



BACHAREL EM CIÊNCIA DA COMPUTAÇÃO

**ANÁLISE DE DADOS A PARTIR DE COMENTÁRIOS EM VÍDEOS
DE ENSINO DE PROGRAMAÇÃO NO YOUTUBE**

MATEUS BARROS MACEDO

Iporá, GO

2023



INSTITUTO FEDERAL GOIANO - CAMPUS IPORÁ
BACHAREL EM CIÊNCIA DA COMPUTAÇÃO

**ANÁLISE DE DADOS A PARTIR DE COMENTÁRIOS EM VÍDEOS
DE ENSINO DE PROGRAMAÇÃO NO YOUTUBE**

MATEUS BARROS MACEDO

Trabalho de Conclusão de Curso apresentado ao Instituto Federal Goiano - Campus Iporá, como requisito parcial para a obtenção do Grau de Bacharel em Ciência da Computação.

Orientador: Prof. Cleon Xavier Pereira Junior

Iporá, GO
dezembro, 2023

Sistema desenvolvido pelo ICMC/USP
Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas - Instituto Federal Goiano

M141a Macedo, Mateus Barros
ANÁLISE DE DADOS A PARTIR DE COMENTÁRIOS EM
VÍDEOS DE ENSINO DE PROGRAMAÇÃO NO YOUTUBE / Mateus
Barros Macedo; orientador Cleon Xavier Pereira
Junior. -- Iporá, 2023.
26 p.

TCC (Graduação em BACHARELADO EM CIÊNCIA DA
COMPURAÇÃO) -- Instituto Federal Goiano, Campus
Iporá, 2023.

1. YouTube. 2. programação. 3. comentários. 4.
gênero. I. Pereira Junior, Cleon Xavier, orient. II.
Título.

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO

PARA DISPONIBILIZAR PRODUÇÕES TÉCNICO-CIENTÍFICAS

NO REPOSITÓRIO INSTITUCIONAL DO IF GOIANO

Com base no disposto na Lei Federal nº 9.610, de 19 de fevereiro de 1998, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia Goiano a disponibilizar gratuitamente o documento em formato digital no Repositório Institucional do IF Goiano (RIIF Goiano), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IF Goiano.

IDENTIFICAÇÃO DA PRODUÇÃO TÉCNICO-CIENTÍFICA

- | | |
|--|---|
| <input type="checkbox"/> Tese (doutorado) | <input type="checkbox"/> Artigo científico |
| <input type="checkbox"/> Dissertação (mestrado) | <input type="checkbox"/> Capítulo de livro |
| <input type="checkbox"/> Monografia (especialização) | <input type="checkbox"/> Livro |
| <input checked="" type="checkbox"/> TCC (graduação) | <input type="checkbox"/> Trabalho apresentado em evento |

Produto técnico e educacional - Tipo:

Nome completo do autor:

Mateus Barros Macedo

Matrícula:

2020105231940006

Título do trabalho:

Análise de dados a partir de comentários em vídeos de ensino de programação no YouTube

RESTRIÇÕES DE ACESSO AO DOCUMENTO

Documento confidencial: Não Sim, justifique:

Informe a data que poderá ser disponibilizado no RIIF Goiano: / /

O documento está sujeito a registro de patente? Sim Não

O documento pode vir a ser publicado como livro? Sim Não

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O(a) referido(a) autor(a) declara:

- Que o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
- Que obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autoria, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia Goiano os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
- Que cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia Goiano.

Ipórá - Goiás

Local

/ /

Data

Documento assinado digitalmente

gov.br

MATEUS BARROS MACEDO

Data: 21/02/2024 14:03:29-0300

Verifique em <https://validar.iti.gov.br>

Assinatura do autor e/ou detentor dos direitos autorais

Ciente e de acordo:

gov.br

Documento assinado digitalmente

CLEON XAVIER PEREIRA JUNIOR

Data: 26/02/2024 10:47:57-0300

Verifique em <https://validar.iti.gov.br>

Assinatura do(a) orientador(a)



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA GOIANO

Ata nº 124/2023 - GE-IP/CMPIPR/IFGOIANO

ATA DA SESSÃO DE JULGAMENTO DO TRABALHO DE CURSO
DE MATEUS BARROS MACEDO

Aos oito dias do mês de dezembro de dois mil e vinte e três, às treze horas e oito minutos, na sala 8 do bloco 2 do Instituto Federal Goiano – Campus Iporá, reuniu-se, em sessão pública, a banca examinadora designada na forma regimental pela Coordenação do Curso para julgar o trabalho de curso intitulado **“Análise de dados a partir de comentários em vídeos de ensino de programação no Youtube”**, apresentado pelo acadêmico **Mateus Barros Macedo** como parte dos requisitos necessários à obtenção do grau de Bacharel em Ciência da Computação. A banca examinadora foi presidida pelo orientador do trabalho de curso, Professor Doutor Cleon Xavier Pereira Júnior, tendo como membros a Professora Mestra Lais Candido Rodrigues da Silva Lopes e a Professora Mestra Livia Mancine Coelho de Campos. Aberta a sessão, o acadêmico expôs seu trabalho. Em seguida, foi arguido pelos membros da banca e:

(**X**) tendo demonstrado suficiência de conhecimento e capacidade de sistematização do tema de seu trabalho de curso, a banca conclui pela **aprovação** do acadêmico, sem restrições.

() tendo demonstrado suficiência de conhecimento e capacidade de sistematização do tema de seu trabalho de curso, a banca conclui pela **aprovação** do acadêmico, **condicionada a satisfazer as exigências** listadas na Folha de Modificação de Trabalho de Curso anexa à presente ata, no prazo máximo de 60 dias, a contar da presente data, ficando o professor orientador responsável por atestar o cumprimento dessas exigências.

() não tendo demonstrado suficiência de conhecimento e capacidade de sistematização do tema de seu trabalho de curso, a banca conclui pela **reprovação** do acadêmico.

Conforme avaliação individual de cada membro da banca, será atribuída a nota **10 (dez)** para fins de registro em histórico acadêmico.

Os trabalhos foram encerrados às treze horas e quarenta e seis minutos. Nos termos do Regulamento do Trabalho de Curso do Bacharelado em Ciência da Computação do Instituto Federal Goiano – Campus Iporá, lavrou-se a presente ata que, lida e julgada conforme, segue assinada pelos membros da banca examinadora.

