proposta editorial especializada, o Canal Energia foi o que mais citou a ANEEL ao longo de 2020.



FIGURA 4: PRINCIPAIS PALAVRAS ASSOCIADAS À ANEEL NAS NOTÍCIAS PUBLICADAS EM 2020



O destaque das palavras “empresa”, “setor”, “governo” e “pandemia” apresenta indícios de alta relação entre a imagem da ANEEL e a imagem do Governo. Em seguida, em uma segunda ordem de relevância, aparecem as palavras: sistema, luz, brasil, federal, presidente. Por fim, em um terceiro nível, as palavras: distribuidoras, valor, serviços, amapá, tarifa, geração, coronavírus, consumidor. Agrupando essas palavras, percebe-se que o nome da ANEEL aparece junto de temas relacionados à política brasileira, ao Estado, à temas sazonais do ano, aos consumidores e ao setor de energia elétrica

4.2. Análise descritiva

Nesta seção, buscou-se entender melhor a linguagem utilizada pela imprensa nacional nas notícias veiculadas sobre a ANEEL. Os bigramas e tri-

gramas estão relacionados com a utilização conjunta de palavras, e o intuito dessa técnica é definir quais são os termos e expressões mais utilizados pela imprensa para se referir à ANEEL. Já com a técnica da etiquetagem (*Part-of-Speech*), buscou-se identificar os verbos, adjetivos e substantivos associados à ANEEL com mais frequência.

A Figura 5 contrasta dois gráficos: o bigrama de todas as notícias que citam a ANEEL na íntegra (5A), e os bigramas apenas das manchetes das notícias que citam a ANEEL (5B). Com o intuito de ter uma visualização ainda mais contundente, o instrumento também mostra os trigramas (Figura 6). À exemplo dos bigramas, aqui também foram realizadas visualizações tanto dos dados das notícias na íntegra (6A) quanto das manchetes (6B).

