Universidade Federal do Piauí
Centro de Ciências da Natureza
Programa de Pós-Graduação em Ciência da Computação

Mineração de Padrões Textuais sobre Segurança Pública: Modelos e Aplicações em Plataformas de Mídias Sociais

Saul Sousa da Rocha

Número de Ordem PPGCC: M001

Teresina-PI, Dezembro de 2025

Saul Sousa da Rocha

Mineração de Padrões Textuais sobre Segurança Pública: Modelos e Aplicações em Plataformas de Mídias Sociais

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFPI (área de concentração: Sistemas de Computação), como parte dos requisitos necessários para a obtenção do Título de Mestre em Ciência da Computação.

Universidade Federal do Piauí – UFPI

Centro de Ciências da Natureza

Programa de Pós-Graduação em Ciência da Computação

Orientador: Glauber Dias Gonçalves

Coorientador: Carlos Henrique Gomes Ferreira

Teresina-PI

Dezembro de 2025

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Biblioteca Comunitária Jornalista Carlos Castello Branco
Divisão de Representação da Informação

R672m Rocha, Saul Sousa da.
Mineração de padrões textuais sobre segurança pública : modelos e aplicações em plataformas de mídias sociais / Saul Sousa da Rocha. -- 2025.
80 f.

Dissertação (Mestrado) – Universidade Federal do Piauí, Centro de Ciências da Natureza, Programa de Pós-Graduação em Ciências da Computação, Teresina, 2025.
"Orientador: Glauber Dias Gonçalves
Coorientador: Carlos Henrique Gomes Ferreira"

1. Aprendizado de máquinas. 2. Reconhecimento de padrões. 3. Mineração de dados. 4. Mídias sociais. I. Gonçalves, Glauber Dias. II. Ferreira, Carlso Henrique Gomes. III. Título.

CDD 006.31

Bibliotecária: Milane Batista da Silva – CRB3/1005

Saul Sousa da Rocha

Mineração de Padrões Textuais sobre Segurança Pública: Modelos e Aplicações em Plataformas de Mídias Sociais

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFPI (área de concentração: Sistemas de Computação), como parte dos requisitos necessários para a obtenção do Título de Mestre em Ciência da Computação.

Documento assinado digitalmente
gov.br GLAUBER DIAS GONCALVES
Data: 13/01/2026 10:27:06-0300
Verifique em https://validar.iti.gov.br

Glauber Dias Gonçalves (UFPI)
Orientador

Documento assinado digitalmente
gov.br CARLOS HENRIQUE GOMES FERREIRA
Data: 13/01/2026 10:59:44-0300
Verifique em https://validar.iti.gov.br

Carlos Henrique Gomes Ferreira(UFOP)
Co-Orientador

Documento assinado digitalmente
gov.br RAIMUNDO SANTOS MOURA
Data: 13/01/2026 14:52:14-0300
Verifique em https://validar.iti.gov.br

Professor/Pesquisador
Raimundo Santos Moura (UFPI)

Documento assinado digitalmente
gov.br VINICIUS DA FONSECA VIEIRA
Data: 15/01/2026 18:04:29-0300
Verifique em https://validar.iti.gov.br

Professor/Pesquisador
Vinícius Fonseca Vieira (UFSJ)

Firmato digitalmente da: Idilio Drago
Organizzazione: UNIVERSITA' DEGLI STUDI DI TORINO/02099650010
Limitazioni d'uso: Explicit Text: I titolari fanno uso del certificato solo per le finalità di lavoro per le quali esso è rilasciato. The certificate holder must use the certificate only for the purposes for which it is issued.
Data: 24/01/2026 22:57:18
Idílio Drago (UNITO)

Teresina-PI
Dezembro de 2025

Resumo

A crescente produção de conteúdo textual e audiovisual em plataformas de mídias sociais tem ampliado as possibilidades de investigação sobre a percepção pública em temas sensíveis, como violência urbana e operações policiais no Brasil, ao mesmo tempo em que introduz desafios associados à informalidade linguística, à ausência de bases anotadas e à forte heterogeneidade entre plataformas. Esta dissertação apresenta um arcabouço integrado para análise dessa percepção pública, estruturado em três eixos complementares que abordam desde a inferência de posicionamento em comentários até a identificação automática de eventos reais em vídeos de violência urbana. No primeiro eixo, desenvolvemos e avaliamos modelos de detecção de posicionamento aplicados a comentários do YouTube, demonstrando que arquiteturas baseadas em *transformers*, especialmente o BERTimbau, superam abordagens tradicionais mesmo diante de textos ruidosos e heterogêneos. No segundo eixo, ampliamos a análise para um cenário interplataformas, utilizando dados do X/Twitter e do YouTube e introduzindo uma estratégia de anotação híbrida que combina rótulos humanos com validação automática por LLMs, garantindo maior consistência semântica e confiabilidade nas anotações. Os experimentos revelam que modelos treinados no X/Twitter generalizam melhor para o YouTube do que o inverso, evidenciando diferenças estruturais entre ecossistemas discursivos e reforçando a importância de considerar especificidades linguísticas no desenvolvimento de modelos de PLN em português brasileiro. No terceiro eixo, avançamos da análise textual para a organização de vídeos por evento real, propondo duas heurísticas não supervisionadas baseadas exclusivamente em atributos textuais, uma semântica e outra temporal. Os resultados mostram que a heurística temporal, fundamentada em janelas de publicação e recorrência de entidades nomeadas, supera tanto a heurística semântica quanto uma linha de base baseada em GPT-4, alcançando acurácia próxima de 0,90 e NMI superior a 0,95 em cenários de monitoramento contínuo. As contribuições desta dissertação incluem novas bases de dados anotadas em português, heurísticas originais para agrupamento de vídeos, um estudo sistemático sobre transferência entre plataformas e um arcabouço completo para monitoramento de percepções sobre segurança pública em larga escala. A principal contribuição deste trabalho é metodológica, por meio de um arcabouço de mineração textual aplicado à percepção sobre segurança pública. Os resultados demonstram que abordagens leves, adaptadas ao domínio e apoiadas por estratégias de anotação híbrida, podem superar métodos genéricos de última geração, reforçando o potencial das mídias sociais como fonte estruturada de conhecimento e fornecendo subsídios metodológicos e práticos para pesquisas futuras e aplicações voltadas à análise e monitoramento de segurança no Brasil.

Palavras-chaves: PLN, Detecção de Posicionamentos, Operações Policiais, Monitoramento, Mídias Sociais.

Abstract

The growing production of textual and audiovisual content on social media platforms has expanded the possibilities for investigating public perception on sensitive topics such as urban violence and police operations in Brazil, while simultaneously introducing challenges related to linguistic informality, the lack of annotated datasets, and strong cross-platform heterogeneity. This dissertation presents an integrated framework for analyzing public perception, structured into three complementary axes that span from stance inference in comments to the automatic identification of real-world events in videos depicting urban violence. In the first axis, we develop and evaluate stance detection models applied to YouTube comments, showing that transformer-based architectures, especially BERTimbau, outperform traditional approaches even in the presence of noisy and heterogeneous texts. In the second axis, we extend the analysis to a cross-platform setting using data from X/Twitter and YouTube, and introduce a hybrid annotation strategy that combines human labels with automatic validation by large language models (LLMs), thereby increasing the semantic consistency and reliability of annotations. Experiments reveal that models trained on X/Twitter generalize better to YouTube than the reverse, highlighting structural differences between discursive ecosystems and reinforcing the importance of accounting for linguistic specificities when developing NLP models for Brazilian Portuguese. In the third axis, we move from textual analysis to the organization of videos by real-world event, proposing two unsupervised heuristics based solely on textual attributes, one semantic and the other temporal. Results show that the temporal heuristic, grounded in publication windows and recurrence of named entities, outperforms both the semantic heuristic and a GPT-4-based baseline, achieving accuracy close to 0.90 and NMI above 0.95 in continuous monitoring scenarios. The contributions of this dissertation include new annotated datasets in Portuguese, original heuristics for video clustering, a systematic study of cross-platform transfer, and a complete framework for large-scale monitoring of public perceptions about security. The main contribution of this work is methodological, through a text mining framework applied to the analysis of public perceptions about security. Overall, the findings demonstrate that lightweight, domain-adapted approaches supported by hybrid annotation strategies can outperform generic state-of-the-art methods, reinforcing the potential of social media as a structured source of knowledge and providing methodological and practical insights for future research and applications focused on security analysis and monitoring in Brazil.

Keywords: NLP, Monitoring, Social Media, Stance Detection, Police Operations.

Lista de ilustrações

Lista de tabelas

Sumário

1 Introdução

A segurança pública é essencial para o bem-estar das pessoas em centros urbanos, e tanto sua manutenção quanto sua melhoria apresentam desafios significativos para os governos (RICARDO; SIQUEIRA; MARQUES, 2013). Em muitos países em desenvolvimento, como o Brasil, onde os índices de criminalidade são particularmente elevados, esses desafios tornam-se ainda mais evidentes. De acordo com o *Global Peace Index* de 2025, o Brasil ocupa a 130ª posição entre 163 países avaliados, destacando-se negativamente por suas altas taxas de criminalidade e homicídios, especialmente em áreas urbanas densamente povoadas, ficando na 9ª posição da América do Sul (GLOBAL. . . , 2025). No enfrentamento da violência urbana, as operações policiais desempenham um papel central, frequentemente gerando debates sobre sua condução e impacto social.

Com o avanço da comunicação digital, as mídias sociais tornaram-se espaços fundamentais para a expressão pública sobre temas relacionados à segurança. O YouTube, em especial, destaca-se no Brasil como uma das principais plataformas para o consumo de notícias, sendo utilizado por 43% dos entrevistados, segundo o Instituto Reuters (NEWMAN et al., 2022). Vídeos de operações policiais recebem milhares de comentários, refletindo posicionamentos divergentes da população, desde apoio irrestrito até críticas severas (BARRETO LEONARDO, 2023; SBTNEWS, 2025). Essa interação oferece uma fonte valiosa para compreender como o público percebe e reage às ações das forças de segurança.

Vídeos oficiais publicados sobre tais eventos costumam receber comentários que expressam aprovação ou desaprovação sobre a atuação policial, permitindo análises sobre a percepção pública em nível nacional ou regional. Paralelamente ao agravamento dos indicadores de criminalidade, observa-se uma crescente digitalização das manifestações de violência (CHEN et al., 2025). A produção massiva de conteúdo em vídeo nas plataformas de mídia social, especialmente no YouTube, consolidou esse formato como meio dominante para comunicação e disseminação de informações (YOUSEFI; SHAIK; AGARWAL, 2024; PEREIRA et al., 2022). A popularidade da plataforma a tornou um repositório valioso de dados sobre eventos sociais e de segurança pública, incluindo confrontos armados, protestos e emergências (OTTONI et al., 2018). Esse acervo possibilita investigações aprofundadas sobre frequência, dinâmica e repercussão de eventos de violência urbana (WEAVER; ZELENKAUSKAITE; SAMSON, 2012).

Nesse sentido, a tarefa de detecção de posicionamento tem ganhado relevância. Essa tarefa visa inferir se um autor expressa uma posição favorável, desfavorável ou neutra em relação a um tópico específico (MOHAMMAD et al., 2016; COSTA et al., 2025), indo além da análise de sentimentos tradicional ao considerar o alvo e o contexto das

opiniões. Estudos recentes aplicaram a detecção de posicionamento e técnicas relacionadas a questões sociopolíticas complexas, incluindo reações à violência policial em Portugal (MATOS; RIBEIRO; FERREIRA, 2025), índices de sentimento ligados à desigualdade em cidades dos EUA (OH; ZHANG; GREENLEAF, 2021), previsão de crimes a partir do X/Twitter (YANG et al., 2024) e debates sobre racismo e brutalidade policial no Facebook (SCOTLAND; THOMAS; JING, 2024). Análises baseadas no YouTube capturaram respostas a casos de grande repercussão, como a morte de George Floyd (NANAYAKKARA, 2024), enquanto estudos brasileiros investigaram posicionamentos em relação a operações policiais no X/Twitter (FEITOSA et al., 2022). Em contextos de segurança pública, essa percepção não é neutra: ela influencia o apoio ou a rejeição a políticas de segurança e pode retroalimentar a própria cobertura midiática de operações policiais, na medida em que reações intensas do público tendem a orientar quais eventos são destacados e como são enquadrados.

Apesar de promissora, a detecção de posicionamento no contexto de violência urbana brasileira enfrenta desafios únicos. Há uma carência de conjuntos de dados rotulados em português, e poucos modelos são adaptados à sua linguagem informal, ruidosa e culturalmente específica (MAIA; SILVA, 2024; HAND; CHING, 2020; CHAPARRO et al., 2020). Sintaxe informal, gírias e sarcasmo complicam ainda mais a análise, especialmente em um cenário político polarizado.

Outra lacuna importante é a investigação limitada do uso de modelos de inferência entre diferentes plataformas. Apesar da crescente importância da transferência de de aprendizagem (*transfer leanrning*), poucos estudos investigam sistematicamente a generalização de modelos de detecção de posicionamento entre plataformas, especificamente, como modelos treinados em um domínio (por exemplo, X/Twitter) se comportam quando implementados em outro (por exemplo, YouTube) (MOZAFARI; FARAHBAKHSH; CRESPI, 2020; YI; ZUBIAGA, 2021; SCHILLER; DAXENBERGER; GUREVYCH, 2021). Isto é particularmente relevante em contextos de poucos recursos, onde práticas de IA sustentáveis, como a reutilização de modelos e a adaptação entre domínios, são essenciais para reduzir os custos de anotação e permitir a implementação escalável (KAPLAN et al., 2025).

O desafio é transformar esse ecossistema em uma fonte estruturada de dados confiáveis. O principal obstáculo está na identificação automática de vídeos que se referem ao mesmo evento real. Vídeos sobre uma mesma operação policial podem ser publicados por diferentes canais, apresentando títulos heterogêneos, descrições incompletas e datas de publicação que não correspondem necessariamente ao momento do ocorrido. Essa ausência de padronização, combinada à informalidade e ao ruído textual, dificulta a associação entre conteúdos correlatos. A literatura especializada tem abordado esse problema com técnicas multimodais que integram informações visuais, auditivas e textuais (WU et al.,

2014; ELHOSEINY et al., 2016; JANSEN et al., 2017). Embora promissoras, essas soluções demandam elevado poder computacional e infraestrutura sofisticada, características que limitam sua adoção em larga escala (MONDAL et al., 2025).

Por outro lado, abordagens baseadas exclusivamente em texto têm demonstrado resultados relevantes. Métodos como o GraphTMT (THIES et al., 2021) mostram que transcrições podem capturar informações suficientes para identificar tópicos, sem a necessidade de processar sinais de áudio ou imagem. Essa tendência é especialmente relevante para o português brasileiro, diante do grande volume de metadados disponíveis no YouTube, como títulos, descrições e datas de publicação. Além disso, a análise de comentários textuais tem sido amplamente utilizada para compreender percepções sobre temas sensíveis. A detecção automática de posicionamento no X/Twitter, por exemplo, já foi estudada no contexto da COVID-19 (HOSSAIN et al., 2020; WEINZIERL; HOPFER; HARABAGIU, 2021), vacinação (D'ANDREA et al., 2019) e operações policiais (FEITOSA et al., 2022). Modelos baseados em *transformers*, como BERT, têm impulsionado esses avanços, alcançando desempenho superior em textos curtos e informais (DEVLIN et al., 2019; SOUZA; NOGUEIRA; LOTUFO, 2020).

Diante disso, esta dissertação aborda três eixos de pesquisa que se complementam para formar um arcabouço de mineração de padrões textuais, adotando o tema de violência urbana como domínio de aplicação. O primeiro eixo refere-se ao desenvolvimento de modelos de detecção de posicionamento ajustados ao contexto brasileiro, explorando arquiteturas para Processamento de Linguagem Natural (PLN) estado da arte, além de empregar uma estratégia híbrida de anotação apoiada por *Large Language Models* (LLMs). O segundo eixo avalia a transferência de aprendizado entre plataformas, analisando a viabilidade de generalizar modelos treinados em uma plataforma (origem) possam ser aplicados a outra plataforma (destino). O terceiro eixo aborda identificar a ocorrência de violência urbana a partir de conteúdos postados em plataformas de mídia social de forma automatizada, em larga escala e com baixo custo computacional, ou seja, uma abordagem baseada em algoritmos que exploram metadados textuais, capazes de agrupar conteúdos pertencentes ao mesmo incidente. Esses três eixos de pesquisa são fundamentais para desenvolver *modelos e aplicações, baseados em mineração de padrões textuais na Web, que permitam monitorar a ocorrência e repercussão de eventos associados à segurança pública.*

Assim, esta dissertação propõe um arcabouço de monitoramento de eventos de segurança pública baseado em fontes da Web que integra inferência de posicionamento, transferência de aprendizado entre plataformas e identificação automática de eventos. Nesse sentido, essa dissertação visa contribuir para o avanço das pesquisas em PLN aplicadas ao contexto de violência urbana. Por conseguinte, essa proposta contribui também para o desenvolvimento de ferramentas computacionais que auxiliam governos ou sociedades civis organizadas a monitorarem de forma automatizada e em larga escala eventos de segurança

pública, assim como opinião das pessoas, no que tange a usuários de plataforma multimídia na Web, sobre as ações das autoridades responsáveis para lidar com esses eventos.

1.1 Objetivos

O presente estudo tem como objetivo geral desenvolver um arcabouço para inferir a opinião de usuários de plataformas de mídia social na Web tendo o tema de violência urbana como domínio de aplicação, com foco na identificação de posicionamentos de aprovação ou desaprovação em comentários sobre mídias de operações policiais postadas nessas plataformas, baseado em modelos de PLN.

O objetivo geral acima pode ser detalhado nos seguintes objetivos específicos:

1 Desenvolver e avaliar modelos de detecção de posicionamentos (*stance detection*) para comentários sobre operações policiais no Brasil postados em plataformas de mídias sociais, explorando arquiteturas baseadas em *transformers* e analisando sua aplicação para monitorar a opinião dos usuários da plataforma sobre esses eventos ao longo do tempo.

2 Avaliar experimentalmente o desempenho dos modelos desenvolvidos para inferir posicionamentos de comentários no cenário interplataformas de mídias sociais, ou seja, comentários da plataforma A utilizados para treinamento de modelos e aplicação desses modelos a comentários da plataforma B e vice-versa.

3 Propor e avaliar métodos automáticos para identificar e agrupar vídeos que descrevem um mesmo evento de violência urbana no YouTube, utilizando exclusivamente metadados textuais (títulos, descrições e transcrições), com foco em soluções escaláveis e de baixo custo computacional.

1.2 Contribuições

Baseados nos objetivos estabelecidos na seção anterior, destacamos, a seguir, as principais contribuições deste trabalho:

- Metodologia para analisar a percepção pública acerca de operações policiais em mídias sociais, com foco na identificação automática de posicionamentos de aprovação, desaprovação ou neutralidade. Para isso, propomos um sistema de monitoramento contínuo de operações policiais no YouTube e uma estratégia de aprimoramento de modelos de PLN baseados em arquiteturas *transformers*, fundamentada em uma abordagem híbrida de anotação que combina validação humana e suporte de LLMs (GPT) para tratar inconsistências semânticas presentes em bases rotuladas da

literatura. A partir do refinamento dessas bases, ajustamos modelos especializados e conduzimos uma aplicação inédita desses modelos em vídeos de grande repercussão no YouTube, analisando tendências de opinião pública ao longo de dois anos consecutivos e em diferentes regiões brasileiras. Os resultados indicam acurácia de 88% e F1-macro de 87% no X/Twitter, com ganhos superiores a 18% e 7% em relação à literatura, bem como desempenho satisfatório em comentários de vídeos que monitoramos no YouTube, com acurácia de 83% e F1-macro de 73%, mesmo sem treinamento prévio nessa plataforma. A análise temporal e regional dos comentários no YouTube revelou variações relevantes nos níveis de aprovação e desaprovação, demonstrando a capacidade do modelo de captar dinâmicas da opinião pública, conforme publicado em (ROCHA et al., 2024).

- Capacidade de transferência de conhecimento dos modelos de detecção de posicionamentos quando aplicados no cenário interplataformas. Especificamente, conduzimos uma avaliação sistemática e estatisticamente fundamentada dos modelos previamente desenvolvidos em cenários intraplataforma (treino e teste na mesma plataforma) e interplataformas (treino e teste em plataformas distintas). Nossa análise concentra-se em comentários do YouTube e do X anotados por meio do pipeline híbrido proposto, garantindo maior consistência na rotulação de posicionamentos em um contexto multilíngue e sensível ao domínio. Os resultados mostram que modelos treinados em dados do X/Twitter apresentam maior transferibilidade quando aplicados ao YouTube do que o inverso, evidenciando que características linguísticas próprias de cada plataforma influenciam significativamente o processo de adaptação entre domínios.

- Duas heurísticas não supervisionadas para o agrupamento de vídeos que se referem a um mesmo evento real: uma **Heurística Temporal (HT)**, fundamentada na proximidade temporal de publicação e na recorrência de entidades âncoras (como localidades e operações policiais), e uma **Heurística Semântica (HS)**, que combina múltiplos atributos textuais, incluindo *embeddings* densos, entidades nomeadas, menções a vítimas e nomes formais de operações, integrados por meio de um grafo de similaridade. Os experimentos demonstram que ambas as heurísticas superam uma linha de base construída com o modelo GPT-4, indicando que abordagens especializadas, de menor custo computacional e adaptadas ao domínio podem apresentar desempenho superior a soluções genéricas de última geração, conforme demonstrado em (ROCHA et al., 2025).

Todas as bases de dados rotuladas e códigos-fonte utilizados no desenvolvimento das contribuições acima descritas estão disponíveis em repositórios públicos.[1]

[1] <https://github.com/LABPAAD/crimes_stance>; <https://github.com/LABPAAD/cross-platform>; <https://github.com/LABPAAD/urban_events_identification>

1.3 Estrutura do Documento

Essa seção apresenta a organização desta dissertação. A Figura 1 ilustra a visão global das principais atividades ou técnicas desenvolvidas nessa dissertação e os capítulos onde são aplicadas. Iniciamos com a etapa de Coleta dos dados aplicado nos capítulos 4, 5 e 6. A seguir, o tratamento dessas informações envolve Pré-processamento, filtragem, normalização e extração de metadados e procede-se a Rotulação do conteúdo multimídia tratado, i.e., comentários de texto e vídeo, também utilizados nos capítulos 4, 5 e 6, atividade essa necessária para treinamento e avaliação de modelos PLN desenvolvidos nesses capítulos. Essas três atividades iniciais nos levam a dois tipos de modelos como mostra a bifurcação na Figura 1. Por um lado, os comentários são utilizados para o treinamento e avaliação dos modelos supervisionados de inferência de posicionamento. Por conseguinte, na atividade seguinte, temos as aplicações desses modelos para o Monitoramento de Opinião e Análises Interplataformas. Ambas as etapas desse lado da bifurcação ocorrem nos Capítulos 4 e 5. Do outro lado, os metadados extraídos dos vídeos, i.e., títulos, descrições e transcrições, são utilizados em modelos não supervisionados, ou seja, Heurísticas de agrupamento de vídeos do mesmo evento.