(Assinado Eletronicamente)

Prof. Dr. Cleon Xavier Pereira Júnior

(Assinado Eletronicamente)

Prof. Ma. Lais Cândido Rodrigues da Silva Lopes

(Assinado Eletronicamente)

Documento assinado eletronicamente por:

- Cleon Xavier Pereira Junior, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 08/12/2023 13:48:06.
- Livia Mancine Coelho de Campos, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 08/12/2023 13:53:56.
- Lais Candido Rodrigues da Silva Lopes, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 11/12/2023 06:58:50.

Este documento foi emitido pelo SUAP em 08/12/2023. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifgoiano.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código Verificador: 555779

Código de Autenticação: a06ff324de



INSTITUTO FEDERAL GOIANO

Campus Iporá

Av. Oeste, Parque União, 350, Parque União, IPORA / GO, CEP 76.200-000

(64) 3674-0400

RESUMO

BARROS MACEDO, MATEUS. **ANÁLISE DE DADOS A PARTIR DE COMENTÁRIOS EM VÍDEOS DE ENSINO DE PROGRAMAÇÃO NO YOUTUBE.** dezembro, 2023. 26 f. Monografia — (Curso de Bacharel em Ciência da Computação), Instituto Federal Goiano - Campus Iporá. Iporá, GO.

No cenário atual, marcado pelo avanço crescente da tecnologia, os paradigmas sociais se adaptaram englobando o uso de plataformas digitais nos contextos do cotidiano. No âmbito educacional, o uso de plataformas como YouTube como ferramenta de compartilhamento e disseminação de conteúdos se tornou popular em diversas áreas inclusive na área de ensino de programação, gerando um arcabouço de dados tanto de conteúdo quanto das interações dos usuários a partir desses conteúdos. Contudo, o volume de dados gerados através dessa interação entre os vídeos e os alunos torna inviável a análise manual desses dados. O presente estudo busca analisar comentários de vídeos de ensino de programação publicados na plataforma YouTube com base em um recorte de gênero, com o objetivo de identificar padrões e discrepâncias entre os dados. Foram realizados três experimentos, um com foco em aspectos quantitativos dos comentários como número médio de caracteres e número médio de perguntas, o segundo focado em aspectos qualitativos, estabelecendo uma análise de sentimentos nos comentários dos vídeos e a busca por aspectos subjetivos presentes nos comentários. Os experimentos apresentaram diferenças mínimas com relação à parte quantitativa nas duas classes selecionadas para a pesquisa. Os dados dos vídeos apresentaram correlação neutra, não permitindo nenhuma conclusão a partir desses resultados. Na análise quantitativa dos comentários, os dois grupos apresentaram uma média e uma mediana de quantidade de caracteres por comentário similares, mas o desvio padrão e o máximo mostraram que os comentários com maior número de caracteres são nos vídeos de homens, estes também apresentaram a maior quantidade de comentários que eram perguntas. A análise de sentimentos mostrou uma porcentagem maior de comentários “positivo” em vídeos de mulheres em comparação aos comentários classificados como “negativo” ou “neutro”. A análise qualitativa, no entanto evidenciou aspectos de deslegitimação e estereotipação em comentários de vídeos de mulheres que mostraram a ocorrência de palavras como “linda” e “gata” que não estão presentes em comentários em vídeos de homens.

Palavras-chave: YouTube, programação, comentários, gênero.

ABSTRACT

BARROS MACEDO, MATEUS. DATA ANALYSIS FROM COMMENTS ON YOUTUBE PROGRAMMING EDUCATION VIDEOS. dezembro, 2023. 26 f. Trabalho de Conclusão de Curso – Bacharel em Ciência da Computação, Instituto Federal Goiano - Campus Iporá. Iporá, GO, dezembro, 2023.

In the current scenario, marked by the increasing advance of technology, social paradigms have adapted by incorporating the use of digital platforms into everyday contexts. In the educational context, the use of platforms such as YouTube as a tool for sharing and disseminating content has become popular in various areas, including the field of programming education, generating a framework of data both in terms of content and user interactions stemming from that content. However, the volume of data generated through the interaction between videos and students makes manual analysis of these data impractical. This study aims to analyze comments on programming education videos published on the YouTube platform based on a gender perspective, with the goal of identifying patterns and discrepancies among the data. Three experiments were conducted: one focused on quantitative aspects of comments, such as the average number of characters and the average number of questions; the second focused on qualitative aspects, establishing a sentiment analysis in video comments and searching for subjective aspects present in the comments. The experiments showed minimal differences in the quantitative part for the two classes selected for the research. The video data exhibited a neutral correlation, not allowing any conclusions to be drawn from these results. In the quantitative analysis of comments, both groups showed similar averages and medians for the number of characters per comment, but the standard deviation and the maximum showed that comments with a higher number of characters are in videos featuring men. These videos also had the highest number of comments that were questions. The sentiment analysis showed a higher percentage of "positive" comments in women's videos compared to comments classified as "negative" or "neutral." However, qualitative analysis highlighted aspects of delegitimization and stereotyping in comments on women's videos, showing the occurrence of words like "beautiful" and "attractive" that are not present in comments on men's videos.

Keywords: YouTube, programming, comments, gender.

LISTA DE FIGURAS

Figura 1 – Etapas da pesquisa.	12
Figura 2 – Mapa de correlação - Dados de vídeos de homens	14
Figura 3 – Mapa de correlação - Dados de vídeos de mulheres	15
Figura 4 – Análise de sentimentos dos comentários (Homens X Mulheres).	17
Figura 5 – Palavras frequentes em vídeos de homens e mulheres.	18

LISTA DE TABELAS

Tabela 1 – Valores estatísticos sobre a quantidade de caracteres em comentários separados pelo gênero do autor do vídeo.	15
Tabela 2 – Comparativo de API para análise de sentimento.	16
Tabela 3 – Comentários que envolvem características da apresentadora ou indicam interesse em relacionamento.	19

LISTA DE ABREVIATURAS E SIGLAS

PLN	Processamento de Linguagem Natural.
SVM	Support Vector Machine.
NB	Naive Bayes
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
TI	Tecnologia da Informação
STEAM	<i>Science, Technology, Engineering, and Mathematics</i>
VBL	<i>Video Based learning</i>
API	<i>Application Programming Interface</i>
CNN	<i>Convolutional Neural Network</i>
RNN	<i>Recurrent Neural Network</i>

SUMÁRIO

1 – INTRODUÇÃO	1
1.1 Objetivos	4
1.1.1 Objetivo geral	4
1.1.2 objetivos específicos	4
1.2 PROBLEMA E JUSTIFICATIVA	5
2 – REVISÃO DE LITERATURA	6
3 – METODOLOGIA	12
4 – RESULTADOS E DISCUSSÕES	14
4.1 Experimento 1: Panorama dos vídeos	14
4.2 Experimento 2: Análise quantitativa	15
4.3 Experimento 3: Análise qualitativa	17
4.4 Discussão dos resultados	18
5 – CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS	21
Referências	22

1 INTRODUÇÃO

No cenário atual, marcado pelo avanço crescente da tecnologia, é possível observar uma clara relação entre os mecanismos de interação social no mundo físico e no mundo virtual. Através das redes sociais, as pessoas têm experienciado novas dinâmicas de relações interpessoais ao se conectarem com pessoas que, de outra maneira, não teriam contato. Essa dinâmica tem resultado no surgimento de novos meios de comunicação e, conseqüentemente, de interação, assim como também tem influenciado significativamente as dinâmicas de troca interpessoal que até então já estavam estabelecidas na sociedade (ROSA; SANTOS; FALEIROS, 2016).

