

Análise da Percepção do Uso de Cigarros Eletrônicos no Brasil por meio de Comentários no YouTube

Aline Dias¹, Richardy R. Tanure¹, Jussara M. Almeida², Helen C. S. C. Lima¹, Carlos H. G. Ferreira¹
{aline.md,richardy.tenure}@aluno.ufop.edu.br,jussara@dcc.ufmg.br,helen@ufop.edu.br,chgferreira@ufop.edu.br

¹Departamento de Computação e Sistemas, Universidade Federal de Ouro Preto, Brasil

²Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Brasil

ABSTRACT

The rise of video platforms such as YouTube has revolutionized information sharing and influenced social habits and product consumption. At the same time, the diverse global regulatory landscape and ongoing studies on the health effects of e-cigarettes have led to intense and controversial debates. Amidst this, the production of content about e-cigarettes on the internet is rapidly increasing, particularly on YouTube, one of the most popular video platforms globally and the most popular in Brazil. Despite the growing body of research focused on understanding online interactions with e-cigarettes, there is a lack of comprehensive and detailed temporal analyses that capture the dynamics of the debate and people's stance towards these products, especially in the Brazilian context. To address this gap, our study aims to investigate the popularity and acceptance of e-cigarettes in Brazil by analyzing YouTube videos and their associated comments. We collected an extensive dataset of videos, channels, comments and their metadata from 2018 to 2023. Our methodology involved analyzing production and engagement metrics, and developing a deep learning-based stance detection model to estimate people's stance (approval or disapproval) based on comments and quantify the temporal dynamics of these attitudes over the years. Our findings reveal a significant increase in content production and user engagement, indicating a growing public interest, with a notable increase in approving comments on the product. This study fills previous research gaps by offering a comprehensive and pioneering overview of e-cigarette use and public perception in Brazil, emphasizing the need for a more informed discussion among society and regulatory bodies.

KEYWORDS

Cigarros Eletrônicos, Aprendizado de Máquina, YouTube.

1 INTRODUÇÃO

A ascensão de plataformas de vídeo, como *YouTube*, *TikTok* e *Kwai*, representa uma revolução na troca de informações, abrangendo vários domínios, inclusive a saúde [16, 19, 43]. Essas plataformas tornaram-se ambientes propícios para a propagação de informações, influenciando os hábitos e o consumo de produtos e serviços pela sociedade [25, 36]. Em paralelo, o cigarro eletrônico (também denominado como *vape*, *pod*, entre outros) apresenta-se como um grande desafio para a saúde pública global. Devido à diversidade de

regulamentações globais e aos efeitos colaterais ainda em estudo, o debate em torno desse produto é intenso e controverso [21, 26].

Muitas organizações e entidades de saúde sugerem que medidas regulatórias devem ser criadas até que haja mais evidências sobre a segurança de sua utilização [9, 27, 34, 39]. A Organização Mundial da Saúde (OMS), por exemplo, enfatiza que onde o uso do cigarro eletrônico não for proibido, precisa ao menos ser regulamentado [35], e insta por "ações urgentes" contra esses produtos, enfatizando que não são eficazes para parar de fumar [18]. No Brasil, este produto continua sem regulamentação, sendo proibida a comercialização, importação e propaganda, conforme orientações estabelecidas pela Agência Nacional de Vigilância Sanitária (ANVISA) desde 2009 [2], reafirmada em 2024 [3]. Pesquisas recentes indicam, contudo, um aumento no uso entre os jovens brasileiros, estimando que um em cada cinco já experimentou esse produto [8]. Essa crescente popularidade gera preocupações devido aos potenciais danos à saúde e ao crescente uso por crianças e adolescentes [7, 13, 35]. Embora alguns estudos indiquem que muitos fumantes utilizam o cigarro eletrônico para cessar o hábito de fumar [15, 21], outros afirmam que os danos do cigarro eletrônico podem ser maiores do que os do tabagismo tradicional [4, 5, 41].

Além disso, observa-se um aumento na produção de conteúdos sobre este produto na Internet, especialmente no YouTube, uma das plataformas de vídeo mais populares do mundo e a mais popular do Brasil [12, 23, 24]. Na literatura, alguns estudos examinam a promoção de truques do cigarro eletrônico, habilidade de criar efeitos visuais com a fumaça, no YouTube [28], enquanto outros utilizam modelagem de rede e técnicas de classificação de texto para entender como diferentes termos de pesquisa sobre o tema afetam os resultados de busca de vídeos [12, 38]. Outros ainda investigam como perfis de usuários (idade e sexo) podem afetar os resultados de pesquisa sobre cigarros eletrônicos no YouTube [33]. No entanto, esses estudos não focam no contexto brasileiro e nem modelam a evolução da popularidade e aceitação deste produto ao longo do tempo. Além disso, exploraram superficialmente um número limitado de vídeos e termos de pesquisa, não investigando um grande número de marcas, termos relacionados ao uso do produto, doenças potencialmente associadas e, consequentemente, o volume de conteúdo gerado, bem como o mapeamento das interações dos usuários ao longo dos anos.

Neste contexto, este trabalho pretende investigar a popularidade e aceitação do cigarro eletrônico no Brasil, utilizando vídeos compartilhados no YouTube e comentários associados como fonte primária de dados para modelagem e análise. Este estudo cobre um volume de vídeos substancialmente maior e um período de tempo mais abrangente do que os estudos anteriores, por meio de uma busca extensa de conteúdos relacionados a um amplo conjunto

de termos específicos que incluem marcas, sabores e outros termos noticiados e relacionados. Investigamos os comentários para compreender a opinião pública sobre o uso do cigarro eletrônico, concentrando-nos na popularidade, interesse, aceitação e rejeição do produto no contexto brasileiro. Dessa forma, a pergunta central deste estudo é: *Como a população brasileira tem percebido o uso do cigarro eletrônico ao longo dos últimos anos pela ótica de comentários associados a vídeos compartilhados no YouTube?* Para isso, coletamos dados associados a 8.932 vídeos, pertencentes a 4.925 canais do YouTube, cobrindo o período de 2018 a 2023. Nossa coleção consiste em 543.600 comentários em português compartilhados por 361.015 usuários únicos e associados a vídeos sobre cigarros eletrônicos. Analisamos métricas de produção e engajamento (# canais, # vídeos, # visualizações, # curtidas, etc.) para caracterizar a evolução das mesmas neste tipo de conteúdo. Em seguida, desenvolvemos e validamos um modelo para inferir o posicionamento (aceitação ou rejeição) dos usuários sobre o tema a partir dos comentários, quantificando a dinâmica do posicionamento ao longo do tempo.

As principais contribuições deste estudo são: i) análise em larga escala da produção de conteúdo associado ao cigarro eletrônico no contexto brasileiro; ii) desenvolvimento e validação de um modelo de aprendizado de máquina para estimar a prevalência da aceitação ou rejeição dos cigarros eletrônicos com base nos comentários associados a vídeos informativos do YouTube; e iii) análise da evolução temporal do engajamento e do posicionamento dos usuários em relação aos cigarros eletrônicos no Brasil. Em suma, nosso estudo fornece uma visão ampla e pioneira sobre o consumo e a percepção pública dos cigarros eletrônicos no Brasil (sob a ótica de usuários do YouTube), ressaltando a importância de uma discussão mais urgente e cuidadosa por parte das entidades responsáveis.

2 TRABALHOS RELACIONADOS

Nos últimos anos, diversos estudos investigaram as interações online com o cigarro eletrônico, focando em aspectos como propaganda online [10, 29], recomendações em mídias sociais [12, 33], e interações dos usuários com o produto em diferentes plataformas [12, 30, 40, 45, 46]. Nesta seção, revisamos os trabalhos mais próximos ao nosso objetivo.

