

Caracterização do debate online sobre cigarro eletrônico no Brasil: Uma análise de tópicos de discussão no YouTube

Richardy R. Tanure¹, Aline M. Dias¹, Lucas A. Camelo¹, Jussara Almeida²,
Helen C. S. C. Lima¹, Carlos H. G. Ferreira¹

¹ Departamento de Computação e Sistemas – Universidade Federal de Ouro Preto (UFOP)
{richardy.tanure, aline.md, lucas.camelo}@aluno.ufop.edu.br
{helen, chgferreira}@ufop.edu.br

² Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)
jussara@dcc.ufmg.br

Resumo. O YouTube é uma das plataformas de compartilhamento de vídeo mais populares do Brasil, proporcionando um ambiente dinâmico para a disseminação de opiniões sobre questões de saúde pública, incluindo os cigarros eletrônicos. Neste contexto, este estudo tem como objetivo caracterizar o debate online sobre cigarros eletrônicos no Brasil, empregando técnicas de processamento de linguagem natural para a análise de tópicos nos comentários de vídeos do YouTube relacionados ao tema. A coleção de dados utilizada abrange mais de 285 mil comentários sobre o tema, cobrindo o período entre o início de 2018 e o fim de 2023. Nossos resultados identificam múltiplos tópicos de discussão, refletindo uma dinâmica diversificada ao longo dos anos. A discussão é polarizada, com predominância de tópicos favoráveis ao uso de cigarros eletrônicos, mas também revela fortes divergências sobre questões de saúde e regulamentação.

Abstract. YouTube is one of the most popular video-sharing platforms in Brazil, providing a dynamic environment for the dissemination of opinions on public health issues, including electronic cigarettes. In this context, this study aims to characterize the online debate about electronic cigarettes in Brazil by employing natural language processing techniques to perform topic analysis on comments from YouTube videos related to the subject. The dataset used includes over 285,000 comments on the topic, covering the period from early 2018 to the end of 2023. Our results identify multiple discussion topics, reflecting a diverse dynamic over the years. The discussion is polarized, with a predominance of topics favorable to the use of electronic cigarettes, but it also reveals strong disagreements on health and regulatory issues.

1. Introdução

O YouTube destaca-se como uma das plataformas de compartilhamento de vídeo mais populares do mundo, especialmente no Brasil, onde serve como um repositório diversificado de conteúdo [Dashtian et al. 2022, IBOPE 2023]. Desde vídeos de fontes jornalísticas até postagens de usuários comuns que oferecem visões pessoais, a plataforma tornou-se uma rica e variada fonte de informação e discussão, abordando uma ampla gama de temas, incluindo questões cruciais de saúde pública [Su 2023]. Essa vasta gama de conteúdo transforma o YouTube em um ambiente dinâmico para a formação e disseminação de opiniões

públicas, refletindo e, ao mesmo tempo, moldando debates sociais sobre tópicos controversos [Porreca et al. 2020, Turenne 2023, Costa et al. 2025].

Um desses tópicos é o cigarro eletrônico, também conhecido como *vape* ou *pod*, que tem gerado intensa controvérsia em escala global [Oncology, The Lancet 2013, Campus et al. 2021, WHO 2023]. No Brasil, a comercialização, importação e propaganda de dispositivos eletrônicos para fumar são proibidas pela Agência Nacional de Vigilância Sanitária (ANVISA) desde 2009 [ANVISA 2009], uma proibição que foi recentemente reafirmada [ANVISA 2024]. O tema ganhou ainda mais relevância em 2024, com a tramitação de projetos de lei antagônicos no legislativo: o Projeto de Lei 2158/2024¹, que propõe a criminalização da fabricação e comercialização de cigarros eletrônicos, e o Projeto de Lei 5008/2023², que visa regulamentar a produção e comercialização desses dispositivos. A discussão sobre os cigarros eletrônicos nas redes sociais e plataformas de compartilhamento de vídeo, como o YouTube, tem sido intensa, refletindo a polarização do tema em vários países e contextos [Hussein et al. 2020, Dashtian et al. 2022, Kong et al. 2019, Sari Kaunang et al. 2021, Murthy et al. 2023].

Apesar da importância crescente desse debate, ainda existem poucos estudos na literatura que abordam a dinâmica do discurso sobre cigarros eletrônicos no Brasil, especialmente por meio de uma análise de tópicos que possa revelar as diferentes perspectivas e a evolução das discussões ao longo do tempo. A maioria das pesquisas existentes foca em aspectos isolados, como a identificação de padrões de consumo e problemas de saúde [Zhan et al. 2017, Wu et al. 2022, Chen et al. 2020, Lyu et al. 2021] ou em contextos internacionais, como ocorre com as pesquisas [Malik et al. 2021, Kong et al. 2019, Donaldson et al. 2023, Vassey et al. 2023], limitando a compreensão do debate que ocorre em outros idiomas, principalmente no cenário brasileiro. Uma exceção consiste em uma recente investigação sobre o posicionamento (e.g., favorável ou não ao uso do cigarro eletrônico) das pessoas no Brasil a partir da análise de comentários associados a vídeos sobre este tema no YouTube [Dias et al. 2024]. O presente estudo avança em uma direção complementar, mas ortogonal a este trabalho anterior, ao realizar uma análise do debate online sobre o uso de cigarros eletrônicos no YouTube brasileiro por meio da análise de tópicos. O objetivo central é investigar como a discussão sobre os cigarros eletrônicos, em diferentes tópicos, evoluiu no Brasil ao longo dos últimos anos.