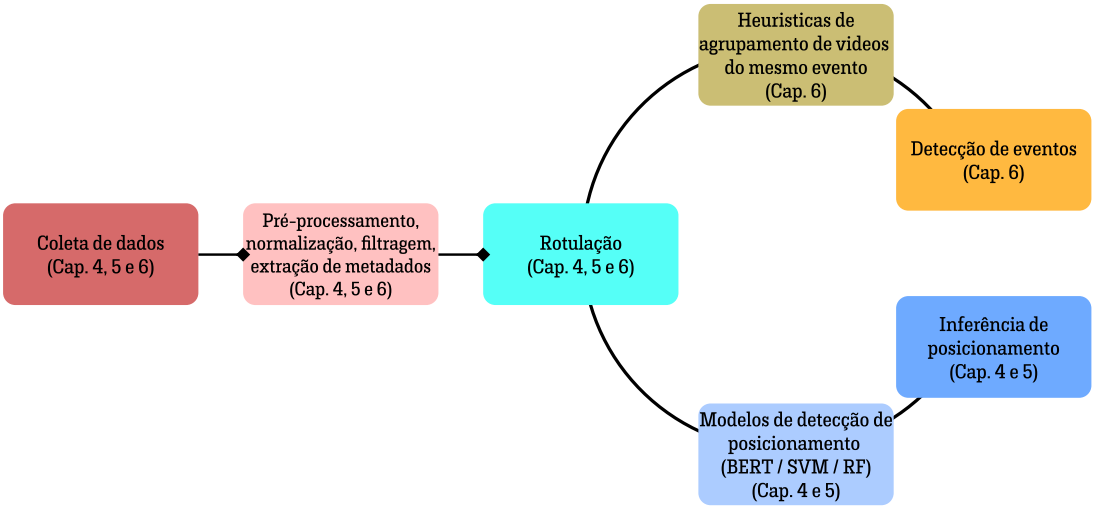


Figura 1 – Estrutura integrada da dissertação.

Visto a estrutura mostrada na Figura 1, os capítulos seguintes tem a seguinte organização: Nesse sentido, os capítulos seguintes estão organizados da seguinte forma. O **Capítulo 2** apresenta os trabalhos relacionados. O **Capítulo 3** descreve, de forma unificada e abstrata, as etapas metodológicas que fundamentam os Capítulos 4, 5 e 6. O **Capítulo 4** apresenta o sistema de monitoramento da opinião pública sobre violência urbana com base em comentários de vídeos do YouTube, abordando o processo de coleta e pré-processamento dos dados, a estratégia híbrida de anotação apoiada por LLMs e os modelos de PLN utilizados para inferência de posicionamentos ao longo do tempo e por região. O **Capítulo 5** investiga os cenários intra- e interplataformas (YouTube e X/Twitter) para detecção de posicionamento, descrevendo as bases de dados utilizadas, os modelos

de aprendizado de máquina avaliados e a análise estatística dos resultados, com ênfase na transferência de aprendizado entre plataformas. O **Capítulo 6** apresenta heurísticas não supervisionadas, fundamentadas em características temporais e semânticas, para o agrupamento de vídeos de violência urbana no YouTube que se referem ao mesmo evento real, detalhando o conjunto de vídeos, a extração de entidades e atributos, as estratégias de agrupamento e a avaliação experimental das heurísticas. Por fim, o **Capítulo 7** apresenta as conclusões desta dissertação, destacando as principais contribuições em relação aos Objetivos Específicos 1–3, bem como direções para trabalhos futuros.

2 Trabalhos Relacionados

Este capítulo apresenta uma revisão da literatura que fundamenta esta pesquisa, destacando trabalhos que empregam técnicas de PLN em diferentes contextos. Para compor essa revisão, foi realizada uma busca estruturada nas bases de dados *Scopus*, *ACM Digital Library*, *IEEE Xplore* e *Springer*. As buscas utilizaram a seguinte combinação: police AND ("sentiment analysis" OR nlp OR "natural language processing").

Inicialmente, discutimos estudos de mineração de dados aplicados a áreas como saúde, política e outras temáticas, que serviram como base metodológica para o desenvolvimento desta dissertação. Em seguida, apresentamos o eixo de trabalhos relacionados à detecção automática de eventos em vídeos de violência urbana no YouTube, um tema emergente que dialoga diretamente com o estudo apresentado no Capítulo 6 e que complementa a discussão sobre análise textual e monitoramento de conteúdos relacionados à violência urbana. Por fim, apresentamos pesquisas voltadas à análise de segurança pública e operações policiais, com ênfase no uso de dados de mídias sociais para inferir percepções públicas.

2.1 Mineração de Dados em Multi Contextos

Nesta seção, organizamos os trabalhos em categorias distintas conforme o contexto em que foram aplicadas as técnicas de mineração de texto. Dessa forma, destacamos os seguintes eixos temáticos: (i) Saúde, (ii) Política e (iii) Outros.

Saúde – D'Andrea et al. (2019) introduziram um sistema para avaliar a opinião pública sobre vacinação no X/Twitter, categorizando *tweets* como favoráveis, neutros ou contrários. Empregando técnicas sofisticadas de mineração de texto e aprendizado de máquina, incluindo *bag-of-words* e SVM, o sistema mostrou-se eficaz em identificar mudanças de opinião ao longo do tempo no contexto italiano do debate sobre vacinação. Miao, Last e Litvak (2022) exploraram a detecção de posicionamentos e monitoramento da opinião pública sobre medidas governamentais durante a pandemia da COVID-19 a partir de *tweets* na cidade de New York. Para contornar a limitação de dados rotulados, os autores aplicaram a técnica de *Data Distillation* como estratégia de aumento de dados (*Data Augmentation*). Essa abordagem é uma forma de aprendizado semissupervisionado que combina dados rotulados e não rotulados para expandir o conjunto de treinamento de forma eficiente. Inicialmente, um modelo denominado *Teacher Model* foi treinado utilizando um pequeno conjunto de dados manualmente rotulado. Em seguida, esse modelo foi utilizado para gerar previsões em um conjunto maior de dados não rotulados. Essa abordagem demonstra a importância do refinamento de dados para aumentar a precisão dos modelos, um aspecto

que é considerado na metodologia dessa dissertação. Bechini et al. (2020) conduziram uma análise detalhada sobre a opinião pública acerca da vacinação na Itália utilizando dados do Twitter. O estudo utilizou uma abordagem baseada em posicionamento (*stance detection*) com técnicas de PLN e aprendizado de máquina, incluindo *bag-of-words* com TF-IDF e SVM com núcleo linear para classificar os *tweets* em três categorias: favorável, contrário e neutro. Além disso, o trabalho destacou a importância de incorporar informações do usuário para fornecer uma estimativa mais precisa da opinião pública, abrangendo uma análise temporal e espacial para identificar variações regionais e eventos sociais que influenciaram as posturas sobre a vacinação.

Política– Políticos brasileiros recentemente começaram a se comunicar com a população por meio de transmissões ao vivo em mídias sociais, em vez de depender apenas da mídia tradicional. Nubila et al. (2023) investigaram como esse tecnopopulismo via YouTube foi utilizado como recurso durante o governo Bolsonaro. Decisões estratégicas foram tomadas na configuração dessas transmissões ao vivo. O estudo apresentou uma análise estatística que mostrou a probabilidade de aumento na manipulação da opinião pública e na disseminação de propaganda governamental. Os eventos eleitorais são momentos-chave para a divulgação de conteúdo político e podem ser influenciados por diversos fatores. Os autores de Kappaun e Oliveira (2023) apresentaram um método para examinar o viés de gênero por meio da análise de engajamento e sentimentos em comentários de vídeos do YouTube que mencionam nomes de homens ou mulheres em seus títulos ou descrições. O estudo abordou as eleições brasileiras de 2018 e 2022, com foco nas disputas presidenciais e estaduais. A análise quantitativa revelou que vídeos com homens tiveram maior engajamento, enquanto vídeos com mulheres apresentaram maior prevalência de sentimentos negativos nos comentários.

O autor Boyd (2014) investigou os diferentes papéis que as pessoas que comentam em vídeos do YouTube podem assumir. O estudo analisou comentários do vídeo do discurso inaugural de Barack Obama na plataforma e mostrou que os pronomes envolvendo o autor do comentário foram os termos lexicais mais populares indicando envolvimento. Quando os comentários foram rotulados como parte de conversas, eles foram principalmente rotulados como construtivos. Os autores também observaram que comentários perturbadores estavam associados com linguagem odiosa ou conspiratória, por exemplo. Karamouzas, Mademlis e Pitas (2022) propuseram um mecanismo de monitoramento automatizado da opinião pública, utilizando análise semântica coletiva de tweets. Esse trabalho quantificou atributos como polaridade, viés e figuratividade em tweets relacionados às eleições presidenciais dos EUA, permitindo prever tendências políticas. Essa abordagem reforça a aplicabilidade de métodos de análise semântica na identificação de padrões de opinião pública.

Outros – Barachi et al. (2021) propõem um *framework* de análise de sentimentos para monitorar a evolução da opinião pública em tempo real, utilizando um classificador LSTM

bidirecional. O estudo foi aplicado no contexto de debates sobre mudanças climáticas e utilizou dados coletados da rede social Twitter. Como estudo de caso, os autores analisaram 278.264 tweets publicados entre julho de 2019 e outubro de 2020, envolvendo a ativista ambiental Greta Thunberg e as reações de seus seguidores. Wang et al. (2020) identificaram períodos críticos em eventos controversos por meio da análise de sentimentos, ressaltando a necessidade de detectar emoções predominantes para prever as tendências da opinião pública. O estudo analisou dados coletados da plataforma *Sina Weibo*, uma das principais redes sociais da China, e investigou quatro eventos polêmicos que geraram grande repercussão pública. Chakraborty e Sharma (2019) conduziram uma análise de sentimentos de *tweets* durante a implementação da política de rodízio veicular em Delhi na Índia, evidenciando a capacidade de extrair opiniões públicas das mídias sociais para compreender a reação do público a medidas governamentais. Ceron e Negri (2016) analisaram a opinião pública sobre as reformas trabalhistas e do sistema educacional na Itália, ocorridas entre 2014 e 2015. Utilizando a técnica de *Supervised Aggregated Sentiment Analysis*, os autores monitoraram semanalmente a opinião pública no Twitter, acompanhando tanto o processo de formulação quanto a implementação das reformas.

O uso de modelos pré-treinados combinados com estratégias de transferência também aparece de forma destacada em (MOZAFARI; FARAHBAKHSH; CRESPI, 2020), que aplicam o BERT para detecção de discurso de ódio no Twitter, mostrando que o ajuste supervisionado com camadas adicionais melhora substancialmente o desempenho em corpora ruidosos. Embora o foco seja toxicidade, a arquitetura explorada (pré-treinamento generalista + fine-tuning específico) é análoga à pipeline utilizada nesta dissertação para detecção de posicionamento em português.Além disso, trabalhos recentes abordam explicitamente o problema da transferência interplataforma. (YI; ZUBIAGA, 2021) propõem um arcabouço baseado em BERT *adversarial* para detecção de perfis de usuários adolescentes em cenários de escassez de rótulos na plataforma destino, transferindo conhecimento entre domínios com diferenças linguísticas e de estilo. Ainda no que se refere à generalização de modelos entre domínios, (SCHILLER; DAXENBERGER; GUREVYCH, 2021) introduzem um benchmarking abrangente de detecção de posicionamento, avaliando robustez e transferibilidade entre múltiplos conjuntos com diferentes tópicos. O estudo mostra que modelos treinados em múltiplos domínios apresentam ganhos de desempenho, mas se tornam mais sensíveis a desvios semânticos.

De forma complementar, Chen et al. (2025) explorou deteção de violência em mídias sociais com foco explícito em comparação entre plataformas. Os autores introduziram o *dataset* InViS, contendo 30.000 postagens anotadas manualmente, e realizaram experimentos cruzados entre dados provenientes do YouTube, Reddit, Parler e fóruns. Os resultados mostram que modelos treinados em uma plataforma ainda obtêm desempenho satisfatório ao serem testados em outra, indicando a presença de padrões linguísticos compartilhados de violência e ameaças online

2.2 Mineração de Dados Multimodal

A detecção automática de eventos em vídeos publicados em plataformas de mídia social tem se tornado um tema de crescente interesse, impulsionado pelo aumento de conteúdos relacionados à violência urbana e operações policiais no YouTube (YOUSAF; NAWAZ, 2022). Diversos estudos têm proposto abordagens não supervisionadas ou com baixo grau de supervisão para identificação de eventos, frequentemente combinando informações visuais, auditivas e textuais (ELHOSEINY et al., 2016; JANSEN et al., 2017; WU et al., 2014). Em geral, essas abordagens exploram características multimodais, como imagem, áudio e legendas, para estimar a similaridade entre vídeos e atribuí-los a categorias ou eventos de interesse público. No entanto, tais métodos costumam exigir alto custo computacional, infraestrutura especializada e dados anotados em grande escala, o que limita sua aplicabilidade em cenários contínuos de monitoramento, especialmente na realidade brasileira (MONDAL et al., 2025).

Por outro lado, trabalhos que utilizam exclusivamente informações textuais têm mostrado resultados promissores, especialmente quando se considera a grande quantidade de metadados disponíveis no YouTube, como títulos, descrições, datas de publicação e, em alguns casos, transcrições automáticas. Abordagens baseadas em PLN, como o GraphTMT, demonstram que transcrições de vídeos podem conter informações suficientes para identificação de tópicos ou padrões recorrentes (THIES et al., 2021), dispensando a necessidade de processamento de áudio e imagem. Entretanto, mesmo essas abordagens enfrentam desafios quando aplicadas ao contexto de violência urbana, marcado por linguagem informal, variações regionais e baixa padronização textual (BERTAGLIA; NUNES, 2017).

Uma contribuição recente que dialoga diretamente com este contexto é o trabalho Tube2Vec, proposto por (BOESINGER et al., 2024), que gera *embeddings* sociais e semânticos para canais do YouTube utilizando relações de coocorrência em compartilhamentos, metadados e recomendações da plataforma. Embora voltado para caracterização social e não para agrupamento de eventos, o estudo demonstra que representações puramente textuais (como descrições e títulos) capturam dimensões semânticas profundas, reforçando a viabilidade de abordagens leves baseadas em metadados — fundamento que motivou nossa heurística textual no agrupamento de vídeos de violência urbana.

Além disso, técnicas de NER (*Named Entity Recognition*) têm sido exploradas como forma de estruturar dados não supervisionados e facilitar a organização de conteúdo (AMER; DAVIES, 2025; KERAGHEL; MORBIEU; NADIF, 2024; LU et al., 2024). Métodos recentes investigam desde pipelines enriquecidos para identificação de entidades em comentários de vídeos (AMER; DAVIES, 2025) até abordagens escaláveis, como o PaDeLLM-NER, que acelera a inferência em grandes volumes de dados (LU et al., 2024). Também há trabalhos que combinam análise textual, visual e comportamental em frameworks de

anotação assistida (STEINER et al., 2011). Ainda assim, esses estudos tendem a focar em tendências gerais, extração de temas ou melhoria da navegabilidade de vídeos, não abordando diretamente a tarefa de identificar quando múltiplos vídeos se referem ao mesmo evento específico, como uma operação policial ou confronto armado.

Assim, a literatura indica que, embora existam diferentes abordagens para organização e caracterização de vídeos em larga escala, poucas tratam especificamente da identificação de eventos utilizando apenas informações textuais, e nenhuma delas aborda as particularidades linguísticas, sociais e informacionais do contexto brasileiro. Nesse sentido, o resultado apresentado no Capítulo 5 contribui para preencher essa lacuna ao propor uma solução leve, escalável e adaptada ao domínio de violência urbana, combinando heurísticas eficientes com técnicas modernas de PLN para agrupar vídeos que se referem ao mesmo evento real.

2.3 Mineração de Dados Sobre Violência Urbana

Para uma visão abrangente e atualizada, foi realizada uma busca por artigos técnicos utilizando bases de dados acadêmicas reconhecidas. Diversos trabalhos têm se dedicado ao estudo da percepção pública sobre a polícia e segurança pública utilizando técnicas de PLN. Para melhor compreensão, organizamos essa discussão em duas categorias principais: (i) Discussão sobre segurança pública no mundo e (ii) Discussão sobre segurança pública no Brasil.

Discussão sobre violência urbana no mundo – Matos, Ribeiro e Ferreira (2025) aplicaram modelagem de tópicos e análise emocional de postagens no Reddit e X/Twitter para compreender como diferentes comunidades interagem e reagem à atuação policial. O estudo analisou dados referentes a eventos de violência policial ocorridos em Portugal entre janeiro de 2018 e dezembro de 2021. Foram destacados episódios como o Caso da Cova da Moura, que envolveu confrontos entre a Polícia de Segurança Pública e moradores locais, especialmente afrodescendentes; o Caso do Aeroporto Humberto Delgado, relacionado à morte de um cidadão ucraniano sob custódia do Serviço de Estrangeiros e Fronteiras; além de incidentes durante o Estado de Emergência, quando medidas de *lockdown* e abordagens policiais severas geraram reações negativas significativas nas redes sociais. O estudo revelou que esses eventos influenciaram diretamente a percepção pública sobre a polícia, com sentimentos como ódio, agressividade e, em menor grau, felicidade sendo predominantes nas postagens analisadas. Essa análise evidenciou uma forte polarização nas opiniões coletadas, refletindo tensões sociais e políticas relacionadas à atuação policial em Portugal.

Outro estudo relevante é o de Oh, Zhang e Greenleaf (2021), que propuseram um índice geográfico de sentimento para correlacionar sentimentos expressos no X/Twitter sobre a polícia com taxas de criminalidade e desigualdade socioeconômica. O estudo analisou

dados de 82 áreas metropolitanas nos Estados Unidos, destacando que regiões com maior desigualdade socioeconômica, taxas elevadas de criminalidade violenta e heterogeneidade racial apresentaram percepções mais negativas sobre a polícia. As cidades de Baltimore, St. Louis e Riverside-San Bernardino foram algumas das áreas com os sentimentos mais negativos, enquanto Des Moines, Greenville e Madison se destacaram com percepções menos negativas. Já Chaparro et al. (2020) investigaram a percepção da segurança pública através da análise de sentimentos em textos publicados na rede social *Twitter*. O estudo foi conduzido na cidade de Bogotá, na Colômbia, e teve como objetivo quantificar a percepção da segurança utilizando dados geolocalizados dessa região.

Yang et al. (2024) desenvolveram um modelo baseado em aprendizado por reforço para previsão de crimes no Twitter, utilizando padrões de postagens para identificar tendências criminais emergentes e prever possíveis áreas de risco. O estudo foi conduzido com base em dois conjuntos de dados distintos: (i) relatórios de incidentes criminais fornecidos pelo Departamento de Polícia de Chicago, abrangendo o período de setembro de 2019 a julho de 2024, e (ii) tweets relacionados a crimes na cidade de Chicago coletados na mesma faixa temporal.

A análise de ativismo digital também tem sido um foco importante na literatura, como evidenciado pelo estudo de Sidharta et al. (2023), que analisaram dados coletados no X/Twitter com foco na hashtag #PercumaLaporPolisi, utilizada na Indonésia para expressar insatisfação pública com a atuação policial. O estudo coletou 100 tweets por meio da API (*Application Programming Interface*) do X/Twitter e aplicou o algoritmo Naïve Bayes com a biblioteca *TextBlob* para classificar as postagens em três categorias: positivo, neutro e negativo. Apesar da *hashtag* implicar um teor negativo, os resultados mostraram que 40% dos tweets apresentaram sentimento positivo, 40% foram classificados como neutros, e apenas 10% foram classificados como negativos. Os autores destacaram que essa discrepância pode estar relacionada à presença de linguagem sarcástica ou limitações do modelo aplicado, o que ressalta a complexidade da análise de sentimentos em contextos sociais delicados.

No campo do monitoramento da opinião pública sobre a polícia, Nanayakkara (2024) analisaram comentários postados no YouTube sobre o vídeo intitulado *"How George Floyd was killed in Police Custody"*, publicado em 1º de junho de 2020. O estudo utilizou técnicas de Exploratory Data Analysis (EDA) e ferramentas de PLN para analisar o comportamento emocional dos usuários e suas reações ao incidente. Foram analisados 60.010 comentários coletados até 10 de julho de 2023. Similarmente, Scotland, Thomas e Jing (2024) analisaram a polarização do discurso sobre o movimento Black Lives Matter na rede social Facebook, identificando padrões de toxicidade em debates envolvendo políticas de policiamento e racismo estrutural em comentários de videos no Facebook sobre a morte de George Floyd, nos EUA.

Os autores Ku, Iriberri e Leroy (2008) propuseram uma metodologia de extração de informações criminais para melhorar a eficácia de investigações policiais. O estudo foi conduzido nos Estados Unidos e utilizou dados provenientes de múltiplas fontes, incluindo relatórios narrativos da polícia, relatos de testemunhas e artigos de notícias. Essas fontes foram selecionadas para abranger tanto registros formais quanto descrições informais de crimes. Os eventos analisados incluíram uma ampla gama de crimes, como assaltos, furtos e casos de violência urbana. A abordagem desenvolvida foi implementada utilizando o framework GATE (General Architecture for Text Engineering), no qual foram criados módulos específicos para segmentação textual, reconhecimento de entidades e extração de informações relevantes.

Discussão sobre violência urbana no Brasil – Barros, Pires e Nascimento (2023) implementaram técnicas de sumarização baseada no modelo BERT para auxiliar investigações criminais, proporcionando uma melhor organização e compreensão de grandes volumes de documentos forenses. O estudo foi conduzido com base em documentos da Polícia Federal Brasileira, especificamente nos chamados *Notitia Criminis*, que representam o ponto inicial das investigações criminais no Brasil. Esses documentos, que variam de uma a mais de 100 páginas, contêm informações detalhadas sobre crimes supostamente ocorridos, incluindo depoimentos, provas e outros registros relevantes. As investigações abordaram uma ampla variedade de crimes, incluindo fraudes financeiras, crimes previdenciários e desvio de recursos públicos. Para processar esse conteúdo extenso e complexo, foi aplicada uma abordagem de sumarização extrativa utilizando o modelo BERTSUM, que se destacou pela capacidade de identificar e condensar as informações mais relevantes para facilitar o trabalho dos agentes envolvidos nas investigações. Essa técnica mostrou-se especialmente eficaz na síntese de dados críticos, permitindo uma análise mais rápida e precisa de casos complexos.

Outro estudo relevante foi conduzido por Souza et al. (2023), que analisaram boletins de ocorrência relacionados à violência contra mulheres registrados entre os anos de 2020 e 2021 no estado de Mato Grosso, Brasil. Esses registros foram disponibilizados pela Superintendência do Observatório de Segurança Pública, vinculada à Secretaria Adjunta de Inteligência da Secretaria de Estado de Segurança Pública de Mato Grosso. O objetivo do estudo foi identificar padrões e associações semânticas presentes nas narrativas textuais dos boletins de ocorrência utilizando a técnica Word2Vec. A metodologia adotou a arquitetura Skip-gram, reconhecida por sua eficácia na captura de relações semânticas mesmo em conjuntos de textos reduzidos.

Em outro estudo, Gusmão, Figueiredo e Brito (2021) propuseram um processo de automatização e classificação de denúncias criminais utilizando técnicas de PLN aplicadas a textos informais. O estudo foi realizado no estado do Rio de Janeiro, Brasil, e utilizou dados de denúncias registradas anonimamente por meio do aplicativo do serviço Disque

Denúncia RJ. As denúncias abrangeram uma variedade de crimes e apresentavam uma linguagem coloquial, caracterizada por erros ortográficos e morfossintáticos, desafiando o uso de técnicas tradicionais de PLN.

Já em um contexto mais direto, Feitosa et al. (2022) analisaram posicionamentos sobre operações policiais que ocorreram no Brasil e que tiveram grande repercussão, utilizando um corpus extraído do X/Twitter e rotulado manualmente. Seu estudo evidenciou a importância de bases de dados anotadas para melhorar a classificação de sentimentos. Nessa dissertação, estendemos amplamente os modelos iniciados por Feitosa et al. (2022), melhorando o desempenho desses significativamente e os aplicando a um arcabouço de monitoramento de eventos de operações policiais em plataformas de mídias sociais.

No contexto brasileiro, observamos uma quantidade ainda limitada de estudos que aplicam PLN especificamente para inferir posicionamentos de usuários de mídias sociais sobre operações policiais. Embora existam contribuições relevantes que utilizam PLN para investigar temas políticos e sociais, como a percepção pública sobre figuras políticas ou debates eleitorais, há uma lacuna significativa no uso dessas técnicas, monitorar eventos relacionados à segurança pública ou analisar comentários dos vários usuários de mídias sociais sobre operações policiais. Essa carência destaca a necessidade de mais investigações voltadas à aplicação de PLN para apoiar estratégias de segurança pública no Brasil.