Massey et al. [30] exploraram a exposição a anúncios de vaporização em *smartphones* e seu impacto em jovens adultos, constatando que o uso de Snapchat, Instagram e Facebook estava ligado a maior exposição. De forma semelhante, Sun et al. [45] analisaram vídeos sobre cigarros eletrônicos no TikTok, identificando temas como comédia, estilo de vida e marketing. Vassey et al. [46] estudaram o conteúdo de cigarros eletrônicos no Instagram e o efeito da intervenção do FDA, observando que, apesar do aumento nos rótulos de advertência, ainda eram insuficientes. Dashtian et al. [12] usaram modelagem de rede para entender como diferentes termos de busca no YouTube afetam os resultados relacionados a cigarros eletrônicos.

Kong et al. [28] avaliaram vídeos de truques com cigarros eletrônicos no YouTube, descobrindo que metade dos vídeos eram patrocinados pela indústria. Bigwanto et al. [6] monitoraram contas do Instagram, destacando a promoção desenfreada dos produtos devido à falta de regulamentação. Sari Kaunang et al. [38] e Adhikari et al. [1] analisaram *tweets* para prever opiniões e identificar tópicos relacionados a cigarros eletrônicos e maconha. Por outro

lado, Hassan et al. [22] usaram o modelo LDA para analisar *tweets* sobre EVALI, enquanto Murthy et al. [33] estudaram como idade e sexo afetam os resultados de busca no YouTube.

Embora esses trabalhos forneçam percepções de grande importância, destaca-se o baixo volume de dados utilizados, seja em número de comentários [22, 38], vídeos analisados [28, 33], termos analisados [28, 33], ou período de cobertura [22]. Além disso, estes estudos não focam no cenário brasileiro. Neste sentido, este trabalho visa preencher a lacuna deixada pelas pesquisas anteriores, com uma cobertura temporal abrangente especificamente para o cenário brasileiro, utilizando um grande volume de dados de vídeos e comentários, com o objetivo de compreender o posicionamento do público em relação ao cigarro eletrônico, focando no interesse, popularidade e aceitação do produto nos últimos anos.

3 METODOLOGIA

Esta seção descreve a coleta e preparação de dados, bem como o desenvolvimento do modelo de detecção de posicionamento.

3.1 Coleta e Pré-processamento dos Dados

Para a coleta de dados da plataforma YouTube, utilizou-se a *Application Programming Interface* (API) de dados do YouTube V3¹ para coletar informações sobre os canais, vídeos e os respectivos comentários e metadados. Inicialmente, os autores realizaram uma ampla busca na *Web* focada nas marcas mais vendidas nos últimos anos, doenças relacionadas, termos e variações, expressões adotadas, entre outros. As palavras-chave foram criteriosamente selecionadas com base em notícias, artigos e conteúdos disponíveis na *Web* relacionados ao tema deste estudo cobrindo o período do estudo. Em resumo, foram mapeadas marcas reconhecidas no mercado do cigarro eletrônico, líquidos saborizados e expressões de interesse sobre potenciais problemas de saúde atrelados ao produto. As expressões foram consideradas para garantir que a coleta de dados incluisse vários pontos de vista, desde a aprovação (incluindo potencial interesse, defesa, recomendação, etc.) até a desaprovação e críticas do produto. Com isso, os seguintes conjuntos de palavras-chave finais foram definidos:

Marcas : "Vaporesso", "Elfbar", "Ignite", "Jim pod", "Puff mamma", "Smok", "Uwell", "Voopoo", "Geekvape", "Oxva", "Aspire", "Lost Vape", "Argus", "Juul", "Nikbar".

Sabores : "Juice vape", "Sabores vape", "Sabores mais populares de juices para cigarro eletrônico", "Sabores mais populares de juices", "Juice Joyetech Salt Nic", "Juice Solace", "Juice Caravela Liquids", "Juice BLVK Unicorn Salt", "Juice Zomo Salt", "Juice Dream Collab", "Juice Firefly E-juice", "Juice Radiola Juice", "Juice Zomo Vape - Blueberry Iceburst", "Juice Nasty - Cush Man Mango", "Juice Mr. Freeze - Strawberry Watermelon Frost", "Juice POP! Vapors - Strawberry Kiwi e Mango Strawberry (Ice)", "Juice Naked - Hawaiian Pog Ice", "Juice Fantasi", "Juice Caravela", "Juice Mr. Freeze", "Juice Naked 100", "Juice Nasty".

Expressões : "Quantos cigarros equivalem a um vape?", "Cigarro eletrônico faz mal?", "vape faz mal?", "Cigarro eletrônico vicia?", "vape vicia?", "Cigarro eletrônico ajuda a parar de fumar", "vape ajuda a parar de fumar", "Cigarro eletrônico tem mais nicotina?", "vape tem mais nicotina?", "Cigarro eletrônico ou cigarro normal?", "vape ou cigarro normal?", "Efeitos do cigarro eletrônico na saúde", "Efeitos do vape na saúde", "Riscos do uso de cigarro eletrônico para a saúde", "Riscos do uso de vape para a saúde", "Cigarro eletrônico causa doenças?", "vape causa doenças?", "Impactos do cigarro eletrônico no

¹<https://developers.google.com/youtube/v3/getting-started>

pulmão", "Impactos do vape no pulmão", "Dependência de cigarro eletrônico", "Dependência de vape", "Cigarro eletrônico causa dependência?", "vape causa dependência?", "Diferenças na dependência de cigarro eletrônico e cigarro tradicional", "Diferenças na dependência de vape e cigarro tradicional", "Cigarro eletrônico é eficaz para largar o cigarro?", "vape é eficaz para largar o cigarro?", "Experiências de parar de fumar com cigarro eletrônico", "Experiências de parar de fumar com vape", "Composição química do cigarro eletrônico", "Composição química do vape", "Concentração de nicotina em cigarros eletrônicos", "Concentração de nicotina em vapes", "Diferença entre nicotina no cigarro eletrônico e no cigarro tradicional", "Diferença entre nicotina no vape e no cigarro tradicional".

Para garantir uma ampla cobertura temporal, cada palavra-chave definida anteriormente foi submetida a uma busca por meio da API do YouTube, sendo restringido o retorno de conteúdos referentes a cada mês entre 2018 e 2023. Dessa forma, um mesmo termo foi considerado individualmente em cada mês, permitindo uma análise temporal mais fiel. Em cada busca mensal, coletamos os cinquenta vídeos mais relevantes conforme o algoritmo de relevância do YouTube, que leva em consideração fatores como tempo de visualização, engajamento (curtidas, descurtidas, comentários, compartilhamentos), novos inscritos gerados pelo vídeo, título, descrições, tags, histórico de visualização do usuário. Esta abordagem foi adotada para que a coleta de dados cubra um período de tempo extenso e forneça uma visão detalhada das tendências e mudanças na popularidade relacionadas ao tema de estudo. Toda a coleta de dados foi realizada entre dezembro de 2023 e março de 2024. É importante destacar que a coleta de dados foi configurada para focar no contexto brasileiro, definindo o idioma de relevância (*relevance-language*) como *português* e o código da região (*region-code*) como Brasil, de acordo com o formato ISO 3166-1. Na prática, esses parâmetros direcionam a coleta para conteúdos em português e recomendados para o Brasil, garantindo que os dados reflitam o contexto cultural e linguístico do público brasileiro. No entanto, também é possível que vídeos em outros idiomas sejam direcionados, já que existe a possibilidade de existência de legendas, o que também atrai comentários em português do público brasileiro.