Para isso, utilizamos uma coleção de dados previamente coletada [Dias et al. 2024], que reúne cerca de 406 mil comentários, feitos por mais de 279 mil usuários distintos, relacionados a aproximadamente 8 mil vídeos publicados em cerca de 4 mil canais do YouTube, no período de 2018 a 2023. A análise foi conduzida utilizando o BERTopic [Grootendorst 2022], um *framework* que integra técnicas de *embedding*³ com algoritmos de agrupamento, permitindo a identificação de tópicos latentes nos comentários. Além de identificar os tópicos, exploramos uma análise temporal desses tópicos e como eles se relacionam com as posturas de defesa ou contrariedade ao uso de cigarros eletrônicos, utilizando o modelo de detecção de posicionamento proposto pelos autores de [Dias et al. 2024]. Em suma, nossos principais resultados evidenciam que as discussões online sobre o cigarro eletrônico são amplas e polarizadas, tendo uma predominância de tópicos que apoiam o uso do produto.

¹<https://www.camara.leg.br/propostas-legislativas/2437521>

²<https://www25.senado.leg.br/web/atividade/materias/-/materia/160523>

³<https://blogs.oracle.com/ai-and-datascience/post/introduction-to-embedding-in-natural-language-processing>

Além disso, questões relacionadas ao uso do cigarro eletrônico para largar o tabagismo, à regulamentação do produto e à saúde geram opiniões divergentes sobre benefícios e riscos do produto, destacando a necessidade de um debate urgente e cuidadoso por parte das entidades responsáveis, especialmente no que diz respeito à formulação de políticas públicas baseadas em evidências sobre os impactos e percepções sociais do produto.

O restante do artigo está organizado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados; a Seção 3 descreve a metodologia utilizada; a Seção 4 discute os resultados obtidos; por fim, a Seção 5 conclui o trabalho e apresenta as direções futuras.

2. Trabalhos Relacionados

O aumento na discussão em torno dos cigarros eletrônicos tem atraído a atenção de pesquisadores. Alguns estudos buscaram identificar os padrões de consumo [Zhan et al. 2017, Wu et al. 2022] e relatos de possíveis problemas de saúde [Chen et al. 2020, Lyu et al. 2021], enquanto outros analisaram o engajamento dos usuários [Donaldson et al. 2023] e a promoção do conteúdo nas redes sociais [Vassey et al. 2023, Shah et al. 2023, Dias et al. 2024]. Nesta seção, revisamos os trabalhos mais próximos ao nosso objetivo, que estão relacionados à análise do debate online sobre esse produto.

Em [Zhan et al. 2017], os autores analisaram os tópicos discutidos em mensagens compartilhadas no Reddit e no Twitter utilizando o modelo *Latent Dirichlet Allocation* (LDA)⁴. Os tópicos encontrados estão relacionados à promoção dos produtos, discussões sobre sabores, compartilhamento de experiências e debates sobre regulamentações. Já em [Wu et al. 2022], os autores analisaram dados extraídos do Twitter e Reddit, publicados durante um surto de lesões pulmonares associadas ao uso de cigarros eletrônicos, observando que os tópicos prevalentes em ambas as plataformas incluíam menções à juventude, marketing, regulamentação e interesse em parar de fumar.

Malik *et al.* analisaram *tweets* para inferir a aprovação (ou desaprovação) dos diferentes sabores de uma marca popular de cigarro eletrônico, a JUUL. [Malik et al. 2021]. Outros autores analisaram vídeos do YouTube com conteúdo relacionado ao cigarro eletrônico, concluindo que esses vídeos são facilmente acessíveis a jovens e frequentemente patrocinados pela indústria [Kong et al. 2019], além de serem muito engajadores [Donaldson et al. 2023]. Por outro lado, Vassey *et al.* desenvolveram um modelo para identificação de produtos de cigarros eletrônicos em imagens e vídeos e analisaram a prevalência desses produtos em conteúdos do Instagram e Tik Tok [Vassey et al. 2023].

Em comum, os trabalhos acima mencionados não investigaram o contexto brasileiro e não examinaram a evolução das discussões por um período temporal abrangente, dificultando a identificação de padrões. Até onde sabemos, o único trabalho que focou no contexto brasileiro foi um recente estudo em que os autores desenvolveram um modelo de detecção de posicionamento (i.e., *stance detection*) sobre o uso de cigarros eletrônicos, aplicando-o a comentários em português associados a vídeos do YouTube [Dias et al. 2024]. Os autores concluíram que houve expressivo aumento dos comentários favoráveis ao produto nos últimos anos. Este artigo apresenta um estudo ortogonal e complementar ao estudo dos autores de [Dias et al. 2024] ao analisar a evolução do debate sobre o uso de cigarros eletrônicos no Brasil, considerando uma tarefa diferente, a saber, a modelagem e análise de tópicos.

