ANÁLISE DE TÓPICOS EM REDES SOCIAIS UTILIZANDO PROCESSAMENTO DE LINGUAGEM NATURAL (NLP)

TOPIC ANALYSIS ON SOCIAL NETWORKS USING NATURAL LANGUAGE PROCESSING (NLP)

Guilherme Henrique Lavezzo – guilherm909088@gmail.com
Faculdade de Tecnologia de Taquaritinga (Fatec) –Taquaritinga –SP –Brasil

Gislaine Cristina da Conceição – gislaine.conceicao@fatec.sp.gov.br
Faculdade de Tecnologia de Taquaritinga (Fatec) –Taquaritinga –SP –Brasil

DOI: 10.31510/infa.v20i2.1750
Data de submissão: 06/09/2023
Data do aceite: 16/11/2023
Data da publicação: 20/12/2023

RESUMO

Este artigo investiga o papel do Processamento de Linguagem Natural (NLP) na análise de tópicos em redes sociais. Discute os fundamentos do NLP, seus principais desafios e técnicas, além de ilustrar como estas contribuem para a compreensão e processamento da linguagem humana por sistemas de Inteligência Artificial. Através de um estudo qualitativo, coletou-se e processou-se dados do *Twitter/X* e do *Reddit* utilizando a API de cada plataforma e técnicas de raspagem de dados. Os dados coletados foram limpos e normalizados, depois analisados usando técnicas de NLP como a lematização, o método *Bag of words* e TF-IDF. O objetivo principal do estudo é desenvolver uma compreensão sólida do NLP e suas técnicas, e aplicá-los em dados relevantes coletados de redes sociais para identificar tendências e tópicos relevantes. Essa investigação destaca a importância do NLP no mundo digital atual, onde a análise de dados em plataformas sociais se tornou crucial para entender tendências e comportamentos. O artigo também enfatiza a relevância das técnicas avançadas de NLP na extração de insights significativos de grandes conjuntos de dados textuais e na superação dos desafios inerentes à linguagem humana.

Palavras-chave: Análise de tópicos. Linguagem Natural. Análise de dados. Inteligência artificial. Aprendizado de máquina.

ABSTRACT

This article investigates the role of Natural Language Processing (NLP) in topic analysis on social networks. It discusses the fundamentals of NLP, its main challenges, and techniques, as well as illustrating how they contribute to the understanding and processing of human language by Artificial Intelligence systems. Through a qualitative study, data from *Twitter/X* and *Reddit* were collected and processed using each platform's API and data scraping techniques. The collected data were cleaned and normalized, then analyzed using NLP techniques such as lemmatization, the *Bag of words* method, and TF-IDF. The main objective of the study is to develop a solid understanding of NLP and its techniques, and to apply them to relevant data collected from social networks to identify trends and relevant topics. This investigation

highlights the importance of NLP in today's digital world, where data analysis on social platforms has become crucial to understand trends and behaviors. The article also emphasizes the relevance of advanced NLP techniques in extracting meaningful insights from large textual datasets and in overcoming the challenges inherent to human language.

Keywords: Topic analysis. Natural Language. Data analysis. Artificial intelligence. Machine learning.

1 INTRODUÇÃO

Inspirado pelas ideias de Manuel Castells em "A Sociedade em Rede", pode-se dizer que as redes sociais se transformaram em uma verdadeira mina de dados e informações, com milhões de usuários expressando suas opiniões, compartilhando conteúdo e interagindo sobre uma miríade de assuntos (CASTELLS, 2013, p. 212). O volume de informações geradas é gigantesco e cresce exponencialmente, tornando-se um desafio para empresas, pesquisadores e instituições acompanhar e compreender as principais tendências e tópicos discutidos. A necessidade de estratégias eficientes para explorar e analisar esses dados é crucial (CASTELLS, 2013, p. 213).