Os conjuntos de palavras-chave inicialmente proposto foi elaborado visando maximizar, ao máximo, a revocação de conteúdo relacionado ao tema de estudo. Entretanto, uma avaliação inicial dos resultados retornados pela coleta apontou um impacto na precisão. Em outras palavras, muitos vídeos não relacionados a cigarro eletrônico foram retornados, seja pelo uso de palavras-chave relacionadas a sabores que podem estar associadas a vídeos sobre comida, pela indisponibilidade de cinquenta vídeos de um dado tema no mês, ou ainda por outros fatores considerados pelo algoritmo de ranqueamento do YouTube, como relevância, histórico de visualizações e interações, e engajamento do público. Assim, foi necessário realizar uma filtragem adicional. Utilizando um conjunto de palavras-chave baseado na consulta anterior, um vídeo e seus respectivos dados (comentários, canal, métricas de engajamento, etc) foram mantidos se seu título contivesse pelo menos uma das seguintes palavras-chave²: "cigarro eletrônico", "cigarros eletrônicos", "vape", "vaper", "vaping", "pod", "mod", "ecigar", "e-cig", "tabaco aquecido", "ejuice", "juice", "nicotina", "nic", "concentração de nicotina", "juices salt nic", "juices freebase", "uwell", "Aphrodite", "vaporesso", "elfbar", "elfbar", "ignite", "jim pod", "puff mamma", "smok", "uwell", "voopoo", "geekvape", "oxva", "aspire", "lost

²Essa comparação foi feita de forma *case-insensitive*.

Tabela 1: Sumário dos dados analisados.

Descrição	Valor
Número de Canais	4.625
Número de Vídeos	8.932
Número de Comentários	543.600
Número de Usuários	361.015

vape", "argus", "juul", "nikbar", "vaporar", "vapor", "puffs", "atomizador", "atty", "bocal", "chain-vape", "coil", "clapton coil", "deck", "dli", "drip-tip", "dripper", "dry-burn", "dry-hit", "estilo velocity", "flood", "flooded", "glicerina", "glicerol", "gunk", "hot-spot", "kick", "mech", "mesh", "mtl", "nicsalt", "zomo", "pen-style", "pg", "piteira", "rba", "rda", "rta", "rdta", "spitback", "blvk", "MR. FREEZE", "starter kit", "throat hit", "twisted coil", "wick", "wicka", "eliquid", "Eliquid", "e-liquid", "E-liquid", "e liquid", "E liquid", "liquid", "nasty cig", "ice apple", "grape ice", "Apple Ice", "strawberry", "mango", "Watermelon", "Kiwifruit", "Ice", "naked", "fantasi", "flavor juice", "atomizer", "ajuda a parar de fumar", "parar de fumar", "eficaz para largar o cigarro".

O próximo passo consistiu em identificar, entre os comentários dos vídeos selecionados, aqueles que estão em português. Essa tarefa foi necessária porque muitos vídeos populares em inglês e espanhol podem ter comentários em português, devido à disponibilização de legendas e ao direcionamento para o público brasileiro. Notamos que essa situação é comum, pois os criadores de conteúdo frequentemente configuraram seus vídeos ou canais para serem recomendados em diferentes países, aumentando a visibilidade e o engajamento em múltiplas regiões. No entanto, essa tarefa apresenta desafios, já que os comentários frequentemente contêm termos em inglês específicos do contexto, como "vape", "juice", "mango" e "e-liquid".

Para lidar com essa complexidade, os comentários foram processados usando dois métodos de detecção de idioma: o modelo RoBERTa (Robustly optimized BERT approach)³, ajustado para a tarefa de detecção de idioma, e a biblioteca *LangDetector*⁴, uma ferramenta baseada em algoritmos de aprendizado de máquina, conhecida por sua eficiência e precisão na identificação de idiomas em textos curtos [47]. Dessa forma, optou-se por utilizar ambos os métodos para mitigar a presença de falsos negativos, mantendo em nossa análise todos os comentários identificados como sendo em português por pelo menos um dos dois modelos.

A Tabela 1 descreve a coleção de dados final, incluindo o número de canais, vídeos, comentários em português e usuários que realizaram tais comentários. Vale ressaltar que, durante a etapa de coleta dos dados, foram recuperados diversos metadados disponíveis, como o número de visualizações e o número de curtidas. Essas métricas, que simbolizam engajamento e popularidade, serão utilizadas como indicadores nos resultados a serem apresentados.

3.2 Modelo de Detecção de Posicionamento

Com a coleção de dados final definida, o próximo passo consiste em desenvolver e validar um modelo de detecção de posicionamento. Este modelo supervisionado exige uma amostra de dados rotulados para treinamento. Dessa forma, o objetivo do modelo é inferir o posicionamento em relação ao cigarro eletrônico, classificando os comentários como *favorável*, *inconclusivo* ou *contrário*.

³<https://huggingface.co/papluca/xlm-roberta-base-language-detection>

⁴<https://pypi.org/project/langdetect/>

O processo envolve quatro etapas principais. Primeiro, realizamos a amostragem da coleção de dados para obter uma quantidade de comentários para compor o conjunto de treinamento do modelo. Em seguida, discutimos e definimos diretrizes claras para a rotulagem dos comentários, assegurando consistência na classificação. Após a apresentação das diretrizes e esclarecimento de eventuais dúvidas, a rotulagem dos comentários foi realizada por voluntários, e um índice de concordância foi calculado para verificar a consistência e a precisão dos rótulos atribuídos. Finalmente, treinamos e validamos o modelo com os dados rotulados, classificando os comentários restantes na coleção conforme os três rótulos definidos, permitindo assim as análises subsequentes.

Neste estudo, para classificar os comentários e identificar o posicionamento como sendo *favorável*, *contrária* ou *inconclusiva* em relação ao tema, foram definidas as seguintes diretrizes de rotulagem:

Favorável: Inclui comentários que explicitamente apoiam, defendem, expressam interesse ou compartilham conhecimento sobre cigarros eletrônicos e produtos associados. Isso abrange, mas não se limita a, discussões sobre sabores, marcas, bateria, métodos de uso, alusões a termos específicos do *vape*, cultura *vaping* e opiniões que refutam alegações sobre possíveis efeitos adversos à saúde decorrentes do uso de tais produtos.

Contrário: Abrange comentários que explicitamente condenam, criticam ou apresentam argumentos contrários ao uso de cigarros eletrônicos e produtos afins. Isso inclui, mas não se limita a, comentários contrários ao uso de tais produtos, cultura *vaping*, ou que evidenciem relação direta do uso destes produtos com riscos à saúde.

Inconclusivo: Destina-se a comentários que não se enquadram nas categorias anteriores ou que desviam do tópico proposto. Inclui manifestações que não permitem uma determinação explícita da postura do usuário em relação aos cigarros eletrônicos, seus sabores, marcas ou variedades, seja por falta de informação, ambiguidade ou irrelevância do comentário para a tarefa em questão.

Após a definição das diretrizes, foi gerada uma amostra aleatória e estratificada de 30 comentários por mês para cada ano, sendo a data definida pela postagem do comentário. Isso é importante para desenvolver um modelo mais generalista, considerando o período analisado⁵. Dessa forma, os comentários amostrados foram rotulados manualmente por três indivíduos distintos após a apresentação das diretrizes de rotulagem, cada um responsável por analisar e atribuir classificações conforme os critérios definidos. Obtivemos 294 comentários rotulados para a classe *contrária*, 930 para a classe *inconclusiva* e 1674 para a classe *favorável*. Observou-se um grande desbalanceamento dos dados, com uma maior produção de conteúdos *favoráveis*, referentes a *reviews*, propagandas e uso de produtos, em comparação com conteúdos *contrários*, como reportagens e vídeos de médicos alertando e explicando os riscos do uso. Esse desbalanceamento desafia o aprendizado de modelos de classificação, uma vez que pode levar a vieses para a classe majoritária.