⁴<https://www.ibm.com/br-pt/think/topics/latent-dirichlet-allocation>

Nesse sentido, buscamos aqui caracterizar os principais tópicos discutidos em comentários em português no YouTube, bem como a evolução desses tópicos ao longo de 6 anos.

3. Metodologia

Esta seção apresenta a metodologia utilizada. A Seção 3.1 descreve a coleção de dados e a Seção 3.2 apresenta o modelo utilizado para extração e análise dos tópicos.

3.1. Coleção de Dados

Os dados analisados foram gentilmente cedidos pelos autores de [Dias et al. 2024], que investigaram a detecção de posicionamento sobre cigarros eletrônicos no Brasil por meio de comentários no YouTube. Os dados foram coletados utilizando a Interface de Programação de Aplicação (API) oficial do YouTube ⁵ e um conjunto amplo de palavras-chave relacionadas a cigarros eletrônicos, incluindo nomes de marcas populares, sabores e expressões associadas aos riscos e benefícios do uso destes produtos (e.g., “cigarro eletrônico faz mal?”, “cigarro eletrônico ajuda a parar de fumar?”). Os termos foram selecionados visando buscar vídeos que apresentem diferentes perspectivas e tópicos relacionados ao tema. Além disto, a coleta foi configurada definindo o idioma de relevância (*relevance-language*) como português, seguindo o formato ISO 639-1, e o código de região (*region-code*) como Brasil, conforme o formato ISO 3166-1 alpha-2. Essas configurações visam maximizar as chances de coleta de comentários no contexto cultural e linguístico brasileiro, já que direcionam a recomendação do YouTube para vídeos que estão em português no Brasil ou em outros idiomas, mas que são também recomendados para o público brasileiro. Como filtro posterior, os autores também aplicaram modelos de NLP como o RoBERTa (Robustly optimized BERT approach)⁶, ajustado para a tarefa de detecção de idioma para restringir os dados a comentários em português.

A coleção de dados cobre o período de janeiro de 2018 a dezembro de 2023. Ela inclui o texto do comentário e todas as informações associadas, como vídeo, canal, número de visualizações, curtidas, autor, etc. Para cada comentário, é também incluída a classificação do posicionamento – *Favorável*, *Contrário* ou *Inconclusivo* – conforme estimado pelo modelo desenvolvido em [Dias et al. 2024]. Para fins de análise de tópicos, foram desconsiderados comentários curtos, especificamente, com menos de 15 caracteres, como foi feito em [da Rosa Jr et al. 2022, Liu and Mengoni 2023, Ma et al. 2024, Linhares et al. 2022]. Ao final, a coleção analisada neste estudo compreende 285.972 comentários, associados a 7.333 vídeos de 3.683 canais diferentes, e produzidos por 192.332 usuários distintos.

3.2. Extração e Análise dos Tópicos

Neste estudo, utilizamos o BERTopic [Grootendorst 2022], um *framework* computacional amplamente empregado na literatura [Capellaro and Caseli 2021, da Rosa Jr et al. 2022, Pereira and da Silva 2023, Venâncio et al. 2024b, Venâncio et al. 2024a]. Ele combina modelos de representação vetorial (*embeddings*) com algoritmos de agrupamento para extrair tópicos do conjunto de comentários coletados.

O processo inicia-se com a geração de representações vetoriais densas (i.e., *embeddings*) capazes de capturar a semântica dos comentários. Para isso, utilizamos o modelo

⁵<https://developers.google.com/youtube/v3/getting-started>

⁶<https://huggingface.co/papluca/xlm-roberta-base-language-detection>

BERTimbau Large, uma variante do modelo BERT ajustada especificamente para o português brasileiro [Souza et al. 2020]. Este modelo é baseado em uma arquitetura de rede neural estado-da-arte, treinada e otimizada em grandes corpora de textos em português. O *BERTimbau Large* mapeia sentenças para um espaço vetorial de alta dimensionalidade (1024 dimensões), no qual comentários com significados similares são mapeados em vetores próximos neste espaço.

A alta dimensionalidade das representações vetoriais traz desafios para o processo de agrupamento, que pode se tornar computacionalmente oneroso e também menos eficiente. Para lidar com isso, o BERTopic utiliza o algoritmo UMAP (*Uniform Manifold Approximation and Projection for Dimension Reduction*)⁷ para reduzir a dimensionalidade do espaço vetorial [McInnes et al. 2018], capaz de preservar as características globais e locais dos dados e manter a integridade da semântica capturada pelos *embeddings*. Em seguida, o BERTopic aplica o algoritmo HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*) para agrupar as representações dos comentários em conjuntos com semântica similar, refletindo tópicos de discussão [Campello et al. 2013]. Para caracterizar esses tópicos, o BERTopic utiliza uma versão ajustada da técnica TF-IDF (*Term Frequency-Inverse Document Frequency*), conhecida como c-TF-IDF (*class-based TF-IDF*) [Grootendorst 2022], que destaca termos frequentes em um dado *cluster*, mas raros nos demais e que, portanto, descrevem adequadamente os tópicos identificados.