A relevância e a profundidade do Processamento de Linguagem Natural (NLP) na gestão da comunicação em linguagem natural entre máquinas e seres humanos são aspectos também sublinhados por *Kasliwal*. O objetivo do NLP é ler, decifrar, entender e fazer sentido da linguagem humana de uma maneira valiosa e escalável. Neste anteprojeto, será explorada a aplicação do NLP na análise de tópicos em redes sociais (BIRD, KLEIN, & LOPER, 2009).

O objetivo principal é desenvolver uma compreensão sólida dos fundamentos do NLP e das técnicas de análise de tópicos, explorando e avaliando diferentes métodos e algoritmos para essa análise, e aplicando-os a dados relevantes coletados das redes sociais. Por fim, será avaliada a eficácia desses métodos na identificação de tópicos relevantes e propostas futuras direções de pesquisa e possíveis melhorias.

Para concluir, o crescente fluxo de informações em redes sociais apresenta um desafio e uma oportunidade sem precedentes para o estudo da dinâmica social, comportamental e mercadológica. Por meio do uso eficaz do Processamento de Linguagem Natural e da análise de tópicos, pode-se decifrar tendências e padrões ocultos que possam fornecer insights valiosos para diversas aplicações. Este estudo não só contribuirá para o avanço acadêmico na área de NLP e ciência de dados, mas também poderá ter implicações práticas significativas, ajudando empresas e instituições a tomar decisões mais informadas baseadas em um entendimento mais

profundo das conversas e comportamentos nas redes sociais. Além disso, este trabalho serve como um ponto de partida para futuras pesquisas, oferecendo a possibilidade de aprimorar as ferramentas e abordagens existentes para enfrentar o contínuo crescimento e evolução das redes sociais.

2 FUNDAMENTOS DO PROCESSAMENTO DE LINGUAGEM NATURAL

O processamento de linguagem natural é uma das ramificações da inteligência artificial, onde concentra seu processamento na interação entre máquina e linguagem humana. Sua principal função é criar processos de análise de grandes quantidades de dados de texto natural, sendo capaz de analisar, entender, interpretar e gerar uma linguagem próxima ou exatamente correspondente com a linguagem humana (BIRD & KLEIN, 2009).

O NLP foi um dos primordiais cases de pesquisa para no avanço das IA's pois é a interface natural entre humanos e máquinas. A implementação inicial do Processamento de Linguagem Natural (NLP) foi evidenciada no experimento de 1954, realizado pela Universidade de Georgetown em colaboração com a IBM. Este projeto pioneiro conseguiu traduzir mais de 60 frases do russo para o inglês de maneira automatizada, demonstrando a viabilidade do NLP (NADKARNI, OHNO-MACHADO & CHAPMAN, 2011).

Atualmente, o Processamento de Linguagem Natural (NLP) está presente no cotidiano, seja em pesquisas realizadas em mecanismos de busca como Google e Bing, em assistentes virtuais como *Alexa* e *Google Assistant*, em sistemas de recomendações personalizadas utilizados por empresas como *Amazon* e *Netflix*, e nos principais tradutores de idiomas contemporâneos. Apesar de sua ampla aplicabilidade, o processamento de linguagem natural torna-se desafiador devido à complexidade da linguagem humana.

Nas seções subsequentes, serão explorados em profundidade os principais conceitos, técnicas e desafios associados ao NLP, e como eles são aplicados na análise de dados textuais, especificamente em relação à análise de tópicos em redes sociais.

2.1 Principais desafios do NLP

Enquanto o processamento de linguagem natural (NLP) se destaca como uma ferramenta promissora no campo da inteligência artificial, sua evolução contínua e acelerada não obscurece os obstáculos significativos que ainda precisam ser superados. Em meio ao seu

desenvolvimento, uma série de desafios são constantemente identificados, estudados e encarados com o objetivo de propor soluções eficientes. Alguns desses desafios incluem:

Ironia e sarcasmo: a linguagem humana é de uma complexidade exacerbada e é rica em figuras de linguagem e expressões que possam conter ironia e sarcasmo, que causa uma dificuldade de os algoritmos de inteligência artificial conseguirem realizar uma compreensão profunda do contexto e do tom dos textos a serem analisadas (LIMA, NUNES & VIEIRA, 2007).