FIGURA 5: BIGRAMAS DAS NOTÍCIAS NA ÍNTEGRA E DAS MANCHETES DAS NOTÍCIAS

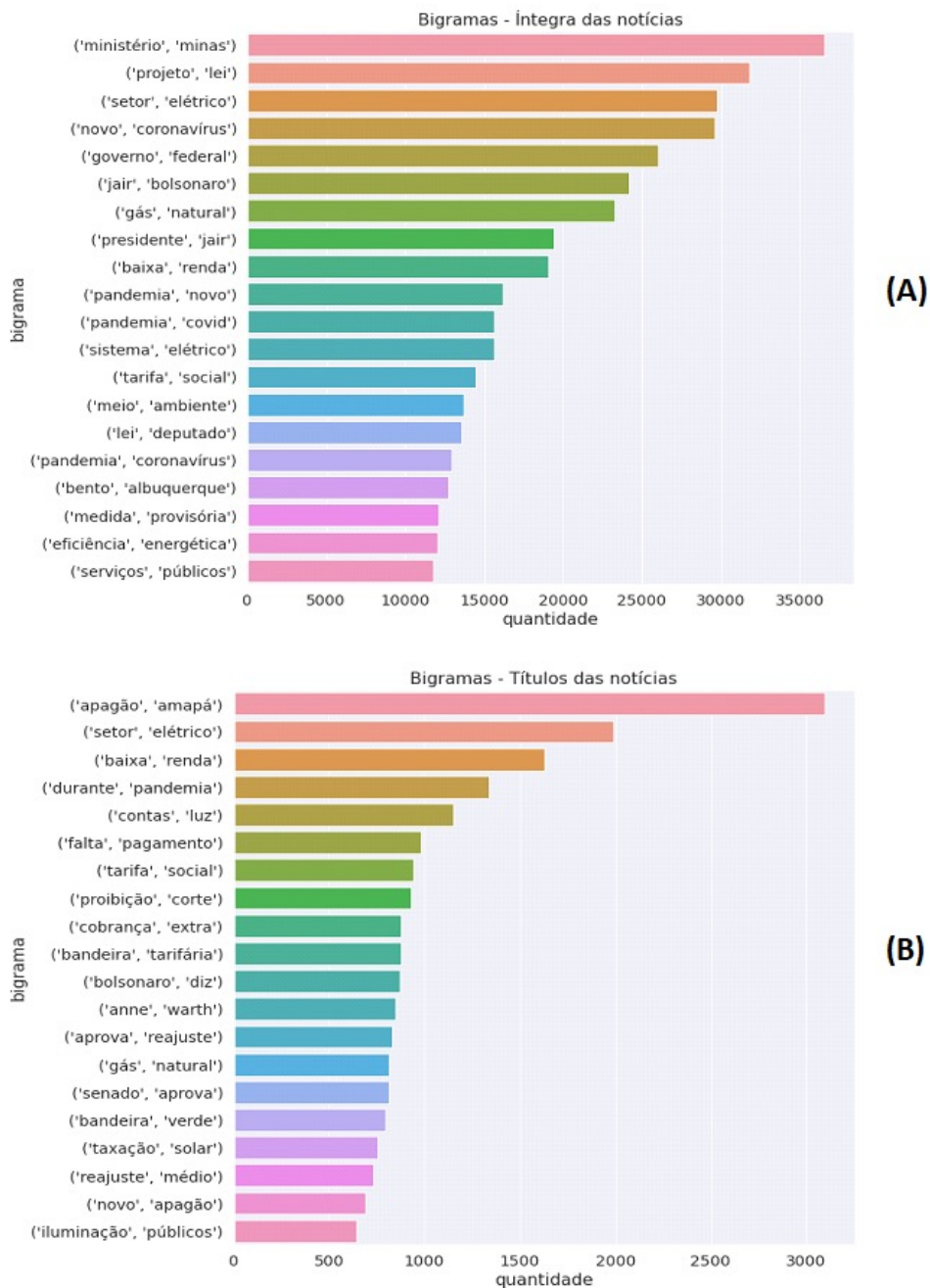


FIGURA 06: TRIGRAMAS DAS NOTÍCIAS NA ÍNTEGRA E DAS MANCHETES DAS NOTÍCIAS



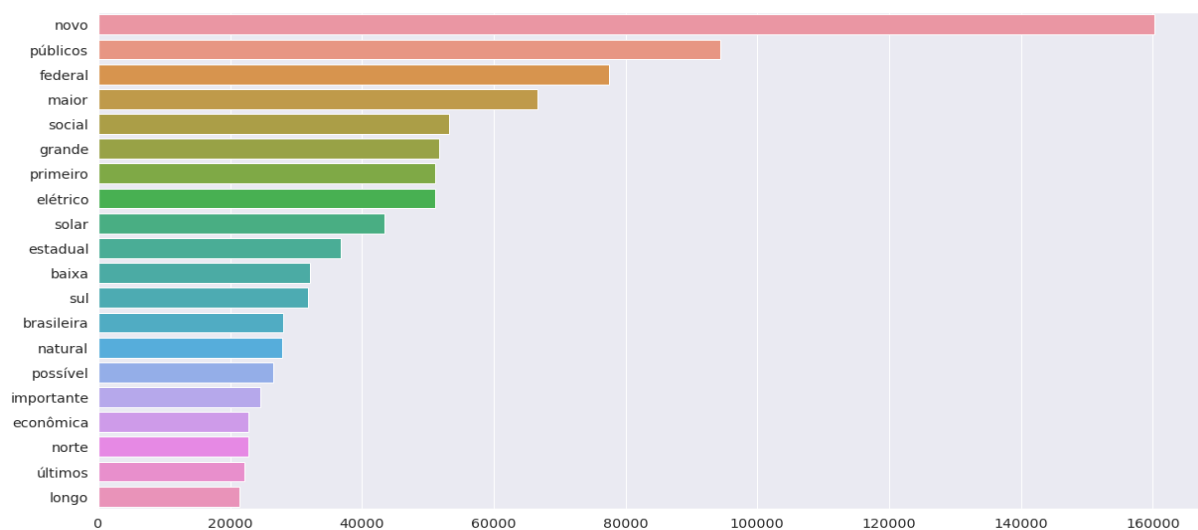
Observa-se que do bigrama referente às notícias na íntegra (Figura 5A), emergem palavras que fortalecem a evidência de associação da ANEEL ao contexto político, Estado, pandemia e setor elétrico. No entanto, considerando que parte do público que acessa as notícias não aprofunda a leitura, ficando somente com a informação evidenciada na manchete, para efeitos de comparação apresenta-se também um bigrama para os títulos das notícias (Figura 5B).

Nele, tem-se uma interpretação distinta das notícias na íntegra, sendo que as palavras mais frequentes são: apagão do Amapá, setor elétrico, baixa renda, durante a pandemia e conta de luz. Ou seja, apesar de a manchete ter o propósito editorial de síntese da notícia completa, percebe-se que as manchetes dão o entendimento menos de governo e Estado, e mais do setor elétrico.

Os trigramas das notícias na íntegra (Figura 6A) confirmam mais uma vez a forte associação da imagem da ANEEL com o governo e o Estado. Nos trigramas das manchetes, observa-se uma associação às questões ligadas ao setor elétrico em geral. Novamente, fica claro que o público que lê as manchetes pode ter uma associação distinta da imagem organizacional da ANEEL em relação ao público que lê as notícias na íntegra, uma vez que as manchetes não refletem o teor das notícias completas.

Outra análise realizada foi uma exploração das principais estruturas gramaticais presentes nas notícias: uso de adjetivos, substantivos e verbos na construção das mensagens. A Figura 7 apresenta os adjetivos mais comuns nas notícias. Destaca-se que nenhum dos principais adjetivos possuem um teor de significado negativo.

FIGURA 7: POS: ADJETIVOS MAIS FREQUENTES



A Figura 8 apresenta os substantivos que mais apareceram nas notícias. Percebe-se que são substantivos ligados a questões de política, pandemia e do setor elétrico. Ao associar os adjetivos mais utilizados com os substantivos mais proeminentes nos textos das notícias, percebe-se as junções de qualificação do governo, empresa e setor.

Por fim, a Figura 9 apresenta os verbos mais utilizados nos textos das notícias. São verbos que indicam ações dos substantivos mais utilizados. Percebe-se que são verbos que indicam ações, em sua maioria, de comunicar, extraídos de entrevistas ou falas.