No campo educacional, onde os vídeos são frequentemente usados como complemento na formação do estudante em uma determinada área do conhecimento (EBIED; KAHOUF; RAHMAN, 2016; MAZIRIRI; GAPA; CHUCHU, 2020; LIN, 2022), observa-se um aumento no uso de plataformas de compartilhamento de vídeos como ferramenta educacional. O formato de aprendizado baseado em vídeos, conhecido como *Video Based Learning* (VBL), tem sido largamente utilizado, principalmente por possuir vantagens que propulsionam o processo de aquisição de conhecimento, como, por exemplo, a possibilidade de repetição e de pausas de um determinado conteúdo enquanto se aprende (SABLIĆ; MIROSAVLJEVIĆ; ŠKUGOR, 2021).

A popularidade progressiva dessas plataformas de compartilhamento de vídeos no contexto educacional se deve, em parte, à flexibilidade e acessibilidade que oferecem. Os vídeos permitem que estudantes aprendam em seu próprio ritmo, revisem o material quantas vezes forem necessários e acessem o conteúdo a qualquer momento e em qualquer lugar, dentro dos requisitos necessários. Tais materiais fornecem uma fonte vasta de informações, tutoriais e demonstrações práticas, que podem auxiliar na compreensão de conceitos simples e complexos, e também auxiliam no desenvolvimento de habilidades técnicas e práticas. Além disso, a natureza visual e auditiva dos vídeos proporciona uma experiência de aprendizado mais imersiva e envolvente, que pode contribuir para uma melhor compreensão e assimilação dos conceitos abordados.

Pesquisas realizadas nessa área mostram que o uso de conteúdo audiovisual direcionado ao ensino se mostra eficaz em promover o engajamento de estudantes no processo de ensino-aprendizagem (CARLISLE, 2010). Através das redes sociais, é possível compartilhar uma gama de informações, desde conteúdo básico até assuntos altamente especializados e complexos (POCHÉ et al., 2017), o que as torna o substrato perfeito para novas formas de ensinar e aprender.

Nesse mesmo sentido, o campo da tecnologia da informação (TI) e da computação se aproveita positivamente com o uso de materiais áudio-visuais em redes sociais como TikTok¹ e YouTube² para compartilhar conteúdos relacionados a diversas tecnologias e também voltados para a programação de computadores (KADRIU et al., 2020; GARCIA; JUANATAS; JUANATAS, 2022).

O campo da TI e da computação é bastante diverso com relação às possibilidades de atuação, indo desde o meio acadêmico até o empresarial (CARDOSO; DAVID, 2016). Essa versatilidade e a procura crescente por conteúdos relacionados à área tem alavancado, por consequência, a criação de materiais com o intuito de facilitar o ensino/aprendizagem de conteúdos da área, com grande predominância em conteúdos no formato audiovisual. O

¹ <https://www.tiktok.com>

² <https://www.youtube.com>

conjunto desses materiais compartilhados nas redes de forma acessível, gerou um arcabouço de informações que impulsiona a disseminação de conhecimento de uma forma acelerada e dinâmica (BARROSO; ANTUNES, 2020).

O YouTube emergiu, nesse cenário de popularização do consumo de vídeos, como uma referência no compartilhamento de conteúdo audiovisual e, conseqüentemente, de materiais relacionados ao ensino de diversas áreas do conhecimento. Além da capacidade de disponibilizar vídeos, essa plataforma oferece recursos interativos que possibilitam aos usuários criar, interagir, discutir e expressar suas opiniões direta ou indiretamente (MACLEOD; STOREY; BERGEN, 2015). Essa dinâmica interativa do YouTube, assim como a facilidade no compartilhamento, acesso e na interação, colocaram a plataforma como um ponto de estudo para diversos pesquisadores (JUNGES; GATTI, 2019).

Como consequência da facilidade de adaptação dos vídeos ao ambiente educacional, a popularização do YouTube como ferramenta de compartilhamento de vídeos e a crescente demanda das áreas de TI e computação, o âmbito do ensino de programação se expandiu vastamente no tocante ao uso da plataforma como ferramenta de disseminação de conteúdo, oferecendo uma variedade de formatos de materiais, como tutoriais, apresentações, palestras e até mesmo vídeos de resolução de problemas em tempo real atendendo assim às diferentes preferências de aprendizado de cada pessoa.

Porém, mesmo com a popularidade do uso de vídeos como ferramenta de ensino e aprendizado de conceitos de programação e as diversas vantagens apresentadas, alguns fatos relacionados ao contexto atual desse cenário devem ser ressaltados. O primeiro se volta para um dado apresentado pela Sociedade Brasileira de Computação (SBC) que mostra uma série temporal que vai dos anos de 2001 a 2020, mostrando a quantidade reduzida de pessoal do gênero feminino que concluem cursos superiores das áreas do conhecimento relacionadas a computação (SBC, 2023). Isso mostra que, mesmo com o crescimento dessas áreas, elas ainda contam com um público que é majoritariamente masculino.

Essa preponderância masculina, ou melhor, a ausência feminina nesse campo do conhecimento, pode ser explicado pela quantidade de estereótipos e estigmas impostos pelo meio social às mulheres (LOUZADA et al., 2014). Além disso, esse desequilíbrio entre a presença masculina e feminina nessas áreas do conhecimento podem também ser motivadas pelos fatores culturais, de modo que pessoas consideradas mulheres são influenciadas desde muito jovens a se voltar com maior afinco para conteúdos que conversam diretamente com cuidado e comunicação (SANTOS; CARVALHO; BARRETO, 2021).