Expressões idiomáticas e gírias: A linguagem, enquanto fenômeno social, se molda de forma única e particular para cada grupo que compartilha um idioma comum, incorporando jargões, gírias e expressões características Essa peculiaridade linguística pode se apresentar como um obstáculo à análise, dado que essas expressões frequentemente carregam significados extremamente específicos que podem não corresponder à interpretação literal das palavras utilizadas (LIMA, NUNES & VIEIRA, 2007).

Suporte a vários idiomas: o NLP teve um avanço significativo na língua inglesa americana, mas para outros idiomas, principalmente os que não utilizam o alfabeto romano para se comunicar e também aqueles que consistem em estruturas gramaticais complexas e que são falados por uma quantidade menor de pessoas. Há uma defasagem no nível de suporte desses idiomas para se conseguir criar processos de NLP para esses tipos de idiomas (LIMA, NUNES & VIEIRA, 2007).

Atualização constante da linguagem: A linguagem está sempre evoluindo, com novas palavras, frases e gírias sendo adicionadas regularmente. Isso cria o desafio de manter os modelos de NLP atualizados com essas mudanças (LIMA, NUNES & VIEIRA, 2007).

Com isso, por mais que o processamento de linguagem natural seja um campo de estudo fundamental para o avanço da inteligência artificial, sua complexidade e ressalta a necessidade de aprimoramento constante. É um esforço contínuo para desenvolver métodos que possam capturar e entender a essência da linguagem humana em toda a sua diversidade e riqueza, enfrentando desafios que vão desde a decodificação de ironia e sarcasmo até a compreensão de idiomas menos representados.

2.2 Principais técnicas do processo de NLP

Serão exploradas as principais técnicas envolvidas no processo de NLP, ilustrando como cada uma delas contribui para a eficácia geral dos sistemas de IA na compreensão e

processamento da linguagem humana. As informações e conceitos apresentados nesta seção foram baseados e adaptados do artigo da Alura sobre NLP (2021):

***Corpus*:** Originada do latim, a palavra "*corpus*" é usada para descrever o conjunto principal de um texto, podendo ser em forma escrita ou verbal, e composto por um ou mais idiomas. Quando se refere a um conjunto de textos, utilizamos o termo "corpora", que é o plural de *corpus*. Essa aglomeração de textos pode abordar um assunto particular ou uma variedade de temas. Exemplos de materiais de texto empregados em Processamento de Linguagem Natural (NLP) incluem: análises de filmes, feedbacks online, avaliações de cursos, opiniões sobre aplicativos na web, correspondências por e-mail, scripts de filmes, transcrições de vídeos, entre outros (PREMEBIDA, 2021, p.35).

Tokenização: conforme descrito por SANTOS (2020, p.35), é o processo de divisão de textos em unidades menores denominadas tokens, que podem ser palavras, caracteres ou frases inteiras. Este conceito fundamental em NLP inclui a separação de frases em palavras ou tokens individuais, removendo caracteres especiais e pontuações. Importante destacar, como evidenciado por Premebida (2021, p.35), que um token não se limita a uma única palavra; pode também englobar termos compostos como “Ano-luz” ou “Arco-íris”.

Normalização: como descrito por SANTOS (2020, p.36), objetiva a padronização do conteúdo textual, removendo variações desnecessárias como pontuações e convertendo para letras minúsculas. Esta etapa, idealmente realizada após a *tokenização*, aumenta a precisão analítica. Exemplifica-se com as palavras “Manga” e “manga”, que possuem significados distintos. A normalização ajuda a diferenciar contextos, mas a conversão entre maiúsculas e minúsculas pode afetar a confiabilidade da aplicação, conforme observado por Premebida (2021, p.36). Cada etapa tem suas vantagens e desvantagens na análise textual.