FIGURA 8: POS: SUBSTANTIVOS MAIS FREQUENTES

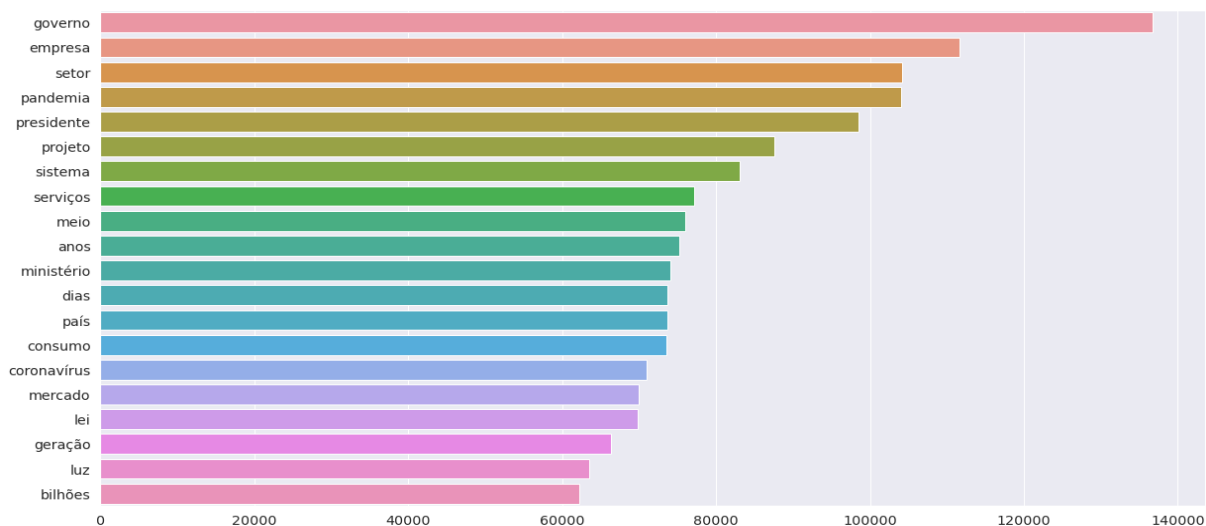
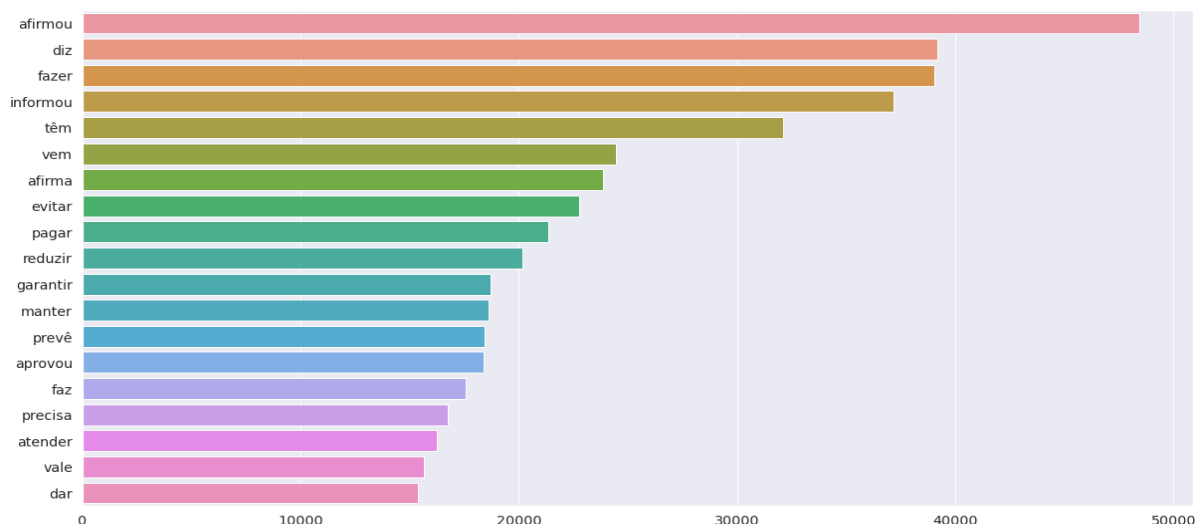