2.4 Resumo

A revisão desses trabalhos considera um volume notável de estudos acadêmicos sobre aplicação de PLN para inferência de posicionamento de usuários sobre a segurança pública e operações policiais. No entanto, ainda há desafios a serem superados, como o viés presente em algumas análises de sentimentos e a dificuldade de transferência de aprendizado entre diferentes plataformas de mídias sociais. Assim, a presente pesquisa busca contribuir para essa discussão ao propor um modelo robusto de inferência de posicionamentos e monitoramento da percepção pública sobre a polícia em mídias sociais.

A Tabela 1 apresenta um comparativo entre os principais trabalhos revisados, considerando quatro aspectos principais relacionados ao projeto dessa dissertação de mestrado. São eles: local, inferência de posicionamento, análise inter-plataforma e identificação automática de eventos.

A contribuição dessa dissertação em relação a esses trabalhos consiste em desenvolvimento de um arcabouço utilizando *BERTimbau* e técnicas avançadas de PLN para analisar percepções sobre violência urbana no Brasil. Assim como detecção de posicionamento em múltiplas plataformas (X/Twitter e YouTube) e identificação automática de eventos no contexto de violência urbana usando características semânticas e temporais.

Tabela 1 – Comparação entre os principais trabalhos da literatura considerando o contexto de violência urbana.

Trabalho Relacionado	**Local**	**Inferência de Posicionamento**	**Análise Inter Plataforma**	**Identificação automática de Eventos**
Oh, Zhang e Greenleaf (2021)	EUA	Sim	Não	Não
Chaparro et al. (2020)	Colômbia	Sim	Não	Não
Matos, Ribeiro e Ferreira (2025)	Portugal	Sim	Sim	Não
Sidharta et al. (2023)	Indonésia	Sim	Não	Não
Nanayakkara (2024)	EUA	Sim	Não	Não
Scotland, Thomas e Jing (2024)	EUA	Sim	Não	Não
Yang et al. (2024)	EUA	Sim	Não	Não
Barros, Pires e Nascimento (2023)	Brasil	Não	Não	Sim
Feitosa et al. (2022)	Brasil	Sim	Não	Não

3 Metodologia Geral

Este capítulo apresenta uma visão geral da metodologia que fundamenta os três eixos de pesquisa desta dissertação e que são desenvolvidos, respectivamente, nos Capítulos 4, 5 e 6. Tais eixos correspondem às seguintes frentes de estudo: (i) detecção e monitoramento de posicionamentos em comentários de mídias sociais; (ii) avaliação intra e interplataformas; e (iii) agrupamento de vídeos que descrevem um mesmo evento real, especificamente operações policiais e episódios de violência urbana.

O objetivo deste capítulo é descrever, de forma unificada, as etapas metodológicas que orientam toda a dissertação, permitindo que diferentes plataformas (YouTube e X/Twitter), diferentes tipos de conteúdo textual (comentários, *tweets*, títulos, descrições, transcrições) e diferentes abordagens de modelagem supervisionada ou não supervisionada possam ser abordados sob um único arcabouço técnico.

A proposta aqui apresentada é suficientemente geral para que pesquisas futuras possam estendê-la, substituindo modelos, estratégias de avaliação estatística, sem comprometer a estrutura conceitual básica do trabalho. Assim, o arcabouço opera como um guia central que conecta as etapas de aquisição, preparação, rotulação, modelagem e análise descritas nessa dissertação.

3.1 Visão Geral do Arcabouço

A metodologia proposta nesta dissertação fundamenta-se no pressuposto de que diferentes plataformas de mídia social podem ser tratadas de forma unificada como fontes heterogêneas de dados textuais ou multimodais. Para fins de exemplificação, consideram-se as plataformas YouTube e X/Twitter, cada uma disponibilizando um conjunto de objetos textuais ou associados a vídeos. Cada objeto pode corresponder a um comentário, a uma postagem ou a um vídeo enriquecido com metadados estruturados, tais como data de publicação, título, descrição, localidade e características do canal responsável pela publicação.

A partir desse conjunto inicial de dados multimodais, o arcabouço metodológico proposto organiza-se em um fluxo contínuo composto por cinco etapas conceituais que permitem a transição desde a aquisição bruta dos objetos até a integração final dos resultados analíticos. Inicialmente, os dados são coletados e filtrados por meio de mecanismos de busca e seleção temática, assegurando que apenas conteúdos efetivamente relacionados ao domínio de violência urbana sejam mantidos. Em seguida, o material textual dessas plataformas passa por etapas de pré-processamento, normalização e enriquecimento semântico,

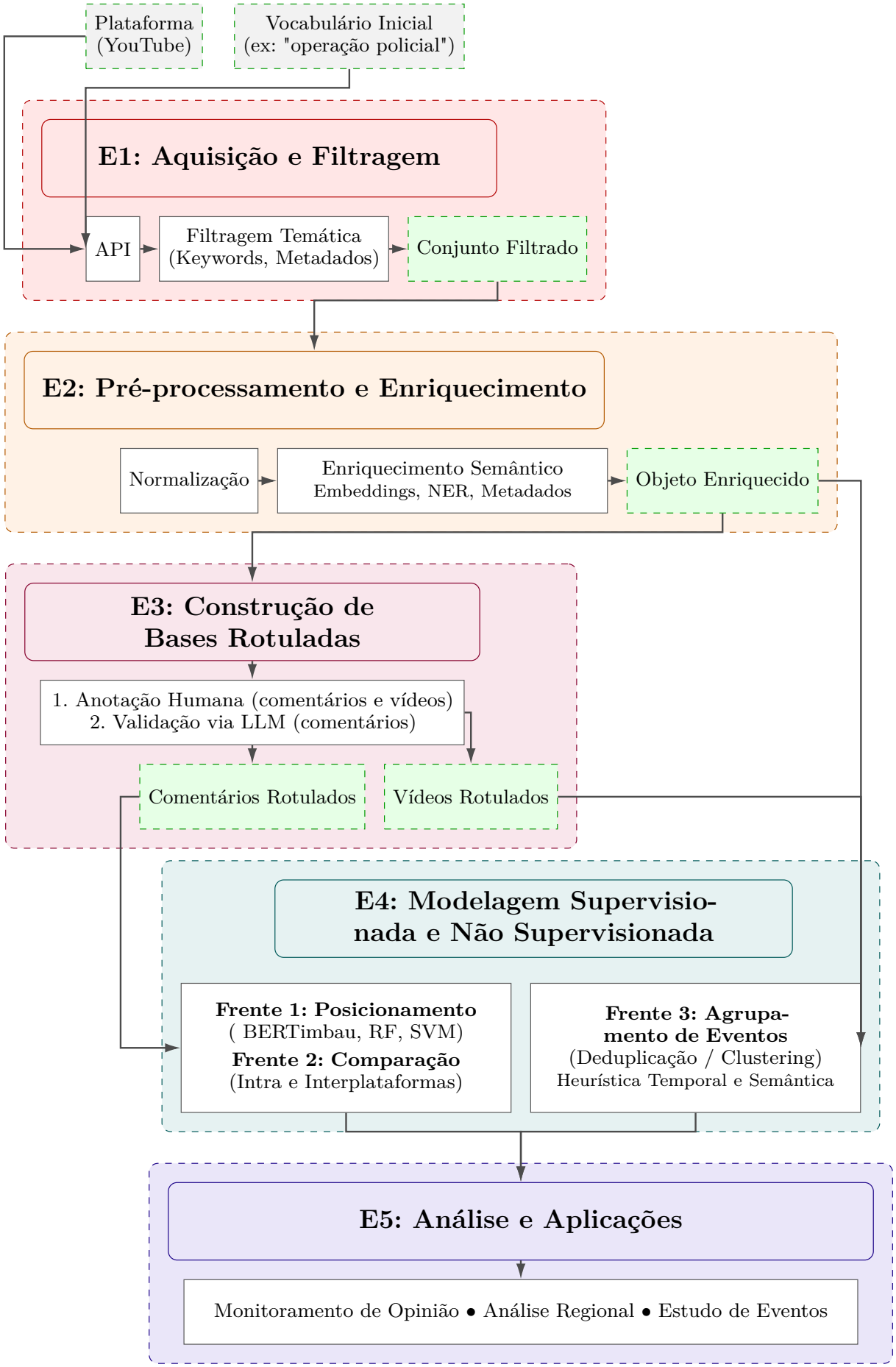


Figura 2 – Fluxograma do Arcabouço Metodológico Geral.

de modo a preparar esses dados para tarefas de modelagem. Posteriormente, constroem-se conjuntos rotulados por meio de uma abordagem híbrida para anotação de comentários, combinando julgamento humano e validação com modelos de linguagem, e uma abordagem manual para vídeos. Essas bases rotuladas sustentam as etapas de modelagem supervisionada aplicadas à detecção de posicionamento, enquanto representações textuais enriquecidas alimentam também abordagens não supervisionadas voltadas à identificação e ao agrupamento de eventos reais. Finalmente, os resultados dessas modelagens possibilitam monitorar tendências temporais, diferenças regionais, padrões discursivos interplataformas e a repercussão pública de eventos específicos, como operações policiais.

O Capítulo 4 utiliza esse arcabouço ao aplicar as quatro primeiras etapas, concentrando-se na tarefa de detecção de posicionamento e em sua evolução temporal e regional. No Capítulo 5, a mesma estrutura metodológica é ampliada para contemplar duas plataformas (X/Twitter e YouTube) de maneira simultânea, permitindo avaliar a capacidade de generalização dos modelos treinados em cenários intra e interplataformas, além de incorporar avaliações estatísticas formais por meio de técnicas como *bootstrapping*[1]. Já o Capítulo 6 explora um desdobramento alternativo do arcabouço, fazendo uso das representações textuais do vídeo coletadas e enriquecidas, para propor heurísticas temporais e semânticas que identificam quando múltiplos vídeos se referem ao mesmo evento real. Esses agrupamentos são posteriormente integrados à análise de posicionamento.

3.2 Etapa E1: Aquisição e Filtragem

A primeira etapa do arcabouço metodológico, representada na Figura 2, consiste na obtenção sistemática de dados via API provenientes de mídias sociais, nesse trabalho, por praticidade, o YouTube[2]. A literatura em aprendizado de máquina aplicado a mídias sociais evidencia que a definição inadequada do vocabulário de busca pode introduzir vieses estruturais e comprometer análises posteriores, sobretudo em domínios sensíveis como violência urbana. Essa preocupação é consistente com estudos que apontam a forte variabilidade linguística, presença de ruído, informalidade e oscilações estilísticas em ambientes digitais (BALUSU; MERGHANI; EISENSTEIN, 2018; BERTAGLIA; NUNES, 2017), o que torna a etapa de aquisição particularmente delicada.

Para mitigar esses riscos, define-se um vocabulário inicial contendo termos temáticos que funcionam como pontos de entrada para o processo de busca, gerando um conjunto bruto de objetos. Estratégias similares têm sido adotadas em estudos que monitoram conteúdos sensíveis em ambientes de vídeo ou de alta repercussão pública, demonstrando

[1] O método de *bootstrapping* é uma técnica de reamostragem não paramétrica que consiste em gerar, a partir dos dados observados, diversas amostras artificiais com reposição e calcular nelas a estatística de interesse, permitindo aproximar sua distribuição amostral e estimar medidas de incerteza, como intervalos de confiança, sem assumir uma forma paramétrica específica para a população.

[2] A API do YouTube é gratuita, o que a torna prática para fins de pesquisa.

que a seleção criteriosa de termos influencia profundamente a composição da base (CHEN et al., 2025). A filtragem subsequente, embora aqui descrita de maneira genérica, complementa a construção da base de dados, sendo uma camada de garantia para obtenção de dados no contexto especificado.

3.3 Etapa E2: Pré-processamento e Enriquecimento

Após a coleta e filtragem, a etapa de pré-processamento busca transformar o material bruto em representações adequadas para algoritmos de aprendizado. Trata-se de um desafio bem documentado na literatura de PLN aplicada a mídias sociais, marcada por informalidade, presença de gírias, erros ortográficos e forte variabilidade discursiva. Esses aspectos tornam o pré-processamento indispensável e convergem com achados que justificam a necessidade de normalização robusta (BERTAGLIA; NUNES, 2017).

Além da remoção de artefatos digitais (*links*, emojis, símbolos), opta-se por preservar *stopwords*, dado que modelos baseados em *Transformers* dependem da estrutura sintática completa para capturar relações contextuais. Esse princípio é coerente com diretrizes recentes para modelos contextualizados, que demonstram que a remoção agressiva de termos funcionais pode prejudicar o desempenho.

Entretanto, sistemas modernos de aprendizado beneficiam-se de mecanismos adicionais de enriquecimento semântico. Assim, cada texto é processado via uma representação expandida, incorporando *embeddings* vetoriais e entidades nomeadas extraídas por NER. A literatura reforça a importância de NER em cenários ruidosos (LI et al., 2022; KERAGHEL; MORBIEU; NADIF, 2024), inclusive em comentários de vídeo (AMER; DAVIES, 2025), além de reconhecer o papel dos *Sentence Transformers* como forma eficaz de capturar similaridade semântica (REIMERS et al., 2019).

Esse enriquecimento é ainda mais relevante quando se considera a heterogeneidade entre plataformas: enquanto *tweets* frequentemente omitem entidades ou recorrem a abreviações, descrições de vídeo podem apresentar narrativas detalhadas. O arcabouço busca ser robusto a essa diversidade textual.

3.4 Etapa E3: Construção de Bases Rotuladas por Anotação Híbrida

A escassez de bases rotuladas no domínio de segurança pública em língua portuguesa constitui um obstáculo central, já observado por (FEITOSA et al., 2022). Modelos supervisionados dependem de rótulos consistentes, mas anotações humanas são suscetíveis a vieses cognitivos, divergências interpretativas e inconsistências.

Para enfrentar esse cenário, esta dissertação adota uma abordagem híbrida de anotação, alinhada às tendências recentes de uso de LLMs como anotadores auxiliares. A

estratégia consiste em: (i) realizar anotação humana inicial; (ii) submeter cada instância ao julgamento de um LLM; (iii) reter apenas os casos com concordância entre ambas as partes. Essa abordagem é apoiada por evidências de que LLMs alcançam desempenho comparável ou superior ao trabalho manual em certas tarefas de anotação (GILARDI; ALIZADEH; KUBLI, 2023), ao mesmo tempo em que apresentam limitações semânticas importantes, especialmente em tarefas de postura, ironia e ambiguidades (KOCOŃ et al., 2023).

3.5 Etapa E4: Modelagem Supervisionada e Não Supervisionada

A quarta etapa formaliza os processos de aprendizado supervisionado e não supervisionado. Do lado supervisionado, a inferência de posicionamento é tratada como um problema de classificação multiclasse. O arcabouço dialoga diretamente com o conceito de posicionamentos, i.e., *stance detection* inaugurado por (MOHAMMAD; SOBHANI; KIRITCHENKO, 2017) e posteriormente ampliado em diversos contextos multilíngues.

Modelos baseados em *Transformers*, como BERT (DEVLIN et al., 2019) e, em especial, BERTimbau (SOUZA; NOGUEIRA; LOTUFO, 2020), têm demonstrado resultados superiores em português e em cenários de alta informalidade. A literatura recente comprova ganhos expressivos desses modelos em tarefas relacionadas a segurança pública, incluindo análises de operações policiais (FEITOSA et al., 2022) e comparações entre plataformas. Trabalhos sobre especialização contextual, como BERTweet (NGUYEN; VU; NGUYEN, 2020) e o uso de *fine-tuning* para domínios sensíveis (MOZAFARI; FARAHBAKHSH; CRESPI, 2019), reforçam essa abordagem.

A relevância desse trabalho para nossa pesquisa está na demonstração de que modelos podem manter desempenho competitivo mesmo quando aplicados a plataformas cuja distribuição textual difere significativamente da fonte — um desafio central neste estudo ao aplicar modelos treinados no X/Twitter para inferir posicionamentos em comentários do YouTube.

Em paralelo, a dimensão não supervisionada desta etapa busca identificar quando múltiplos vídeos se referem ao mesmo evento, um problema clássico em estudos multimodais. Abordagens anteriores recorrendo a análise de áudio, imagem ou metadados complexos (WU et al., 2014; ELHOSEINY et al., 2016; JANSEN et al., 2017) mostram que a detecção de eventos pode ser altamente custosa. Em contraste, métodos baseados em grafos e tópicos extraídos de transcrições e descrições (THIES et al., 2021), bem como heurísticas de similaridade baseadas em (JACCARD, 1912), apontam para alternativas mais leves e escaláveis.

O arcabouço aqui proposto adota heurísticas temporais e semânticas para construção de grafos de similaridade, permitindo clusterizações capazes de identificar grupos de vídeos

relacionados ao mesmo fato sem depender de modalidades adicionais, i.e., áudio e vídeo. Essa combinação permite integrar problemas supervisionados e não supervisionados.

3.6 Etapa E5: Análise e Aplicações

A etapa final sintetiza os resultados das fases anteriores para investigar como eventos de violência urbana são discutidos e percebidos em ambientes digitais. Nesse sentido, a análise da opinião pública deixa de ser apenas um desdobramento opcional da modelagem e passa a cumprir um papel central: ela permite observar como operações policiais e incidentes específicos afetam a legitimidade das instituições, a confiança na atuação do Estado e a forma como diferentes grupos sociais enquadram esses eventos. Ao articular séries temporais de posicionamento, padrões regionais e diferenças discursivas entre plataformas, essa etapa transforma manifestações espontâneas em mídias sociais em um indicador sistemático das dinâmicas de apoio, crítica e neutralidade que cercam eventos de violência urbana.

A literatura mostra que redes sociais funcionam como indicadores sensíveis de opinião pública diante de políticas, crises ou eventos de grande impacto, como revelam estudos em temas diversos, desde políticas de mobilidade urbana (CHAKRABORTY; SHARMA, 2019) até vacinação (D'ANDREA et al., 2019) e mudanças climáticas (BARACHI et al., 2021). Trabalhos que analisam janelas críticas de reação pública (WANG et al., 2020) ajudam a explicar fenômenos observados nas séries temporais geradas nesta dissertação.

Além disso, pesquisas que integram opinião pública digital ao ciclo de políticas públicas (CERON; NEGRI, 2016) sustentam a relevância sociotécnica da análise produzida nesta etapa. Quando combinadas com o agrupamento automático de eventos e a inferência de posicionamento, essas análises permitem identificar padrões regionais, temporalidades relevantes e divergências discursivas entre plataformas.

3.7 Resumo

A Metodologia Geral aqui apresentada estabelece uma estrutura unificada capaz de conectar, de maneira coerente, os eixos de pesquisa desta dissertação. Da aquisição de dados em plataformas heterogêneas ao enriquecimento semântico, passando pela anotação híbrida e pelos módulos de modelagem supervisionada e não supervisionada, constrói-se um fluxo contínuo que transforma objetos textuais dispersos em insumos analíticos consistentes.

Cada etapa se apoia em desafios e resultados documentados na literatura, desde a variabilidade linguística e ruído das mídias sociais (BALUSU; MERGHANI; EISENSTEIN, 2018), à necessidade de representações densas (REIMERS et al., 2019), à integração de NER em ambientes complexos (LI et al., 2022), ao uso de LLMs como validadores semânticos

(GILARDI; ALIZADEH; KUBLI, 2023), até a detecção de eventos por heurísticas leves (THIES et al., 2021).

Ao articular problemas supervisionados e não supervisionados em um mesmo arcabouço, as etapas aqui descritas preparam o terreno para as análises empreendidas nos capítulos seguintes. O Capítulo 4 instância principalmente a porção supervisionada do arcabouço para investigar padrões de posicionamento no YouTube; o Capítulo 5 explora a dimensão comparativa ao submeter modelos às variações inerentes a diferentes plataformas; e o Capítulo 6 estende o fluxo metodológico para identificar e agrupar eventos reais a partir de múltiplos vídeos.

Assim, a síntese deste arcabouço destaca não apenas a função organizadora das etapas E1 a E5, mas também sua natureza extensível, permitindo que futuras pesquisas substituam ou ampliem modelos, técnicas e representações sem comprometer a lógica central do processo. Trata-se, portanto, de um alicerce metodológico flexível, que possibilita compreender de forma integrada como conteúdos relacionados à violência urbana circulam, são percebidos e se conectam em diferentes ambientes de mídia social.

4 Análise de Posicionamentos no YouTube sobre Operações Policiais

Este capítulo tem o objetivo é analisar o posicionamento de usuários em comentários do YouTube sobre operações policiais. Está organizado da seguinte forma: Seção 4.1 apresenta a metodologia abordada, descrevendo os dados, modelos e métricas utilizados; Seção 4.2 mostra a avaliação dos modelos e sua aplicação assim como os resultados obtidos.

4.1 Metodologia

Nesta Seção é descrita a metodologia para inferir posicionamentos de comentários de usuários de mídias sociais na Web sobre operações policiais no Brasil. O capítulo está organizado em três seções principais: monitoramento (Seção 4.1.1), onde são apresentados os métodos de coleta dos comentários, seguido pelo pré-processamento (Seção 4.1.2) desses dados, e por conseguinte, os métodos para inferência (Seção 4.1.3) desses comentários.

A organização da metodologia, conforme as seções acima descritas, está ilustrada na Figura 3. O YouTube foi utilizado como fonte de vídeos e comentários sobre notícias relacionadas a operações policiais devido à sua ampla popularidade e acesso no Brasil, além da disponibilidade gratuita de dados através da API do YouTube versão 3. Assim, foi desenvolvido um programa coletor e classificador de comentários automatizado. Os códigos fontes, bem como a base de dados obtida das etapas descritas a seguir estão disponíveis em repositório público.[1]

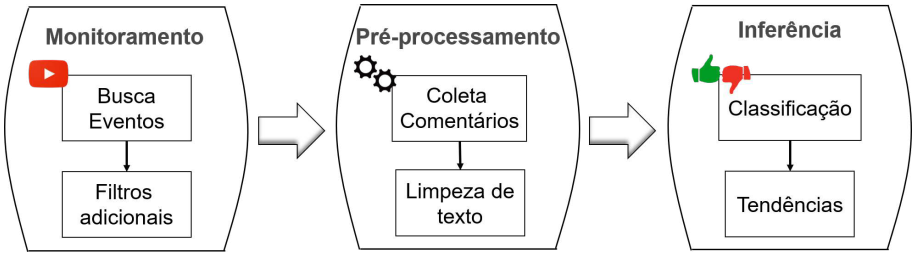


Figura 3 – Módulos do sistema proposto enfatizando o fluxo contínuo de monitoramento, pré-processamento e inferência de comentários do YouTube.

4.1.1 Monitoramento

A plataforma de mídia social YouTube foi monitorada para coletar vídeos sobre notícias relacionadas a operações policiais entre os anos de 2021 e 2022. Primeiramente,

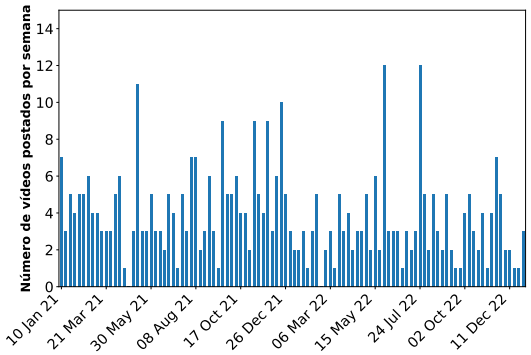
[1] <https://github.com/LABPAAD/crimes_stance>

foi realizada uma busca por vídeos via API do YouTube, que utilizou a princípio apenas as palavras-chave “assassinato”, “morte”, “roubo” e “furto” , combinadas com o termo “operação policial”. Posteriormente, foram incorporadas informações de localização geográfica para monitorar regiões específicas. Cada busca foi delimitada por um período de sete dias, i.e., unidade de tempo semanal. Várias buscas foram realizadas de modo a cobrir os anos monitorados com unidades de tempo consecutivas, resultando em 104 unidades de tempo (semanas) nos anos de 2021 e 2022. Uma busca semanal também teve um limite de vinte resultados, que foi o melhor compromisso encontrado para garantir a relevância dos resultados retornados pela API do YouTube.