Stopwords: são palavras comuns filtradas em processos de análise de linguagem natural. Em textos em português, categorizam-se artigos, pronomes e preposições como stopwords, dada a sua baixa relevância semântica para modelos de aprendizado de máquina. Esta prática é corroborada por estudos de Teja (2020, p.12) e Premebida (2021, p.36).

***Stemming* e lematização**: essas duas técnicas, consistem na redução de palavras a sua forma principal para que, o ruído no texto seja diminuído exponencialmente. Na tabela abaixo, demonstra como funciona as diferentes técnicas (PREMEBIDA, 2021, p.36).

Tabela 1 - Comparação Entre Stemming e Lematização

Original	Stemming	Lematização
Correndo	Correnđ	Correr
Menores	Menor	Menor
Voando	Voand	Voar

Falaram	Fal	Falar
Caminharam	Caminh	Caminhar

Fonte: Autoria própria

Bag of words*:* A técnica "*Bag of words*" (BoW), como ilustrada por Fonseca (2020, p.12), representa textos baseando-se na ocorrência das palavras, sem considerar sua ordem ou estrutura. Este método, amplamente utilizado na extração de características, converte textos em informações numéricas através da frequência de cada palavra. Embora os textos não sejam dados estruturados, o BoW permite representá-los de maneira organizada, como se todas as palavras estivessem distribuídas aleatoriamente dentro de um recipiente, conforme Premebida (2021, p.37) descreve. BoW destaca-se por sua simplicidade e eficácia na identificação de palavras significativas com base na frequência de uso.

TF-IDF: TF-IDF, acrônimo para *Term Frequency - Inverse Document Frequency*, é uma técnica estatística detalhada por Fonseca (2020, p.21), que avança em relação ao método *Bag of words*. Este método calcula a relevância de uma palavra em um texto através do produto de duas métricas: a frequência do termo no documento e a sua importância inversa em relação à frequência nos documentos. Semelhante à BoW em alguns aspectos, o TF-IDF se distingue ao avaliar a relevância de uma palavra não apenas pela sua frequência no texto, mas também pelo seu peso no contexto geral dos documentos, como ressaltado por Premebida (2021, p.22). A técnica enfatiza palavras que são frequentes em um texto, mas raras nos outros, conferindo-lhes maior relevância.

O Processamento de Linguagem Natural (NLP) incorpora diversas técnicas como *tokenização*, normalização, remoção de *stopwords, Bag of words* e TF-IDF, essenciais para que sistemas de IA processem a linguagem humana. Essas técnicas juntas transformam textos em dados numéricos para análise. Cada uma tem suas particularidades e eficácia variável conforme o contexto (PREMEBIDA, 2021).

3 PROCEDIMENTOS METODOLÓGICOS

Nesse estudo, foi realizado uma metodologia de pesquisa qualitativa utilizando as técnicas citadas no tópico acima para o processamento de linguagem natural (NLP). Os dados realizados para estudos foram retirados de mídias sociais, mais especificamente *Twitter/X* e *Reddit*, para realização de análise de tópicos abrangente.

Primeiramente foi realizado a coleta dos dados utilizando a API das redes sociais estudadas nesse artigo e foram realizadas técnicas de raspagem de dados (*web scrapping)*. A

partir desses dados, foi realizado o tratamento padrão de uma análise de dados, nesse processo foi realizado a limpeza dos nomes de usuários do texto, a retirada de postagens duplicadas, correção de erros gramaticais, ortografia e a de palavras irrelevantes para análise (*stopwords*).