FIGURA 9: POS: VERBOS MAIS FREQUENTES

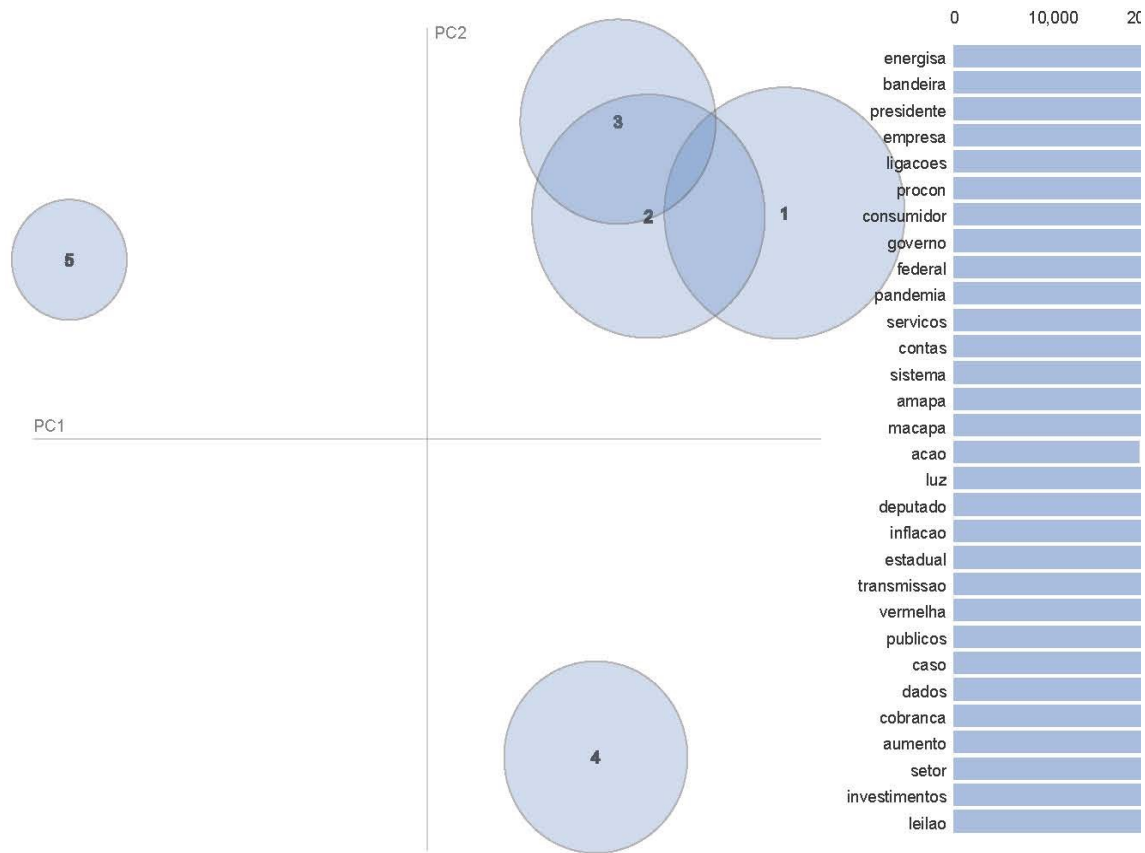


4.3. Modelagem de tópicos

Dados os resultados da seção anterior, para a análise de tópicos, as notícias na íntegra e as manchetes também foram analisadas separadamente. O intuito desta técnica é visualizar como os termos se agrupam em torno tópicos, representados pelos círculos, considerando também as diferentes estratégias utilizadas pela imprensa para veiculação de informações. As Figuras 10 e 11 apresentam o conjunto de todos os tópicos, respectivamente, para o texto das notícias na íntegra (10) e apenas para as manchetes (11).

O tamanho do círculo (tópico), mostra a sua relevância no conjunto dos dados analisados. Quanto maior for o círculo, maior a relevância daquele tópico abordado pela a imprensa nacional. Percebe-se também que os círculos se sobrepõem, ou seja, há alguns termos que são comuns a mais de um tópico. Quanto maior a área de sobreposição, maior a correlação entre os tópicos abordados. Outro aspecto a ser observado na visualização dos tópicos é a distância entre um círculo e outro. Quanto mais distantes forem, maior é a discricionariedade entre um tópico e outro, ou seja, os tópicos tratam de temáticas distintas.