Além dessa diferença relacionada a gênero, é possível ser notado também (ou até mesmo por consequência) a discrepância entre as quantidades de criadores e de criadoras de conteúdo e canais que voltam a produção dos seus vídeos para a área da computação. Na área de Ciência, Tecnologia, Engenharia e Matemática (STEM³) é possível encontrar pouquíssimos canais que são produzidos/apresentados por mulheres dentre os mais populares das plataformas citadas (AMARASEKARA; GRANT, 2019). Em consonância a isso, estudos mostram que, mesmo em canais voltados principalmente para o entretenimento, os canais apresentados/produzidos por pessoas do gênero masculino são a maioria entre os com maior número de visualizações e com maior quantidade de inscritos (REGUEIRA; ALONSO-FERREIRO; DA-VILA, 2020). A partir disso, é esperada uma quantidade ainda menos significativa de mulheres que assumam um protagonismo no que se refere ao compartilhamento de conteúdos via vídeos.

Ao fazer essa ponderação sobre questões relacionadas a gênero, as redes sociais se

³ Sigla oriunda do inglês (*Science, Technology, Engineering and Mathematics*) para representar o currículo das quatro áreas em uma abordagem interdisciplinar e aplicada.

mostram como um espelho, podendo refletir os processos discriminatórios que permeiam a sociedade, seja de forma direta ou indiretamente, mostrando o assim um potencial mecanismo de perpetuação dessas problemáticas sociais através dos meios digitais (MOURA, 2016).

O público feminino, principalmente em se tratando do ambiente virtual, é cercado por processos de menosprezo, xingamentos, ofensas, assédio, dentre outras formas de violência simbólica veiculadas por interações nos ambientes virtuais (SANTANA et al., 2022), que geralmente passam despercebidas no caos do volume de informações presentes nessas redes.

Outro ponto a ser ressaltado com relação a essa dinâmica de trocas é que, ao se tratar de redes sociais com dezenas de milhares de usuários, também lida-se com um grande conjunto de dados. O crescente volume desses dados gerados pelos conteúdos e interações nas redes e a velocidade com que esses conteúdos e interações são publicados, torna inviável a análise individual de cada dado (COUTINHO; MALHEIROS, 2020). Essa dificuldade pode resultar na perda de informações valiosas que poderiam contribuir positivamente tanto para a produção de conteúdo quanto para o acesso mais assertivo de informações pelos usuários. Diante desse desafio, destaca-se a necessidade da análise automatizada de dados como um mecanismo para lidar com essa questão. Ao analisar dados como os comentários deixados nos vídeos, é possível obter informações valiosas sobre a percepção do público, identificar pontos fortes e fracos do material, descobrir dúvidas recorrentes e até mesmo identificar possíveis erros ou lacunas de informação.

Assim, uma análise de dados gerados por usuários desempenha um papel crucial na compreensão da eficácia desses materiais, na identificação de tendências e na melhoria contínua do conteúdo, bem como a possibilidade de descoberta de informações que não seriam percebidas com a análise manual. Ao analisar os comentários deixados nos vídeos, é possível obter percepções do público, identificar pontos fortes e fracos do material, descobrir dúvidas recorrentes e até mesmo assinalar possíveis erros ou lacunas de informação (POCHÉ et al., 2017). A análise dos comentários em específico, pode permitir uma compreensão mais profunda das necessidades e expectativas dos usuários, assim como as reações dos mesmos frente a determinados tipos de didática, conteúdo ou sobre quem está apresentando.

Nessa direção, o YouTube proporciona a possibilidade de coletar dados gerados a partir dos vídeos postados em sua plataforma, esses dados podem gerar análises de comportamento referente aos usuários e como eles se relacionam com o vídeo. Desse modo, é possível realizar comparações entre os tipos de conteúdos, como, por exemplo, entre os vídeos apresentados por mulheres e os vídeos apresentados por homens. O YouTube em específico apresenta diversos dados relacionados a comentários, curtidas, número de visualizações, dentre outros, que podem originar discussões relacionadas a esses materiais hospedados na plataforma (FERREIRA et al., 2020).

Dado que existem públicos diversos que disponibilizam seu conteúdo para disseminar conhecimentos sobre diferentes tópicos, nesse caso especificamente tópicos relacionados a computação, e as problemáticas que podem estar refletidas nos meios de interação presentes em plataformas como o YouTube, esta pesquisa visa realizar uma análise exploratória em vídeos hospedados na plataforma YouTube, buscando identificar padrões e/ou novas informações contidas nesses dados.

1.1 Objetivos

1.1.1 Objetivo geral

Analisar dados de interação em vídeos do YouTube que envolvam o ensino de programação.

1.1.2 objetivos específicos

- Fazer uma revisão da literatura a respeito das ferramentas de classificação de textos;
- Realizar o levantamento dos dados para a criação de uma base com dados a respeito dos conteúdos audiovisuais sobre programação (duração do vídeo, formato de apresentação, comentários, etc.);
- Usar técnicas de correlação para analisar formatos de vídeos e as interações;
- Usar APIs de classificação de texto para analisar sentimentos;

1.2 PROBLEMA E JUSTIFICATIVA

Ao compreender melhor o potencial das plataformas de compartilhamento de vídeos e como elas podem ser integradas de maneira eficaz no ambiente educacional, adaptando-as de maneira adequada às necessidades e demandas dos estudantes, surge a possibilidade de explorar todo o seu poder transformador. A partir disso, promove-se um ensino mais dinâmico, participativo e alinhado com as tendências tecnológicas da sociedade contemporânea, visando a uma formação mais completa e abrangente dos indivíduos.

Identificar as correlações entre as interações e o conteúdo, em específico no contexto do YouTube, podem direcionar a produção de conteúdos mais envolventes e personalizados com base na compreensão mais detalhada das preferências, opiniões e comportamentos das pessoas usuárias. Isso permite que os criadores de conteúdo identifiquem rapidamente os pontos fortes e fracos de seus vídeos, bem como as áreas em que podem melhorar, aumentando a eficácia do ensino.

De mesmo modo, explorar os principais aspectos dessa dicotomia existente entre o mundo físico e o mundo virtual, examinando as implicações dessas mudanças na forma como as pessoas se relacionam e interagem, tanto no âmbito pessoal quanto profissional, e identificar os benefícios e desafios trazidos por essa nova realidade social vai permitir entender tanto as consequências dessas mudanças quanto nos proporcionar ideias de formas eficazes e eficientes de lidar com essa nova realidade híbrida.