Para minimizar esse efeito, optou-se por estender a amostra para rotulagem, com foco em aumentar o número de comentários *contrários*. Para isso, foi realizada uma nova amostragem, incluindo mais 15 comentários por mês, também estratificados e aleatórios. Nessa abordagem, definimos que os comentários deveriam conter palavras-chave relacionadas a um posicionamento *contrário*, sendo elas: "*adolescente*", "*água no pulmão*", "*agente cancerígeno*", "*ajuda*",

"*alerta*", "*armadilha*", "*aroma*", "*ave*", "*consequências*", "*consumo*", "*criancinha*", "*criança*", "*câncer*", "*dependência*", "*desenfreado*", "*doente*", "*doença*", "*edema pulmonar*", "*embolia pulmonar*", "*epidemia*", "*escola*", "*evali*", "*falta de ar*", "*faz mal*", "*febre*", "*fogo*", "*fumantes*", "*hospital*", "*infantil*", "*infarto*", "*inflamação*", "*internado*", "*internação*", "*intubado*", "*intubação*", "*jovem*", "*jovens*", "*largar o cigarro*", "*larguei*", "*lesionado*", "*lesão*", "*livre*", "*mal*", "*malefícios*", "*matar*", "*morrer*", "*morte*", "*nunca mais fumei*", "*não fumem*", "*não é saudável*", "*parar de fumar*", "*pare de fumar*", "*parem de fumar*", "*perfuração pulmonar*", "*perigo*", "*perigoso*", "*porcaria*", "*proibir*", "*proibição*", "*pulmonar*", "*pulmão*", "*raiva*", "*reportagem*", "*repórter*", "*risco*", "*saúde*", "*tontura*", "*tosse*", "*tossindo*", "*veneno*", "*vicia*", "*vomitando*", "*vomitei*", "*vomitou*", "*vício*", "*vômito*".

Da mesma forma, seguindo as mesmas diretrizes, a rotulagem foi feita nessa amostra adicional. Essa estratégia permitiu aumentar a representação de comentários que expressam uma posição *contrária* em relação ao uso do cigarro eletrônico, permitindo o treinamento de um modelo com maior capacidade de identificar e classificar adequadamente o posicionamento dos comentários. É importante destacar que os comentários selecionados possuem identificadores únicos, desta forma não houve a inclusão de comentários duplicados, entretanto pode ocorrer que existam mensagens semelhantes, o que é comum nas mídias sociais, mas isso não compromete a integridade da nossa coleção detectar os posicionamentos. Em suma, foram rotulados 3369 comentários, cuja distribuição de rótulos consiste em 755 *contrários*, 915 *inconclusivos* e 1699 *favoráveis*. Em casos em que o comentário foi rotulado de forma diferente por pelo menos um dos voluntários, definimos o rótulo através da classe majoritária, ou seja, a classe que recebeu mais votos entre os voluntários. Posteriormente, o coeficiente de concordância Kappa de Fleiss foi calculado [17], resultando em um valor de 0.76, indicando um nível de *concordância substancial* entre os rotuladores.

Para a tarefa de detecção de posicionamentos, utilizamos o modelo BERTimbau *Large*, pré-treinado em português e considerado um dos estados da arte para o idioma, realizando o ajuste fino para esta tarefa específica com base nos dados rotulados [42]. O ajuste fino é um processo iterativo de treinamento que envolve treinar um modelo BERT pré-treinado em um conjunto de dados rotulado. Durante esse processo, o modelo incorpora seu conhecimento linguístico existente com o conhecimento específico da tarefa, ajustando seus parâmetros. Esse procedimento melhora o desempenho do modelo na tarefa alvo ao aproveitar tanto seu conhecimento linguístico geral quanto as adaptações específicas da tarefa. Neste caso, o objetivo é melhorar a separabilidade das classes de posicionamentos nos comentários e criar uma rede capaz de inferir se o posicionamento de um comentário é *favorável*, *contrário* ou *inconclusivo* em relação ao uso de cigarros eletrônicos.

Seguindo as recomendações da literatura [11, 20], realizamos uma busca de hiperparâmetros para ajustar o modelo [44]. Os parâmetros variáveis incluíram taxa de aprendizado (1e-5, 2e-5, 3e-5, 4e-5, 5e-5), probabilidade de *dropout* (0.1, 0.2, 0.3, 0.4), decaimento de peso (0.01, 0.1, 1) e até 10 épocas. Utilizamos validação cruzada de 5 partições (*5-fold cross-validation*) para determinar os melhores parâmetros e avaliar a habilidade de generalização do modelo treinado. Através desse processo, identificamos que os melhores parâmetros eram uma taxa de aprendizado de 2e-5, probabilidade de *dropout* de 0.3, e decaimento de peso de 0.1, 2 épocas, considerando lotes de tamanho 4. O modelo final foi então treinado utilizando essas

⁵Não é possível garantir que os comentários selecionados não sejam relacionados a um mesmo canal ou vídeo, pois estratificar em dois níveis, como data e vídeo, resultaria em um número excessivo de comentários a serem rotulados.

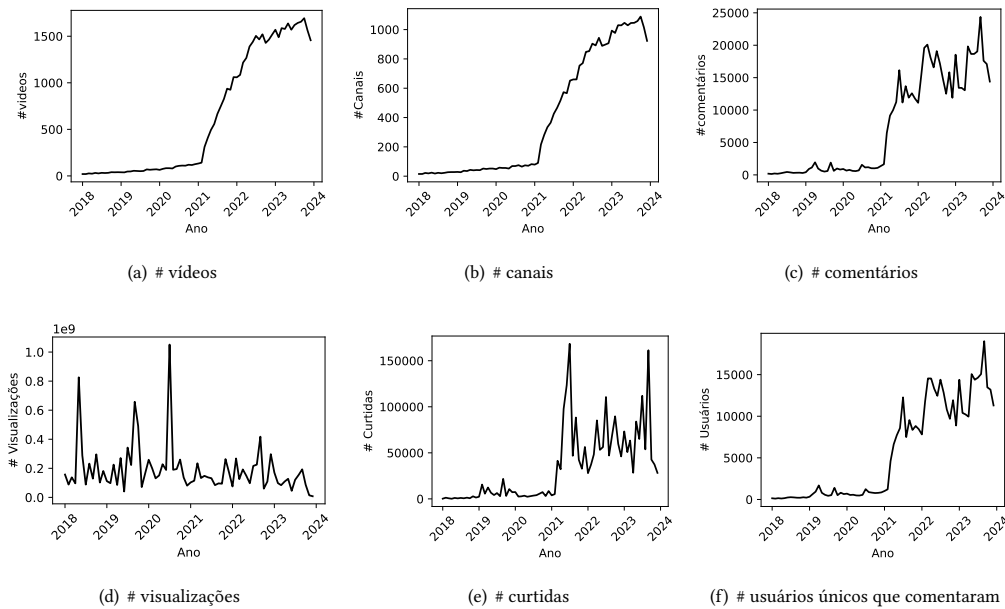


Figura 1: Caracterização das métricas de popularidade e engajamento ao longo do tempo.

configurações, e reportamos a média e o intervalo de confiança de 95% das métricas de validação cruzada ao longo das 5 partições.

4 RESULTADOS

Nesta seção, apresentamos nossos resultados, incluindo a caracterização do conteúdo, engajamento, e a análise de posicionamento.

4.1 Produção de Conteúdo e Engajamento

Iniciamos nossa análise mensurando a dinâmica da produção de conteúdo (Figuras 1(a), 1(b) e 1(c)), bem como o engajamento (Figuras 1(d), 1(e) e 1(f)) durante o período analisado. Observando o número de vídeos criados (Figura 1(a)), notamos um aumento expressivo na quantidade de vídeos produzidos sobre cigarros eletrônicos a partir de 2021. Esse crescimento contínuo reflete um aumento no interesse pelo tema, com um pico notável em 2023. A tendência ascendente sugere que a discussão e o compartilhamento de informações sobre cigarros eletrônicos se tornaram mais comuns ao longo do tempo. Da mesma forma, o número de canais (Figura 1(b)) que publicam conteúdo relacionado ao tema também apresentou um crescimento consistente, especialmente a partir de 2021. Esse aumento no número de canais indica que mais criadores de conteúdo têm explorado o tópico, contribuindo para uma maior diversidade de opiniões e informações disponíveis sobre cigarros eletrônicos. Já o número de comentários (Figura 1(c)) segue a mesma tendência de crescimento, com um aumento substancial a partir de 2021, indicando um maior engajamento dos espectadores com o conteúdo. Esse crescimento no número de comentários reflete uma maior interação entre os usuários e os vídeos, sugerindo que o tema tem gerado mais discussões e debates entre o público. Vale ressaltar, que os resultados apresentados nesta seção, referem-se ao conjunto completo dos dados, englobando todas as informações coletadas durante o período analisado.