O UMAP foi parametrizado para usar 10 vizinhos e 10 dimensões, distância mínima de 0.01 e métrica de cosseno para preservar a similaridade angular no espaço vetorial [da Rosa Jr et al. 2022, McInnes et al. 2018]. Já o HDBSCAN utilizou um tamanho mínimo de *cluster* igual a 100 e adotou a distância euclidiana como métrica de agrupamento [Campello et al. 2013]. Por fim, para caracterização dos tópicos, termos com ocorrência em menos de 1% e mais de 80% dos tópicos foram desconsiderados, a fim de filtrar termos que poderiam prejudicar a discriminação entre tópicos.

4. Resultados

Esta seção apresenta os resultados encontrados. A Seção 4.1 descreve os tópicos analisados e os padrões de engajamento; a Seção 4.2 analisa o posicionamento dos comentários por tópico e, por fim, a Seção 4.3 apresenta a evolução temporal dos tópicos.

4.1. Tópicos e Padrões de Engajamento

A metodologia descrita na Seção 3.2 resultou na identificação de 15 tópicos distintos, cada um representando diferentes aspectos das discussões em torno dos cigarros eletrônicos. Vale ressaltar que os comentários examinados são autônomos e não integram diálogos constantes. Portanto, a determinação dos tópicos se baseou unicamente na semelhança semântica entre os textos, como modelado pelos *embeddings* aplicados no BERTopic. O algoritmo HDBSCAN é capaz de isolar *outliers*, correspondendo a comentários ruidosos, com erros gramaticais ou fora do escopo do tema analisado. Esses *outliers* não estão incluídos nos tópicos identificados. A grande maioria (76,9%) dos comentários da coleção está coberta pelos 15 tópicos, que estão apresentados na Tabela 1, juntamente com os principais termos e uma breve descrição. A Tabela 2 apresenta as métricas de engajamento associadas a cada tópico, incluindo o número de canais, vídeos e usuários com comentários envolvidos, além do número de respostas e de curtidas recebidas por esses comentários.

⁷<https://umap-learn.readthedocs.io/en/latest/>

Tabela 1. Descrição dos Tópicos

ID	Breve descrição e principais termos
1	Discute sobre modelos de cigarro eletrônico e acessórios variados. <i>Termos:</i> carregador, loja, navarro, atomizador, comprou, valor, maravilhoso, dicas, preço, adorei, trocar, modelo, sensacional
2	Debate sobre o uso do cigarro eletrônico como forma de largar o tabagismo. <i>Termos:</i> maço, tentei, prejudicial, larguei, morreu, tradicional, redução, tabagismo, abstinência, reduzindo, adesivo, maldito
3	Mensagens de agradecimento que incluem desde <i>review</i> de produtos, dicas, mensagens de apoio à cessação, etc. <i>Termos:</i> salvou, amém, beijo, parceiro, gratidão, inscrito, agradeço, obrigado
4	Discussão sobre os sabores e aromas de essências usadas em cigarros eletrônicos. <i>Termos:</i> glicerina, ice, banana, leite, sabores, maçã, substituir, zomo, nasty, essências, THC
5	Argumentação em torno da legalização dos cigarros eletrônicos e suas consequências comparado a outras drogas lícitas e ilícitas. <i>Termos:</i> álcool, <i>narguilhe</i> , coca cola, cocaína, dependência, vício, cannabis, legalizar, bebida, vícios, tráfico, consumo, erva
6	Debate sobre os <i>juices</i> , relato de problemas causados pelo cigarro eletrônico e reclamação de vazamentos. <i>Termos:</i> algodão, vaporizador, ervas, queimado, lavar, coloco, pulmão, essência, vazamento, queimar
7	Mensagens de fé, geralmente associadas à cessação do tabagismo e à cura de doenças relacionadas ao tema. <i>Termos:</i> Cristo, guarde, amém, voltando, chip, creio, rei, humanidade, viva, morreu, Bíblia, alma, diabo, cruz
8	Dicas sobre manutenção, uso e manipulação dos cigarros eletrônicos. <i>Termos:</i> abrir, tampa, testar, ansioso, recarregar, ligar, irei, ensina, pino, demora
9	Discussão sobre a idade e uso de cigarros. <i>Termos:</i> 17, 16, reais, 2021, 2022, 23, fumei, luta, tomo, deixo, custa, aniversário, idade, adolescentes
10	Debate sobre cheiro, sabores e melhoria de experiências com os cigarros eletrônicos em relação aos tradicionais. <i>Termos:</i> cheiro, gostoso, manga, morango, piores, experimentei, ANVISA, nojo, fumei, sensação, bafo, ruins, palha
11	Discussão sobre as consequências dos cigarros eletrônicos e tabagismo para a saúde. <i>Termos:</i> morri, querem, viva, chorando, susto, fumem, pulmão, menino, cachorro, humanidade
12	Argumento de que reportagens e entrevistas de médicos trazem desinformação sobre o cigarro eletrônico. <i>Termos:</i> fake, combinado, namorada, vergonha, rindo, médica, contrário, mundial, vergonha, falam, contrário
13	Relatos de problemas de vazamento relacionados a cigarros eletrônicos. <i>Termos:</i> vazamento, modelo, limpar, resolvi, ligar, moço, grato, substituir, limpar, puxada, funcionar, adaptador
14	Problemas com queimaduras, gosto de queimado durante o uso de cigarros eletrônicos. <i>Termos:</i> troquei, essência, rápido, recarregar, puffs, abri, durar, vermelha, queimado, fogo, quente, esquentar
15	Repúdio e xingamentos a quem critica o uso dos cigarros eletrônicos. <i>Termos:</i> médico, regulamentação, humanidade, álcool, mãe, fumem, burro, idiota, acesso, governo, vicia, pendrive