Assim, com o aumento do uso de VBL como metodologia de aprendizado, juntamente com as problemáticas relacionadas a gênero que continuam estabelecidas na sociedade, se mostra pertinente o desenvolvimento de pesquisas que busquem voltar o olhar para o comportamento das relações interpessoais no meio virtual, e identificar como e através de quais mecanismos esses processos mostram o comportamento das pessoas no mundo real.

Diante disso, este estudo pretende avaliar a relação entre os elementos presentes nas interações em vídeos de ensino de programação postados no YouTube, fazendo um recorte de gênero, ou seja, comparando e analisando canais apresentados por homens e canais apresentados por mulheres, com o intuito de entender os padrões e dinâmicas das interações dos usuários presentes na plataforma, assim como também entender como os estudantes interagem com esse conteúdo, seus pontos positivos e negativos bem como as possíveis diferenças entre esses processos quando se tratando de vídeos produzidos por homens e vídeos produzidos por mulheres. Ao final, espera-se contribuir para um maior entendimento dessas relações e transformações nas dinâmicas sociais e fornecer reflexões relevantes para o contexto contemporâneo, visando a uma adaptação consciente e saudável ao mundo cada vez mais híbrido em que vivemos.

2 REVISÃO DE LITERATURA

O YouTube é uma plataforma onde são compartilhados em larga escala vídeos *online* e que permite que sejam extraídos diversos dados correlacionados a esses vídeos através de uma *Application Programming Interface* (API). Além disso, esses metadados, que popularmente são definidos como dados sobre dados (PAL et al., 2019), podem fornecer informações sobre a entidade em questão. Devido a amplitude dos mecanismos de interação presentes na plataforma como curtidas, comentários, entre outros, o YouTube é hoje considerado uma mídia social (HOU, 2019). Assim sendo, é interessante a possibilidade de analisar o comportamento dos usuários que interagem com os conteúdos presentes na plataforma (ALHUKAILI; YAFOOZ, 2021).

O rápido desenvolvimento da tecnologia da informação resultou na geração constante de um grande volume de dados. No entanto, essa abundância de informações apresenta um desafio significativo: como sintetizar e extrair informações relevantes eficientemente. Os métodos convencionais de análise humana muitas vezes são inadequados para lidar com esse grande volume de dados. A análise manual requer tempo, recursos e especialização, o que pode ser inviável diante da escala e da velocidade com que os dados são gerados. Do mesmo modo, o simples ato de gerar dados não é mais, nesse cenário de produção massiva de dados, sinônimo de gerar informações, sendo apenas uma pequena parcela desses dados de fato utilizada para obter percepções relevantes (LAUSCH; SCHMIDT; TISCHENDORF, 2015a). Como resultado, informações valiosas que já existem podem permanecer ocultas e não serem adequadamente aproveitadas (ISOTANI; BITTENCOURT, 2015).

Frente a essa realidade em lidar com a explosão de dados, foram desenvolvidas abordagens automatizadas para a análise, como a inteligência artificial. Essa técnica permite a aplicação de algoritmos para extrair padrões, tendências e *insights* dos dados, ajudando a identificar informações relevantes de forma mais eficiente.

O uso de técnicas de Processamento de Linguagem Natural (PLN) surge como uma opção promissora para lidar com a análise de grandes conjuntos de dados, por exemplo, de textos, revelando informações ocultas, que não são tão claras quando feitas por meio de uma análise manual (SANTOS et al., 2015). O PLN é uma área da inteligência artificial que se concentra na interação entre computadores e linguagem humana (PEZZINI, 2017). É uma disciplina que visa capacitar as máquinas a compreender, interpretar, analisar e gerar linguagem natural de maneira semelhante à dos seres humanos (FALCÃO; LOPES; SOUZA, 2022).

Alguns trabalhos na área de mineração de textos e PLN mostram que diversas técnicas são utilizadas para o processamento de textos (NADKARNI; OHNO-MACHADO; CHAPMAN, 2011; SANTOS et al., 2015). A classificação e identificação de textos é também aplicada em áreas do conhecimento que vão desde ciências sociais (SCARPA, 2017), o meio empresarial (BORDIGON et al., 2018), até áreas da medicina (CASTILLA, 2007).

A partir disso, trabalhos como o de Tontodimamma et al. (2021), fazem um panorama da situação da área de estudos de PLN nos últimos trinta anos. Mais especificamente, o autor trata da utilização de técnicas de PLN na classificação e detecção de textos que possuam discurso de ódio, visto que esses discursos têm se aproveitado da névoa produzida pela quantidade dos dados e o anonimato que as redes sociais proporcionam. Por fim, ainda é citada a utilização de técnicas de aprendizado de máquina para a viabilização da classificação.

No trabalho de Lausch, Schmidt e Tischendorf (2015b), são utilizadas técnicas para a mineração de informações e PLN com o intuito de extrair informações de fontes diversas que possuem um escopo muito grande, tentando trazer a tona novas formas de lidar com a informação e gerar conhecimento a partir de dados brutos.

Na área de classificação de textos, os *Support Vector Machines* (SVMs) têm sido amplamente utilizados em estudos. Em Plath et al. (2022), as pessoas autora ressaltam uma problemática com relação a pesquisas em português, onde faltam bases de dados para o treinamento e pesquisas relacionadas à detecção de discursos de ódio em textos. E sobre isso propõem a criação de uma base de dados com textos em português brasileiro, assim como faz testes nos modelos de classificação SVM, NB, *Random Forest Classifier* (RFC). Os resultados variam conforme a metodologia de pré-processamento do texto, o SVM apresentou resultado melhor quando utilizado com a técnica de *stemming*.

A análise de sentimentos é uma das linhas de pesquisa investigada pela área da mineração de textos (MEDHAT; HASSAN; KORASHY, 2014). Dentre as possibilidades que esta linha de pesquisa proporciona, a principal está concentrada na investigação para classificar textos, verificando se o mesmo carrega consigo uma mensagem que contenha um sentimento positivo, negativo ou neutro. Essa investigação é popularmente utilizada em mídias sociais (RAHMAN; ALOTAIBI; ALSHEHRI, 2019), que possibilitam o “comentário” ou publicações textuais. Além disso, a partir dessa análise de sentimentos, surge a possibilidade de investigar a aprendizagem por transferência, fazer a detecção de emoções e construir recursos relacionados ao contexto analisado (MEDHAT; HASSAN; KORASHY, 2014).

No trabalho de Mullen e Collier (2004), as pessoas autoras exploraram o algoritmo SVM para realizar a classificação de sentimentos em textos. Eles investigaram como o SVM poderia ser aplicado para identificar se um texto possui uma conotação positiva, negativa ou neutra, o que pode ser útil em diversas aplicações, como análise de sentimentos em mídias sociais ou revisões de produtos.