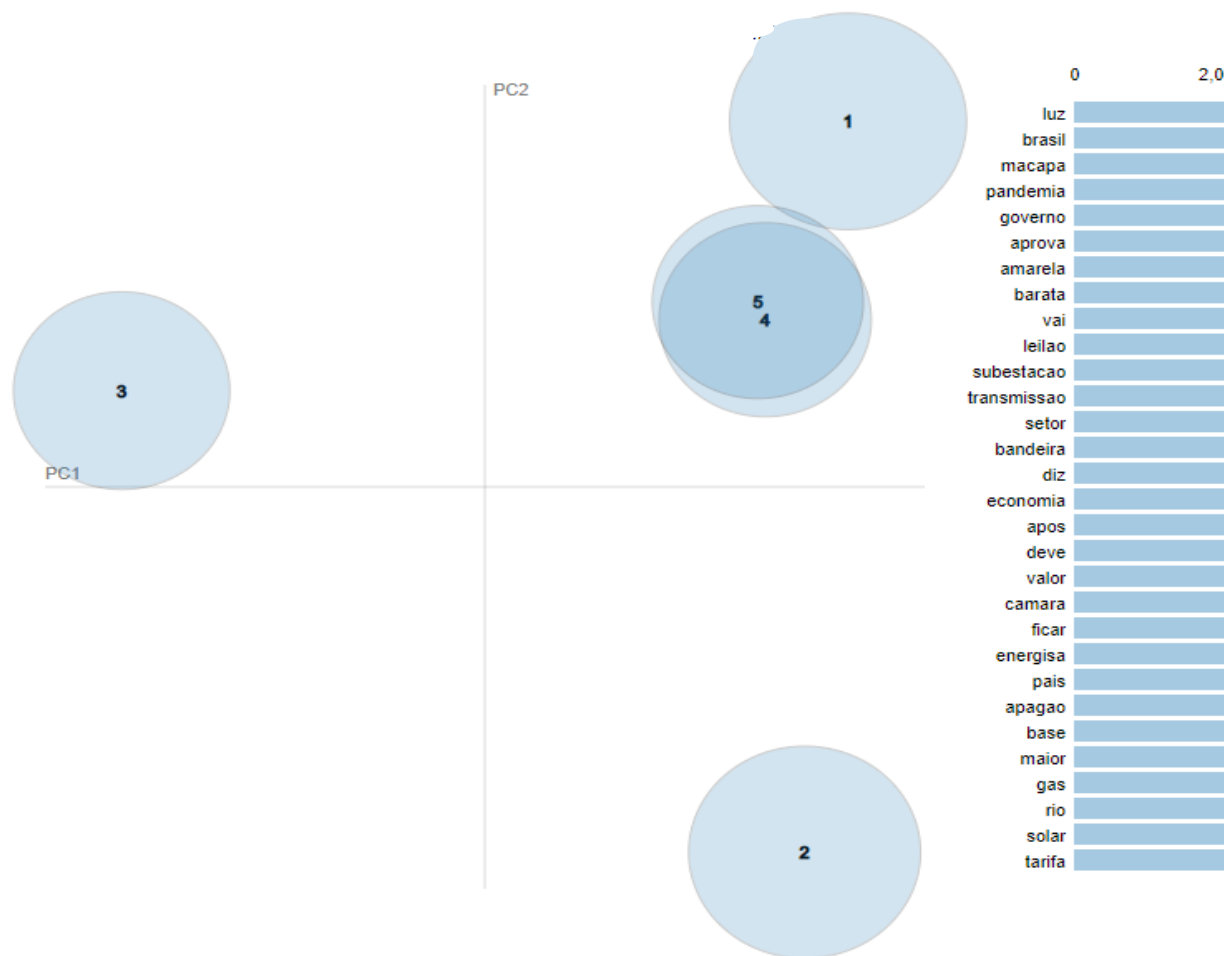
FIGURA 10: TÓPICOS DO TEXTO INTEGRAL DAS NOTÍCIAS QUE CITAM A ANEEL



As visualizações dos tópicos das notícias completas mostram cinco tópicos (Figura 10), em que os tópicos 1, 2 e 3 apresentam sobreposições de termos. Já os tópicos 4 e 5 estão mais distantes entre si e entre os demais. O tópico 1 apresenta uma forte associação com temas relacionados ao setor de energia elétrica: setor, transmissão, geração, empresa, Brasil, investimentos, mercado, país, usinas, elétrico. O tópico 2, apesar das sobreposições, apresenta uma associação forte com “impacto financeiro”: pandemia, luz, contas, valor, consumidor, inflação, cobrança, alta, aumento, redução, taxa, custo, tarifa, preços, IPCA, pagamento. O tópico 3, também com sobreposições com o 1 e principalmente com o tópico 2, apresenta forte asso-

ciação com o sistema bandeiras tarifárias: bandeira(s), sistema/programa, dados, vermelha, amarela, verde, água, reservatórios, economia, consumo, municípios, kwh. O tópico 4, bem distinto dos demais, têm forte relação com o governo, política e o Estado: governo, presidente, federal, Amapá, Macapá, deputado, ministério, lei, CPI, órgãos, ministro, Bolsonaro, apagão, justiça, Jair, Câmara. O tópico 5, também distinto dos demais, porém com pequena relevância se comparado com os demais (menor tamanho do círculo), agrupa termos que têm relação com jurídico e consumo: Energisa, ligações, Procon, ação, serviços, estadual, empresa, consumidor, reclamações, suspensão, fiscalização, fornecimento.

FIGURA 11: TÓPICOS DAS MANCHETES DAS NOTÍCIAS QUE CITAM A ANEEL



As visualizações dos tópicos das manchetes das notícias mostram também cinco tópicos (Figura 11), em que os tópicos 4 e 5 apresentam sobreposições de termos, estando bem próximos do tópico 1. Já os tópicos 2 e 3 estão mais distantes entre si e entre os demais. Não foi observada grande distinção entre o tamanho dos círculos.

O tópico 1 têm forte relação com o Estado e eventos específicos: Macapá, Energisa, país, apagão, pública, Covid, Norte, aumento, problema, Amazônia. O tópico 2, sem sobreposições e distante dos demais, apresenta uma forte associação com temas de governo, política e o Estado: pandemia, governo, economia, Câmara, Amapá, projeto, alta, Bolsonaro, Ministério, privatizações, petróleo. O tópico 3, também sem sobreposições e distante dos demais, apresenta relação com temas de economia: leilão, subestação, gás, inflação, Senado, IPCA, contas, recursos, comer-

cialização, sobe, ministro, demanda. O tópico 4, com sobreposição ao tópico 5 e bastante próximo ao tópico 1, apresenta temas voltados para o setor de energia elétrica: setor, investimentos, ambiental, elétrico, Procon, metas, leilões, usina, venda, geração, Cuiabá, obras, hidrelétricas, biodiesel, privatização. O tópico 5, apesar da sobreposição, apresenta forte associação com o sistema de bandeiras tarifárias: luz, Brasil, amarela, bandeira, barata, valor, tarifa/tarifária, crise, mercado, Brasília, preço, reajuste, corte.

5. DISCUSSÃO

Ao noticiar sobre a ANEEL, determinado veículo de imprensa pode ou não, de forma consciente ou inconsciente, refletir uma imagem organizacional na linguagem e na escolha das palavras, sentenças e temas

de suas notícias. Essas escolhas rotulam a realidade e, portanto, contribuem para formar a imagem organizacional da ANEEL.

Apesar da ampla quantidade de veículos de imprensa nacional que publicaram notícias mencionando a ANEEL em 2020, grande parte dessas notícias está concentrada em canais que não possuem uma linha editorial especializada no setor de energia elétrica. Observou-se que existem diferenças de teor entre as notícias veiculadas em canais especializados versus não especializados: veículos de imprensa especializados tendem a associar a imagem da ANEEL às questões mais técnicas do setor, como as atividades de geração, distribuição e transmissão de energia elétrica; enquanto veículos de imprensa não-especializados tendem a associar a imagem organizacional da ANEEL a questões políticas, de governo e do Estado. Esta associação da imprensa não-especializada merece atenção especial, pois representa um risco ao associar a imagem de políticos e do governo à imagem da ANEEL.