Os Tópicos 1, 3, 4, 6, 8, 9, 10, 13 e 14 abrangem discussões sobre experiências práticas, problemas no uso, revisões sobre produtos e sabores e dicas de manuseio dos dispositivos. O Tópico 1, por exemplo, foca em discussões sobre modelos e acessórios variados, sendo o mais popular, com 141.651 comentários em 5.672 vídeos publicados em 2.847 canais (vide Tabela 2). Exemplos de comentários neste tópico são: “*E como faz para carregar a bateria?*” e “*Pode ser que você consiga mais desconto na net?*”. Já o Tópico 8, que aborda dicas sobre manutenção e uso, reflete um interesse contínuo dos usuários em melhorar suas experiências com os produtos, com 4.031 comentários que totalizam 11.101 curtidas. Problemas operacionais são frequentes, como evidenciado no Tópico 13, associado a relatos de vazamentos, e no Tópico 14, com menções a queimaduras e gosto de queimado durante o uso. Um exemplo é “*Comprei o meu hoje e ele está com gosto de queimado muito forte, o que eu faço?*”. Esses são exemplos de como os tópicos encontrados evidenciam o grande engajamento dos usuários em discutir e resolver problemas relacionados ao uso de cigarros eletrônicos no YouTube.

Tabela 2. Métricas de Engajamento dos Tópicos Identificados.

Tópico	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
# Canais	2847	1027	998	776	567	671	438	694	618	532	425	310	381	307	208
# Vídeos	5672	1856	1881	1496	887	1319	589	1249	990	851	644	437	636	561	264
# Comentários	141651	24421	8047	8125	8382	8072	2241	4031	3317	3197	2740	1511	1567	1731	888
# Usuários	102825	20121	5462	6736	7713	6854	2010	3637	3116	3069	2657	1388	1463	1578	857
# Respostas	38220	6920	3292	2603	2287	2506	605	1139	987	856	668	357	588	534	320
# Curtidas	488700	163136	15760	28463	29487	15260	6784	11101	9753	33303	14802	2137	3128	3571	2483

Movendo para os Tópicos 2, 5, 12 e 15, observa-se que eles abordam um debate sobre a regulamentação dos cigarros eletrônicos e seu uso como alternativa ao tabagismo, além de comparações com outras substâncias. O Tópico 2, em particular, que debate o

uso do produto como alternativa ao tabagismo, é um dos mais comentados, com 24.421 comentários em 1.856 vídeos publicados em 1.027 canais. Comentários como *“Evidência anedótica: parei de fumar cigarros há 4 anos, migrando para os vaporizadores. Minha saúde geral melhorou absurdamente...”* refletem a percepção positiva de muitos usuários. Já comentários como *“fraco é vc que acredita que vape é melhor que cigarro. tem centenas de estudos que comprovam que vape é centenas de vezes pior que cigarro e causa problemas pulmonares muito piores em um tempo muito menor de uso.”* refletem a percepção negativa de alguns usuários. Por outro lado, o Tópico 5, que trata da legalização dos produtos em comparação com outras drogas, também muito comentado, exemplifica a polarização das opiniões, como em *“Álcool tb é droga”*. O Tópico 12 discute a desinformação percebida em reportagens e entrevistas médicas, enquanto o Tópico 15 reflete repúdio e insultos, com frases como *“O seu velho idiota, cigarro eletrônico não faz fumaça!!! é vapor...”*. Esses tópicos mostram um grande engajamento em discussões que envolvem tanto a defesa quanto a regulamentação dos cigarros eletrônicos.

Por outro lado, os Tópicos 7 e 11 abordam críticas ao uso dos cigarros eletrônicos. O Tópico 11, por exemplo, com 2.740 comentários, revela uma forte preocupação com os riscos à saúde, exemplificado em frases como *“Com ou sem nicotina eu não fumo mais nem que me pagassem, eu prefiro meu bem-estar e minha saúde do que morrer cedo.”*. Já o Tópico 7 contém mensagens de fé associadas à cessação do tabagismo e cura de doenças relacionadas, com 2.241 comentários e 6.784 curtidas, destacando uma visão crítica do uso dos produtos em um contexto mais espiritual e religioso. Em suma, o Tópico 1, que possui o maior número de comentários e curtidas, tende a envolver discussões práticas de quem usa o cigarro eletrônico, como modelos e acessórios. Tópicos mais polêmicos, como aqueles relacionados ao uso do cigarro eletrônico como forma de largar o tabagismo, à regulamentação do produto e à saúde (tópicos 2, 5 e 11), revelam debates polarizados.