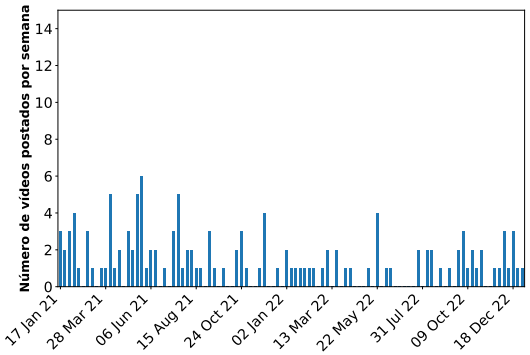
O monitoramento por região consistiu na definição de um centro por coordenada geográfica e um raio de circunferência em metros de modo a cobrir toda a área de uma região de interesse. Caso nenhuma coordenada geográfica seja especificada, a API do YouTube retorna vídeos de qualquer região, baseando-se exclusivamente nas palavras-chave da busca. Foram monitoradas operações policiais, desconsiderando regiões, e também as regiões Nordeste e Sudeste do Brasil, para fins de avaliação do sistema proposto. Para o monitoramento regional, foram definidos centros geográficos em duas áreas específicas: a Região Nordeste (próxima a São João do Rio do Peixe, Paraíba) e a Região Sudeste (entre Bom Jardim de Minas e Lima Duarte, Minas Gerais). Para garantir uma abrangência adequada, utilizaram-se raios de aproximadamente 1 milhão de metros e 590 mil metros, respectivamente. A busca desconsiderando regiões retornou um total de 1.223 vídeos, enquanto a busca por região retornou 1.328 e 1.313 vídeos no Nordeste e Sudeste, respectivamente.

Após a busca inicial, foi aplicado um filtro sobre os vídeos coletados para selecionar apenas aqueles relacionados com operações policiais. Primeiramente, os vídeos da categoria “notícia e política” identificados pela API do YouTube foram selecionados, visando remover vídeos sobre blogs e documentários policiais. A seguir, foram selecionados os vídeos que continham no título as palavras “pm”, “pms”, “policia”, “policial”, “policias”, “policiais”, “operacao”, “operacoes”, considerando transformação dos textos em minúsculo, remoção de acentos e sinais ortográficos. Assim, foram selecionados 752 vídeos ao total, sendo que 327 desses corresponderam a buscas por regiões, especificamente 135 da Região Nordeste e 192 da Região Sudeste.

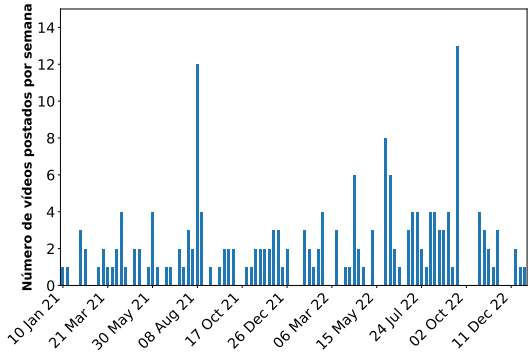
A Figura 4 mostra a distribuição da quantidade de vídeos ao longo do tempo sobre operações policiais resultantes do monitoramento acima descrito. Consideramos como Brasil (Figura 4a) o monitoramento desconsiderando regiões, i.e., vídeos sobre operações policiais em várias regiões brasileiras, ao passo que as regiões Nordeste e Sudeste (Figuras 4b e 4c) são monitoramentos por região. Notavelmente, os picos semanais de vídeos no Brasil não coincidem com os picos regionais, por exemplo, em 18 de setembro de 2022 na região sudeste. Isso porque alguns vídeos de menor relevância que atendiam



(a) Geral



(b) Região nordeste do Brasil



(c) Região sudeste do Brasil

Figura 4 – Distribuição da quantidade de vídeos coletados sobre operações policiais nos anos de 2021 e 2022 em unidades de tempo semanal.

aos critérios de filtragem entram nas coletas regionais e não entram na coleta do Brasil, porém, isso não gera grande impacto na quantidade de comentários do período específico.

4.1.2 Pré-processamento

O pré-processamento visa coletar comentários de todos os vídeos selecionados na etapa anterior e adequá-los para construção de modelos de inferência de posicionamentos.

A coleta dos comentários com a API do YouTube foi codificada, i.e., automatizada, via linguagem Python 3.0 recebendo o identificador do vídeo como parâmetro. Coletamos informações de todos os comentários disponíveis para cada vídeo, sem aplicar nenhum filtro ou critério de seleção. As informações de cada comentário consistem em um identificador único, data da postagem e o texto do comentário. O código coletor armazenou essas informações associando-as aos seus respectivos vídeos. Foram coletados um total de 257.025 comentários, sendo 11.063 na Região Nordeste e 31.419 na Região Sudeste.

Com esses comentários, conduzimos uma série de pré-processamentos nos comentários coletados via a biblioteca NLTK[2] na linguagem Python 3.0, removendo *links*, *emojis* e quebras de linha, mas mantendo *stop words* para inferências com a mesma sequência

[2] https://www.nltk.org/

em que as palavras aparecem nos comentários. Além disso, foram removidos comentários compostos por uma única palavra, pois seu baixo conteúdo semântico dificulta a inferência de posicionamento.

4.1.3 Inferência

Nesta etapa modelos de processamento de linguagem natural (PLN) são aplicados via aprendizagem supervisionada e análise de tendência de posicionamentos de usuários sobre os vídeos e comentários obtidos nas etapas anteriores.

A abordagem de aprendizagem supervisionada adotada consiste na construção de modelos para classificar comentários em *Aprova, Desaprova* ou *Neutro* considerando os posicionamentos dos usuários sobre os vídeos de operações policiais. Os comentários *Aprova* e *Desaprova* significam, respectivamente, que o usuário concorda e discorda com a operação policial do vídeo, ao passo que *Neutro* significa que o usuário é indiferente ao assunto ou não manifesta claramente seu posicionamento. Logo, é necessário rotular um conjunto de comentários suficiente para treinar modelos com essas três classes. Contudo, a rotulação é realizada por humanos, como usual na literatura, o que é uma tarefa exaustiva, propensa a erros e inconsistência. Por outro lado, a rotulação é fundamental para o desempenho dos modelos e por conseguinte qualidade da inferência.

Para lidar com essa questão propomos uma estratégia híbrida, que utiliza a rotulação manual por humanos e reavaliação dos rótulos por *Large Language Models* (LLM). Nesse sentido, Utilizamos o conjunto de 4467 *tweets* sobre operações policiais de grande repercussão no Brasil rotulados manualmente com as classes acima descritas no trabalho de (FEITOSA et al., 2022). A seguir, reavaliamos esses comentários e seus rótulos com a API do *Generative Pre-trained Transformer* (GPT) versão 3.5 Turbo, utilizando o formato de requisição para inferência de posicionamento (*stance*) no GPT proposto em (KOCOŃ et al., 2023).

A reavaliação foi um passo importante para selecionar comentários com rotulação consistente. Isso porque o LLM GPT é capaz de inferir posicionamentos de todos os comentários seguindo um mesmo padrão. Por outro lado, ele apresenta altas taxas de erro devido baixa especificidade, e.g., F1-Macro em 52% para posicionamentos (KOCOŃ et al., 2023). Logo, selecionamos os comentários rotulados igualmente entre os avaliadores em (FEITOSA et al., 2022) e o LLM GPT, o que resultou em um subconjunto de 2671 comentários (tweets), i.e. 60% do total. A distribuição das classes desse subconjunto é desbalanceada, seguindo as características do conjunto principal, com comentários neutros, aprovação e desaprovação com 44%, 26% e 30% respectivamente.

Os modelos para classificação foram construídos a partir desses rótulos utilizando BERT (*Bidirectional Encoder Representations from Transformers*) (DEVLIN et al., 2019),

que é um arcabouço de aprendizado profundo desenvolvido pelo *Google* para processamento de linguagem natural. Um recurso importante do BERT é sua construção com representações contextuais de modelos pré-treinados, baseado na arquitetura *transformers*. Neste trabalho utilizamos dois modelos pré-treinados que são o *Multilingual* (DEVLIN et al., 2019) e o *BERTimbau* (SOUZA; NOGUEIRA; LOTUFO, 2020). Nesse caso, esses modelos foram retreinados com os comentários rotulados como uma nova camada de neurônios na rede, o que denominamos como *modelos ajustados*, i.e., *fine tunning* para o contexto.

Outra forma de aplicar o BERT é gerando uma matriz de *word embeddings*, que é uma representação quantitativa das características das palavras contidas em um comentário via processamento de linguagem natural. Utilizamos essa matriz para treinar dois modelos de classificação populares, baseados em aprendizagem de máquina: *Random Forest* (RF) e *Support Vector Machine* (SVM) com e sem balanceamento de classes via *smote* (MAHESH, 2020) e hiper-parâmetros padrões.

Finalmente para inferência de opinião pública o modelo PLN previamente treinado é aplicado para classificar e identificar a tendência ao longo do tempo para posicionamento de usuários no serviço monitorado. Definimos tendências como picos de comentários em unidades de tempo que indicam claramente a percepção pública via os posicionamentos de aprovação, desaprovação ou neutralidade. Esses picos estão relacionados a eventos específicos relativos a incidentes de segurança polêmicos que demandam operações policiais com grande repercussão na plataforma YouTube.

4.2 Resultados

Nessa seção apresentamos os resultados e as discussões acerca do trabalho desenvolvido. Primeiramente, avaliamos o desempenho dos modelos considerando plataformas de testes distintas (Seção 4.2.1). A seguir na Seção 4.2.2, discutimos os resultados da inferência do melhor modelo nos comentários do YouTube coletados com o monitoramento ao longo dos anos 2021 e 2022.

4.2.1 Comparando Modelos de Inferência

Dado o cenário acima descrito. Nosso principal objetivo nessa seção é analisar o impacto da abordagem híbrida para seleção de instâncias anotadas e comparar o desempenho do BERTimbau com classificadores treinados com e sem estratégia de balanceamento de cargas. É importante mencionar que os classificadores são métodos mais simples e de menor custo computacional para inferir posicionamentos. Logo é importante analisar quão próximos são seus desempenhos comparados a métodos mais complexos como BERTimbau, que é baseado em redes neurais.

Nos testes, avaliamos o desempenho dos modelos com cinco métricas: acurácia, precisão, revocação, f1-score e f1-macro para os comentários do Twitter e YouTube como mostra a Tabela 2. A acurácia indica o percentual de classificações corretas, i.e., a soma acertos de todas as classes dividido pelo número total de comentários testados. Já a precisão é calculada para cada classe individualmente e evidencia o percentual de textos corretamente classificados para aquela classe. A revocação é calculada justamente pelo total de textos corretamente classificados para uma classe sobre o total de textos dessa classe. F1-score é a média harmônica entre precisão e revocação para cada classe, ao passo que o F1-macro é a média do F1-score considerando todas as classes.

Tabela 2 – Desempenho dos modelos BERTimbau (PtBr) usando as estratégias de ajuste fino da rede neural (RN) e da matriz de *embedding* com os classificadores (SVM e RF). Avaliando com as seguintes métricas: Precisão (P), Revocação (R), F1-score (F1), Acurácia (Acc) e F1-macro.

Teste	**Modelos**	**Desaprova**			**Aprova**			**Neutro**			**Acc**	**F1-macro**
		P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**		
Twitter	SVM-PtBr-smote	0,76	0,77	0,77	0,73	0,73	0,73	0,88	0,86	0,87	0,80	0,79
	SVM-PtBr	0,70	0,71	0,71	0,67	0,70	0,69	0,84	0,81	0,82	0,75	0,74
	RF-PtBr-smote	0,75	0,74	0,75	0,81	0,58	0,68	0,78	0,90	0,84	0,78	0,75
	RF-PtBr	0,75	0,68	0,71	0,83	0,56	0,67	0,76	0,94	0,84	0,77	0,74
	RN-PtBr	0,83	0,85	0,84	0,83	0,87	0,85	0,94	0,90	0,92	0,88	0,87
YouTube	SVM-PtBr-smote	0,22	0,67	0,33	0,84	0,83	0,83	0,92	0,70	0,79	0,76	0,65
	SVM-PtBr	0,20	0,89	0,32	0,89	0,70	0,78	0,84	0,66	0,74	0,69	0,62
	RF-PtBr-smote	0,23	0,44	0,30	0,91	0,70	0,79	0,77	0,86	0,81	0,76	0,64
	RF-PtBr	0,32	0,61	0,42	0,95	0,71	0,82	0,78	0,90	0,83	0,79	0,69
	RN-PtBr	0,33	0,72	0,46	0,91	0,88	0,89	0,90	0,80	0,85	0,83	0,73

Primeiramente, discutimos o desempenho dos modelos ao aplicá-los em uma plataforma diferente, i.e., o impacto de treiná-los em um conjunto de comentários do X/Twitter e testá-los com comentários do YouTube. Observa-se que o desempenho dos modelos é melhor quando aplicados à plataforma onde foram treinados. Nesse caso, a acurácia na plataforma YouTube teve impactos negativos que variam de 2-7%, ao passo que o F1-macro (reflexos da precisão, revocação e f1) teve também impactos negativos variando de 6-17%. O classificador SVM obteve as maiores perdas de desempenho e RF as menores perdas em ambas as métricas. A despeito dessa observação, o desempenho dos modelos na plataforma YouTube, em geral, foram maiores que os observados em (FEITOSA et al., 2022). Em especial, a rede neural BERTimbau (RN), que é o melhor modelo, alcançou aumentos de acurácia e F1-macro superiores a 18% e 7% respectivamente em relação à esse trabalho, o que significa que a metodologia de treinamento proposta levou a ganhos relevantes no desempenho.

Em mais detalhes, a rede neural BERTimbau alcançou F1-macro de 87(73)% e acurácia de 88(83)% nas plataformas X/Twitter (YouTube), o que é um desempenho relevante para posicionamentos na literatura (MOHAMMAD; SOBHANI; KIRITCHENKO, 2017). Em comparação aos classificadores, o desempenho da rede neural é superior ao melhor classificador (SVM-PtBr-smote) na plataforma X/Twitter com ganho de 8 pontos percentuais em termos de F1-macro. Na plataforma YouTube, o desempenho da rede

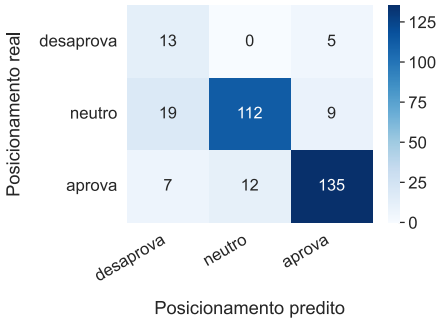
neural é superior ao melhor classificador (RF-PtBr) com ganho de 4 pontos percentuais. Importante observar que, considerando o custo computacional para ajustar a rede neural BERTimbau, o modelo RF apresenta um bom compromisso entre desempenho e custo de treinamento.

Em seguida, analisamos o efeito do tratamento de desbalanceamento das classes para os classificadores RF e SVM. Esse tratamento visa aumentar a precisão e a revocação para as classes minoritárias, evitando que apenas comentários com posicionamentos claros de aprovação ou desaprovação sejam classificadas corretamente e tendencia a posicionamentos neutros. O balanceamento com *smote* teve impacto positivo no desempenho de SVM e RF nos comentários do Twitter, aumentando o F1 e, por consequência, aumentando a acurácia e F1-macro para as posturas de aprovação e desaprovação. Nesse caso, SVM teve o melhor desempenho alcançando F1-macro melhores (79%), superando as marcas de F1-macro do classificador RF (75%). Contudo, observa-se menor impacto do balanceamento com smote para os comentários do YouTube, e RF teve o melhor desempenho nessa plataforma alcançando F1-macro melhor (69%).

(a)

Comentário	Real	Predito
vergonha para o pais mesmo	-1	1
prato cheio para o trafico e seus simpatizantes	-1	1
abordagem foi execucao mesmo	-1	1
o importante e que morreu	1	0
sim pq a policia nao pode subir nos morros	1	0
bandido bom e bandido	1	0
kkkkkkkkkkkkk discurso dentro da cartilha parabens	0	1
nenhum adv para lutar pela ajudar essa mulher	0	-1
isso e um tiro no proprio pe	0	-1

(b)



(c)

Comentário	Real	Predito
Ban... bom é ban... morto!!!!!	1	0
Que na próxima vez, ele venha o dobro.	1	0
Eu sou brasileiro e estou a favor da polícia	1	0
O câncer do Brasil e os fardados , apenas 1 morto foi pouco !	-1	1
Poderiam usar armas de choque, tranquilizantes e etc, não precisava ser letal a menos que a PM quisessem esconder algo.	-1	1
Como são ruim de Tiro esses pms de SP ein quase toda troca de tiro eles leva a pior mesmo usando colete e nunca acerta o bandido.	-1	1
s t f. tem ser preso vagabundos	0	1
Petistas defendendo os seus bandidos de estimação	0	-1
Cdd é cemitério de polícia	0	1

(d)

Figura 5 – Erros do modelo BERTimbau (RN-PtBr): (a) matriz de confusão e (b) exemplos do Twitter, (c) matriz de confusão e (d) exemplos do YouTube.

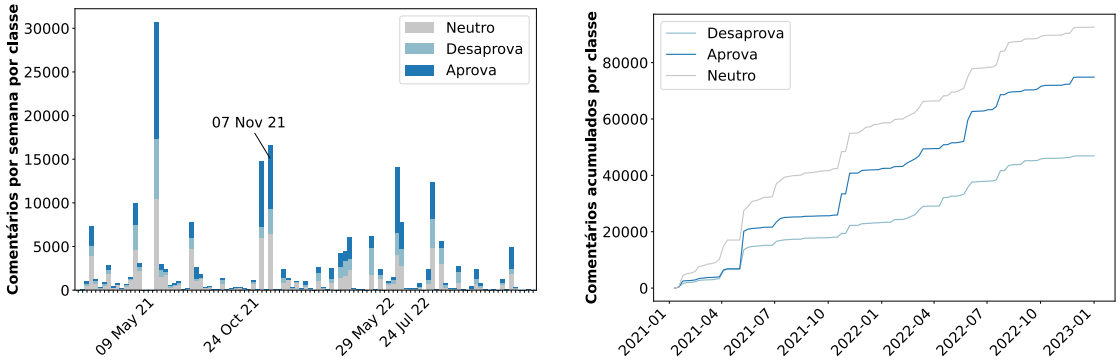
As Figuras 5a e 5c reportam matrizes de confusão para as plataformas X/Twitter e YouTube no modelo RN-PtBr, que obteve o melhor desempenho, como mostra a Tabela **??**. Cada linha representa os comentários em uma classe real, enquanto cada coluna representa

os comentários em uma classe prevista, o que nos permite analisar onde o classificador mais erra e como isso acontece em função da classe. De forma geral, é possível notar que, em termos absolutos, o maior desafio na tarefa aqui endereçada está em diferenciar as classes *Neutro* da *Desaprova* e *Desaprova* da *Aprova*. Por sua vez, os exemplos apresentados nas Figuras 5b e 5d mostram como comentários curtos, contendo algum tipo de sarcasmo ou que não apontam um posicionamento claro em relação à ação policial, dificultam a tarefa de inferência aqui endereçada. Nós observamos que, de fato, esses casos são representativos dos comentários classificados incorretamente.

4.2.2 Monitorando Posicionamentos de Usuários

Nesta subseção, discutimos os resultados da inferência do melhor modelo (RN-PtBr) nos comentários do YouTube coletados com o monitoramento ao longo dos anos 2021 e 2022.

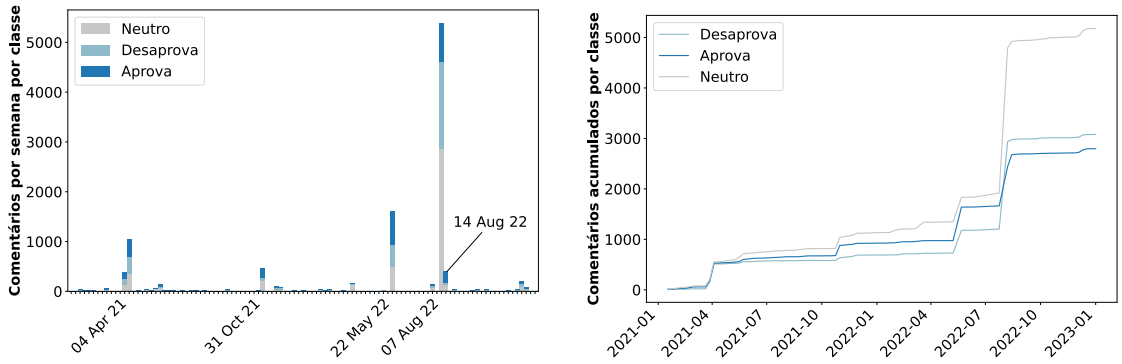
A Figura 6 mostra a distribuição de comentários sobre operações policiais no Brasil em geral nesse período em unidades de tempo semanal e acumulado para as três classes de posicionamentos inferidas pelo sistema. Observa-se majoritariamente neutralidade, em especial nos picos de comentários cujas datas são indicadas na Figura 6a. Isso ocorre devido a interferência de temas externos ao contexto, em especial politização de discussões sobre operações policiais. Contudo, aprovações dominam notavelmente desaprovações em quatro dos cinco picos indicados. Na Figura 6b a tendência de aprovação sob desaprovação está mais evidente. Considerando o acumulado total no período obtivemos posicionamentos de aprovação 35%, desaprovação 22% e neutralidade 43%. Observamos que comentários não neutros aumentam significativamente no mês 05-2021: aprovações, inicialmente pouco superior a 6.400 em Abril, aumenta para 20.204 a partir de 9 de Maio de 2021, tornando esse o mês com os maiores percentuais de opiniões de aprovação, impactando no acumulado notavelmente.



(a) (b)

Figura 6 – Distribuição de comentários sobre operações policiais no Brasil em geral em 2021 e 2022: (a) unidades de tempo semanal e (b) acumulado.

Quanto aos picos semanais de comentários na Figura 6a, é importante destacar que eles estão relacionados com vídeos das seguintes operações: "*Operação no Jacarezinho: moradores registram ação policial que deixou 25 mortos no Rio (09 mai. 21)*", "*Bandidos são mortos após tentarem fugir da PM - SBT Brasil (24 out. 21)*", "*Combate ao "Novo Cangaço": 26 mortos em operação policial (07 nov. 2021)*", "*Operação policial faz 22 mortos no Rio de Janeiro (29 maio 2022)*", "*Policiais do BOPE são encurralados por bandidos, no Complexo do Alemão (24 jul. 2022)*".



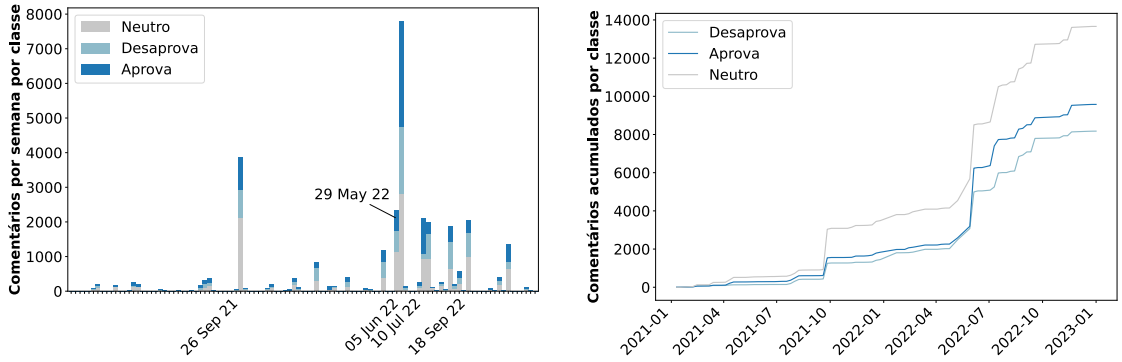
(a) (b)

Figura 7 – Distribuição de comentários sobre operações policiais na região Nordeste do Brasil em 2021 e 2022: (a) unidades de tempo semanal e (b) acumulado.