Após esse processo, foi dado a sequência da utilização das técnicas de NLP, como o processo de lematização, para que fosse possível corroborar com uma análise mais efetiva referente aos dados reduzindo as palavras a sua raiz, permitindo que variações da mesma palavra fossem tratadas como uma única entidade. Depois foi montado o processo de *Bag of words* para fazer a representação do texto de maneira matemática, utilizando-se das técnicas de vetores que serão ferramenta importante para análise de tópicos.

Com as etapas anteriores finalizadas, avançou-se para a aplicação da técnica TF-IDF. Essa técnica, por meio do uso de vetores e do vocabulário de palavras analisadas, permitiu identificar e destacar as palavras de maior relevância no texto. Em seguida, foram conduzidos dois processos de aprendizado de máquina para criar os tópicos finais para a análise. Como ferramentas, foram utilizadas a linguagem de programação *PySpark*, juntamente com a plataforma de nuvem *Databricks Community* para a execução deste estudo.

4 APLICAÇÃO DE TÉCNICAS DE NLP NA ANÁLISE DE TÓPICOS

O Processamento de Linguagem Natural (NLP) desempenha um papel vital na extração de insights significativos de grandes volumes de texto (BISHOP, 2011, p.10). Uma das aplicações mais destacadas do NLP é a análise de tópicos, que visa identificar tópicos ou temas predominantes em um conjunto de documentos ou textos. No ambiente dinâmico das redes sociais, onde milhões de mensagens são compartilhadas diariamente, a capacidade de filtrar e categorizar essas informações é inestimável. A análise de tópicos, auxiliada por NLP, permite que pesquisadores, empresas e profissionais de comunicação identifiquem tendências emergentes, compreendam sentimentos públicos e reajam proativamente a narrativas em desenvolvimento.

Continuando no contexto das redes sociais, a análise de tópicos pode ser usada para descobrir os principais temas que estão sendo discutidos pelos usuários em um determinado período. Para ilustrar essa aplicação, houve a coleta de dados de redes sociais, principalmente dos canais de comunicação *Twitter/X* e *Reddit*, disponíveis no site *Kaggle.* Estes dados, oriundos de redes sociais, refletem discussões e perspectivas variadas sobre questões geopolíticas globais. Para analisar e categorizar esses vastos conjuntos de informações, foi empregada a técnica de *Atribuição Latente de Dirichlet* (LDA) para a topificação.

A tabela abaixo apresenta uma amostra dos resultados obtidos após a aplicação do LDA:

Tabela 2 - Distribuição de Peso das Palavras-Chave em Discussões Geopolíticas

Tópico	Palavra	Peso
0	*trump*	00006
1	*bitcoin*	0,004
2	*qanon*	0,011
3	*realdonaldtrump*	0,003
4	*mccabe*	0,004
5	*british*	0,003