Outro resultado que chama atenção são as diferenças entre os conteúdos das notícias na íntegra e das manchetes. Ao comparar as Figuras 10 e 11, fica evidente que a distribuição dos tópicos e as principais palavras sofrem variação, ainda que, em tese, a mensagem passada deveria ter certa similaridade. Apesar de a manchete ter o propósito editorial de síntese da notícia completa, percebe-se que estas possuem mais termos sobre governo e Estado, e menos sobre o setor elétrico. Considerando que parte do público que acessa as notícias não aprofunda a leitura e que a imprensa utiliza de técnicas de linguagem para chamar a atenção do público nos títulos das notícias, o nome da ANEEL associado à estas manchetes pode criar na mente do público uma imagem mais associada à opinião pública sobre o governo (e seus governantes) em exercício, o que pode causar flutuações de percepções dependendo da aprovação (ou não) dos cidadãos.

Foram identificados também picos de veiculação de notícias quando o tema tem caráter de associação negativo à imagem da ANEEL, como: “aumento de tarifas de energia elétrica”, “apagão no Amapá”, “projeto de lei com taxa de energia solar”. Essa situação requer atenção, pois a repetição da mídia sobre uma questão específica é chamada de *agenda-setting*, e a repetição da mesma notícia por veículos de imprensa diferentes passa ao público uma ideia de importância relativa do assunto repetido em relação a outros as-

suntos e eventos. Isso concentra a atenção do público nessas questões particulares e fortalece a associação de temas negativos à imagem organizacional da ANEEL.

O agrupamento de termos proporcionado pela modelagem em tópicos (seção 4.3) também evidencia a utilização de estratégias de comunicação pelos veículos de imprensa (seção 2.2). Por exemplo, para os tópicos das notícias na íntegra (Figura 10): para termos evidenciados no tópico 1, não foi identificada nenhuma estratégia evidente, uma vez que ele apresenta termos informativos do setor elétrico. No entanto, para o tópico 2, que apresenta uma forte associação com economia e finanças, é possível identificar as estratégias de *agenda-setting*, *framing* (modelar a informação para que o público a compreenda de forma recortada) e *linkage* (vincular um acontecimento a outro). No tópico 3, cujo principal tema é o sistema de bandeiras tarifárias, também foi possível identificar o uso do *framing* e *agenda-setting*. No tópico 4, sobre governo, política e o Estado, é possível observar a estratégia do *linkage*. Por fim, no tópico 5, que agrupa termos que têm relação com jurídico e consumo, identificou-se o uso de rotulagem (associação de uma organização a uma ação, fato ou tema) e *framing*.

Para os tópicos associados às manchetes (Figura 11): para os termos evidenciados no tópico 1, que têm forte relação com o Estado e eventos específicos, foi possível identificar as estratégias de *agenda-setting* e *linkage*. No tópico 2, sobre governo, política e o Estado, também foi identificado o uso de *linkage*. No tópico 3, que apresenta relação com temas de economia e finanças, observou-se o uso do *agenda-setting*. No tópico 4, com temas voltados para o setor de energia elétrica, não foram identificadas estratégias evidentes. Por último, no tópico 5, cujo tema é o sistema de bandeiras tarifárias, foi identificado o uso do *framing*.

Entende-se que o uso dessas ferramentas, principalmente nos veículos não-especializados, tem como objetivo ajudar o público a entender uma realidade social – no âmbito deste estudo, essa realidade social está intrinsecamente relacionada à imagem organizacional. No entanto, seu uso indevido e não justificado favorece uma composição textual mais dramática do que realista. E, considerando que os jornalistas estão cada vez mais injetando seus pontos de vista nos seus trabalhos (Kuypers, 2002), pode acabar se tornando difícil para o leitor cuidadoso separar as informações verdadeiras daquelas que são tendenciosas.

6. CONCLUSÕES

O objetivo deste estudo foi utilizar técnicas de Ciência de Dados para construir um instrumento capaz de analisar e oferecer insights sobre a imagem organizacional de uma organização pública, conforme veiculado em portais de notícias e jornais. Para validação da proposta, foi utilizado o caso da Agência Nacional de Energia Elétrica – ANEEL.

Dentre os principais resultados, observou-se uma forte relação entre a imagem da ANEEL e a imagem do Governo. Ainda, evidenciou-se a existência de uma discrepância entre o conteúdo das manchetes e da íntegra das notícias publicadas sobre a ANEEL. Considerando que parte do público que acessa as notícias não aprofunda a leitura, a imagem organizacional da ANEEL transmitida pelas manchetes pode não estar compatível com aquela pretendida pela organização – que é balizada em equilíbrio, transparência e comprometimento com o interesse público (ANEEL, 2021). Também foi possível identificar que os veículos de imprensa utilizam estratégias de comunicação ao publicarem sobre a ANEEL, sendo as mais frequentes *agenda-setting*, *linkage* e *framing*.

Este estudo apresentou algumas limitações. A base de dados apresentou alguns problemas no que diz respeito a filtros e *tags* desajustados, o que oca-

sionou a presença de algumas notícias pouco relacionadas com o setor ou à Agência, devidamente filtradas. Além disso, os dados utilizados são de 2020, ano bastante atípico, marcado pelo desencadeamento de uma pandemia global e por episódios pontuais relacionados ao setor de energia, o que pode ter influenciado os resultados. Ainda, os portais de notícias e jornais não são as únicas fontes de informações que podem influenciar a imagem organizacional. Logo, para estudos futuros, sugere-se a utilização de outras fontes de dados e a validação a partir de outros casos.