A Figura 7 mostra a distribuição de comentários sobre operações policiais na região Nordeste do Brasil entre 2021 e 2022. Observam-se picos semanais, porém em uma quantidade menor de comentários comparada ao cenário Brasil geral dado que esses eventos foram monitorados em uma área específica. Cinco picos são identificados, mas o destaque é para a semana de 07 de agosto de 2022, quando a operação reportada no vídeo "*Policiais embriagados são presos após bater em carro e gerar confusão na Paraíba*" provoca uma mudança na tendência da opinião pública acumulada no período, i.e., desaprovações ultrapassam aprovações, como mostra a Figura 7b. Assim, a tendência de aprovação não é evidente para os vídeos reportados na região nordeste, e considerando o acumulado temos aprovação 25%, desaprovação 28% e neutralidade 47%.

Finalmente, para os comentários na região sudeste, observa-se picos com volumes superiores à região nordeste e mais concentrados em 2022, como mostra a Figura 8a. Não há evidência de posicionamentos de aprovação ou desaprovação dominantes. Porém, o pico destacado na semana de 5 Junho de 2022 (operação "*Intenso tiroteio e policiais encurralados no Complexo da Penha*") impacta nos comentários de aprovação a tornando levemente superior às desaprovações. Portanto, observa-se uma tendência de aprovação para os vídeos reportados na região sudeste, ainda que leve, indicado pelos percentuais de posicionamentos acumulados em aprovação 31%, desaprovação 26% e neutralidade 43%.

A análise dos comentários sobre vídeos de ações policiais indica que, de maneira



(a) (b)

Figura 8 – Distribuição de comentários sobre operações policiais na região sudeste do Brasil em 2021 e 2022: (a) unidades de tempo semanal e (b) acumulado.

geral, há uma tendência de aprovação das ações policiais pela população brasileira. No entanto, observa-se uma divergência regional significativa devido aos eventos abordados. Enquanto no nordeste há uma percepção negativa das ações policiais, justificado pelo caso específico destacado na Figura 8a. No sudeste essa percepção se alinha mais positivamente com a tendência geral. Esse contraste reflete a importância de considerar contextos diferentes na avaliação da opinião pública sobre segurança pública.

4.3 Resumo

Neste Capítulo foi proposto um sistema para coleta, pré-processamento e inferência de opinião sobre operações policiais no Brasil em geral e por regiões, baseado em comentários sobre vídeos da plataforma YouTube nesse contexto. Considerando os desafios em PLN, esse trabalho contribui na melhoria do desempenho de modelos para inferência de posicionamentos baseados em arquiteturas *transformers* com uma estratégia que utiliza LLM na redução de conflitos de rotulação entre anotadores. Foram conduzidos experimentos que mostram o impacto positivo dessa estratégia no desempenho de modelos para inferir posicionamentos de comentários em plataformas de mídias sociais distintas. Especificamente, comentários do X/Twitter foram utilizados para treinos e testes de modelos, ao passo que utilizamos comentários do YouTube para testes, seguidos de demonstrações extensivas de aplicações do sistema proposto.

5 Detecção de Posicionamento entre Plataformas

Este capítulo tem como foco investigar a capacidade de modelos de PLN generalizarem entre diferentes plataformas de mídias sociais. Aqui ampliamos a análise do capítulo anterior para o cenário interplataformas, avaliando comentários provenientes tanto do YouTube quanto do X/Twitter, plataformas que diferem substancialmente em estilo, estrutura discursiva e padrões linguísticos. Está organizado da seguinte forma: Seção 5.1 aborda a metodologia utilizada descrevendo os dados, a abordagem de anotação híbrida e os modelos e métricas; Seção 5.2 mostra o desempenho intra e interplataforma e uma discussão dos resultados obtidos.

5.1 Metodologia

Esta seção apresenta a abordagem metodológica adotada para avaliar o desempenho e a capacidade de generalização de modelos de classificação de posicionamento aplicados a comentários de usuários sobre operações policiais no Brasil. O estudo foca em duas plataformas de mídias sociais distintas: X/Twitter e YouTube. Dada a natureza informal, ruidosa e altamente subjetiva do conteúdo em mídias sociais, bem como a escassez de corpora rotulados em português, adotamos um desenho experimental comparativo. Nossa metodologia inclui tanto algoritmos tradicionais de aprendizado supervisionado quanto modelos baseados em *transformers*, com o objetivo de explorar seu comportamento sob restrições de mundo real e em condições de uso interplataforma.

O pipeline metodológico está organizado em cinco componentes principais: (i) coleta e descrição dos dados; (ii) pré-processamento de texto com atenção ao contexto linguístico; (iii) anotação híbrida, combinando julgamento humano com validação baseada em LLMs; (iv) treinamento de modelos utilizando diferentes arquiteturas, incluindo BERTimbau ajustado (*fine-tuned*) e classificadores tradicionais; e (v) um protocolo de avaliação baseado em métricas robustas, *bootstrapping* para confiança estatística e experimentos de generalização entre plataformas. O conjunto de dados, bem como os códigos-fonte para implementar todos os componentes metodológicos mencionados acima, estão disponíveis para a comunidade no repositório: https://github.com/LABPAAD/cross-platform.

5.1.1 Coleta e Descrição dos Dados

Empregamos dois conjuntos de dados compostos por comentários de usuários sobre operações policiais no Brasil, coletados nas plataformas X/Twitter e YouTube. Essas

plataformas foram selecionadas devido às suas dinâmicas comunicativas distintas: usuários do X/Twitter tendem a publicar comentários curtos e diretos, muitas vezes acompanhados de links para notícias, enquanto usuários do YouTube geralmente reagem ao conteúdo em vídeo com comentários mais longos e elaborados.

X/Twitter: Utilizamos o conjunto de dados compilado por (FEITOSA et al., 2022), composto por comentários de usuários sobre reportagens jornalísticas acerca de operações policiais. As notícias originais foram obtidas em grandes portais de mídia brasileiros, como *G1*[1], *UOL*[2] e *Folha de S. Paulo*[3], amplamente reconhecidos como veículos nacionais de referência. O conjunto inclui comentários vinculados a oito notícias de grande repercussão publicadas entre 2019 e 2021. Utilizando a API do Twitter, os autores originais recuperaram um total de 16,276 tweets a partir de consultas envolvendo títulos e URLs das matérias selecionadas. Esse conjunto é particularmente valioso devido à sua relevância temática, amplitude temporal e curadoria manual em torno de eventos de significativo interesse público.

YouTube: Também utilizamos o conjunto de dados descrito em (ROCHA et al., 2024), que reúne comentários em vídeos noticiosos sobre operações policiais publicados entre 2021 e 2022. Os dados foram coletados por meio da API do YouTube, com base em buscas por palavras-chave relacionadas a ações policiais. Os autores aplicaram filtragem temática para reter apenas vídeos diretamente relevantes ao tópico, resultando em 425 vídeos e um total de 257,025 comentários. Essa coleção fornece uma amostra rica e diversa do discurso público sobre a atuação das forças de segurança no Brasil, expressa em textos espontâneos gerados por usuários em centenas de fontes multimídia.

5.1.2 Pré-processamento dos Dados

Aplicamos uma série de etapas de pré-processamento aos dados textuais de ambas as plataformas utilizando a biblioteca NLTK[4] implementada em Python. Especificamente, removemos hiperlinks, emojis e quebras de linha, elementos comuns em conteúdo de mídias sociais, mas que tipicamente introduzem ruído sem contribuir semanticamente para a tarefa de classificação de posicionamento.

Importante destacar que optamos por manter as *stopwords*, termos frequentes como preposições, artigos e pronomes, de forma a preservar o contexto linguístico completo de cada sentença. Essa decisão é especialmente relevante quando se utilizam modelos baseados em *transformers*, como o BERT, que dependem da compreensão das relações sintáticas e semânticas entre as palavras em uma sequência. Remover *stopwords* poderia prejudicar a capacidade do modelo de capturar essas dependências contextuais, potencialmente

[1] <https://g1.globo.com/>
[2] <https://www.uol.com.br/>
[3] <https://www.folha.uol.com.br/>
[4] <https://www.nltk.org/>

degradando o desempenho. Por fim, descartamos comentários compostos por apenas uma palavra, pois oferecem baixo conteúdo semântico e contexto insuficiente para uma classificação confiável. Essas instâncias frequentemente não expressam um posicionamento discernível e podem introduzir ruído ou viés no treinamento dos modelos ao incentivar padrões de decisão excessivamente simplificados.

5.1.3 Estratégia Híbrida de Anotação

Para garantir rótulos de classe confiáveis e semanticamente consistentes, empregamos uma estratégia híbrida de anotação combinando julgamento humano com validação automatizada. A tarefa de classificação de posicionamento foi formulada com três classes mutuamente exclusivas:

- **Aprova**: o comentário expressa explicitamente apoio à ação policial ou aos seus resultados. Isso inclui endosso moral (por exemplo, elogios à atuação policial) ou aprovação pragmática (por exemplo, considerar a ação necessária ou eficaz).
- **Desaprova**: o comentário expressa crítica, rejeição ou preocupação em relação à operação policial. Isso inclui oposição moral (por exemplo, condenação da violência ou de abusos) ou crítica política (por exemplo, questionamento da legitimidade ou dos motivos da intervenção).
- **Neutro**: o comentário não transmite um posicionamento claro, expressa ambiguidade ou se concentra em aspectos não relacionados (por exemplo, gramática, humor ou fatos periféricos), tornando-o inadequado para identificação de posicionamento.

Inicialmente, três anotadores estudantes da área de Sistemas de Informação rotularam manualmente uma amostra dos comentários de acordo com essas diretrizes. Para mitigar inconsistências comuns em tarefas de rotulação subjetiva, especialmente em contextos politicamente polarizados, adotamos uma abordagem híbrida inspirada em (GILARDI; ALIZADEH; KUBLI, 2023; HUANG; HUANG, 2024). Especificamente, submetemos cada comentário anotado manualmente à API GPT-3.5 Turbo[5] utilizando um *prompt* elaborado para classificar posicionamento de forma consistente e baseada em regras. Apenas os comentários para os quais os rótulos humanos e do GPT concordaram foram mantidos no conjunto final. Esse filtro conservador visa aumentar a confiabilidade da anotação ao triangulá-la por meio do acordo entre agentes independentes (humano e LLM). Ele também minimiza casos ruidosos ou limítrofes que poderiam prejudicar o treinamento dos modelos. Embora essa estratégia reduza o número total de amostras utilizáveis, ela assegura maior fidelidade semântica, essencial em tarefas que envolvem detecção de opinião

[5] <https://platform.openai.com>

sutil. Os *prompts* GPT usados para classificar posicionamento foram adaptados de Kocoń et al. (KOCOŃ et al., 2023), e estão disponíveis no repositório mencionado acima.

X/Twitter: Após o pré-processamento, o conjunto do X/Twitter continha 4.467 comentários. Após o procedimento de validação híbrida, 2.671 foram retidos. Esses comentários abrangem oito eventos distintos e compreendem um vocabulário de 7.669 termos únicos. O tamanho médio dos comentários é de 115,51 caracteres (DP: 102,66).

YouTube: A partir de uma coleção maior de 257.025 comentários do YouTube, realizamos uma amostragem aleatória estratificada para selecionar 4.000 comentários para anotação manual. Após aplicar o filtro de concordância híbrida, 2.867 foram retidos. Esses comentários abrangem 102 vídeos distintos e contêm um vocabulário de 8.163 termos únicos. O tamanho médio dos comentários é de 106,11 caracteres (DP: 139,48), conforme resumido na Tabela 3.

Tabela 3 – Estatísticas descritivas dos conjuntos de dados finais rotulados após filtragem híbrida.

Plataforma	# Comentários	% Aprova	% Desaprova	% Neutro	Vocabulário	# Vídeos	Tam. Médio (Desvio padrão)
Twitter / X	2,671	26.43%	29.61%	43.95%	7,669	8	115.51 (102.66)
YouTube	2,867	36.34%	25.18%	38.47%	8,163	102	106.11 (139.48)

5.1.4 Modelagem e Arquiteturas Utilizadas

Para avaliar a eficácia e a capacidade de generalização de diferentes modelos de detecção de posicionamento, adotamos uma estratégia de modelagem comparativa que abrange tanto algoritmos tradicionais de aprendizado de máquina quanto arquiteturas baseadas em *transformers*. Essa abordagem dupla permite explorar os trade-offs entre complexidade do modelo, eficiência de treinamento e acurácia, especialmente relevantes em aplicações de mundo real nas quais restrições computacionais e disponibilidade de dados rotulados variam amplamente.

Empregamos a arquitetura *Bidirectional Encoder Representations from Transformers* (BERT) (DEVLIN et al., 2019), especificamente o modelo *BERTimbau-large* (SOUZA; NOGUEIRA; LOTUFO, 2020), pré-treinado em grandes corpora de português brasileiro. Essa variante foi selecionada devido à sua comprovada capacidade de capturar dependências contextuais profundas em textos curtos e informais, como aqueles encontrados em plataformas de mídias sociais como X/Twitter e YouTube, tornando-o particularmente adequado para nossa tarefa de classificação. O ajuste fino (*fine-tuning*) foi realizado adicionando-se uma camada densa de saída para classificação multiclasse de posicionamento, permitindo que o modelo se adapte às nuances específicas do discurso sobre segurança pública.

Além do modelo baseado em *transformers*, exploramos uma rota de modelagem mais leve. Embeddings textuais foram extraídos utilizando o modelo *Legal-BERTimbau-*

large (SOUZA; NOGUEIRA; LOTUFO, 2020), uma variante ajustada a textos jurídicos e formais em português, para codificar cada comentário em uma representação vetorial de comprimento fixo. Esses embeddings foram então utilizados como atributos de entrada para dois classificadores amplamente adotados em aprendizado de máquina: *Random Forest* (RF) e *Support Vector Machine* (SVM) (MINING, 2006). Essa estratégia complementar atende a vários propósitos: (i) comparar o desempenho de modelos clássicos com abordagens de aprendizado profundo; (ii) avaliar a robustez dos modelos em cenários de poucos recursos; e (iii) simular cenários realistas de implantação, nos quais o ajuste fino de grandes modelos de linguagem pode não ser computacionalmente viável. Embora RF e SVM não capturem dependências sequenciais ou sintáticas tão efetivamente quanto modelos neurais, eles oferecem vantagens em interpretabilidade, treinamento rápido e desempenho estável quando aplicados a vetores de atributos bem estruturados. Além disso, esses modelos podem atuar como fortes baselines em tarefas de detecção de posicionamento, especialmente quando aplicados a embeddings de alta qualidade extraídos de modelos de linguagem poderosos.

Conduzimos uma extensa busca de hiperparâmetros para cada classificador utilizando uma estratégia de busca baseada em validação. Para RF, exploramos parâmetros como *n_estimators*, *max_depth*, *min_samples_split*, *min_samples_leaf* e *max_features*. Para SVM, variamos *C*, tipo de *kernel*, *gamma* e *degree*. Para o BERTimbau, o ajuste envolveu *batch size*, *learning rate*, *dropout* e *weight decay*. Uma vez identificados os melhores valores com base nos dados de validação, os modelos foram re-treinados em 90% do conjunto de dados (treino + validação) e avaliados em uma partição de teste reservada de 10%.

Todos os experimentos foram conduzidos em ambiente Google Colaboratory[6] com aceleração por GPU. Os dados foram divididos usando uma estratégia *hold-out* estratificada: 80% para treinamento, 10% para validação e 10% para teste. A Tabela 4 apresenta os intervalos de parâmetros testados e as melhores configurações de hiperparâmetros obtidas para cada modelo e plataforma. Notavelmente, as diferenças nos valores ótimos entre os conjuntos do X/Twitter e do YouTube revelam a heterogeneidade textual e estilística entre plataformas e destacam a importância de um ajuste sensível à plataforma.

5.1.5 Protocolo de Avaliação Experimental

Para avaliar o desempenho dos modelos, utilizamos três métricas padrão amplamente empregadas na literatura de classificação: Precisão, Revocação e F1-score (MINING, 2006). A Precisão reflete a proporção de previsões corretas dentre todas as previsões positivas feitas pelo modelo. A Revocação mede a capacidade do modelo de identificar corretamente todas as instâncias relevantes de uma dada classe. O F1-score, por sua vez, é a

[6] https://colab.research.google.com/

Tabela 4 – Intervalos testados e melhores valores de hiperparâmetros por modelo e plataforma.

Modelo	Plataforma	Parâmetro	Intervalo Testado	Melhor Valor
RF	X/Twitter	n_estimators	{100, 200, 300}	100
		max_depth	{None, 10, 20}	None
		min_samples_split	{2, 5, 10}	5
		min_samples_leaf	{1, 2, 4}	1
		max_features	{'auto', 'sqrt', 'log2'}	auto
	YouTube	n_estimators	{100, 200, 300}	200
		max_depth	{None, 10, 20}	20
		min_samples_split	{2, 5, 10}	2
		min_samples_leaf	{1, 2, 4}	4
		max_features	{'auto','sqrt', 'log2'}	auto
SVM	X/Twitter	C	{0.1, 1, 10, 100}	100
		kernel	{'linear', 'rbf', 'poly', 'sigmoid'}	rbf
		gamma	{'scale', 'auto', 0.001, 0.01, 0.1}	scale
		degree	{3, 4, 5}	3
	YouTube	C	{0.1, 1, 10, 100}	0.1
		kernel	{'linear', 'rbf', 'poly', 'sigmoid'}	linear
		gamma	{'scale', 'auto', 0.001, 0.01, 0.1}	scale
		degree	{3, 4, 5}	3
BERTimbau	X/Twitter	batch size	{4}	4
		learning rate	{1e-5, 2e-5, 3e-5, 4e-5, 5e-5}	3e-05
		dropout	{0.05, 0.1, 0.2, 0.4}	0.2
		weight decay	{0.001, 0.01, 0.1, 1}	1.0
	YouTube	batch size	{4}	4
		learning rate	{1e-5, 2e-5, 3e-5, 4e-5, 5e-5}	3e-05
		dropout	{0.05, 0.1, 0.2, 0.4}	0.1
		weight decay	{0.001, 0.01, 0.1, 1}	0.1

média harmônica entre Precisão e Revocação, oferecendo uma visão equilibrada da eficácia da classificação. No contexto específico da detecção de posicionamento em comentários de usuários sobre operações policiais, uma tarefa caracterizada por desequilíbrio entre classes e implicações sociopolíticas sensíveis, damos ênfase particular à Revocação e ao F1-score. A Revocação é essencial para garantir que expressões críticas de aprovação ou desaprovação não sejam ignoradas, o que é crucial para análises que buscam monitorar a percepção pública. O F1-score, ao integrar Revocação e Precisão, fornece uma medida mais robusta e holística do desempenho do modelo.

Para fortalecer a confiabilidade estatística dos resultados, aplicamos a técnica de *bootstrapping* aos dados de teste (TSAMARDINOS; GREASIDOU; BORBOUDAKIS, 2018). Esse método de reamostragem não paramétrico gera intervalos de confiança (ICs) sem exigir suposições de normalidade, uma vantagem importante quando se lida com conjuntos de dados pequenos ou não normalmente distribuídos, como frequentemente ocorre em análises de mídias sociais. Para cada modelo (RF, SVM e BERTimbau) e plataforma (Twitter e YouTube), realizamos 5,000 iterações de *bootstrapping*. Em cada iteração, uma amostra com reposição foi sorteada a partir do conjunto de teste, e as métricas de avaliação foram recalculadas. A partir desses resultados, computamos média, desvio padrão e intervalos de confiança de 95% para todas as métricas.

Além das avaliações intra-plataforma, conduzimos experimentos de generalização inter-plataformas para avaliar a capacidade de transferência dos modelos entre diferentes ambientes linguísticos. Especificamente, modelos treinados em dados do X/Twitter foram testados em comentários do YouTube, e vice-versa. Essa configuração simula um cenário real de implantação em que dados anotados podem estar disponíveis em um domínio, mas

ausentes em outro. Ao avaliar a robustez inter-plataformas, buscamos investigar em que medida modelos de detecção de posicionamento podem generalizar entre contextos discursivos e estilísticos distintos, uma questão importante para a escalabilidade de aplicações de PLN em ecossistemas online heterogêneos.

5.2 Resultados

Esta seção apresenta os resultados da classificação de posicionamento em comentários de usuários sobre operações policiais, com foco em dois cenários principais: (i) *intra-plataforma*, em que o treinamento e o teste são realizados na mesma plataforma de mídia social, e (ii) *inter-plataformas*, em que o treinamento ocorre em uma plataforma e o teste em outra. Relatamos Precisão, Revocação e F1-score, métricas amplamente adotadas para avaliação de classificadores, especialmente em situações de desequilíbrio entre classes. Além disso, fornecemos intervalos de confiança de 95% para cada métrica e classe por meio de *bootstrapping*, conforme descrito na Seção 5.1, a fim de identificar diferenças estatisticamente relevantes entre modelos. A Tabela 5 resume os resultados completos, e a Figura 9 compara visualmente os F1-scores para todas as condições avaliadas. Para facilitar a comparação entre cenários, destacamos em negrito o maior F1-score por classe em cada configuração de treinamento–teste.

5.2.1 Desempenho Intra-Plataforma

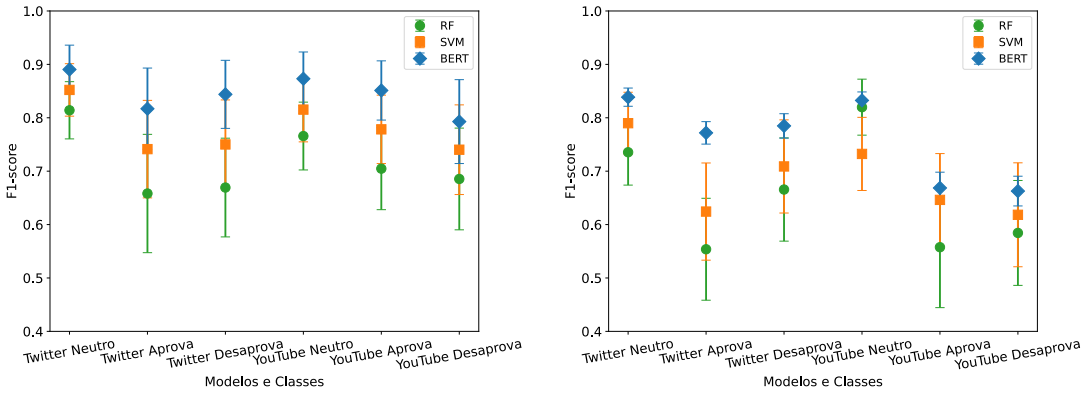
Iniciamos analisando o cenário intra-plataforma, em que modelos são treinados e testados na mesma fonte de dados. Os resultados para YouTube e X são apresentados nas linhas superiores da Tabela 5. Entre todos os modelos avaliados, o BERTimbau obteve as maiores médias de F1-score em ambas as plataformas, variando de 0,79 a 0,89 nas três classes de posicionamento (Neutro, Aprova e Desaprova). O SVM também apresentou desempenho competitivo, frequentemente produzindo escores próximos aos do BERTimbau, enquanto o Random Forest (RF) exibiu maior variação de desempenho e se mostrou menos robusto, especialmente para classes polarizadas.

Embora a Figura 9(a) mostre intervalos de confiança sobrepostos entre os modelos em todas as configurações intra-plataforma, uma análise mais detalhada revela fragilidades específicas por classe. Em X, a classe Aprova apresentou desempenho reduzido, com F1-scores entre 66% e 82%, enquanto no YouTube a classe Desaprova foi a mais desafiadora, com escores de 69% a 79%. Essas diferenças são consistentes com a distribuição de classes, em que certos posicionamentos aparecem sub-representados. O desequilíbrio resultante provavelmente dificultou a capacidade dos modelos de aprender representações informativas para as classes menos frequentes, comprometendo o desempenho de classificação.

Análises adicionais indicam que essas disparidades entre classes são impulsionadas

Tabela 5 – Desempenho dos modelos RF, SVM e BERTimbau em termos de Precisão (P), Revocação (R) e F1-score (F1), nos cenários intra-plataforma e inter-plataformas. Valores em negrito indicam os maiores valores médios por classe em cada configuração treino–teste.