Visando analisar a dispersão de tópicos entre diferentes canais e usuários, utilizou-se o coeficiente de Jaccard (j) para medir a sobreposição entre os conjuntos de canais associados a cada par de tópicos. Os resultados para canais, mostrados na Figura 1(a) indicam uma similaridade, no máximo, moderada ($j \leq 0.45$). Observa-se que certos tópicos, como o tópico 12 (que discute desinformação médica sobre o cigarro eletrônico) e o tópico 6 (que trata de problemas e vazamentos), apresentam uma similaridade relativamente alta entre si, comparada com outros tópicos, sugerindo que os mesmos canais frequentemente produzem conteúdo que gera comentários sobre esses temas. Da mesma forma, é possível notar uma alta similaridade entre os tópicos 4 (discussão sobre sabores e aromas de essências usadas) e 2 (defesa do cigarro eletrônico para largar o tabagismo), bem como os tópicos 11 (consequências do tabagismo para a saúde) e 7 (mensagens de fé associadas à cessação do tabagismo). Esses exemplos demonstram como os canais tendem a focar em temas inter-relacionados, criando nichos específicos de conteúdo que abordam tanto questões técnicas quanto argumentos pró e contra o uso de cigarros eletrônicos, o que contribui para a formação de comunidades de discussão com interesses comuns.

4.2. Análise de Posicionamento dos Comentários por Tópico

Como mencionado na Seção 3, os comentários presentes na coleção de dados disponibilizada pelos autores foram previamente classificados por um modelo de detecção de posicionamento quanto ao uso de cigarros eletrônicos em três categorias: *favorável*, *contrário* ou *inconclusivo*. A Figura 1(b) apresenta, para cada tópico, as frações de comentários em cada

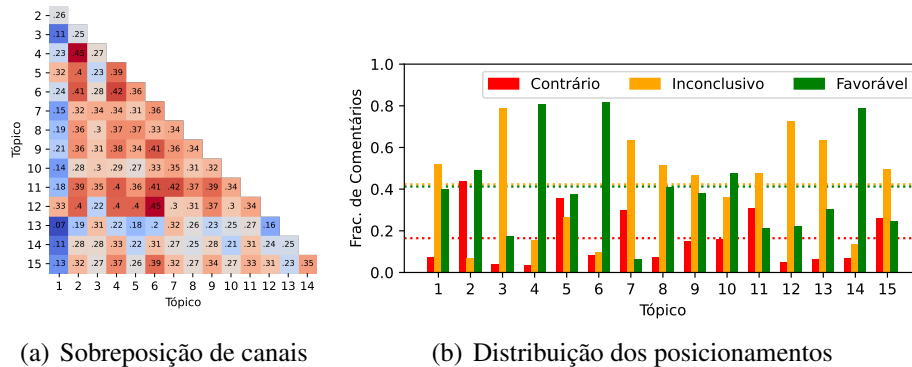


Figura 1. Caracterização dos tópicos: (a) sobreposição de canais com comentários nos mesmos pares de tópicos (índice de Jaccard) e (b) distribuição de posicionamentos entre comentários de cada tópico

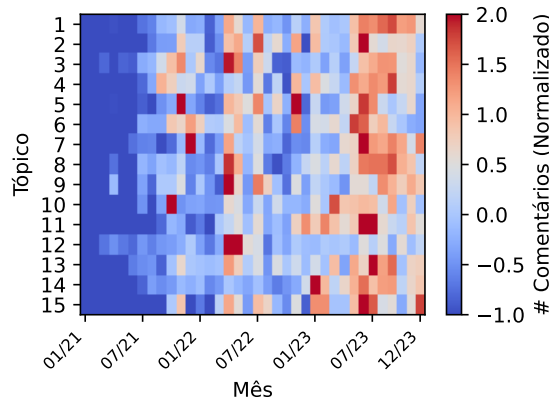


Figura 2. Evolução temporal dos tópicos entre 01/2021 e 12/2024.

uma das três classes, bem como as médias gerais de cada classe (linhas pontilhadas). Nos tópicos relacionados a experiências, problemas e manuseio (1, 3, 4, 6, 8, 9, 10, 13 e 14) observa-se uma prevalência de discussões *favoráveis* ao uso do produto. Destacam-se os tópicos 4, 6 e 14 com frações de comentários favoráveis bem acima da média, refletindo uma aceitação muito positiva dos usuários.

Nos tópicos que abordam o uso do cigarro eletrônico como alternativa ao tabagismo e a sua regulamentação (2, 5, 15), observa-se uma maior polarização. O tópico 2, que debate o uso como alternativa ao tabagismo, tem o percentual de comentários favoráveis e contrários similar (49% e 44%, respectivamente). O mesmo ocorre nos tópicos 5 e 15, que discutem a legalização do produto e comparações com outras substâncias. Já nos tópicos que criticam o uso (7 e 11), especialmente o tópico 7, relacionado a mensagens de fé associadas à cessação do tabagismo, há uma fração de comentários contrários bem acima da média para esse posicionamento e maior do que o percentual de comentários favoráveis no tópico. Essas observações sugerem que, a despeito das discussões polarizadas sobre regulamentação e impacto na saúde, as discussões relacionadas ao uso dos dispositivos geralmente recebem mais apoio favorável dos usuários.

4.3. Análise Temporal dos Tópicos

A Figura 2 apresenta a evolução temporal dos tópicos de discussão sobre cigarros eletrônicos no YouTube, com foco no período de ascensão entre janeiro de 2021 e dezembro de 2023.