Além da produção de conteúdo, também analisamos métricas de popularidade e engajamento. O número de visualizações (Figura 1(d)) apresenta picos periódicos, especialmente em 2018 e 2020, indicando momentos de maior interesse pelo tema. Esses picos podem estar associados a eventos específicos, como lançamento de produtos, cobertura midiática ou outras questões relacionadas aos cigarros eletrônicos. O número de curtidas (Figura 1(e)) já mostra uma tendência de crescimento, com um aumento expressivo a partir de 2021. Isso sugere que os vídeos passaram a atrair mais avaliações positivas dos espectadores, evidenciando uma aprovação crescente do conteúdo. Por fim, o número de usuários únicos que comentam (Figura 1(f)) também cresceu substancialmente a partir de 2021. Esse aumento indica que mais indivíduos estão se interessando e participando mais das discussões sobre cigarros eletrônicos, ampliando o alcance e a diversidade das opiniões expressas nos comentários.

Comparados aos dados das pesquisas da Covitel [31] e da Inteligência em Pesquisa e Consultoria Estratégica (IPEC) [14], que mostram um crescimento no uso, experimentação e conhecimento do cigarro eletrônico em regiões específicas do país, nossos resultados sugerem que o aumento do conhecimento sobre o tema pode ser ainda maior do que o estimado. Todas as métricas de popularidade e engajamento analisadas mostram um crescimento maior do que o apresentado pela pesquisa ao longo do tempo, especialmente a partir de 2021. Esse aumento na produção de conteúdo e no engajamento do público, evidencia a crescente relevância do tema dos cigarros eletrônicos na plataforma YouTube. A análise dessas métricas sugere que a discussão sobre cigarros eletrônicos têm se intensificando, com uma quantidade maior de vídeos, canais, visualizações, curtidas e comentários, refletindo um interesse crescente e uma interação mais ampla com o tema.

Visando compreender o interesse pelas marcas de cigarro eletrônico e produtos relacionados, analisamos o número de menções nos comentários ao longo do tempo. No mapa de calor da Figura

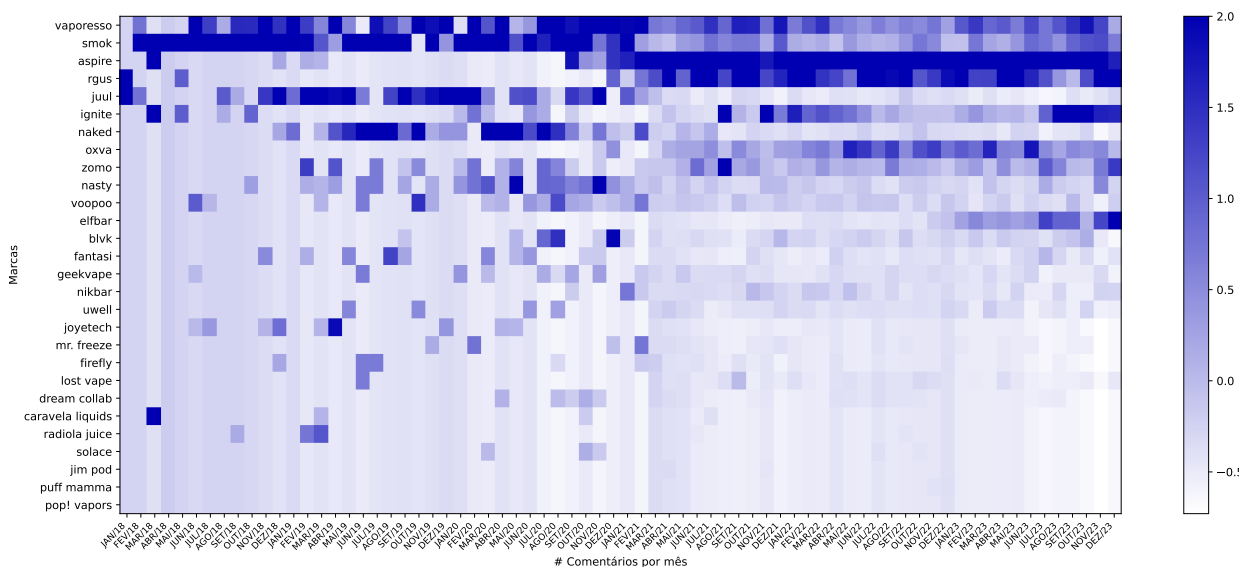


Figura 2: Presença de marcas ao longo dos anos ordenado pela popularidade total.

Tabela 2: Desempenho do modelo ajustado.

Classe	P	R	Macro F1
Contrário	0.71 ± 0.02	0.81 ± 0.03	0.76 ± 0.01
Inconclusivo	0.79 ± 0.03	0.73 ± 0.02	0.76 ± 0.02
Favorável	0.85 ± 0.01	0.83 ± 0.03	0.84 ± 0.01

2, o eixo x representa os meses ao longo do período de 2018 a 2023, enquanto o eixo y lista as diferentes marcas relacionadas ao cigarro eletrônico, ordenadas pela popularidade, ou seja, o número de comentários que fazem menções a elas. Cada célula indica o número de comentários mensais, normalizados por coluna por meio da normalização z -score⁶. Dessa forma, nosso objetivo é mostrar, a cada mês, quais foram as marcas mais populares e mencionadas. A intensidade da cor em cada célula reflete a quantidade de comentários normalizada, em que células mais escuras indicam um maior número de comentários.

Em geral, é possível notar diferentes padrões caracterizados por marcas constantemente populares, outras emergentes e outras pouco populares. Por exemplo, as marcas *Vaporesso* e *Smok* mantêm uma presença consistente no mercado desde 2018, enquanto marcas como *Aspire*, *Argus*, *Ignite* e *Elfbar* começam a intensificar sua popularidade a partir de 2021. Isso sugere que existe um mercado estabelecido e uma crescente diversidade de interesse por diferentes produtos ao longo do tempo. Esses padrões distintos revelam a longevidade de certas marcas, bem como a dinâmica de entrada e crescimento de novas marcas no mercado de cigarros eletrônicos.

4.2 Detecção e Análise de Posicionamentos

Relembre que um dos nossos objetivos é inferir os posicionamentos dos usuários e a percepção da população por meio dos comentários. Seguindo a metodologia explicada na Seção 3, o modelo foi ajustado para identificar o posicionamento com base nos dados rotulados. A Tabela 2 mostra o desempenho final do modelo considerando as métricas Precisão (P), Revocação (R) e Macro F1-escore (F1) para cada classe, com o intervalo de confiança (I.C.) de 95% nas 5 partições.

O modelo apresenta um bom desempenho na classe *contrária*, com uma precisão média de 0.71, revocação de 0.81 e F1-escore de 0.76. Isso indica que o modelo é mais confiável na detecção de comentários contrários, com alta revocação sugerindo que a grande maioria dos comentários contrários foram identificados corretamente. É importante destacar que identificar comentários contrários é mais difícil, pois a maioria do conteúdo no YouTube sobre o mundo *vape* é voltada para propaganda, análises e uso de produtos.

Para a classe *inconclusiva*, o modelo alcançou uma precisão média de 0.79, revocação de 0.73 e F1-escore de 0.76. Embora a precisão seja alta, a revocação ligeiramente menor sugere que alguns comentários inconclusivos podem não ter sido capturados, o que é esperado devido à natureza ambivalente desses comentários. Essa classe é inerentemente desafiadora para o modelo, pois os comentários frequentemente não apresentam um posicionamento claro. Na classe *favorável*, o modelo teve o melhor desempenho, com uma precisão média de 0.85, revocação de 0.83 e F1-escore de 0.84, indicando uma alta capacidade do modelo para detectar comentários favoráveis com equilíbrio entre precisão e revocação. Em resumo, o modelo demonstrou um desempenho sólido em todas as classes, com métricas de F1-escore indicando boa precisão e revocação, e

⁶https://en.wikipedia.org/wiki/Standard_score

intervalos de confiança pequenos, com pouca variação nos resultados, garantindo a confiabilidade para inferir os posicionamentos dos usuários. Ressalta-se, ainda, que estudos de detecção de posicionamento, especialmente, em português, têm reportado valores comumente inferiores [32, 37].