Treino/Teste	Modelo	Neutro			Aprova			Desaprova		
		P	R	F1	P	R	F1	P	R	F1
YouTube/YouTube	RF	0.68 ± 0.08	0.87 ± 0.07	0.77 ± 0.06	0.80 ± 0.09	0.63 ± 0.09	0.70 ± 0.08	0.74 ± 0.11	0.64 ± 0.11	0.69 ± 0.10
	SVM	0.83 ± 0.08	0.80 ± 0.08	0.82 ± 0.06	0.76 ± 0.08	0.80 ± 0.08	0.78 ± 0.06	0.75 ± 0.11	0.74 ± 0.11	0.74 ± 0.08
	BERTimbau	**0.87 ± 0.06**	**0.87 ± 0.07**	**0.87 ± 0.05**	**0.82 ± 0.08**	**0.88 ± 0.07**	**0.85 ± 0.06**	**0.84 ± 0.09**	**0.75 ± 0.11**	**0.79 ± 0.08**
X/X	RF	0.72 ± 0.07	0.94 ± 0.05	0.81 ± 0.05	0.88 ± 0.11	0.53 ± 0.12	0.66 ± 0.11	0.71 ± 0.11	0.63 ± 0.11	0.67 ± 0.09
	SVM	0.79 ± 0.07	0.92 ± 0.05	0.85 ± 0.05	0.81 ± 0.11	0.68 ± 0.11	0.74 ± 0.09	0.80 ± 0.10	0.71 ± 0.11	0.75 ± 0.08
	BERTimbau	**0.89 ± 0.06**	**0.89 ± 0.06**	**0.89 ± 0.05**	**0.84 ± 0.09**	**0.80 ± 0.10**	**0.82 ± 0.08**	**0.83 ± 0.09**	**0.86 ± 0.08**	**0.84 ± 0.06**
YouTube/X	RF	0.75 ± 0.07	0.90 ± 0.06	0.82 ± 0.05	0.69 ± 0.13	0.47 ± 0.12	0.56 ± 0.11	0.59 ± 0.11	0.58 ± 0.11	0.58 ± 0.10
	SVM	0.79 ± 0.08	0.68 ± 0.08	0.73 ± 0.07	0.54 ± 0.10	**0.80 ± 0.10**	0.65 ± 0.09	0.70 ± 0.12	0.56 ± 0.11	0.62 ± 0.10
	BERTimbau	**0.78 ± 0.02**	**0.89 ± 0.02**	**0.83 ± 0.02**	**0.70 ± 0.04**	0.64 ± 0.04	**0.67 ± 0.03**	**0.72 ± 0.03**	**0.62 ± 0.03**	**0.66 ± 0.03**
X/YouTube	RF	0.61 ± 0.08	0.92 ± 0.06	0.74 ± 0.06	0.78 ± 0.12	0.43 ± 0.09	0.55 ± 0.10	0.72 ± 0.12	0.62 ± 0.12	0.67 ± 0.10
	SVM	0.70 ± 0.08	0.90 ± 0.06	0.79 ± 0.06	0.84 ± 0.10	0.50 ± 0.10	0.62 ± 0.09	0.66 ± 0.10	0.76 ± 0.10	0.71 ± 0.09
	BERTimbau	**0.83 ± 0.02**	**0.85 ± 0.02**	**0.84 ± 0.02**	**0.83 ± 0.02**	**0.72 ± 0.03**	**0.77 ± 0.02**	**0.73 ± 0.03**	**0.84 ± 0.03**	**0.78 ± 0.02**



((a)) Avaliação na mesma plataforma

((b)) Avaliação interplataformas (a plataforma indica a origem do treinamento).

Figura 9 – Desempenho em F1-score (eixo x escalado de 0,4 a 1,0) dos modelos RF, SVM e BERTimbau nos cenários avaliados.

principalmente por valores reduzidos de Revocação. Por exemplo, a Revocação da classe Aprova em X variou de 53% a 80%, e para Desaprova no YouTube de 64% a 75%. Isso sugere que muitas instâncias verdadeiras dessas classes não foram corretamente identificadas. Embora os intervalos de confiança obtidos por *bootstrapping* se sobreponham na maioria dos casos, o que torna a significância estatística inconclusiva, os padrões observados destacam a influência da frequência de classes e das tendências específicas de cada plataforma. Esses achados reforçam a importância de incorporar distribuições de dados sensíveis à plataforma no desenvolvimento de modelos de classificação de posicionamento.

5.2.2 Generalização Inter-Plataformas

Voltamo-nos agora ao cenário inter-plataformas, em que modelos treinados em uma plataforma são avaliados em dados de outra fonte. Esses resultados são apresentados nas duas últimas linhas da Tabela 5. Como esperado, houve queda de desempenho nesse contexto, confirmando os desafios de mudança de domínio e generalização. No entanto, modelos treinados em X demonstraram maior robustez ao serem transferidos para o

YouTube, com reduções de F1-score entre 1% e 17%. Em contraste, treinar no YouTube e testar em X levou a quedas mais acentuadas, com perda de desempenho de até 21%.

Esses resultados assimétricos sugerem que a natureza do texto em X, tipicamente mais curto, direto e sintaticamente simples, pode oferecer padrões linguísticos mais consistentes para aprendizado, o que por sua vez melhora a capacidade de transferência. Em contrapartida, comentários do YouTube tendem a ser mais longos e variados em estilo, o que pode introduzir complexidade adicional durante o treinamento e reduzir a capacidade de generalização.

A Figura 9(b) reforça esses achados. Quando treinado em X (painéis à esquerda), o BERTimbau atingiu os maiores F1-scores em todas as classes e apresentou desempenho estatisticamente superior na classe Aprova, como indicado pelos intervalos de confiança não sobrepostos em relação a RF e SVM. Isso sugere que modelos baseados em *transformers* conseguem abstrair e preservar representações semânticas mais transferíveis, especialmente quando treinados em dados mais limpos ou regulares.

Por outro lado, o mesmo modelo treinado no YouTube (painéis à direita) apresentou desempenho inferior tanto em Aprova quanto em Desaprova ao ser testado em X. As diferenças de desempenho, particularmente nessas classes polarizadas, revelam que estruturas discursivas específicas de cada plataforma afetam fortemente a capacidade do modelo de generalizar. Isso enfatiza que, além da arquitetura do modelo, a origem dos dados de treinamento desempenha papel crucial na detecção de posicionamento entre plataformas. Em cenários de poucos recursos, plataformas com padrões textuais mais estáveis e discriminativos, como X, podem ser mais adequadas para construir modelos destinados a uso em múltiplas plataformas.

5.2.3 Síntese dos Resultados

De modo geral, os resultados confirmam que o BERTimbau supera consistentemente os classificadores tradicionais (RF e SVM) em cenários intra- e inter-plataformas. Embora todos os modelos sofram degradação de desempenho sob mudança de domínio, a intensidade dessa queda depende fortemente do cenário, sendo influenciada pela plataforma de origem, pela distribuição das classes de posicionamento e pelo grau de polarização das classes. Os intervalos de confiança obtidos por *bootstrapping* fornecem base estatística para essas observações, ainda que a sobreposição frequente desses intervalos muitas vezes impeça afirmações definitivas de significância. Para facilitar a comparação entre cenários e destacar os principais padrões, a Tabela 6 apresenta um resumo do melhor modelo em cada caso, do F1-score médio entre as classes de posicionamento e da principal observação qualitativa para cada configuração de treinamento–teste.

O conjunto de evidências sugere que, para além da arquitetura do modelo, a

Tabela 6 – Resumo do desempenho nos diferentes cenários de avaliação. O F1-score médio é macro-médio sobre as classes de posicionamento.

Cenário	**Melhor Modelo**	**F1 Médio**	**Principal Observação**
YouTube/YouTube	BERTimbau	0.84	Revocação consistentemente alta entre as classes de posicionamento
X/X	BERTimbau	0.85	Forte desempenho em opiniões polarizadas
YouTube → X	BERTimbau	0.67	Queda significativa, especialmente em Aprova/Desaprova
X → YouTube	BERTimbau	0.78	Menor degradação, melhor generalização entre domínios
Média geral	BERTimbau	0.79	Modelo mais estável e acurado entre os cenários

plataforma utilizada no treinamento desempenha papel crítico na determinação do sucesso da generalização. Em particular, X parece oferecer padrões linguísticos mais transferíveis, possivelmente devido à estrutura textual mais curta e padronizada, tornando-o uma fonte mais confiável para treinamento em aplicações multiplataforma em cenários de poucos recursos. Esses achados enfatizam a necessidade de estratégias de seleção e avaliação de dados sensíveis à plataforma ao se construir sistemas de detecção de posicionamento para domínios sensíveis e socialmente relevantes, como o da segurança pública no Brasil.

5.3 Resumo

Este Capítulo realizou uma avaliação sistemática de modelos de detecção de posicionamento aplicados a comentários de usuários sobre operações policiais, comparando classificadores tradicionais e arquiteturas baseadas em *transformers* em cenários intra- e interplataformas. Foram utilizados comentários do YouTube e do X/Twitter, anotados por meio de uma estratégia híbrida que combina julgamento humano e validação por LLMs em três classes de posicionamento (aprova, desaprova e neutro), e avaliados com métricas padrão (precisão, revocação e F1) e intervalos de confiança obtidos por *bootstrapping*. Os resultados mostraram que o BERTimbau superou os classificadores tradicionais em cenários intraplataforma, com F1-macro entre 0,79 e 0,89, mas houve queda de desempenho em cenários interplataformas devido ao desvio de domínio, ainda que modelos treinados no X tenham generalizado melhor para o YouTube do que o inverso. Esses achados ressaltam a influência das características linguísticas específicas de cada plataforma no desempenho dos modelos e reforçam o potencial de estratégias híbridas de anotação para reduzir custos de rotulação mantendo a confiabilidade em contextos de poucos recursos. Entre as limitações,

destaca-se o foco em um único tema e em apenas duas plataformas, além da dificuldade em capturar posicionamentos mais sutis ou ambíguos, especialmente em casos de ironia e sarcasmo.

6 Detecção de Eventos em Vídeos de Violência Urbana

Enquanto os Capítulos 4 e 5 investigam a inferência de posicionamento em comentários de mídias sociais e sua generalização entre plataformas, aqui o foco desloca-se da análise de comentários para a *identificação automática de eventos* em vídeos de violência urbana publicados no YouTube. Em particular, propomos e avaliamos duas heurísticas não supervisionadas para agrupar vídeos que se referem ao mesmo evento real (como uma operação policial ou um confronto armado), explorando exclusivamente atributos textuais (títulos, descrições e, opcionalmente, transcrições) enriquecidos com NER. Esse componente completa o arcabouço desta dissertação ao permitir que a análise de posicionamento seja organizada em torno de eventos concretos, viabilizando monitoramento e estudos de repercussão em larga escala.

6.1 Metodologia

Nesta seção, apresentamos a metodologia de avaliação da abordagem proposta, contemplando as heurísticas de agrupamento semântico e temporal. Inicialmente, descrevemos os conjuntos de dados de vídeos coletados da plataforma YouTube para a avaliação. Em seguida, apresentamos as métricas de avaliação e as configurações empregadas em cada heurística.

6.1.1 Bases de Dados

Utilizamos duas bases de dados distintas para avaliar a proposta, contemplando diferentes perfis de eventos de violência urbana reportados por usuários no YouTube. Optamos pelo YouTube como fonte devido à sua ampla popularidade e penetração no Brasil, bem como à disponibilidade gratuita de dados por meio da API do YouTube (versão 3).

A primeira base, denominada Eventos esparsos, representa um cenário de monitoramento contínuo da violência urbana com alta variabilidade de eventos. Nesse contexto, a coleta parte de buscas por palavras-chave genéricas como “vítimas”, “morte”, “roubo”, “furto” e termos relacionados a “operação policial”, de modo a identificar vídeos potencialmente relevantes. Dessa forma, a maior parte dos eventos capturados é de menor repercussão, geralmente com poucos vídeos, embora alguns casos de grande impacto social também estejam presentes.

Para compor essa base, seguimos a metodologia de monitoramento descrita em (ROCHA et al., 2024), coletando vídeos relacionados a operações policiais no período de 2019 a 2023. As buscas semanais são realizadas de forma consecutiva, totalizando 104 unidades de tempo (semanas) ao longo dos anos cobertos. Cada busca é limitada a 20 resultados, conforme indicado por (ROCHA et al., 2024) como um bom compromisso entre relevância e diversidade de vídeos retornados pela API.

Após a coleta, aplicamos um filtro para reter apenas vídeos diretamente relacionados a operações policiais. Primeiramente, selecionamos vídeos classificados pela API na categoria "Notícias e Política", a fim de excluir blogs e documentários policiais. Em seguida, mantemos apenas aqueles cujo título contém termos como "pm", "pms", "policia", "policial", "policias", "policiais", "operacao" ou "operacoes", considerando textos transformados para minúsculas e com remoção de acentos e sinais ortográficos. Esse processo resulta em 960 vídeos, correspondentes a 834 eventos distintos, dos quais 65 possuem múltiplos vídeos (191 no total) e 769 são representados por apenas um vídeo.

A segunda base, denominada Alta repercussão, concentra-se em eventos de grande repercussão nas mídias sociais e na imprensa, contemplando exclusivamente casos reportados em múltiplos vídeos. Para sua construção, identificamos sete incidentes de violência urbana que resultaram em operações policiais de ampla divulgação no Brasil em 2024, por meio de busca manual em mídias sociais e veículos jornalísticos. Em seguida, utilizamos palavras-chave específicas de cada incidente para coletar vídeos via API do YouTube. Após curadoria manual, selecionamos 465 vídeos, todos pertencentes a eventos com múltiplos registros.

A Tabela 7 resume as principais características de ambas as bases. Observa-se que Eventos esparsos é majoritariamente composta por eventos com apenas um vídeo (média de 2,9 vídeos/evento), enquanto a Alta repercussão reúne eventos com elevado número de vídeos por evento (média de 66,4 vídeos/evento), configurando cenários de complexidade distinta para as heurísticas.

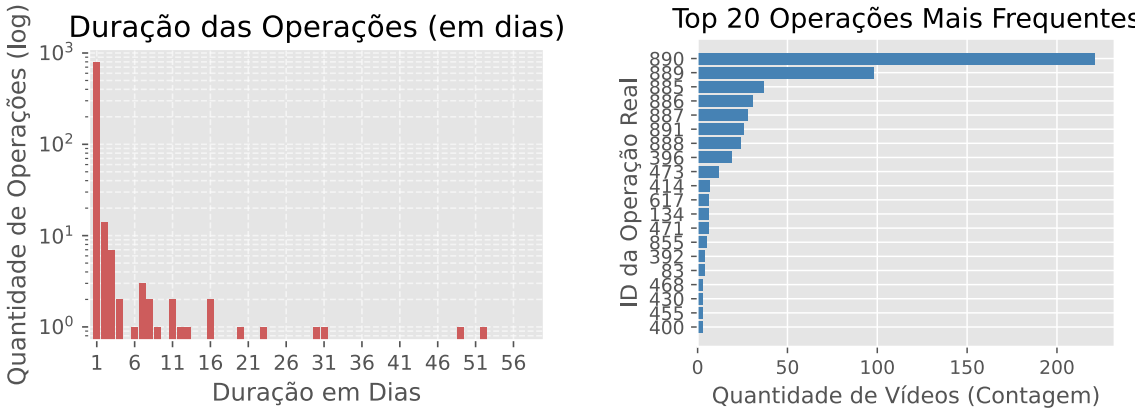
Tabela 7 – Principais características dos conjuntos de dados.

Métrica	**Ev. esparsos**	**Al. repercussão**	**Total**
Total de vídeos	960	465	1 425
Total de eventos distintos	834	7	841
Ev. com apenas 1 vídeo	769	0	769
Eventos com ≥ 2 vídeos	65	7	72
Vídeos em ev. múltiplos	191	465	656
Média de vídeos por ev.	2,9	66,4	9,1

A Figura 10a mostra a distribuição da duração dos eventos em dias. Nota-se que a maioria concentra-se em apenas um dia, com queda acentuada após dois ou três dias. Menos de 5% das operações se estendem por mais de sete dias, sugerindo que eventos

prolongados são exceção e possivelmente requerem atenção especial das heurísticas para evitar fusões indevidas com outros eventos temporalmente próximos.

A Figura 10b apresenta as vinte operações com maior número de vídeos. Observa-se um padrão concentrado, no qual poucos eventos concentram grande volume de material (por exemplo, a operação 890 com 225 vídeos), enquanto a maioria possui cobertura substancialmente menor. Essa distribuição em "degrau" implica que, nos eventos de maior repercussão, as heurísticas devem lidar com grande diversidade narrativa e de perspectivas sobre o mesmo fato; já nos demais, a tarefa exige identificar correspondências a partir de sinais mais sutis.



(a) (b)

Figura 10 – Distribuição e duração dos eventos: (a) duração dos eventos em dias e (b) vinte operações com maior número de vídeos.

Para avaliar a eficácia das heurísticas de agrupamento, construímos um conjunto de referência com rótulos confiáveis para cada evento. Esse *ground truth*[1] é obtido por três voluntários estudantes da área de Sistemas de Informação e Ciência da Computação, que associam manualmente cada vídeo a um identificador numérico de evento, permitindo mapear múltiplos vídeos para um mesmo evento real. O processo de rotulação considera título, descrição, transcrição e data de postagem; quando necessário, realizamos a visualização completa do vídeo. Os voluntários seguem um protocolo padronizado: (i) leitura integral do título e descrição e, quando aplicável, visualização do vídeo; (ii) comparação com vídeos já rotulados, a fim de evitar duplicidades ou erros; e (iii) atribuição de um novo rótulo caso nenhuma correspondência seja encontrada. Esse procedimento resulta em um conjunto de dados consistente e confiável, adequado ao desenvolvimento e validação das heurísticas propostas.

A Tabela 8 apresenta cinco exemplos de vídeos rotulados manualmente, ilustrando a heterogeneidade de formatos e conteúdos. Há títulos objetivos e descritivos, bem como

[1] Neste contexto, o termo *ground truth* refere-se ao conjunto de rótulos de referência (ou "padrão-ouro") utilizado como base para avaliar quantitativamente o desempenho das heurísticas de agrupamento.

descrições incompletas ou pouco informativas. Tal cenário reforça a necessidade de heurísticas robustas, capazes de lidar com lacunas textuais e variações linguísticas na identificação de vídeos referentes ao mesmo evento.

Tabela 8 – Exemplos de eventos rotulados manualmente.

Rótulo	**Trecho do Título**	**Trecho da Descrição**
2	"PM realiza operação em aglomerado contra tráfico..."	"Militares ocuparam durante toda a manhã um aglomerado ..."
11	"Número de mortos na operação policial..."	"O número de mortos na operação policial..."
11	"Operação policial em Jacarezinho deixa..."	"Uma operação policial realizada nesta quinta-feira..."
11	"#OABRJDebate a operação policial no Jacarezinho"	"Neste programa, nossos convidados avaliam a operação..."
88	"Operação policial no Complexo do Alemão..."	"Uma operação conjunta realizada entre..."

6.1.2 Métricas e Configurações

Avaliamos o desempenho da abordagem proposta (heurísticas de agrupamento semântico e temporal) utilizando métricas de acurácia, precisão, revocação e F1-score adaptadas para avaliação de agrupamentos (MANNING; RAGHAVAN; SCHüTZE, 2008), bem como métricas específicas para comparação de partições: NMI (Normalized Mutual Information): e AMI (Adjusted Mutual Information) (VINH; EPPS; BAILEY, 2009; ROMANO et al., 2016). Para referência de desempenho, aplicamos os mesmos agrupamentos com o modelo de linguagem GPT-4 (HUANG; HUANG, 2024) como linha de base. Os parâmetros das heurísticas seguem as definições da Seção 6.2.

As métricas de acurácia, precisão, revocação e F1-score foram adaptadas para o contexto de agrupamento, utilizando os rótulos manuais como referência e mapeando grupos previstos para grupos reais de forma ótima, a fim de maximizar o número de acertos. Para isso, empregamos o método Húngaro (KUHN, 1955) sobre a matriz de confusão gerada pelos agrupamentos, obtendo a correspondência entre rótulos prevista e real. A partir desse mapeamento, a acurácia representa a fração de vídeos corretamente atribuídos; a precisão é a proporção de pares agrupados que realmente pertencem ao mesmo evento; a revocação mede a capacidade do método de recuperar todos os pares corretos; e o F1-score é a média harmônica entre precisão e revocação, equilibrando ambos. Tal abordagem, mapeamento ótimo seguido de métricas macro, é padrão em avaliação de algoritmos de clustering (MANNING; RAGHAVAN; SCHüTZE, 2008).

O NMI mede a quantidade de informação compartilhada entre os agrupamentos previstos e os rótulos reais, normalizada pela média das entropias de ambas as partições, assumindo valores entre 0 (nenhuma relação) e 1 (correspondência perfeita) (VINH; EPPS; BAILEY, 2009). Já o AMI corrige o valor da informação mútua pela expectativa de acordo ao acaso, mantendo o intervalo $[-1, 1]$ e permitindo comparações mais justas (ROMANO et al., 2016). Por essa razão, o AMI é menos sensível a flutuações aleatórias e ao número de grupos, penalizando situações como granularidade excessiva (muitos clusters minúsculos)

ou superfragmentação de grupos grandes em vários clusters pequenos, aspectos observados em alguns cenários e discutidos na Seção 6.3.

No agrupamento semântico, os pesos atribuídos a cada atributo foram $w_e = 0{,}07$ para *embeddings*, $w_l = 0{,}39$ para entidades de localização e pessoa, $w_v = 0{,}25$ para número de vítimas, $w_o = 0{,}18$ para nomes de operações policiais e $w_t = 0{,}11$ para tópicos. A similaridade final (S) é a média ponderada desses atributos, e a formação de grupos ocorre conectando vídeos com $S \geq \tau = 0,55$, valor ajustado para privilegiar grupos menores e mais homogêneos. No agrupamento temporal, adotamos uma janela de $w = 15$ dias (total de 30 dias de abrangência), cobrindo 15 dias antes e depois da data de cada vídeo. Essa configuração foi definida a partir da análise exploratória das bases, que indicou que a maioria dos eventos de violência urbana repercute no YouTube por cerca de duas a três semanas.

Por fim, o GPT-4 foi avaliado a partir de um *prompt* estruturado, fornecendo a lista de vídeos em formato CSV com título, descrição, data e transcrição (quando disponível). O modelo foi instruído a usar o conteúdo textual e a proximidade temporal para atribuir rótulos numéricos sequenciais a cada evento, garantindo que vídeos do mesmo evento recebessem o mesmo rótulo. O objetivo foi testar a capacidade do LLM em realizar agrupamento com base em interpretação semântica e contextual, sem ajustes finos ou treinamento adicional.

Por fim, elaboramos um *prompt* fornecido ao modelo de linguagem GPT-4.0, com o objetivo de rotular vídeos do YouTube relacionados a operações policiais. O *prompt* especifica as entradas no formato CSV, delimitadas por ponto e vírgula, contendo as colunas título, descrição, data de postagem e transcrição (quando disponível). O modelo foi instruído a utilizar o conteúdo textual, complementado pela proximidade temporal, para determinar se diferentes vídeos correspondiam a um mesmo evento. Atribuir um identificador numérico sequencial, iniciando em 1, para cada evento distinto; e garantir que todos os vídeos de um mesmo evento recebessem o mesmo rótulo, retornando como saída um CSV contendo a coluna “evento” na mesma ordem das entradas originais.

Prompt Utilizado

Você é um assistente de IA especialista em identificar vídeos do YouTube sobre operações policiais.
Sua tarefa é ler um CSV (separador ';') com colunas: título, descrição, data de postagem e transcrição (quando disponível).
Considerar todos os vídeos, mesmo com campos ausentes.
Utilizar o conteúdo textual (título, descrição, transcrição) e a proximidade temporal.
Atribuir rótulo numérico único (iniciando em 1) para cada evento, garantindo que vídeos de uma mesma operação policial tenham o mesmo rótulo. Retornar um CSV apenas com a coluna `evento` na ordem original.