Uma boa imagem organizacional, construída em tempo oportuno e alinhada com a estratégia de comunicação constituirá um referencial positivo para a organização em todos os aspectos. Por isso, para estudos futuros, também há a oportunidade de realizar o monitoramento do horizonte, no sentido de identificar, compilar e analisar os vários sinais de mudança (Hines *et al.*, 2018). A proposta de analisar a imagem organizacional a partir de notícias veiculadas em diversos canais dá abertura para que a organização, a partir destes mesmos dados, realize uma nova análise, mas agora com o intuito de coletar *insights* e identificar tendências disruptivas que, em algum momento, podem impulsionar mudanças e questões emergentes que impactam a organização.

REFERÊNCIAS

- Aggarwal, C. C. & Zhai, C. X. (2012). A survey of text classification algorithms. In *Mining Text Data* (pp. 163-222). Springer.
- Altszyler, E., Ribeiro, S., Sigman, M., & Slezak, D. F. (2017). The interpretation of dream meaning: Resolving ambiguity using Latent Semantic Analysis in a small corpus of text. *Consciousness and Cognition: An International Journal*, 56, 178–187.
- Agência Nacional de Energia Elétrica (2021). A ANEEL. <https://www.aneel.gov.br/a-aneel>.
- Ashton, T., Evangelopoulos, N. and Prybutok, V. (2014), *Extending monitoring methods to textual data: a research agenda*, *Quality and Quantity*, Vol. 22, No. 4, pp. 2277-2294.
- Barich, H. & Kotler, P. (1991). A framework for marketing image management. *Sloan Management Review*, 32 (2), 94–104.
- Bartoluzzio, A. I. S. S., Dos Anjos, L. C. M., Dos Santos, M. R. L. & Freire, R. S. (2019). Percepção dos Cidadãos sobre os Portais da Transparência do Governo Estadual e das Gestões Municipais em Pernambuco. *Revista da CGU*, 11 (18), 1115-1133.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.

- Bromley, D. B. (1993). Reputation, image and impression management. John Wiley & Sons.
- Cezar, L. C. (2019). Comunicação e marketing no setor público: diferentes abordagens para a realidade brasileira. Enap.
- Chuprina, S., Postanogov, I., & Kostareva, T. (2017). A way how to impart data science skills to computer science students exemplified by obda-systems development. *Procedia Computer Science*, 108, 2161- 2170.
- Cielen, D. & Meysman, A. (2016). Introducing data science: *big data*, machine learning, and more, using Python tools. Simon and Schuster.
- Crain, S. P., Zhou, K., Yang, S. H. & Zha, H. (2012). Dimensionality reduction and *topic modeling*: From latent semantic indexing to latent dirichlet allocation and beyond. In *Mining Text Data* (pp. 129-161). Springer.
- Da Silva, M. S. P. & Gutierrez, R. H. (2013). Avaliação da visão interna da imagem de autarquias federais: Estudo de caso da Comissão Nacional de Energia Nuclear. *Sustainable Business International Journal*, 30, 01-25.
- Da Fonseca, L. C. L. (2020). Uma análise da reputação das empresas do PIM na mídia. *Revista Eletrônica – Arte, Educação, Comunicação e Design*, 1 (3), 1-30.
- De Sousa, C. M., Ardigo, J. P. & Santos, A. R. (2021). A produção acadêmica nacional e internacional sobre imagem corporativa no setor público. *Gestão e Sociedade*, 15 (43), 4466-4496.
- Dowling, G. R. (1986). Managing your corporate image. *Industrial Marketing Management*, 15(2), 109–115.
- Eickhoff, M. & Neuss, N. (2017, June 5-10). Topic modelling methodology: its use in information systems and other managerial disciplines. *Proceedings of the 25th European Conference on Information Systems (ECIS)*, Guimarães, Portugal.
- Fan, W. & Gordon, M. D. (2014). The Power of Social Media Analytics. *Communications of the ACM*, 57 (6), 74-81.
- Hines, A., Bengston, D. N., Dockry, M. J. & Cowart, A. (2018). Setting Up a Horizon Scanning System: A U.S. Federal Agency Example. *World Futures Review*, 10(2), 136-151.
- Hjarvard, S. (2012). Mídiação: teorizando a mídia como agente de mudança social e cultural. *Matrizes*, 5 (2), p. 53-91.
- Kuyppers, J. A. (2002). *Press Bias and Politics: How the Media Frame Controversial Issues*. Praeger.
- Landauer, T. K. (2007). LSA as a theory of meaning. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 3–34). Lawrence Erlbaum Associates Publishers.
- Langley, A. & Klag, M. (2019). Being where? Navigating the involvement paradox in qualitative research accounts. *Organizational Research Methods*, 22(2), 515-538.
- Lima Jr., A. V. D. (2020), Metric for selecting the number of topics in the LDA Model, Doctoral Dissertation, Federal University from Rio Grande do Sul, Porto Alegre, RS, Brazil.
- Maia, R. C. M. (2009). Mídia e diferentes dimensões da Accountability. *Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação*, 2.
- Manning, C. D., Raghavan, P. & Schütze H. (2010). *Introduction to information retrieval*. Cambridge University Press.
- Marcolin, C. B., Becker, J. L., Wild, F., Behr, A. & Schiavi, G. (2021). Listening to the voice of the guest: A framework to improve decision-making processes with text data. *International Journal of Hospitality Management*, 94, 1-12.
- Martins, S. A. (2014). Mídia e Opinião Pública: Estudo de Caso sobre o Mensalão nas Ópticas dos jornais Folha de São Paulo e o Estado de São Paulo. *Universitas: Arquitetura e Comunicação Social*, 11 (2), p. 47-58.