O uso de redes neurais tem ganhado popularidade como um método eficaz na classificação de textos. Nesse sentido, um estudo realizado por Yin et al. (2017) compara o desempenho de duas arquiteturas de redes neurais: a Rede Neural Convolutiva (Convolutional Neural Network - CNN) e a Rede Neural Recorrente (Recurrent Neural Network - RNN). Nesse estudo, as pessoas autoras investigaram a capacidade de ambas as redes neurais em realizar a classificação de textos, mostrando que cada uma opera melhor dependendo da tarefa que a mesma deve cumprir. A CNN é uma arquitetura que se destaca na extração de recursos por meio de camadas convolucionais e de pooling, sendo aplicada principalmente em problemas de visão computacional. Já a RNN é conhecida por sua capacidade de modelar sequências de dados, sendo especialmente adequada para tarefas de processamento de linguagem natural.

Com base nos aspectos já citados de velocidade e volume de produção de comentários na internet, há conseqüentemente uma inviabilidade de recursos humanos para a moderação dos conteúdos desses comentários. O que pode ocasionar, por conseqüência, a disseminação de ideologias pejorativas e discriminatórias por meio de comentários. O uso de técnicas de PLN para possibilitar a identificação e classificação automática de textos que contenham conteúdo de cunho odioso tem se expandido. Em Souza, Nakamura e Nakamura (2022) é apresentada uma análise de dados onde tweets foram retirados da plataforma usando como método de filtragem de palavras que são geralmente utilizadas em discursos homofóbicos. As pessoas autoras utiliza quatro modelos de predição, sendo eles: SVM Linear, XGBoost (XGB), *Logistic Regression* (LR) e RFC e dentre estes o autor cita que o RFC obteve melhor desempenho com F1 score de 0,94. O trabalho separa a

classificação dos tweets em homofóbico, não homofóbico e incerto.

Já em Ribeiro e Silva (2019), é utilizada uma CNN. O intuito era distinguir, dentre um conjunto de textos do Twitter, quais textos se caracterizavam como sendo ou não discurso de ódio, e também identificar se esse discurso de ódio é voltado a um grupo específico de pessoas. O texto se refere especificamente à detecção de discurso de ódio voltado para mulheres e imigrantes. Uma característica a se pontuar no método utilizado pelos autores é a utilização da técnica de N-gramas. Na maioria dos trabalhos aqui citados, na etapa de pré-processamento são removidas palavras que não possuem uma conotação semântica significativa na identificação de um discurso de ódio, contudo Ribeiro e Silva (2019) através da técnica de N-gramas analisa as palavras usadas como um conjunto e não apenas isoladamente.

No estudo realizado por Alshamrani et al. (2021), foi abordada a classificação de discursos problemáticos ou inadequados em comentários de vídeos de shows de cartoon no YouTube, considerando a faixa etária apropriada. As pessoas autoras coletaram uma amostra de 50 comentários para cada vídeo selecionado, utilizando o nome de cada show como critério de busca. Para identificar os discursos inapropriados, as pessoas pesquisadoras empregaram técnicas baseadas em redes neurais.

Voltando para o contexto de ensino de programação, assim como as mídias sociais proporcionam novas formas de comunicação, trouxeram também novas formas de ensino. O uso de vídeos como material de ensino tem proporcionado resultados positivos no que se refere ao aprendizado de estudantes. No trabalho de Santos, Silva e Meireles (2020), são mostrados os resultados do uso de vídeos produzidos especificamente para o ensino de programação como material de apoio em aulas do curso de engenharia, assim como mostra também os seus resultados no aprendizado. Os resultados da pesquisa destacam tanto os pontos positivos do uso de vídeos no ensino de programação, quanto a necessidade de reforçar os aspectos específicos que impulsionam esses vídeos como meio de aprendizagem.

Ainda no trabalho de Santos, Silva e Meireles (2020), fica clara a importância dos recursos de métricas de feedback, como os comentários, na personalização da produção de conteúdo. Através do conteúdo dos mesmos é possível identificar os reflexos daquele conteúdo e seus impactos no usuário. Nesse sentido, o uso de técnicas para extrair informações de dados, tendo em vista novamente a quantidade massiva de dados existentes e que são geradas através desses conteúdos, facilita a produção e a entrega de conteúdo personalizado, quando proporciona ao criador um meio de estabelecer, de algum modo, o diálogo com o seu interlocutor.

No trabalho de Hosseini e Brusilovsky (2017) as pessoas autoras propõem uma recomendação de exemplos para estudantes, baseados nos problemas que os próprios estudantes possuem. O PLN é usado para reconhecer as dificuldades dos estudantes, assim como identificar textos que possuam relação com as dúvidas e assim tornar mais eficiente a recomendação de conteúdo com informações relacionadas aos problemas de estudantes de forma mais eficaz e eficiente.

A crescente popularização das plataformas de streaming proporciona o compartilhamento de vídeos com conteúdos variados, tais como lives de jogos, de artistas da área da música, da área da programação e também da saúde. Em Maraza-Quispe et al. (2019) mostra o uso do YouTube como uma ferramenta aquisição de informações que servem de base para pesquisas científicas. Por possuir uma grande quantidade de informação, o autor mostra a possibilidade do uso de vídeos do YouTube na fase da pesquisa exploratória de uma pesquisa científica.

No sentido de análise de textos postados em redes sociais é possível extrair uma grande quantidade de dados a respeito da Thread (linha de comentários e posts relacionados

a um mesmo tópico). Trabalhos como o de Jannah e Hermawan (2022) que usa comentários postados em vídeos do YouTube para ver as opiniões das pessoas com relação ao programa de vacinação aplicado pelo governo. O trabalho citado faz o comparativo entre diversas técnicas utilizadas no processamento de linguagem natural, e mostra as diferenças das mesmas em fazer a classificação dos sentimentos contidos nos comentários.

Em um sentido parecido, onde as pessoas autoras buscam por uma classificação, dessa vez não dos comentários, mas daqueles que os fizeram, Ochoa-Hernández e Alemán (2022) estudam a identificação alguns aspectos dos usuários com base em seus comentários. As pessoas autoras usaram diversas técnicas de processamento de linguagem natural, buscando identificar, principalmente, a opinião política dos alvos do estudo, subdividindo-os em duas (esquerda e direita) ou em quatro (direita, esquerda, direita-moderado, esquerda-moderado). Os resultados mostraram uma porcentagem significativa de acertos dos modelos.