Para mostrar a dificuldade dessa tarefa, alguns comentários classificados incorretamente são apresentados na Tabela 3. Esses exemplos apresentam situações onde os comentários exibem sarcasmo ou não indicam uma opinião clara, dificultando a classificação correta pelo modelo. Por exemplo, o primeiro comentário foi classificado como *contrário* pelo modelo devido ao uso de palavras de baixo calão e menção a sintomas negativos relacionados ao uso do cigarro ("pulmão preto"), embora o comentário sugira que o cigarro eletrônico não faça mal à saúde em comparação ao cigarro tradicional, expressando assim uma opinião *favorável*. O segundo comentário mostra um usuário interessado em adquirir um produto, mas é pouco discriminativo, pois apenas a menção da marca indica que o comentário se refere à aquisição, e por tanto, interesse no produto. Já o terceiro comentário não deixa claro se o usuário está defendendo o uso do cigarro eletrônico ou do cigarro convencional, sendo classificado como *contrário* pelo modelo, possivelmente devido aos termos "médico" e "câncer". O próximo exemplo consiste em uma pergunta que não deixa claro se é uma ironia, resultando em um rótulo de *inconclusivo*. No entanto, o modelo o classificou como *favorável*, possivelmente devido ao grande volume de comentários argumentando que o cigarro eletrônico é similar ao processo de inalação por nebulizadores. O último exemplo é um comentário que possui pouco significado semântico e contém ironia, o que contribuiu para sua classificação incorreta.

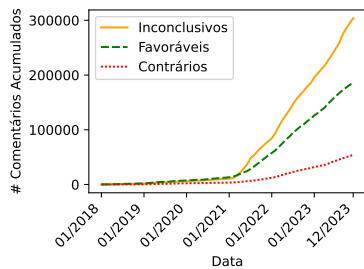


Figura 3: Análise temporal dos comentários por classe.

Com o modelo ajustado e validado por meio do ajuste fino e validação cruzada, estimamos a tendência temporal acumulada dos comentários para as três classes de posicionamento previstas pelo modelo ao longo do período analisado, como apresentado na Figura 3. A partir de 2021, observa-se um aumento expressivo nos comentários de todas as classes, indicando uma intensificação do interesse e participação do público brasileiro. A grande quantidade de comentários *inconclusivos*, apesar de não representarem um posicionamento claro, indica que muitas pessoas estão interessadas na discussão sobre o tema. Nota-se ainda um aumento muito maior e mais rápido de comentários *favoráveis* do que *contrários*. Esse fenômeno pode sugerir uma aceitação crescente dos cigarros eletrônicos entre os usuários do YouTube, que é a principal plataforma de conteúdo de vídeo do Brasil.

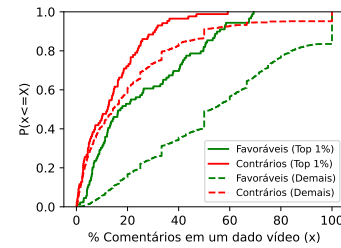


Figura 4: CDF do percentual de comentários favoráveis e contrários nos vídeos mais (Top 1%) e menos populares (Demais).

Para analisar a prevalência de *favoráveis* e *contrários* nos vídeos mais populares em termos do número de comentários recebidos (top 1%) e nos demais, calculamos a Função de Distribuição Acumulada do percentual de comentários *favoráveis* e *contrários* nos vídeos, apresentada na Figura 4. Observamos que, para os vídeos mais populares (Top 1%), o percentual de comentários *favoráveis* tende a ser maior do que o percentual de comentários *contrários*. Focando nos top 1%, observamos que para 70% dos vídeos, até 40% dos comentários são *favoráveis*, enquanto o percentual de comentários *contrários* é de aproximadamente 20%. Para os vídeos menos populares (Demais), essa diferença é ainda maior, com 70% dos vídeos apresentando até 80% de comentários como *favoráveis*, enquanto o percentual de comentários *contrários* para os mesmos 70% dos vídeos é de no máximo 30%. Em geral, os vídeos mais populares, que recebem milhares de comentários, tendem a receber comentários *favoráveis*. Esses vídeos podem ser tanto de conteúdo originalmente favorável ao tema, como por exemplo, *reviews* de produtos relacionados, quanto reportagens, onde conteúdo com argumentos contrários são apresentados. Nos vídeos menos populares, que representam 99% da base de dados, essa discrepância é ainda maior, com uma predominância expressiva de comentários *favoráveis*.

Investigamos também os canais que mais produziram vídeos sobre o cigarro eletrônico e identificamos os cinco principais: *Mike Vapes*, *VapersBrazil*, *El Mono Vapeador*, *Wendy V Reviews* e *VOOPOO*. Em geral, esses canais compartilham experiências com o produto. Em relação aos canais que mais geraram comentários *favoráveis*, destacam-se: *Zona do Vapor*, *VapersBrazil*, *Olá, Ciência!* e *Paulo Jubilut*. Os canais *Zona do Vapor* e *VapersBrazil* são conhecidos por apresentar *reviews* de diferentes modelos e sabores, atraindo um público que busca informações positivas e recomendações. Já os canais *Olá, Ciência!* e *Paulo Jubilut* abordam o cigarro eletrônico com uma perspectiva científica e técnica, recebendo comentários que frequentemente defendem o produto e, portanto, *favoráveis*. Por outro lado, os canais que mais receberam comentários *contrários* são: *Paulo Jubilut*, *Olá, Ciência!*, *Zona do Vapor*, *Drauzio Varella* e *João Caetano*. Esses comentários *contrários* vêm geralmente de usuários que acessam os vídeos para criticar o uso do cigarro eletrônico e/ou discutir os riscos à saúde. É importante destacar também que alguns desses canais, como *Paulo Jubilut* e *Olá, Ciência!*, recebem tanto comentários *favoráveis* quanto *contrários*, refletindo uma batalha de debates sobre o uso do produto.

Para compreender os termos mais usados em comentários *favoráveis* e *contrários*, analisamos nuvens de palavras dos cinco vídeos com maior percentual de cada classe. A Figura 5(a) mostra a nuvem

Tabela 3: Exemplos de comentários classificados incorretamente.

Comentário	Real	Predita
Vai tomar no ** bando de burro os cara da Record eu uso cigarro eletrônico a 2 anos e meu pulmão ta de boa e meu amigo usa cigarro já ta com pulmão preto	Favorável	Contrário
Gente a entrega no site da ignite Brasil demora muito?	Favorável	Inconclusivo
Eu ja não ligo mais, tenho câncer que ta em ultimo estados, os médicos me deram 2 meses, vou usar ate não conseguir mais andar	Inconclusivo	Contrário
E se eu colocar soro em vez de nicotina vai ser como se eu tivesse fazendo inalação?	Inconclusivo	Favorável
O povo só inventa merda pra se matar kkkkkkkk	Contrário	Inconclusivo