6.2 Abordagens Não Supervisionadas para Identificação de Eventos

A identificação automática de vídeos referentes a um mesmo evento em plataformas como o YouTube é desafiadora devido à heterogeneidade linguística, à variação na menção de entidades e à inconsistência nas datas de publicação. Mesmo conteúdos sobre um único episódio de violência urbana ou operação policial podem divergir quanto ao vocabulário, nível de detalhe e estilo narrativo, refletindo a multiplicidade de fontes e a informalidade dos relatos. Esses fatores dificultam a associação entre vídeos correlatos e demandam abordagens capazes de capturar simultaneamente: (i) a similaridade semântica presente nos textos (títulos, descrições e transcrições); e (ii) a coesão temporal e contextual das publicações.

Propomos duas heurísticas não supervisionadas e complementares para o agrupamento de vídeos: a primeira baseada em semelhança textual global, combinando representações vetoriais densas e medidas de similaridade contextual; a segunda fundamentada na proximidade temporal, utilizando janelas de tempo iterativas para identificar eventos comuns. Ambas exploram atributos extraídos automaticamente por NER, viabilizando aplicação em larga escala sobre acervos informais e não curados, com baixo custo computacional em comparação a trabalhos anteriores (ver Seção 2.2).

Ambas as heurísticas compartilham um fluxo inicial de processamento textual que converte o conteúdo multimídia do YouTube em informações estruturadas para análise automática. São extraídos metadados públicos (título, descrição) e, quando disponíveis, transcrições automáticas geradas pela biblioteca Vosk[2] com o modelo *vosk-model-small-pt-0.3*. Em seguida, aplica-se o NER diretamente sobre o texto original, preservando capitalização e acentuação, utilizando o modelo *pt_core_news_lg* da biblioteca spaCy[3].

[2] <https://alphacephei.com/vosk/models>
[3] <https://spacy.io/models/pt>

Essa etapa identifica, de forma automática, menções a localidades, datas, instituições e outras entidades relevantes.

Após a extração das entidades, cada heurística aplica uma estratégia distinta: a primeira explora similaridades semânticas entre vídeos por meio de representações vetoriais. A segunda organiza eventos considerando janelas temporais e a recorrência de entidades. A seguir, detalhamos cada abordagem proposta.

6.2.1 Agrupamento Semântico

A primeira heurística proposta explora múltiplas dimensões de similaridade para agrupar vídeos de um mesmo evento, combinando informações semânticas, entidades nomeadas, sinais quantitativos e contexto temático. A ideia central é que, embora dois vídeos sobre o mesmo incidente possam divergir amplamente em vocabulário ou estrutura narrativa, eles tendem a compartilhar pelo menos algumas dessas pistas. Ao integrar atributos heterogêneos, reduzimos a dependência de coincidências literais e aumentamos a robustez frente à informalidade e fragmentação dos relatos.

O processo, resumido no Algoritmo 1, é dividido em três etapas: (i) extração e representação de atributos, (ii) cálculo de similaridades par a par, e (iii) agrupamento por grafo.

Etapa 1 — Extração e Representação

Partimos da concatenação do título, descrição e, quando disponível, transcrição do vídeo. Em seguida, aplicamos NER diretamente sobre o texto original, preservando capitalização e acentuação, de modo a maximizar a acurácia na detecção de localidades, pessoas, datas e instituições. A escolha do modelo *pt_core_news_lg* da biblioteca spaCy[4] se deve à sua cobertura e desempenho reportados para o português. Além disso, extraímos dois sinais estruturados: (a) número de vítimas, identificado por expressões regulares que buscam termos como "vítima(s)" ou "morte(s)"; e (b) nomes de operações policiais, detectados por padrões do tipo "operação [nome]". Para capturar a semântica global, normalizamos levemente o texto (remoção de ruído e conversão para minúsculas) e geramos *embeddings* densos com o modelo *paraphrase-multilingual-mpnet-base-v2* (REIMERS et al., 2019) da biblioteca Sentence-Transformers[5]. O uso de modelos multilíngues e recursos textuais genéricos torna a abordagem facilmente adaptável a outros domínios e idiomas com mínima reconfiguração. Por fim, extraímos tópicos latentes com BERTopic[6], capazes de revelar padrões temáticos não explícitos no vocabulário.

[4] <https://spacy.io/models/pt>
[5] <https://huggingface.co/sentence-transformers/paraphrase-multilingual-mpnet-base-v2>
[6] <https://maartengr.github.io/BERTopic>

Algoritmo 1: Heurística de Agrupamento Semântico por Similaridades Múltiplas.

Entrada: Conjunto de vídeos $\mathcal{V}$; para cada $v \in \mathcal{V}$: título, descrição, (opcional) transcrição. Hiperparâmetros: pesos $(w_e, w_l, w_v, w_o, w_t)$; limiar τ.
Saída: Partição $\mathcal{C}$ de $\mathcal{V}$ em grupos (eventos).

Etapa 1 — Extração e Representação
para cada $v \in \mathcal{V}$ **faça**
  $x_v \leftarrow$ concatenar {título, descrição, (transcrição)}
  `// (1) NER direto no texto original: preserva capitalização/acentos`
  $\mathrm{ENT}_v \leftarrow \mathrm{NER}(x_v)$ `// locais, pessoas, datas, instituições`
  `// (2) Sinais quantitativos/nomes estruturados`
  $D_v \leftarrow \mathrm{VictimCount}(x_v)$ `// regex em "vítima(s)", "morte(s)"`
  $N_v \leftarrow \mathrm{OpNames}(x_v)$ `// padrão "operação [nome]"`
  `// (3) Normalização leve só para matching/tópicos`
  $\tilde{x}_v \leftarrow \mathrm{Normalize}(x_v)$ `// minúsculas; remover ruído`
  `// (4) Sinais de semântica global e tema`
  $e_v \leftarrow \mathrm{Embed}(\tilde{x}_v)$ `// Sentence-Transformers`
  $T_v \leftarrow \mathrm{Topics}(\tilde{x}_v)$ `// BERTopic`

Etapa 2 — Similaridade Par-a-Par
para cada *par não ordenado* (i, j), $i \neq j$ **faça**
  `// Semântica global (paráfrases/sinônimos)`
  $s_e(i,j) \leftarrow \mathrm{cosine}(e_i, e_j)$
  `// Entidades explícitas (local/pessoa) com Jaccard aproximado`
  $L_i, P_i \leftarrow$ filtrar ENT_i por {local,pessoa}
  $L_j, P_j \leftarrow$ filtrar ENT_j por {local,pessoa}
  $s_l(i,j) \leftarrow \mathrm{FuzzyJaccard}(L_i \cup P_i,\ L_j \cup P_j)$
  `// Magnitude factual (contagem de vítimas)`
  $$s_v(i,j) \leftarrow \begin{cases} 1, & D_i = D_j \\ 0, & \{D_i, D_j\} = \{0, >0\} \\ 1 - \frac{|D_i - D_j|}{\max(D_i, D_j)}, & \text{c.c.} \end{cases}$$
  `// Identificadores oficiais (nome de operação)`
  $s_o(i,j) \leftarrow \mathbf{1}(\exists\ n \in N_i,\ m \in N_j :\ \mathrm{FuzzyMatch}(n,m) \geq 0.8)$
  `// Contexto temático (coocorrência de termos)`
  $s_t(i,j) \leftarrow \mathrm{Jaccard}(T_i, T_j)$
  `// Combinação ponderada das pistas`
  $S(i,j) \leftarrow w_e s_e + w_l s_l + w_v s_v + w_o s_o + w_t s_t$

Etapa 3 — Construção do Grafo e Agrupamento
Crie grafo $G = (\mathcal{V}, \mathcal{E})$ com nós $\mathcal{V}$ e arestas $\mathcal{E} = \{(i,j) \mid S(i,j) \geq \tau\}$
$\mathcal{C} \leftarrow$ componentes conectados de G (via BFS/DFS)
retorna $\mathcal{C}$

Etapa 2 — Similaridade Par-a-Par

Para cada par de vídeos, calculamos cinco métricas complementares:

1. **Similaridade de *embedding*** (s_e): mede proximidade no espaço semântico, sendo eficaz para capturar paráfrases e sinônimos mesmo com redação divergente.

2. **Similaridade de entidades de localização e pessoa** (s_l): obtida via variação do índice de Jaccard (JACCARD, 1912) com correspondência aproximada (*Difflib SequenceMatcher*[7]), útil para unir vídeos que compartilham cenário geográfico ou protagonistas.

3. **Similaridade no número de vítimas** (s_v): captura magnitude factual do evento, com similaridade máxima quando as contagens coincidem ou ambos omitem a

[7] <https://docs.python.org/3/library/difflib.html>

informação.

4. **Similaridade de nomes de operação (s_o)**: funciona como identificador quase-único; atribuímos similaridade 1,0 se ao menos um par de nomes extraídos tem correspondência $\geq 0{,}8$.

5. **Similaridade de tópicos latentes (s_t)**: medida pela sobreposição de tópicos extraídos com BERTopic, indicada pela similaridade de Jaccard, revelando afinidade temática mesmo sem entidades ou termos coincidentes.

Essas métricas são combinadas por média ponderada, última equação do algoritmo, onde pesos $(w_e, w_l, w_v, w_o, w_t)$ calibram a contribuição de cada atributo.

Etapa 3 — Agrupamento por Grafo

Construímos um grafo onde cada vídeo é um nó e há uma aresta entre dois nós se a similaridade final S excede o limiar τ. Os grupos são formados pelos componentes conectados desse grafo, obtidos via busca em largura (BFS). Esse modelo garante que conexões indiretas (*e.g.*, A similar a B e B similar a C) resultem no agrupamento conjunto de A, B e C. O valor de τ controla a coesão de forma que valores altos produzem grupos menores e mais homogêneos, enquanto valores baixos favorecem maior abrangência.

6.2.2 Agrupamento Temporal

A heurística temporal explora a proximidade no tempo como principal evidência de que dois vídeos tratam do mesmo evento, apoiando-se em entidades nomeadas para identificar e rastrear incidentes ao longo da linha do tempo. A motivação é que, em plataformas como o YouTube, repercussões de eventos de violência urbana e operações policiais concentram-se em um intervalo relativamente curto após a ocorrência. Assim, mesmo que o vocabulário varie entre vídeos, a coesão temporal e a recorrência de certas entidades, como nomes de localidades ou operações, podem servir como âncoras robustas para agrupamento.

O processo é dividido em três etapas principais, resumidas no Algoritmo 2:

Etapa 1 — Extração de entidades e definição de janelas

Para cada vídeo, extraímos entidades com NER (Seção 6.2), aplicando normalização com conversão para minúsculas, remoção de acentos e exclusão de caracteres especiais. Definimos então uma *janela temporal* centrada na data de publicação do vídeo, com tamanho $2w + 1$ dias, sendo w um parâmetro fixo (e.g., $w = 15$ dias). Essa janela reflete a duração típica da repercussão de eventos desse tipo na plataforma.

Algoritmo 2: Heurística de Agrupamento Temporal.

Entrada: Conjunto de vídeos $\mathcal{V}$; para cada $v \in \mathcal{V}$: data de publicação, título, descrição e (opcional) transcrição.
Parâmetro: meia-janela w (em dias).
Saída: Partição $\mathcal{C}$ de $\mathcal{V}$ em grupos (eventos).

Etapa 1 — Extração e preparação de entidades
para cada $v \in \mathcal{V}$ **faça**
  $x_v \leftarrow$ concatenar {título, descrição, (transcrição)}
  Aplicar normalização: minúsculas, remoção de acentos e caracteres especiais
  $\text{ENT}_v \leftarrow \text{NER}(x_v)$ `// locais, operações, datas`

Etapa 2 — Iteração por ordem cronológica
Ordenar $\mathcal{V}$ por data de publicação
para cada $v \in \mathcal{V}$ **faça**
  $\mathcal{W}_v \leftarrow$ vídeos com data $\in [\text{data}(v) - w,\ \text{data}(v) + w]$
  cand $\leftarrow$ entidade mais frequente em $\mathcal{W}_v$ presente em ENT_v
  Etapa 3 — Associação a grupos existentes ou criação de novos grupos
  se cand *não é identificador de grupo existente* **então**
    Criar novo grupo G com nome cand e abrangência $\mathcal{W}_v$
    Adicionar v a G
  senão se v *dentro da abrangência de algum grupo* G *com nome* cand **então**
    Adicionar v a G
    Expandir abrangência de G para incluir $\mathcal{W}_v$
  senão
    Criar novo grupo G' com nome cand (com sufixo, se necessário) e abrangência $\mathcal{W}_v$
    Adicionar v a G'

retorna $\mathcal{C}$

Etapa 2 — Seleção da entidade candidata

Dentro da janela do vídeo corrente, calculamos a frequência de cada entidade única extraída dos vídeos presentes nesse intervalo. A *entidade candidata* é definida como a mais frequente na janela *e* presente no vídeo analisado. Essa escolha se justifica por duas razões: (i) entidades frequentes no intervalo tendem a representar o evento central; e (ii) exigir que a entidade esteja no vídeo evita associações por coincidência temporal.

Etapa 3 — Associação a grupos existentes ou criação de novos grupos

Se não existir um grupo com o nome da entidade candidata, criamos um novo grupo, atribuindo a ele o vídeo e registrando sua abrangência temporal inicial com base na janela do vídeo. Se já existir grupo(s) com esse nome, verificamos se o vídeo está dentro da abrangência temporal de algum deles:

- Caso positivo, o vídeo é adicionado ao grupo correspondente, e a abrangência temporal desse grupo é atualizada para englobar todo o intervalo combinado — desde a data mais antiga até a mais recente entre seus vídeos. Esse ajuste dinâmico garante que a janela de um grupo possa se expandir progressivamente.
- Caso negativo, interpretamos como um evento distinto ou fase posterior, criando um novo grupo com o mesmo identificador (e.g., “-2”, “-3” para fases subsequentes).

Tabela 9 – Resultados consolidados em todos os subcenários (com e sem transcrição).

Cenário	**Técnica**	**Acu**	**Pre**	**Rev**	**F1**	**NMI**	**AMI**
Todos eventos (com Transcr.)	HS	0.57	0.69	0.74	0.68	0.86	0.22
	HT	**0.87**	**0.85**	**0.84**	**0.84**	**0.97**	**0.86**
	GPT	0.49	0.60	0.63	0.59	0.82	0.31
Todos eventos (sem Transcr.)	HS	0.57	0.69	0.74	0.68	0.86	0.20
	HT	**0.87**	**0.85**	**0.84**	**0.84**	**0.97**	**0.86**
	GPT	0.49	0.60	0.63	0.59	0.82	0.31
Eventos esparsos (com Transcr.)	HS	0.74	0.81	0.80	0.80	0.93	0.05
	HT	**0.90**	**0.91**	**0.90**	**0.90**	**0.98**	**0.48**
	GPT	0.59	0.58	0.62	0.58	0.90	0.10
Eventos esparsos (sem Transcr.)	HS	0.74	0.81	0.80	0.80	0.93	0.05
	HT	**0.90**	**0.91**	**0.89**	**0.89**	**0.98**	**0.47**
	GPT	0.59	0.58	0.62	0.58	0.90	0.10
Alta repercussão (com Transcr.)	HS	0.25	0.02	0.01	0.01	0.45	0.18
	HT	**0.87**	**0.17**	**0.14**	**0.15**	**0.85**	**0.84**
	GPT	0.24	0.03	0.00	0.01	0.35	0.13
Alta repercussão (sem Transcr.)	HS	0.25	0.02	0.01	0.01	0.45	0.18
	HT	**0.87**	**0.17**	**0.14**	**0.15**	**0.85**	**0.84**
	GPT	0.24	0.03	0.00	0.01	0.35	0.13

6.3 Avaliação e Resultados

Nesta seção analisamos o desempenho das abordagens avaliadas (Heurística Temporal (HT), Heurística Semântica (HS) e GPT-4) em seis variações de cenários, considerando a presença ou ausência de transcrições, consolidados na Tabela 9. Foram avaliadas as métricas acurácia (Acu), precisão (Pre), revocação (Rev), F1-score (F1), *Normalized Mutual Information* NMI e *Adjusted Mutual Information* (AMI). O melhor resultado em cada métrica dentro de cada subcenário está destacado em **negrito**.

Observa-se que, em praticamente todos os cenários, a **HT** apresentou desempenho superior, alcançando as maiores acurácias, F1-scores e medidas de informação mútua. No cenário **todos os eventos**, tanto com quanto sem transcrição, a HT atingiu **0.87** de acurácia e **0.84** de F1, superando a HS (0.57 Acu, 0.68 F1) e o GPT (0.49 Acu, 0.59 F1). Isso confirma a importância do critério temporal na identificação de eventos, dado que a alta similaridade lexical entre diferentes eventos relacionados à violência urbana e operações policiais compromete a eficácia de métodos puramente semânticos.

No cenário de **eventos esparsos**, a inclusão das transcrições também não alterou o padrão geral de desempenho. Contudo, HS mostrou um ganho expressivo nesse cenário (0.74 Acu, 0.80 F1) com e sem transcrição. Esse comportamento sugere que, em bases esparsas, atributos textuais extraídos de títulos e descrições são suficientes para melhorar a discriminação de eventos, enquanto o excesso de detalhes da transcrição não traz informações novas e discriminativas para agrupar eventos. Ainda assim, a HT manteve

o melhor desempenho (até **0.90** Acu e **0.89** F1 sem transcrição), embora o AMI tenha caído para 0.47, refletindo a predominância de grupos unitários. Essa queda no AMI está relacionada à estrutura muito fina da partição (834 grupos reais × 884 previstos) na base de dados de eventos esparsos, onde predominam eventos únicos e há fragmentação de alguns grupos maiores. O ajuste por acaso do AMI penaliza mais nesses casos, pois até partições aleatórias produzem MI elevado, resultando em um valor ajustado baixo.

Já no cenário de **alta repercussão**, independentemente da presença de transcrições, a HT manteve **0.87** de acurácia e **0.84** de AMI, mas com quedas acentuadas em precisão e F1 (0.17 e 0.15, respectivamente). Isso indica que, apesar de preservar a coerência temporal dos agrupamentos, a sobreposição semântica e temporal entre eventos de grande cobertura midiática gera confusões residuais. A HS e o GPT apresentaram desempenho muito baixo (acurácia em torno de 0.25 e F1 praticamente nulo), evidenciando a limitação de métodos sem componente temporal explícito nesse contexto. Com a base de dados Alta repercussão, essa queda de precisão, revocação e F1 macro ocorre devido à superfragmentação de grupos grandes em muitos rótulos previstos. Por exemplo, um grupo com 221 vídeos foi dividido em 20 subgrupos, mas apenas um é mapeado corretamente via Hungarian. Os demais contam como erros, derrubando as médias macro. O AMI permanece alto (0.84) porque os grupos resultantes são internamente puros (*homogeneity* = 1.0) e o número de grupos reais é pequeno.

De forma geral, os resultados permitem concluir que: (i) a proximidade temporal é o discriminador mais eficaz para este contexto; (ii) atributos semânticos ganham relevância em bases esparsas, mas não sustentam desempenho em cenários densos; (iii) mesmo modelos de linguagem de última geração, como o GPT-4, não superam heurísticas simples, porém ajustadas ao contexto, quando o problema exige sensibilidade temporal e resiliência à alta sobreposição lexical; e (iv) métricas como o AMI e o F1 macro podem apresentar quedas expressivas não por baixa pureza dos agrupamentos, mas por efeitos estruturais, como granularidade excessiva e fragmentação de eventos em muitos grupos pequenos, o que eleva o linha de base aleatório de MI ou penaliza fortemente a média macro.

6.4 Resumo

O Capítulo tratou da identificação automática de vídeos do YouTube que descrevem o mesmo evento de violência urbana, um problema difícil devido à alta similaridade lexical entre ocorrências distintas e à forte concentração temporal das coberturas. Para isso, foram propostas duas heurísticas não supervisionadas e complementares: uma heurística temporal, baseada na proximidade de publicação e em entidades recorrentes como âncoras, e uma heurística semântica, que integra *embeddings*, entidades nomeadas, número de vítimas, nomes de operações e tópicos em um grafo de similaridade, além de uma linha de

base com GPT-4. O estudo se destaca por combinar pistas temporais e semânticas em um conjunto de dados inédito com mais de 1.400 vídeos anotados ao longo de cinco anos. Os resultados mostraram que a heurística temporal superou de forma consistente as demais abordagens, alcançando acurácia próxima de 0,90 e NMI próxima de 0,98, enquanto a heurística semântica teve melhor desempenho em cenários mais esparsos e o GPT-4 não superou as heurísticas especializadas. Observou-se ainda que a inclusão de transcrições pouco contribuiu para o desempenho, indicando que títulos e descrições concentram as informações mais relevantes para o agrupamento dos vídeos.

7 Conclusão e Trabalhos Futuros

Neste Capítulo apresentamos as considerações finais sobre os resultados alcançados nessa dissertação, incluindo as suas limitações. Por fim, discutimos os possíveis trabalhos futuros. Primeiramente, é importante colocar que essa dissertação tratou de mineração de padrões textuais sobre segurança pública, propondo modelos e aplicações baseados em PLN. Especificamente, foi mostrado como dados textuais provenientes de mídias sociais podem ser utilizados para compreender e monitorar a percepção das pessoas, particularmente usuários de plataformas de mídia sociais, no contexto de operações policiais e eventos de violência urbana no Brasil. Dentro desse contexto, consideramos três eixos de pesquisa que são complementares: (i) a detecção de posicionamento em comentários, (ii) a avaliação da transferência de conhecimento de modelos para detectar posicionamentos entre plataformas distintas e (iii) a detecção de eventos reais em vídeos do YouTube. Esses eixos, quando integrados, compõem um arcabouço de ferramentas de software para monitoramento de eventos de violência urbana a partir de conteúdo postado na Web, especificamente em plataformas de mídias sociais como YouTube, X/Twitter, dentre outras. Os resultados mostrados nos três capítulos anteriores dessa dissertação evidenciam as contribuições alcançadas nesses três eixos, que também estão alinhados aos objetivos específicos estabelecidos na introdução da dissertação.

O primeiro objetivo específico consiste em desenvolver e avaliar modelos de detecção de posicionamentos (*stance detection*) para comentários sobre operações policiais no Brasil postados em plataformas de mídias sociais. Nesse sentido, desenvolvemos uma metodologia abrangente para analisar a percepção pública sobre operações policiais amplamente repercutidas em mídias sociais, utilizando comentários do X/Twitter e do YouTube. Essa contribuição incluiu a criação de um sistema de monitoramento contínuo de operações no YouTube, o aprimoramento de modelos de PLN baseados em arquiteturas *transformers* por meio de uma estratégia híbrida de anotação com validação humana e apoio de LLMs, e a aplicação inédita desses modelos na análise de tendências de opinião ao longo de dois anos e em diferentes regiões brasileiras. Os modelos alcançaram acurácia de 88% e F1-macro de 87% no X/Twitter, superando a literatura em mais de 18% e 7%, respectivamente, além de desempenho satisfatório no YouTube (83% de acurácia e 73% de F1-macro), demonstrando capacidade de captar dinâmicas de opinião em tempo quase real.

Por sua vez, o segundo objetivo específico consistiu em avaliar experimentalmente o desempenho dos modelos desenvolvidos para inferir posicionamentos de comentários no cenário intraplataforma e interplataformas de mídias sociais. Para tratá-lo, conduzimos uma investigação sistemática sobre a capacidade de generalização entre plataformas, avaliando cenários intraplataforma e interplataformas e aplicando a mesma estratégia

híbrida de anotação para garantir consistência e confiabilidade dos rótulos. A comparação entre modelos tradicionais e modelos baseados em *transformers* revelou que classificadores treinados no X/Twitter podem generalizar (i.e., transferir conhecimento) melhor para o YouTube do que o inverso, evidenciando o impacto das características discursivas próprias de cada plataforma no processo de adaptação entre domínios.