Também no sentido político, o trabalho de Park, Bier e Park (2021) mostra a utilização de comentários de vídeos do YouTube para identificar a opinião e a percepção do público a respeito de notícias da Coreia do Norte. As pessoas executoras da pesquisa usam também técnicas de processamento de linguagem natural, buscando identificar as percepções dos telespectadores, mostrando que algumas características na produção dos vídeo os podem interferir na recepção da notícia, bem como no engajamento do telespectador.

É possível perceber uma vasta utilização de vídeos do YouTube em consonância com outras redes sociais como ferramentas para compartilhar, também, conteúdos relacionados ao ensino de programação. Em Ahmad e Obeidallah (2019), é mostrado o relato da utilização de vídeos do YouTube e grupos do Facebook para auxílio em um laboratório de programação, com foco no ensino da matéria de Programação Orientada a Objetos II. Os resultados revelaram uma maior aderência do conteúdo como também relatos que expressavam uma melhora na dinâmica de ensino, aprendizado, com um ambiente de ensino mais tranquilo e diversificado.

O uso de Aprendizagem baseada em Vídeos é observado em diversas áreas. Em Steinbeck, Zobel e Meinel (2022), é utilizada a metodologia de *Massive Online Open Course* (MOOC) através do YouTube para ensinar conceitos de programação. O trabalho citado compara a diferença existente na percepção dos alunos com relação à presença de um palestrante e o conteúdo aprendido. Os resultados mostraram que a presença de um “professor” no vídeo aumenta a concentração dos participantes por proporcionar, segundo os mesmos, uma sensação de estar sendo observado, além de impulsionar a percepção de qualidade do curso quando ha a presença de um palestrante quando em comparação a vídeos de apresentação de slides. Nos testes de recordação com relação ao conteúdo, o grupo ensinado com a presença de um instrutor no vídeo também apresentou uma variação ligeiramente melhor, mostrando que diferentes metodologias de vídeos podem obter resultados diferentes com relação à aprendizagem.

No trabalho de Nissenon et al. (2019), é relatada a criação de uma base de vídeos tutoriais utilizados no departamento de engenharia de uma universidade da califórnia. Os resultados da utilização mostram uma quantidade grande de acessos aos vídeos no período de aulas, e uma crescente com relação à quantidade de tempo de vídeos assistidos pelos usuários com o tempo, além de mostrar um feedback positivo por parte dos usuários com relação ao uso e impacto dos vídeos em suas graduações.

Pessoas que programam usam plataformas como o YouTube para transmitir o que é chamado de “live coding” onde o transmissor da live programa enquanto transmite o que está fazendo para os telespectadores. Esse tipo de relação se mostra potencialmente favorável com relação às dinâmicas de ensino-aprendizagem quando os dois lados estabelecem uma

troca de conhecimentos. Diversos fatos levam um/uma programador/programadora a transmitir ao vivo enquanto programa, no trabalho de Kokinda e Rodeghero (2023) é relatada a presença do senso de autorresponsabilização, onde o “contrato social” já pre estabelecido e a ideia de outros observadores vendo o seu desenvolvimento promovem um maior compromisso por parte do transmissor a continuar focado em fazer o que se propôs. Outro ponto destacado é o processo de autoeducação, onde a vontade de compartilhar e aprender algo novo é impulsionado através das interações da plataforma e o senso de ensinar outras pessoas, conceitos e conteúdos em que se tem domínio são algumas das principais motivações dos streamers quando compartilham seus códigos e seu processo de programar ao vivo.

No trabalho de Chattopadhyay, Ford e Zimmermann (2021), outro estudo direcionado a programadores/programadoras que postam vídeos na internet, olha para os/as desenvolvedores/desenvolvedoras a partir de uma perspectiva ligeiramente diferente. Nesse trabalho, é focado nos desenvolvedores que compartilham seus vídeos no formato de vlogs, onde passam coisas relacionados ao dia a dia aos telespectadores. O trabalho mostra motivações relacionadas e desmistificação de pessoas que programam como “mestre da matemática” ou até o “nerd da cadeira”, mas também mostra a posição de interesse em criar uma comunidade para a troca de informações. Uma análise a partir dos comentários deixados nos vídeos também revela informações relacionadas as motivações das pessoas para consumir esse conteúdo, mostrando que essa atitude pode ser tanto para aquisição de informação quanto para descobrir novas comunidades que se interessam pelos mesmos temas.

Durante o isolamento social ocorrido entre 2020 e 2023, diversas metodologias de ensino foram aplicadas com o intuito de mitigar os efeitos da inviabilidade do modelo convencional de ensino com segurança. No estudo de (AHMAD; OBEIDALLAH, 2022) analisam o cenário de estudo de estudantes de programação, bem como o uso de uma metodologia apoiada também no uso de materiais em vídeos. Para a pesquisa foi relatado que uma porcentagem significativa de estudantes usa palestrar gravadas como ferramenta de aprendizado, principalmente no cenário da COVID-19, como o trabalho citado.

Tendo em vista que estudantes costumam preferir meios de aprendizado como o YouTube (BRINGULA et al., 2014), o estudo de Bringula et al. (2020) analisa os feedbacks deixados nos vídeos de ensino de programação de duas linguagens distintas. A pesquisa mostrou tanto que as pessoas usuárias se mostram contentes com relação ao uso dos vídeos que estão disponibilizados, que pode ser verificado pela alta presença de palavras de agradecimento nos comentários dos dois grupos de vídeos analisados quanto que os meios de interação proporcionados pela rede impulsionam tanto o aprendizado quando os consumidores dos vídeos fazem perguntas e sugestões, quanto proporciona meios me melhorar os conteúdos criados, através de sugestões expressas nos comentários.

Alguns trabalhos não apresentam o uso do YouTube como ponto principal da metodologia, colocando o mesmo como parte do processo e como uma ferramenta auxiliar que serve como apoio a outras ferramentas. Com isso surgem alguns problemas como os relatados em Luchaninov et al. (2021) onde o aspecto de desatualização, confiabilidade e possibilidade de aplicação dos materiais do vídeo atrapalham o processo de utilização da ferramenta com uma maior potência no quesito transmissão de conteúdo.

A partir desses vídeos, outras possibilidades de interação surgem, como, por exemplo, os comentários. Os mesmos podem ser fonte de uma elevada quantidade de informações, como, por exemplo, a satisfação daqueles que assistem os vídeos ou até novas possibilidades quanto a utilização das técnicas e conteúdos disponíveis nos mesmos. Como, por exemplo, o trabalho de Sarkar et al. (2022) que mostra uma análise da utilização de

funções lambda em planilhas, com base em comentários de usuários retirados das redes sociais, em especial (com maior porcentagem de comentários) o YouTube.