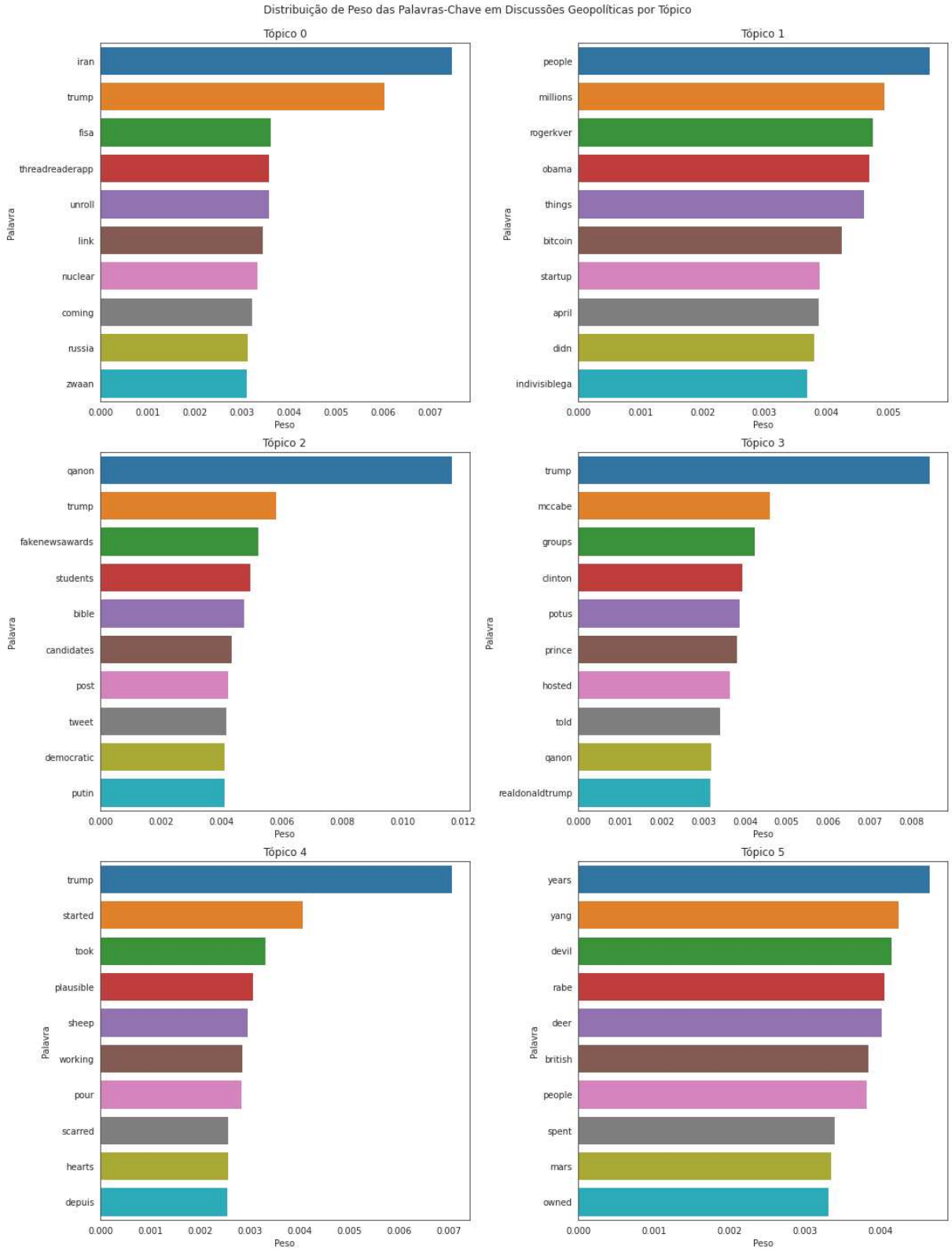
Fonte: Autoria própria

Os resultados coletados, oferecem uma visão detalhada das palavras-chave predominante em discussões sobre geopolítica nas redes sociais. Estas palavras-chave são vitais para entender os temas e assuntos que são amplamente discutidos e considerados de alta relevância no contexto geopolítico. Cada palavra é associada a um tópico e tem um peso específico, o que indica sua importância ou prevalência dentro desse tópico.

Analisando os resultados, é possível inferir os temas centrais de cada tópico com base nas palavras-chave apresentadas. Por exemplo, o tópico 0 destaca a palavra "*trump*", sugerindo discussões centradas em torno de políticas ou acontecimentos relacionados ao ex-presidente dos EUA, *Donald Trump*. Por outro lado, o tópico 1 com a palavra "bitcoin" sugere discussões focadas em tecnologia financeira e criptomoedas. Já "*qanon*", destacado no tópico 2, refere-se a uma teoria da conspiração de extrema-direita originada nos EUA, cujos seguidores foram responsáveis por uma invasão ao Congresso dos EUA após a derrota de Trump nas eleições de 2021 (*WENDLING*, 2021). É intrigante observar como as palavras dos outros tópicos, como "*mccabe*" ou "*realdonaldtrump*", sinalizam discussões focalizadas ou ligadas a eventos atuais e relevantes.