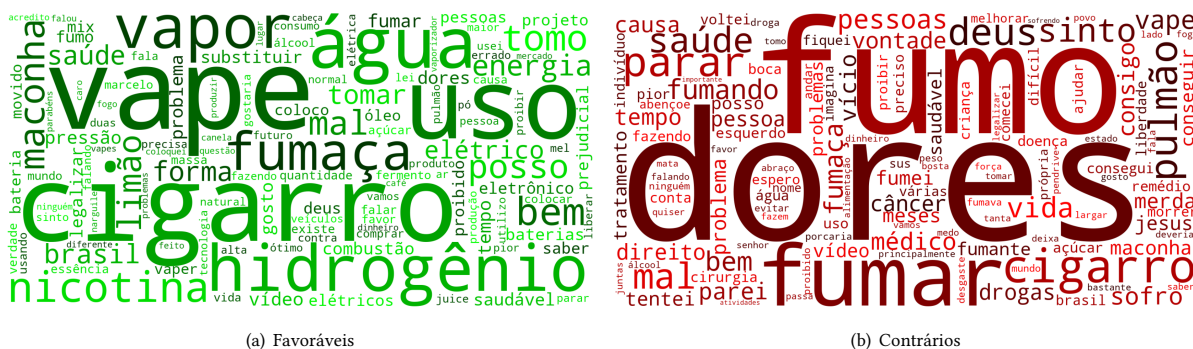


Figura 5: Nuvens de palavras com os top 5 vídeos com maior percentual de comentários de cada classe.

de palavras de comentários *favoráveis*, onde se destacam termos como "uso", "vape", "bem", "fumaça" e "vapor". Esses termos indicam uma associação direta entre a defesa do produto e sua prática de consumo, sugerindo que o uso do cigarro eletrônico é benéfico à saúde porque não produz fumaça, mas sim vapor. Por outro lado, a Figura 5(b) apresenta a nuvem de palavras dos cinco vídeos com maior percentual de comentários *contrários*. Entre os termos mais frequentes estão "dores", "vício", "pulmão", "câncer", "mal" e "saúde", refletindo os relatos e preocupações com os efeitos adversos à saúde. Esses termos indicam que os usuários que expressam posições contrárias ao uso dos cigarros eletrônicos estão cientes dos potenciais danos à saúde, contrastando com as opiniões favoráveis observadas anteriormente.

Em resumo, a análise de posicionamentos revela importantes implicações sociais sobre a percepção pública do cigarro eletrônico. Observamos um crescimento notável na produção de conteúdo e no engajamento dos usuários, refletindo um crescente interesse pelo tema. A predominância de comentários *favoráveis* sugere um forte movimento de defesa e promoção dos cigarros eletrônicos, mesmo em canais informativos que também recebem muitos comentários *contrários*. Esses canais, muitas vezes destinados a discutir os riscos à saúde, acabam sendo alvos de defensores do produto, mostrando uma divisão acirrada nas opiniões. As nuvens de palavras evidenciam essa divisão, com termos relacionados à defesa do uso do cigarro eletrônico, como "vapor", "uso" e "bem", contrastando com termos associados a preocupações com a saúde, como "dores", "pulmão" e "mal".

5 CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho teve como objetivo investigar a popularidade e a aceitação dos cigarros eletrônicos no Brasil por meio da análise de comentários em vídeos conteúdos no YouTube. Considerando a literatura existente, este trabalho oferece uma análise temporal mais abrangente e detalhada do debate em torno desse produto no contexto brasileiro. Nossos principais resultados revelam um aumento significativo na produção de conteúdo e no engajamento dos usuários, refletindo um crescente interesse do público pelos cigarros eletrônicos. O modelo de detecção de posicionamento desenvolvido identificou uma prevalência maior de comentários *favoráveis* do que *contrários* nos vídeos coletados, indicando uma aceitação crescente desses produtos entre os usuários da plataforma. Além disso, a análise temporal destacou a intensificação do debate a partir de 2021, com um aumento substancial nos comentários de todas as classes analisadas, sugerindo um maior engajamento, interesse e polarização das opiniões nos últimos anos. A importância desses achados consiste em uma compreensão mais profunda e inédita das percepções públicas sobre cigarros eletrônicos no Brasil em plataformas de vídeos online.

Como trabalhos futuros, planejamos replicar e comparar o estudo com dados de outros países, a fim de identificar possíveis diferenças e semelhanças em função de aspectos culturais e regulamentários. Além disso, investigar a prevalência de desinformação sobre cigarros eletrônicos junto à uma equipe multidisciplinar.