Por fim, o terceiro objetivo específico concerne a identificar eventos de violência urbana a partir de conteúdos postados em plataformas de mídia social de forma automatizada, em larga escala e com baixo custo computacional. Para esse objetivo, propusemos duas heurísticas não supervisionadas para detecção e agrupamento de vídeos que se referem a um mesmo evento real de violência urbana no YouTube: a Heurística Temporal (HT) e a Heurística Semântica (HS). A HT explora proximidade temporal e recorrência de entidades âncoras, enquanto a HS integra múltiplos atributos textuais, incluindo *embeddings* densos, entidades nomeadas e menções a operações e vítimas — por meio de um grafo de similaridade. Os experimentos demonstraram que ambas as heurísticas superam uma linha de base construída com GPT-4, mostrando que abordagens especializadas, de menor custo computacional e adaptadas ao domínio, podem superar soluções genéricas de última geração.

Embora os resultados alinhados com cada objetivo específico sejam encorajadores, limitações devem ser reconhecidas. A estratégia híbrida de anotação, embora reduza ruídos, diminui o conjunto final de dados e pode introduzir vieses ao descartar casos ambíguos. O desempenho dos modelos de detecção de posicionamento depende fortemente do domínio das operações policiais e do contexto temporal, podendo demandar reanotação ou ajuste contínuo. Além disso, a análise de generalização entre plataformas foi restrita ao X/Twitter e ao YouTube, não contemplando a variedade discursiva de outras redes sociais emergentes. No caso da detecção de eventos, a dependência de atributos textuais limita a precisão em cenários com alta similaridade lexical, e a Heurística Temporal assume proximidade cronológica que pode não se sustentar em casos de repercussão prolongada.

Ainda considerando as limitações mencionadas, os resultados desta dissertação contribuem em termos de metodologia e resultados empíricos para a área de mineração de dados dentro do contexto proposto. Nesse sentido, disponibilizamos novas bases anotadas em português, desenvolvemos modelos e heurísticas originais e avaliamos sistematicamente a transferência entre plataformas. Todos esses itens compõem o nosso arcabouço integrado para monitoramento de percepções sobre violência urbana. Os resultados demonstram que abordagens leves, adaptadas ao domínio e apoiadas por estratégias de anotação híbrida, podem superar soluções genéricas de última geração em condições reais, reforçando o potencial das mídias sociais como fonte estruturada de conhecimento sobre dinâmicas sociais e segurança pública.

Como desdobramento direto desta dissertação, estamos desenvolvendo uma apli-

cação de monitoramento automatizado de violência urbana denominada *Crime Stance*[1]. A plataforma busca monitoramento em larga escala e com baixo custo computacional, baseado nos modelos e análises apresentados neste trabalho, permitindo acompanhar de forma contínua como eventos de segurança pública repercutem nas mídias sociais. Ao integrar inferência de posicionamento, detecção de eventos e agregação temporal de indicadores, o sistema visa oferecer uma ferramenta prática para pesquisadores, gestores públicos, jornalistas e profissionais de diferentes áreas interessados em compreender as repercussões na Web sobre eventos associados à violência urbana.

Diversas direções futuras emergem a partir deste trabalho. Do ponto de vista metodológico, novas investigações podem explorar modelos mais recentes e técnicas de ajuste fino eficientes que preservem desempenho à medida que novos eventos e padrões linguísticos surgem. O aprimoramento da estratégia híbrida de anotação com LLMs mais modernos também representa uma oportunidade para reduzir ainda mais o custo de rotulação juntamente à investigação de discordância entre LLMs e humanos. Análises temporais ao longo de eventos específicos também podem avançar a compreensão das dinâmicas da opinião pública, revelando padrões de mobilização, polarização e repercussão social. Análise da caracterização dos grupos encontrados pelas heurísticas e entender que tipos de grupos as heurísticas são melhores ou piores.

Além dessas direções, pretendemos expandir o escopo do monitoramento desenvolvido nesta dissertação para outros contextos relacionados a eventos reais e que podem ser identificados por meio de dados disponíveis na Web. No âmbito do projeto *C-MonAnalys*[2], buscamos estender nossa abordagem para abarcar não apenas eventos de violência urbana, mas também fenômenos associados ao entretenimento, bem como condições estruturais relacionadas à saúde, educação, renda, saneamento e meio ambiente. Essa ampliação visa investigar como diferentes dimensões se manifestam no debate público digital e como padrões discursivos podem refletir desigualdades, necessidades sociais e variações contextuais ao longo do tempo.

Em síntese, esta dissertação demonstrou que é possível extrair conhecimento estruturado e relevante de grandes volumes de dados gerados espontaneamente nas mídias sociais, contribuindo tanto para a pesquisa acadêmica quanto para o fortalecimento da discussão pública sobre segurança. Espera-se que as contribuições apresentadas inspirem novas investigações que aproximem mineração de dados, políticas públicas e sociedade.

[1] <https://labpaad.github.io/crimes_stance_site/home>
[2] <https://labpaad.github.io/C-MonAnalys/>

Referências

AMER, Z.; DAVIES, M. D. Context-enriched named entity recognition (ner) for identifying emerging trends in video comments. *University of California, Berkeley*, 2025. Citado 2 vezes nas páginas 11 e 20.

BALUSU, M. R. B.; MERGHANI, T.; EISENSTEIN, J. *Stylistic Variation in Social Media Part-of-Speech Tagging*. 2018. Disponível em: <https://arxiv.org/abs/1804.07331>. Citado 2 vezes nas páginas 19 e 22.

BARACHI, M. E. et al. A novel sentiment analysis framework for monitoring the evolving public opinion in real-time: Case study on climate change. *Journal of Cleaner Production*, Elsevier, v. 312, p. 127820, 2021. Citado 2 vezes nas páginas 9 e 22.

BARRETO LEONARDO, G. J. *Caso Genivaldo: um ano após homem ser morto asfixiado pela PRF, viúva diz que filho ainda não sabe que pai foi torturado*. 2023. Acesso em: 4 nov. 2025. Disponível em: <https://g1.globo.com/...> Citado na página 1.

BARROS, T. S.; PIRES, C. E. S.; NASCIMENTO, D. C. Leveraging bert for extractive text summarization on federal police documents. *Knowledge and Information Systems*, v. 65, n. 11, p. 4873–4903, 2023. Citado 2 vezes nas páginas 14 e 16.

BECHINI, A. et al. Stance analysis of twitter users: the case of the vaccination topic in italy. *IEEE Intelligent Systems*, IEEE, v. 36, n. 5, p. 131–139, 2020. Citado na página 9.

BERTAGLIA, T. F. C.; NUNES, M. d. G. V. Exploring word embeddings for unsupervised textual user-generated content normalization. *arXiv preprint arXiv:1704.02963*, 2017. Citado 3 vezes nas páginas 11, 19 e 20.

BOESINGER, L. et al. Tube2vec: Social and semantic embeddings of youtube channels. In: *Proceedings of the International AAAI Conference on Web and Social Media*. [S.l.: s.n.], 2024. v. 18, p. 2084–2090. Citado na página 11.

BOYD, M. S. (new) participatory framework on youtube? commenter interaction in us political speeches. *Journal of Pragmatics*, Elsevier, v. 72, p. 46–58, 2014. Citado na página 9.

CERON, A.; NEGRI, F. The "social side" of public policy: Monitoring online public opinion and its mobilization during the policy cycle. *Policy & Internet*, Wiley Online Library, v. 8, n. 2, p. 131–147, 2016. Citado 2 vezes nas páginas 10 e 22.

CHAKRABORTY, P.; SHARMA, A. Public opinion analysis of the transportation policy using social media data: a case study on the delhi odd–even policy. *Transportation in Developing Economies*, Springer, v. 5, p. 1–9, 2019. Citado 2 vezes nas páginas 10 e 22.

CHAPARRO, L. F. et al. Sentiment analysis of social network content to characterize the perception of security. In: IEEE. *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. [S.l.], 2020. p. 685–691. Citado 3 vezes nas páginas 2, 13 e 16.

CHEN, C. et al. Cross-platform violence detection on social media: A dataset and analysis. In: *Proceedings of the 17th ACM Web Science Conference 2025*. New York, NY, USA: Association for Computing Machinery, 2025. (Websci '25), p. 494–498. ISBN 9798400714832. Disponível em: <https://doi.org/10.1145/3717867.3717877>. Citado 3 vezes nas páginas 1, 10 e 20.

COSTA, J. et al. 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*. [S.l.: s.n.], 2025. p. 42–51. Citado na página 1.

D'ANDREA, E. et al. Monitoring the public opinion about the vaccination topic from tweets analysis. *Expert Systems with Applications*, Elsevier, v. 116, p. 209–226, 2019. Citado 3 vezes nas páginas 3, 8 e 22.

DEVLIN, J. et al. Bert: Pre-training of deep bidirectional transformers for language understanding. In: *Association for Computational Linguistics: Human Language Technologies*. [S.l.: s.n.], 2019. p. 4171–4186. Citado 5 vezes nas páginas 3, 21, 27, 28 e 37.

ELHOSEINY, M. et al. Zero-shot event detection by multimodal distributional semantic embedding of videos. In: *Proceedings of the AAAI conference on artificial intelligence*. [S.l.: s.n.], 2016. v. 30, n. 1. Citado 3 vezes nas páginas 3, 11 e 21.

FEITOSA, M. P. F. et al. Análise da percepção das pessoas no twitter sobre ações policiais. In: SBC. *Anais do XI Brazilian Workshop on Social Network Analysis and Mining*. [S.l.], 2022. p. 73–84. Citado 9 vezes nas páginas 2, 3, 15, 16, 20, 21, 27, 29 e 35.

GILARDI, F.; ALIZADEH, M.; KUBLI, M. Chatgpt outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, National Academy of Sciences, v. 120, n. 30, p. e2305016120, 2023. Citado 3 vezes nas páginas 21, 23 e 36.

GLOBAL Peace Index 2025: Measuring Peace in a Complex World. 2025. Acesso em: julho de 2025. Disponível em: <https://www.visionofhumanity.org>. Citado na página 1.

GUSMÃO, C.; FIGUEIREDO, K.; BRITO, W. A. Técnicas de processamento de linguagem natural em denúncias criminais: Automatizaçao e classificaçao de texto em português coloquial. In: SBC. *Seminário Integrado de Software e Hardware (SEMISH)*. [S.l.], 2021. p. 172–182. Citado na página 14.

HAND, L. C.; CHING, B. D. Maintaining neutrality: A sentiment analysis of police agency facebook pages before and after a fatal officer-involved shooting of a citizen. *Government Information Quarterly*, Elsevier, v. 37, n. 1, p. 101420, 2020. Citado na página 2.

HOSSAIN, T. et al. Covidlies: Detecting covid-19 misinformation on social media. In: *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*. [S.l.: s.n.], 2020. Citado na página 3.

HUANG, Y.; HUANG, J. X. Exploring chatgpt for next-generation information retrieval: Opportunities and challenges. *Web Intelligence*, v. 22, n. 1, p. 31–44, 2024. Disponível em: <https://journals.sagepub.com/doi/abs/10.3233/WEB-230363>. Citado 2 vezes nas páginas 36 e 48.

JACCARD, P. The distribution of the flora in the alpine zone. *New Phytologist*, v. 11, n. 2, p. 37–50, 1912. Disponível em: <https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8137.1912.tb05611.x>. Citado 2 vezes nas páginas 21 e 52.

JANSEN, A. et al. Large-scale audio event discovery in one million youtube videos. In: IEEE. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.], 2017. p. 786–790. Citado 3 vezes nas páginas 3, 11 e 21.

KAPLAN, A. et al. Responsible and sustainable ai: Considering energy consumption in automated text classification evaluation tasks. In: IEEE COMPUTER SOCIETY. *2025 IEEE/ACM 9th International Workshop on Green and Sustainable Software (GREENS)*. [S.l.], 2025. p. 76–83. Citado na página 2.

KAPPAUN, A.; OLIVEIRA, J. Análise sobre viés de gênero no youtube: Um estudo sobre as eleições presidenciais de 2018 e 2022. In: SBC. *Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)*. [S.l.], 2023. p. 127–138. Citado na página 9.

KARAMOUZAS, D.; MADEMLIS, I.; PITAS, I. Public opinion monitoring through collective semantic analysis of tweets. *Social Network Analysis and Mining*, Springer, v. 12, n. 1, p. 91, 2022. Citado na página 9.

KERAGHEL, I.; MORBIEU, S.; NADIF, M. A survey on recent advances in named entity recognition. *arXiv preprint arXiv:2401.10825*, 2024. Citado 2 vezes nas páginas 11 e 20.

KOCOŃ, J. et al. Chatgpt: Jack of all trades, master of none. *Information Fusion*, Elsevier, v. 99, p. 101861, 2023. Citado 3 vezes nas páginas 21, 27 e 37.

KU, C. H.; IRIBERRI, A.; LEROY, G. Natural language processing and e-government: crime information extraction from heterogeneous data sources. In: *Proceedings of the 2008 international conference on Digital government research*. [S.l.: s.n.], 2008. p. 162–170. Citado na página 14.

KUHN, H. W. The hungarian method for the assignment problem. *Naval research logistics quarterly*, Wiley Online Library, v. 2, n. 1-2, p. 83–97, 1955. Citado na página 48.

LI, J. et al. A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*, Institute of Electrical and Electronics Engineers (IEEE), v. 34, n. 1, p. 50–70, jan. 2022. ISSN 2326-3865. Disponível em: <http://dx.doi.org/10.1109/TKDE.2020.2981314>. Citado 2 vezes nas páginas 20 e 22.

LU, J. et al. Padellm-ner: Parallel decoding in large language models for named entity recognition. arxiv e-prints, pages arxiv–2402. *Advances in Neural Information Processing Systems*, 2024. Citado na página 11.

MAHESH, B. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, v. 9, p. 381–386, 2020. Citado na página 28.

MAIA, D. F.; SILVA, N. F. F. da. Enhancing stance detection in low-resource brazilian portuguese using corpus expansion generated by GPT-3.5. In: GAMALLO, P. et al. (Ed.). *Proceedings of the 16th International Conference on Computational Processing of Portuguese, PROPOR 2024, Santiago de Compostela, Galicia/Spain, March 12-15, 2024, Volume 1*. Association for Computational Linguistics, 2024. p. 503–508. Disponível em: <https://aclanthology.org/2024.propor-1.51>. Citado na página 2.

MANNING, C. D.; RAGHAVAN, P.; SCHüTZE, H. *Introduction to Information Retrieval*. USA: Cambridge University Press, 2008. ISBN 0521865719. Citado na página 48.

MATOS, K.; RIBEIRO, R.; FERREIRA, J. C. Mining population opinion about local police. *Multimedia Tools and Applications*, Springer, p. 1–27, 2025. Citado 3 vezes nas páginas 2, 12 e 16.

MIAO, L.; LAST, M.; LITVAK, M. Tracking social media during the covid-19 pandemic: The case study of lockdown in new york state. *Expert Systems with Applications*, Elsevier, v. 187, p. 115797, 2022. Citado na página 8.

MINING, W. I. D. Data mining: Concepts and techniques. *Morgan Kaufinann*, v. 10, n. 559-569, p. 4, 2006. Citado na página 38.

MOHAMMAD, S. et al. SemEval-2016 task 6: Detecting stance in tweets. In: *Proceedings of the 10th International Workshop on Semantic Evaluation*. [S.l.: s.n.], 2016. Citado na página 1.

MOHAMMAD, S. M.; SOBHANI, P.; KIRITCHENKO, S. Stance and sentiment in tweets. *ACM Transactions on Internet Technology (TOIT)*, ACM New York, NY, USA, v. 17, n. 3, p. 1–23, 2017. Citado 2 vezes nas páginas 21 e 29.

MONDAL, M. et al. A survey of multimodal event detection based on data fusion. *The VLDB Journal*, Springer, v. 34, n. 1, p. 9, 2025. Citado 2 vezes nas páginas 3 e 11.

MOZAFARI, M.; FARAHBAKHSH, R.; CRESPI, N. A bert-based transfer learning approach for hate speech detection in online social media. In: SPRINGER. *International Conference on Complex Networks and Their Applications*. [S.l.], 2019. p. 928–940. Citado na página 21.

MOZAFARI, M.; FARAHBAKHSH, R.; CRESPI, N. A bert-based transfer learning approach for hate speech detection in online social media. In: SPRINGER. *Proceedings of the Eighth International Conference on Complex Networks and Their Applications*. [S.l.], 2020. Citado 2 vezes nas páginas 2 e 10.

NANAYAKKARA, A. C. Exploratory analysis of the black lives matter movement's visibility on youtube. In: *2024 4th International Conference on Advanced Research in Computing (ICARC)*. [S.l.: s.n.], 2024. p. 161–166. Citado 3 vezes nas páginas 2, 13 e 16.

NEWMAN, N. et al. Reuters institute digital news report 2022. *Reuters Institute for the study of Journalism*, 2022. Citado na página 1.

NGUYEN, D. Q.; VU, T.; NGUYEN, A. T. Bertweet: A pre-trained language model for english tweets. *arXiv preprint arXiv:2005.10200*, 2020. Citado na página 21.

NUBILA, K. D. et al. Technopopulism and politainment in brazil: Bolsonaro government's weekly youtube broadcasts. *Media and Communication*, PRT, v. 11, n. 2, p. 137–147, 2023. Citado na página 9.

OH, G.; ZHANG, Y.; GREENLEAF, R. G. Measuring geographic sentiment toward police using social media data. *American Journal of Criminal Justice*, Springer, p. 1–17, 2021. Citado 3 vezes nas páginas 2, 12 e 16.

OTTONI, R. et al. Analyzing right-wing youtube channels: Hate, violence and discrimination. In: *Proceedings of the 10th ACM conference on web science*. [S.l.: s.n.], 2018. p. 323–332. Citado na página 1.

PEREIRA Érica et al. Analyzing youtube videos shared on whatsapp and telegram political public groups. In: *Proceedings of the 28th Brazilian Symposium on Multimedia and the Web*. Porto Alegre, RS, Brasil: SBC, 2022. p. 29–38. ISSN 0000-0000. Disponível em: <https://sol.sbc.org.br/index.php/webmedia/article/view/22088>. Citado na página 1.

REIMERS, N. et al. Sentence-bert: Sentence embeddings using siamese bert-networks. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. [S.l.], 2019. Citado 3 vezes nas páginas 20, 22 e 51.

RICARDO, C. d. M.; SIQUEIRA, P. P. de; MARQUES, C. R. Estudo conceitual sobre os espaços urbanos seguros. *Revista brasileira de segurança pública*, v. 7, n. 1, p. 200–216, 2013. Citado na página 1.

ROCHA, S. et al. Uso de características temporais e semânticas para detectar eventos em vídeos de violência urbana. In: *Proceedings of the 31st Brazilian Symposium on Multimedia and the Web*. Porto Alegre, RS, Brasil: SBC, 2025. p. 464–472. ISSN 0000-0000. Disponível em: <https://sol.sbc.org.br/index.php/webmedia/article/view/37992>. Citado na página 5.

ROCHA, S. S. da et al. Monitorando a opinião pública sobre operações policiais no brasil via comentários de vídeos no youtube. In: SBC. *Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)*. [S.l.], 2024. p. 158–171. Citado 3 vezes nas páginas 5, 35 e 46.

ROMANO, S. et al. Adjusting for chance clustering comparison measures. *Journal of Machine Learning Research*, v. 17, n. 134, p. 1–32, 2016. Disponível em: <http://jmlr.org/papers/v17/15-627.html>. Citado na página 48.

SBTNEWS. *Homem com esquizofrenia é morto por policiais durante surto no RS*. 2025. YouTube. Acesso em: 17 nov. 2025. Disponível em: <https://www.youtube.com/watch?v=mR3U_BnGzHw>. Citado na página 1.

SCHILLER, B.; DAXENBERGER, J.; GUREVYCH, I. Stance detection benchmark: How robust is your stance detection? *KI-Künstliche Intelligenz*, Springer, v. 35, n. 3, p. 329–341, 2021. Citado 2 vezes nas páginas 2 e 10.

SCOTLAND, J.; THOMAS, A.; JING, M. Public emotion and response immediately following the death of george floyd: A sentiment analysis of social media comments. *Telematics and Informatics Reports*, Elsevier, v. 14, p. 100143, 2024. Citado 3 vezes nas páginas 2, 13 e 16.

SIDHARTA, S. et al. Sentiment analysis of hashtag activism on social media twitter. In: *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*. [S.l.: s.n.], 2023. p. 73–77. Citado 2 vezes nas páginas 13 e 16.

SOUZA, F.; NOGUEIRA, R.; LOTUFO, R. BERTimbau: pretrained BERT models for Brazilian Portuguese. In: *Brazilian Conference on Intelligent Systems*. [S.l.: s.n.], 2020. Citado 5 vezes nas páginas 3, 21, 28, 37 e 38.

SOUZA, M. d. F. M. de et al. Análise de padrões e associações em crimes relacionados à violência contra mulheres utilizando word2vec para processamento em linguagem natural. In: SBC. *Escola Regional de Informática de Mato Grosso (ERI-MT)*. [S.l.], 2023. p. 196–200. Citado na página 14.

STEINER, T. et al. Crowdsourcing event detection in youtube videos 58-67. In: CEUR WORKSHOP PROCEEDINGS. *DeRiVE 2011 Detection, Representation, and Exploitation of Events in the Semantic Web*. [S.l.], 2011. p. 58–67. Citado na página 12.

THIES, J. et al. Graphtmt: unsupervised graph-based topic modeling from video transcripts. In: IEEE. *2021 IEEE Seventh International Conference on Multimedia Big Data (BigMM)*. [S.l.], 2021. p. 1–8. Citado 4 vezes nas páginas 3, 11, 21 e 23.

TSAMARDINOS, I.; GREASIDOU, E.; BORBOUDAKIS, G. Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation. *Machine learning*, Springer, v. 107, p. 1895–1922, 2018. Citado na página 39.

VINH, N. X.; EPPS, J.; BAILEY, J. Information theoretic measures for clusterings comparison: is a correction for chance necessary? In: *Proceedings of the 26th Annual International Conference on Machine Learning*. New York, NY, USA: Association for Computing Machinery, 2009. (ICML '09), p. 1073–1080. ISBN 9781605585161. Disponível em: <https://doi.org/10.1145/1553374.1553511>. Citado na página 48.

WANG, M. et al. Identifying critical outbreak time window of controversial events based on sentiment analysis. *Plos one*, Public Library of Science San Francisco, CA USA, v. 15, n. 10, p. e0241355, 2020. Citado 2 vezes nas páginas 10 e 22.

WEAVER, A. J.; ZELENKAUSKAITE, A.; SAMSON, L. The (non) violent world of youtube: Content trends in web video. *Journal of Communication*, Oxford University Press, v. 62, n. 6, p. 1065–1083, 2012. Citado na página 1.

WEINZIERL, M.; HOPFER, S.; HARABAGIU, S. Misinformation adoption or rejection in the era of covid-19. In: *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM), AAAI Press*. [S.l.: s.n.], 2021. Citado na página 3.

WU, S. et al. Zero-shot event detection using multi-modal fusion of weakly supervised concepts. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2014. p. 2665–2672. Citado 3 vezes nas páginas 3, 11 e 21.

YANG, L. et al. A reinforcement learning approach combined with scope loss function for crime prediction on twitter (x). *IEEE Access*, v. 12, p. 149502–149527, 2024. Citado 3 vezes nas páginas 2, 13 e 16.

YI, P.; ZUBIAGA, A. Weakly supervised cross-platform teenager detection with adversarial bert. In: *ACM Conference on Hypertext and Social Media*. [S.l.: s.n.], 2021. Citado 2 vezes nas páginas 2 e 10.

YOUSAF, K.; NAWAZ, T. A deep learning-based approach for inappropriate content detection and classification of youtube videos. *IEEE Access*, v. 10, p. 16283–16298, 2022. Citado na página 11.

YOUSEFI, N.; SHAIK, M.; AGARWAL, N. Characterizing multimedia information environment through multi-modal clustering of youtube videos. In: SPRINGER. *International Conference on Smart Multimedia*. [S.l.], 2024. p. 295–309. Citado na página 1.