Para visualizar melhor a distribuição do peso das palavras em cada tópico, foram desenvolvidos gráficos de barras horizontais. Estes gráficos exibem as palavras mais significativas para cada tópico, oferecendo uma compreensão mais clara dos temas em discussão.

Figura 1 - Distribuição de Peso das Palavras-Chave em Discussões Geopolíticas por Tópico



Fonte: Autoria própria

O gráfico visualiza claramente a distribuição de peso das palavras em vários tópicos. Em cada tópico, palavras com maior peso destacam-se, indicando sua prevalência e importância

nas discussões. Além disso, a distinção entre os anos sugere uma evolução nas conversas ao longo do tempo. Estes insights são cruciais, pois permitem identificar não apenas os tópicos predominantes em um determinado momento, mas também compreender a dinâmica e a evolução das discussões nas redes sociais.

Em conclusão, a aplicação de técnicas de NLP, como a análise de tópicos usando LDA, oferece uma ferramenta valiosa para compreender as principais discussões e tendências em redes sociais. Esta análise pode ser de grande valor para empresas, pesquisadores e outros profissionais interessados em monitorar e responder a conversações online relevantes.

5 FUTURO DA ANÁLISE DE TÓPICOS EM REDES SOCIAIS

Com base no que foi dito ao decorrer do artigo, pode-se dizer que as redes sociais têm se consolidado como um pilar fundamental da comunicação moderna, e a crescente quantidade de dados gerados diariamente nestas plataformas exige ferramentas analíticas cada vez mais sofisticadas. A análise de tópicos, em particular, tem se mostrado crucial para navegar nesse mar de informações. À medida que a tecnologia avança, espera-se que os métodos de análise de tópicos se tornem mais precisos, identificando não apenas os tópicos principais, mas também as nuances e subtendências em discussões.

A integração de diferentes técnicas de Processamento de Linguagem Natural, como a análise sentimental e reconhecimento de entidades nomeadas, com a análise de tópicos, pode fornecer insights multidimensionais sobre os dados. Por exemplo, no futuro, não apenas será possível identificar que um tópico político é popular, mas também compreender o sentimento geral associado a ele, identificando, por exemplo, regiões geográficas ou grupos demográficos específicos que possam ter opiniões divergentes.

Outro aspecto promissor é a crescente personalização das análises. Com a capacidade de analisar tópicos em tempo real, as plataformas podem personalizar o conteúdo para os usuários com base em seus interesses e interações, criando uma experiência mais envolvente e relevante. Por último, à medida que questões de privacidade e ética ganham destaque, é imperativo que a análise de tópicos seja conduzida respeitando os direitos e a privacidade dos usuários. O futuro da análise de tópicos em redes sociais não depende apenas de avanços tecnológicos, mas também de uma abordagem ética e responsável para garantir que os dados dos usuários sejam tratados com o devido cuidado e respeito.

6 CONSIDERAÇÕES FINAIS

O crescente volume de dados em redes sociais, evidenciado neste estudo através da análise de informações específicas do *Twitter/X* e *Reddit,* destaca a importância do Processamento de Linguagem Natural. É importante ressaltar que, embora este trabalho tenha oferecido *insights* significativos sobre comportamento, opiniões e tendências sociais, as conclusões são baseadas nos dados destas duas plataformas específicas. Portanto, enquanto os resultados revelam tendências notáveis nessas redes, a generalização para todas as redes sociais deve ser feita com cautela. Futuras pesquisas poderiam expandir essa análise para incluir uma variedade mais ampla de plataformas de mídia social, proporcionando uma visão mais holística do cenário digital

Neste trabalho, foram discutidos os fundamentos do NLP, seus desafios e técnicas, destacando sua relevância e aplicabilidade no contexto das redes sociais. Através do estudo qualitativo apresentado, foi possível observar como a coleta, limpeza e análise dos dados, quando apoiadas por técnicas robustas de NLP, podem desvendar tendências e tópicos de destaque em plataformas sociais.