A popularidade dos tópicos foi normalizada utilizando o z -score⁸, o que permite comparar a popularidade relativa de cada tópico ao longo do tempo.

É possível notar que, embora os tópicos apresentem um crescimento similar, podendo apresentar picos ao longo do período, observa-se a existência de tópicos que ganharam força a partir de 2023, como, por exemplo, os Tópicos 11, 13 e 14, que discutem consequências do uso dos cigarros eletrônicos para a saúde e, também, os problemas ao utilizar os dispositivos (13, 14). Em geral, observa-se uma dinâmica variada nas discussões, com tópicos emergindo e desaparecendo, enquanto outros permanecem como temas recorrentes e centrais na discussão pública. Além disso, nota-se que tanto as preocupações quanto os interesses dos usuários evoluíram ao longo do tempo, com um foco crescente em questões técnicas, regulamentares e suas consequências para a saúde à medida que o uso dos dispositivos vem se popularizando.

5. Conclusões e Trabalhos Futuros

Este estudo explorou o debate online sobre cigarros eletrônicos no Brasil, utilizando comentários em vídeos do YouTube para identificar e analisar tópicos de discussão. A análise revelou que o debate é amplamente polarizado, com uma predominância de discussões favoráveis ao uso dos cigarros eletrônicos. Os tópicos mais comentados incluem modelos e acessórios, experiências práticas e a comparação dos cigarros eletrônicos com outras substâncias, como o tabaco tradicional. Além disso, questões de regulamentação e saúde emergiram como temas centrais e polarizados, com opiniões divergentes entre os usuários.

A análise temporal evidenciou que, enquanto alguns tópicos mantêm relevância ao longo dos anos, outros surgem e desaparecem, refletindo as mudanças no interesse público e na cobertura midiática. Tópicos técnicos e de manutenção dos dispositivos, por exemplo, ganharam força mais recentemente, possivelmente indicando um aumento no número de novos usuários e na complexidade das discussões. Como futuras pesquisas, recomenda-se uma análise mais detalhada de como as discussões evoluem em outras plataformas e a comparação com contextos internacionais para uma maior compreensão do fenômeno.

Agradecimentos. Os autores agradecem ao CNPq, à CAPES e à FAPEMIG pelo apoio financeiro e pelo suporte à realização desta pesquisa.

Referências

- ANVISA (2009). Cigarro eletrônico. <https://www.gov.br/anvisa/pt-br/assuntos/tabaco/cigarro-eletronico>. Acesso em: 24 de abril de 2024.
- ANVISA (2024). Anvisa atualiza regulação de cigarro eletrônico e mantém proibição. <https://www.gov.br/anvisa/pt-br/assuntos/noticias-anvisa/2024/anvisa-atualiza-regulacao-de-cigarro-eletronico-e-mantem-proibicao>. Acesso em: 21 de maio de 2024.
- Campello, R. J., Moulavi, D., and Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining*, pages 160–172. Springer.
- Campus, B., Fafard, P., Pierre, J. S., and Hoffman, S. J. (2021). Comparing the regulation and incentivization of e-cigarettes across 97 countries. *Social science & medicine*, 291:114187.

⁸Calculado para cada valor x subtraindo-se a média μ dos valores da mesma linha e, em seguida, dividindo pelo desvio padrão σ dessa linha, conforme a fórmula: $z = \frac{x - \mu}{\sigma}$.

- Capellaro, L. and Caseli, H. (2021). Análise de polaridade e de tópicos em tweets no domínio da política no brasil. In *Anais do XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pages 47–55, Porto Alegre, RS, Brasil. SBC.
- Chen, L., Lu, X., Yuan, J., Luo, J., Luo, J., Xie, Z., and Li, D. (2020). A social media study on the associations of flavored electronic cigarettes with health symptoms: observational study. *Journal of medical Internet research*, 22(6):e17496.
- Costa, J., Oliveira, G., Fonseca, G., Reis, D., Teixeira, G., Cunha, W., Rocha, L., and Ferreira, C. H. (2025). Characterizing youtube’s role in online gambling promotion: A case study of fortune tiger in brazil. In *Proceedings of the 17th ACM Web Science Conference 2025*.
- da Rosa Jr, J. M., Linhares, R. S., Ferreira, C. H. G., Nobre, G. P., Murai, F., and Almeida, J. M. (2022). Uncovering discussion groups on claims of election fraud from twitter. In *International conference on social informatics*, pages 320–336. Springer.
- Dashtian, H., Murthy, D., and Kong, G. (2022). An exploration of e-cigarette–related search items on youtube: network analysis. *Journal of Medical Internet Research*, 24(1):e30679.
- Dias, A., Tanure, R. R., Almeida, J., Lima, H., and Ferreira, C. (2024). Análise da percepção do uso de cigarros eletrônicos no brasil por meio de comentários no youtube. In *Brazilian Symposium on Multimedia and the Web*.
- Donaldson, S. I., Dormanesh, A., Perez, C., Zaffer, M. O., Majmundar, A., Unger, J. B., and Allem, J.-P. (2023). Monitoring the official youtube channels of e-cigarette companies: a thematic analysis. *Health Education & Behavior*, 50(5):677–682.
- Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.
- Hussein, E., Juneja, P., and Mitra, T. (2020). Measuring misinformation in video search platforms: An audit study on youtube. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–27.
- IBOPE (2023). Video audience share percentage in brazil. <https://kantariibopemedia.com/conteudo/relatorios/april-2023/>.
- Kong, G., LaVallee, H., Rams, A., Ramamurthi, D., and Krishnan-Sarin, S. (2019). Promotion of vape tricks on youtube: Content analysis. *Journal of medical Internet research*, 21(6):e12709.
- Linhares, R. S., Rosa, J. M., Ferreira, C. H., Murai, F., Nobre, G., and Almeida, J. (2022). Uncovering coordinated communities on twitter during the 2020 us election. In *2022 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*, pages 80–87. IEEE.
- Liu, Q. and Mengoni, P. (2023). What do airbnb users care about before, during and after the covid-19? an analysis of online reviews. In *2023 IEEE International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 409–414. IEEE.
- Lyu, J. C., Luli, G. K., and Ling, P. M. (2021). Vaping discussion in the covid-19 pandemic: An observational study using twitter data. *PloS one*, 16(12):e0260290.
- Ma, B., Wong, Y. D., Teo, C.-C., and Wang, Z. (2024). Enhance understandings of online food delivery’s service quality with online reviews. *Journal of retailing and consumer services*, 76:103588.
- Malik, A., Khan, M. I., Karbasian, H., Nieminen, M., Ammad-Ud-Din, M., and Khan, S. A. (2021). Modeling public sentiments about juul flavors on twitter through machine learning. *Nicotine and Tobacco Research*, 23(11):1869–1879.
- McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29):861.