REFERÊNCIAS

- [1] Shishir Adhikari, Akshay Uppal, Robin Mermelstein, Tanya Berger-Wolf, and Elena Zheleva. 2021. Understanding the dynamics between vaping and cannabis legalization using Twitter opinions. In *Proceedings of the International AAAI Conference on Web and Social Media*.
- [2] ANVISA. 2009. Cigarro eletrônico. (2009). <https://www.gov.br/anvisa/pt-br/assuntos/tabaco/cigarro-eletronico> Acesso: 24-04-2024.
- [3] ANVISA. 2024. Anvisa atualiza regulação de cigarro eletrônico e mantém proibição. (2024). <https://www.gov.br/anvisa/pt-br/assuntos/noticias-anvisa/2024/anvisa-atualiza-regulacao-de-cigarro-eletronico-e-mantem-proibicao> Acesso: 21-05-2024.
- [4] Emily Banks, Amelia Yazidjoglou, Sinan Brown, Mai Nguyen, Melonie Martin, Katie Beckwith, Amanda Daluwatta, Sai Campbell, and Grace Joshy. 2023. Electronic cigarettes and health outcomes: umbrella and systematic review of the global evidence. *Medical Journal of Australia* 218, 6 (2023), 267–275.
- [5] Jayesh Mahendra Bhatt, Manisha Ramphul, and Andrew Bush. 2020. An update on controversies in e-cigarettes. *Paediatric respiratory reviews* 36 (2020), 75–86.
- [6] Mouhamad Bigwanto, Farandi Agesti Ramadhan, and Xenia Angelica Wijayanto. 2024. Online e-cigarette promotion in Indonesia. *Tobacco Control* 33, 3 (2024), 406–409.
- [7] Michael Bourke, Naseem Sharif, and Omendra Narayan. 2021. Association between electronic cigarette use in children and adolescents and coughing a systematic review. *Pediatric Pulmonology* 56, 10 (2021), 3402–3409.
- [8] CNN Brasil. 2023. Sem regulamentação, 1 a cada 5 jovens consome cigarro eletrônico no Brasil, diz pesquisa. (2023). <https://www.cnnbrasil.com.br/saude/sem-regulamentacao-1-a-cada-5-jovens-consome-cigarro-eletronico-no-brasil-diz-pesquisa/> Acesso: 19-02-2024.
- [9] Brooke Campus, Patrick Fafard, Jessica St Pierre, and Steven J Hoffman. 2021. Comparing the regulation and incentivization of e-cigarettes across 97 countries. *Social science & medicine* 291 (2021), 114187.
- [10] Lauren Collins, Allison M Glasser, Haneen Abudayyeh, Jennifer L Pearson, and Andrea C Villanti. 2019. E-cigarette marketing and communication: how e-cigarette companies market e-cigarettes and the public engages with e-cigarette information. *Nicotine and Tobacco Research* 21, 1 (2019), 14–24.
- [11] Washington Cunha, Celso França, Guilherme Fonseca, Leonardo Rocha, and Marcos André Gonçalves. 2023. An effective, efficient, and scalable confidence-based instance selection framework for transformer-based text classification. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 665–674.
- [12] Hassan Dashtian, Dhiraj Murthy, and Grace Kong. 2022. An exploration of e-cigarette-related search items on YouTube: network analysis. *Journal of Medical Internet Research* 24, 1 (2022), e30679.
- [13] Katherine East, Sara C Hitchman, Ioannis Bakolis, Sarah Williams, Hazel Cheeseman, Deborah Arnott, and Ann McNeill. 2018. The association between smoking and electronic cigarette use in a cohort of young people. *Journal of Adolescent Health* 62, 5 (2018), 539–547.
- [14] Inteligência em Pesquisa e Consultoria Estratégica. 2023. Cigarros Eletrônicos. (2023). <https://static.poder360.com.br/2024/01/ipecc-pesquisa-cigarros-eletronicos-2023.pdf> Acesso: 05-05-2024.
- [15] Konstantinos E Farsalinos and Raymond Niaura. 2020. E-cigarettes and smoking cessation in the United States according to frequency of e-cigarette use and quitting duration: analysis of the 2016 and 2017 National Health Interview Surveys. *Nicotine and Tobacco Research* 22, 5 (2020), 655–662.
- [16] Widia Resti Fitriani, Arif Budi Mulyono, Achmad Nizar Hidayanto, and Qorib Munajat. 2020. Reviewer's communication style in YouTube product-review videos: does it affect channel loyalty? *Heliyon* 6, 9 (2020).
- [17] Joseph L Fleiss, Bruce Levin, and Myunghee Cho Paik. 2013. *Statistical methods for rates and proportions*. John Wiley & Sons.
- [18] O Globo. 2023. OMS pede ação urgente contra cigarros eletrônicos. (2023). <https://oglobo.globo.com/saude/noticia/2023/12/14/oms-pede-acao-urgente-contra-cigarros-eletronicos-e-diz-que-produtos-nao-sao-eficazes-para-parar-de-fumar.ghtml> Acesso: 19-02-2024.
- [19] Xun Gong, Bo Dong, Li Li, Danping Shen, and Zhiyi Rong. 2023. TikTok video as a health education source of information on heart failure in China: a content analysis. *Frontiers in Public Health* 11 (2023), 1315393.
- [20] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep learning*. MIT press.
- [21] Igor Grabovac, Moritz Oberdorfer, Jimsy Fischer, Winfried Wiesinger, Sandra Haider, and Thomas Ernst Dorner. 2021. Effectiveness of electronic cigarettes in smoking cessation: a systematic review and meta-analysis. *Nicotine and Tobacco Research* 23, 4 (2021), 625–634.
- [22] Lamiece Hassan, Mohab Elkaref, Geeth de Mel, Ilze Bogdanovica, and Goran Nenadic. 2022. Text mining tweets on e-cigarette risks and benefits using machine learning following a vaping related lung injury outbreak in the USA. *Healthcare Analytics* 2 (2022), 100066.
- [23] Eslam Hussein, Prerna Juneja, and Tanushree Mitra. 2020. Measuring misinformation in video search platforms: An audit study on YouTube. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–27.
- [24] IBOPE. 2023. Video Audience Share Percentage in Brazil. <https://kantaribopemedia.com/conteudo/relatorios/april-2023/>.
- [25] Jonas Kaiser, Adrian Rauchfleisch, and Yasodara Córdova. 2021. Comparative approaches to mis/disinformation| fighting Zika with honey: An analysis of YouTube's video recommendations on Brazilian YouTube. *International Journal of Communication* 15 (2021), 19.
- [26] Ryan David Kennedy, Ayodeji Awopegba, Elaine De León, and Joanna E Cohen. 2017. Global approaches to regulating electronic cigarettes. *Tobacco control* (2017).
- [27] D Eisenkraft Klein, Michael Chaiton, A Kundu, and R Schwartz. 2020. A literature review on international e-cigarette regulatory policies. *Current Addiction Reports* 7 (2020), 509–519.
- [28] Grace Kong, Heather LaVallee, Alissa Rams, Divya Ramamurthi, and Suchitra Krishnan-Sarin. 2019. Promotion of vape tricks on YouTube: Content analysis. *Journal of medical Internet research* 21, 6 (2019), e12709.
- [29] Juhan Lee, Sakinah C Suttiratana, Isha Sen, and Grace Kong. 2023. E-cigarette marketing on social media: a scoping review. *Current Addiction Reports* 10, 1 (2023), 29–37.
- [30] Zachary B Massey, Laurel O Brockenberry, and Paul T Harrell. 2021. Vaping, smartphones, and social media use among young adults: Snapchat is the platform of choice for young adult vapers. *Addictive behaviors* 112 (2021), 106576.
- [31] Ana Maria Baptista Menezes, Fernando C Wehrmeister, Luciana Monteiro Vasconcelos Sardinha, Pedro do Carmo Baumgratz de Paula, Tainá de Almeida Costa, Pedro Augusto Crespo, and Pedro C Hallal. 2023. Use of electronic cigarettes and hookah in Brazil: a new and emerging landscape. The Covitel study, 2022. *Jornal Brasileiro de Pneumologia* 49 (2023), e20220290.
- [32] Lin Miao, Mark Last, and Marina Litvak. 2022. Tracking social media during the COVID-19 pandemic: The case study of lockdown in New York State. *Expert Systems with Applications* 187 (2022), 115797.
- [33] Dhiraj Murthy, Juhan Lee, Hassan Dashtian, Grace Kong, et al. 2023. Influence of User Profile Attributes on e-Cigarette-Related Searches on YouTube: Machine Learning Clustering and Classification. *JMIR infodemiology* 3, 1 (2023), e42218.
- [34] The Lancet Oncology. 2013. Time for e-cigarette regulation. , 1027 pages.
- [35] World Health Organization et al. 2023. WHO report on the global tobacco epidemic, 2023: protect people from tobacco smoke. (2023).
- [36] Wael Osman, Fatma Mohamed, Mohamed Elhassan, and Abdulhadi Shoufan. 2022. Is YouTube a reliable source of health-related information? A systematic review. *BMC Medical Education* 22, 1 (2022), 382.
- [37] Saul Rocha, Carlos Silva, Carlos Ferreira, Glauber Gonçalves, and Jussara Almeida. 2024. Monitorando a Opinião Pública sobre Operações Policiais no Brasil via Comentários de Vídeos no YouTube. In *Anais do XIII Brazilian Workshop on Social Network Analysis and Mining (Brasilina/DF)*.
- [38] Cristin Purnama Sari Kaunang, Fitria Amastini, and Rahmad Mahendra. 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)*.
- [39] Dean E Schraufnagel, Francesco Blasi, M Bradley Drummond, David CL Lam, Ehsan Latif, Mark J Rosen, Raul Sansores, and Richard Van Zyl-Smit. 2014. Electronic cigarettes. A position statement of the forum of international respiratory societies. *American journal of respiratory and critical care medicine* (2014).
- [40] Marissa J Smith, Christina Buckton, Chris Patterson, and Shona Hilton. 2023. User-generated content and influencer marketing involving e-cigarettes on social media: a scoping review and content analysis of YouTube and Instagram. *BMC public health* 23, 1 (2023), 530.
- [41] Samir S Soneji, Hai-Yen Sung, Brian A Primack, John P Pierce, and James D Sargent. 2018. Quantifying population-level health benefits and harms of e-cigarette use in the United States. *PLoS one* 13, 3 (2018), e0193328.
- [42] Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. 2020. BERTimbau: pretrained BERT models for Brazilian Portuguese. In *9th Brazilian Conference on Intelligent Systems, BRACIS, Rio Grande do Sul, Brazil, October 20-23 (to appear)*.
- [43] Chunmeizi Su. 2023. *Douyin, TikTok and China's online screen industry: The rise of short-video platforms*. Taylor & Francis.
- [44] Chi Sun, Xipeng Qiu, Yige Xu, and Xuanjing Huang. 2019. How to Fine-Tune BERT for Text Classification?. In *Chinese Computational Linguistics*.
- [45] Tianze Sun, Carmen CW Lim, Jack Chung, Brandon Cheng, Lily Davidson, Calvert Tisdale, Janni Leung, Coral E Gartner, Jason Connor, Wayne D Hall, et al. 2023. Vaping on TikTok: a systematic thematic analysis. *Tobacco Control* (2023).
- [46] Julia Vassey, Catherine Metayer, Chris J Kennedy, and Todd P Whitehead. 2020. # Vape: measuring e-cigarette influence on Instagram with deep learning and text analysis. *Frontiers in communication* 4 (2020), 75.
- [47] Paola Zola, Costantino Ragno, and Paulo Cortez. 2020. A Google Trends spatial clustering approach for a worldwide Twitter user geolocation. *Information Processing & Management* 57, 6 (2020), 102312.