Contudo, como toda área em rápida evolução, o NLP enfrenta desafios constantes, especialmente no que diz respeito à precisão, interpretabilidade e escalabilidade. À medida que as redes sociais continuam a crescer e evoluir, novas técnicas e abordagens serão necessárias para manter a eficácia da análise de tópicos.

Este estudo contribui para a literatura ao demonstrar uma aplicação prática do NLP no cenário atual de redes sociais, mas ainda há um vasto território a ser explorado. Pesquisas futuras podem se aprofundar em técnicas mais avançadas, considerar outras redes sociais ou focar em desafios específicos do NLP no contexto das redes sociais.

REFERÊNCIAS

ALVES, Igor Nascimento. ***Lemmatization vs. stemming****:* quando usar cada uma?. quando usar cada uma?. 2021. Disponível em: https://www.alura.com.br/artigos/lemmatization-vs-stemming-quando-usar-cada-uma. Acesso em: 17 jun. 2023.

BIRD, Steven; KLEIN, Ewan. ***Natural Language Processing with Python***: *analyzing text with the natural language toolkit*. S.I: O'Reilly Media, 2009.

BISHOP, Christopher M. ***Pattern Recognition and Machine Learning***. 2. ed. S.I: Springer, 2011. 738 p.

BLEI, David M.; NG, Andrew Y.; JORDAN, Michael I.. ***Latent Dirichlet Allocation. Journal Of Machine Learning***. S.I, p. 993-1022. 03 jan. 2003. Disponível em: https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf. Acesso em: 15 jun. 2023.

CASTELLS, Manuel. **A sociedade em rede.** 24. ed. São Paulo: Paz e Terra, 2013. 630p.

FONSECA, Camilla. **Introdução a *Bag of Words* e TF-IDF**. 2020. Disponível em: https://medium.com/turing-talks/introdução-a-bag-of-words-e-tf-idf-43a128151ce9. Acesso em: 20 nov. 2023.

JURAFSKY, Daniel; MARTIN, James H.. ***Speech and Language Processing****: an introduction to natural language processing, computational linguistics, and speech recognition*. S.I: Prentice Hall, 2006.

LIMA, Vera Lúcia Strube de; NUNES, Maria das Graças Volpe; VIEIRA, Renata. **Desafios do Processamento de Línguas Naturais**. Semish, São Carlos, p. 1-15, 2007.

NADKARNI, Prakash M; OHNO-MACHADO, Lucila; CHAPMAN, Wendy W. Natural *language processing: an introduction. Jamia:* ***Journal of the American Medical Informatics Association***. S.I, p. 544-551. 05 set. 2011. Disponível em: https://academic.oup.com/jamia/article/18/5/544/829676?login=false. Acesso em: 01 set. 2023.

PREMEBIDA, Sthefanie Monica. **Guia de NLP**: conceitos e técnicas. conceitos e técnicas. 2021. Disponível em: https://www.alura.com.br/artigos/guia-nlp-conceitos-tecnicas. Acesso em: 17 jun. 2023.

SANTOS, Dheiver. **Introdução ao Pré-processamento de Texto em NLP: uma visão prática**. 2020. Disponível em: https://medium.com/@dheiver.santos_10420/título-introdução-ao-pré-processamento-de-texto-em-nlp-uma-visão-prática. Acesso em: 20 nov. 2023.

TEJA, Sai. ***Stop Words in NLP***. 2020. Disponível em: https://medium.com/@saitejaponugoti/stop-words-in-nlp-5b248dadad47. Acesso em: 20 nov. 2023.

WENDLING, Mike. **QANON**: o que é e de onde veio o grupo que participou da invasão ao congresso dos eua. o que é e de onde veio o grupo que participou da invasão ao Congresso dos EUA. 2021. Disponível em: https://www.bbc.com/portuguese/internacional-55577322. Acesso em: 17 set. 2023.