- Murthy, D., Lee, J., Dashtian, H., Kong, G., et al. (2023). Influence of user profile attributes on e-cigarette-related searches on youtube: Machine learning clustering and classification. *JMIR infodemiology*, 3(1):e42218.
- Oncology, The Lancet (2013). Time for e-cigarette regulation. [https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045\(13\)70468-6/fulltext](https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(13)70468-6/fulltext). Acesso em: 30 de agosto de 2024.
- Pereira, P. and da Silva, T. (2023). Uso de modelagem de tópicos para agrupamento de notícias: uma abordagem usando bertopic. In *Anais do XIV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pages 398–402, Porto Alegre, RS, Brasil. SBC.
- Porreca, A., Scozzari, F., and Di Nicola, M. (2020). Using text mining and sentiment analysis to analyse youtube italian videos concerning vaccination. *BMC Public Health*, 20:1–9.
- Sari Kaunang, C. P., Amastini, F., and Mahendra, R. (2021). Analyzing stance and topic of e-cigarette conversations on twitter: Case study in indonesia. In *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*.
- Shah, N., Nali, M., Bardier, C., Li, J., Maroulis, J., Cuomo, R., and Mackey, T. K. (2023). Applying topic modelling and qualitative content analysis to identify and characterise ends product promotion and sales on instagram. *Tobacco control*, 32(e2):e153–e159.
- Souza, F., Nogueira, R., and Lotufo, R. (2020). Bertimbau: pretrained bert models for brazilian portuguese. In *Intelligent Systems: 9th Brazilian Conference, BRACIS 2020, Rio Grande, Brazil, October 20–23, 2020, Proceedings, Part I 9*, pages 403–417. Springer.
- Su, C. (2023). *Douyin, TikTok and China’s online screen industry: The rise of short-video platforms*. Taylor & Francis.
- Turenne, N. (2023). Net activism and whistleblowing on youtube: a text mining analysis. *Multimedia Tools and Applications*, 82(6):9201–9221.
- Vassey, J., Kennedy, C. J., Herbert Chang, H.-C., Smith, A. S., and Unger, J. B. (2023). Scalable Surveillance of E-Cigarette Products on Instagram and TikTok Using Computer Vision. *Nicotine & Tobacco Research*, 26(5):552–560.
- Venâncio, O. R., Ferreira, C. H., Almeida, J. M., and da Silva, A. P. C. (2024a). Unraveling user coordination on telegram: A comprehensive analysis of political mobilization during the 2022 brazilian presidential election. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, pages 1545–1556.
- Venâncio, O. R., Gonçalves, G. H., Ferreira, C. H., and da Silva, A. P. C. (2024b). Evidências de disseminação de conteúdo no telegram durante o ataque aos órgãos públicos brasileiros em 2023. In *Brazilian Symposium on Multimedia and the Web (WebMedia)*, pages 385–389. SBC.
- WHO (2023). World health organization report on the global tobacco epidemic, 2023: protect people from tobacco smoke. {<https://www.who.int/publications/i/item/9789240077164>}.
- Wu, D., Kasson, E., Singh, A. K., Ren, Y., Kaiser, N., Huang, M., and Cavazos-Rehg, P. A. (2022). Topics and sentiment surrounding vaping on twitter and reddit during the 2019 e-cigarette and vaping use-associated lung injury outbreak: Comparative study. *Journal of medical Internet research*, 24(12):e39460.
- Zhan, Y., Liu, R., Li, Q., Leischow, S. J., and Zeng, D. D. (2017). Identifying topics for e-cigarette user-generated contents: a case study from multiple social media platforms. *Journal of medical Internet research*, 19(1):e24.