- Maulud, H. D., Zeebaree, S. R. M., Jacksi, K., Sadeeq, M. A. M. & Sharif, K. H. (2021). State of Art for Semantic Analysis of Natural Language Processing. *Qubahan Academic Journal*, 1(2), 21–28.
- Nikolenko, S. I., Koltcov, S., & Koltsova, O. (2017). Topic modelling for qualitative studies. *Journal of Information Science*, 43(1), 88-102.
- Otto, L. P., Thomas, G., Glogger, I. & De Vreese, C. H. (2022). Linking Media Content and Survey Data in a Dynamic and Digital Media Environment – Mobile Longitudinal *Linkage* Analysis. *Digital Journalism*, 10 (1), 200-215.
- Stevens, K., Kegelmeyer, P., Andrzejewski, D. and Buttler, D. (2012), Exploring topic coherence over many models and many topics, in Proceedings of the 2012 EMNLP-IJCNLP, Association for Computational Linguistics, Stroudsburg, PA, United States, pp. 952-961.
- Schmiedel, T., Müller, O. & Brocke, J. (2019). *Topic modeling* as a strategy of inquiry in organizational research: A tutorial with an application example on organizational culture. *Organizational Research Methods*, 22(4), 941-968.
- Schuler, M. (2004). Management of the organizational image: a method for organizational image configuration. *Corporate Reputation Review*, 7 (1), 37-53.
- Steppacher, D. (2012). Pesquisa de imagem organizacional de um órgão público federal: a Escola de Administração da Universidade Federal do Rio Grande do Sul sob a ótica de seus alunos de graduação. 2012. [Especialização em Marketing, Universidade Federal do Rio Grande do Sul - UFRGS].
- Stieglitz, S., Mirbabaie, M., Ross, B. & Neuberger, C. (2018). Social media analytics: Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, 39, 156–168.
- Tabesh, P., Mousavidin, E. & Hasani, S. (2019). Implementing *big data* strategies: A managerial perspective. *Business Horizons*, 62(3), 347-358.
- Valerio, D. M. & Pizzinatto, N. K. (2003). Análise da Imagem Organizacional de Universidades por Meio da Matriz Familiaridade-Favorabilidade. *Revista de Administração Mackenzie*, 4 (1), 26-37.
- Zhang, J., Wu, B., Morrison, A. M., Tseng, C. & Chen, Y. C. (2016). How country image affects tourists' destination evaluations. *Journal of Hospitality & Tourism Research*, 42(6), 904–930.



Carolina Coelho da Silveira

Universidade Federal de Uberlândia

carolinasilveira@ufu.br

<https://orcid.org/0000-0001-6365-9871>

Mestranda no Programa de Pós-Graduação em Administração na Universidade Federal de Uberlândia (UFU), na área de Gestão Organizacional e Regionalidade. Graduada em Administração em 2022 pela UFU.



Carla Bonato Marcolin

Universidade Federal de Uberlândia

carla@ufu.br

<http://orcid.org/0000-0003-0260-5073>

Doutora em Administração (Linha Pesquisa Operacional) e Mestre em Administração (Linha Gestão de Sistemas e TI) pelo Programa de Pós-Graduação em Administração da UFRGS. Possui especialização em Finanças pela mesma instituição, e graduação pela Pontifícia Universidade Católica do Rio Grande do Sul (2010). Atualmente é professora da Faculdade de Gestão e Negócios (FAGEN) da Universidade Federal de Uberlândia (UFU) e realizou pós-doutoramento na Fundação Getúlio Vargas (EAESP/FGV). Trabalhou com análise e desenvolvimento de sistemas personalizados para PMEs e com gestão de contratos com órgãos públicos. Trabalha com R e possui familiaridade com SQL e estrutura de dados, bem como com técnicas de Processamento de Linguagem Natural (NLP) e de Aprendizado de Máquina (Machine Learning). Principais interesses de pesquisa incluem inteligência artificial, analytics, mineração de dados e principalmente mineração de texto.

**Carlos Henrique Rodrigues**

Agência Nacional de Energia Elétrica (ANEEL) e Faculdade SENAC-DF
caiique@aneel.gov.br

Mestre em Administração (2012) pela Universidade de Brasília, Especialista em Gestão Empresarial (2003) e Graduado em Administração (2000), pela Universidade Federal de Uberlândia. Chefe Adjunto de Assessoria de Gestão na Agência Nacional de Energia Elétrica (ANEEL), atua na coordenação de atividades ligadas ao planejamento estratégico, gestão de processos, projetos e orçamento na Assessoria Institucional da Diretoria. Possui experiência profissional e na docência de ensino superior e pós-graduação, e em atividades de instrutoria na área de Administração e Gestão de Pessoas, atuando em temas como: cargos e salários, avaliação de desempenho, cultura organizacional, carreira profissional, qualidade de vida, treinamento, desenvolvimento e educação corporativa, satisfação, qualidade de vida no trabalho e desenvolvimento.