3 METODOLOGIA

Este trabalho apresenta uma pesquisa exploratória visando análises quantitativas e qualitativas a partir das questões de gênero e ensino de programação utilizando VBL, onde busca-se fazer a coleta de dados do YouTube e as possíveis relações com outros metadados que permitem extração de informações nesse âmbito. Para isso, foram estabelecidas etapas para alcançar os objetivos propostos. A Figura 1 apresenta as etapas da pesquisa.

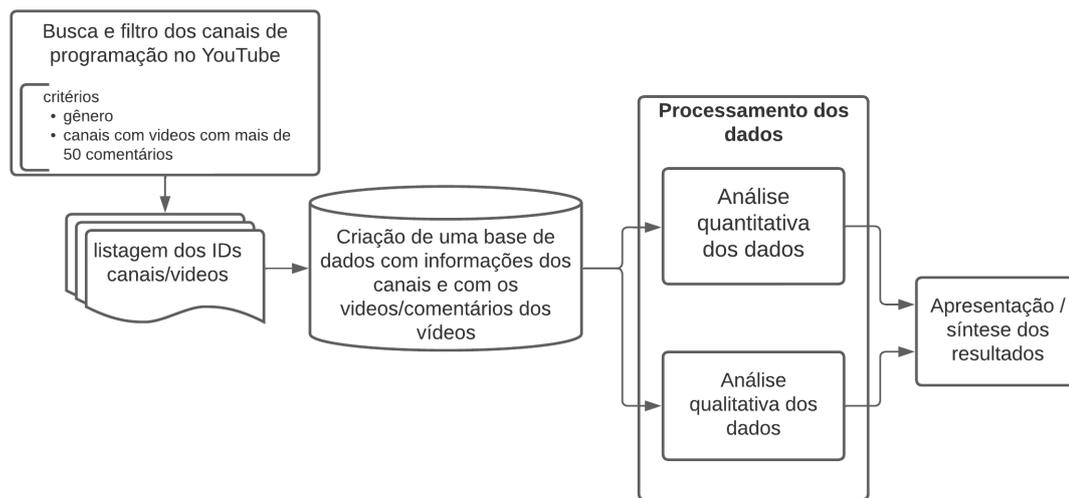


Figura 1 – Etapas da pesquisa.

Na primeira etapa da pesquisa, foi realizada uma revisão bibliográfica a respeito da utilização do YouTube como ferramenta de ensino, e também voltada as principais técnicas de análise de interações que estão sendo utilizadas em plataformas de compartilhamento de conteúdos audiovisuais. Essa etapa buscou esclarecer o estado geral da área de ensino de programação através de vídeos e buscar formas de identificar padrões e divergências nas interações dos usuários nessas plataformas.

Posteriormente, foram definidos os parâmetros de busca na plataforma YouTube com o intuito de delimitar o escopo da pesquisa. Conforme apresentado nas Seções 1 e 2, os canais cujas pessoas apresentadoras são do gênero feminino possuem, em geral, menos inscrições. Neste aspecto, foi feita uma busca exploratória por canais de mulheres que falavam sobre tecnologia¹. A partir desses canais, foi possível obter a quantidade de inscrições por canal. Com isso, foram selecionados os 5 canais produzidos por pessoas do gênero feminino com o maior quantidade de inscrições. Outros critérios de inclusão foram estabelecidos com o intuito de estipular um padrão na seleção dos canais e possibilitar certa proporcionalidade entre a quantidade de comentários dos dois grupos, sendo eles: i) os canais deveriam possuir no mínimo 6 vídeos com mais de 50 comentários em cada vídeo, e ii) os vídeos deveriam ter como conteúdo principal o ensino de conteúdos de programação. Após a aplicação dos critérios, foi realizada a busca por vídeos em canais de pessoas do gênero masculino, visando manter quantidades próximas as de pessoas inscritas e que cumpriam os critérios de inclusão. Também foram selecionados 5 canais.

¹ A lista completa dos canais analisados pode ser conferida clicando aqui

Ao final da etapa de seleção, baseada nos critérios explicitados anteriormente, foi criada uma base com 5 canais apresentados por homens e 5 canais apresentados por mulheres. Para cada canal, foram selecionados manualmente 6 vídeos dos canais selecionados considerando os critérios de inclusão. Após definir os canais que atendiam os critérios do estudo, foram pesquisados os IDs dos canais, posteriormente utilizados como parâmetro de requisição na API do YouTube. Para a extração dos comentários dos vídeos, foi utilizada a API do YouTube, onde, para cada um dos vídeos selecionados, efetuou-se uma requisição na API com o parâmetro de no máximo 50 comentários, e os resultados dessas requisições formularam a base de dados utilizada. Vale ressaltar que em alguns casos, mesmo o vídeo possuindo mais de 50 comentários, a requisição retornou menos de 50 comentários.

A partir disso, foram realizadas coletas de dados extraídos dos vídeos. Essa etapa envolveu a obtenção de informações relevantes a respeito das métricas de criação e de interação com os conteúdos, tais como o número de visualizações, o número de curtidas, a duração dos vídeos, a presença ou não dos apresentadores ou das apresentadoras nos vídeos, a data de criação dos vídeos e o teor dos sentimentos contidos nos comentários. Dados esses que são fundamentais para a construção da base de dados que será utilizada na análise das características dos vídeos e na identificação de possíveis padrões.

A coleta de dados foi conduzida de forma sistemática, considerando critério pré-estabelecidos, como o tema dos vídeos, a quantidade mínima de comentários nos vídeos e o número de pessoas inscritas dos canais. Após a coleta, foi necessário realizar processos de tratamento e organização do conjunto, a fim de facilitar a análise automatizada. Esse processo exigiu a definição de critérios claros que seriam relevantes para o objetivo da pesquisa. Ao final destas etapas, foi construída uma base de dados estruturada, que foi utilizada para análises posteriores das características dos vídeos e a identificação de padrões relevantes.

Após a criação da base de dados, foram aplicadas técnicas com intuito de identificar padrões e tendências nos comentários coletados, como também para fazer cálculos estatísticos com relação aos mesmos. Essa fase envolveu a utilização de técnicas de PLN por meio do uso da interface do Chat GPT (*Generative Pre-trained Transformer*), buscando analisar os sentimentos expressos nos comentários. A análise de sentimentos visa identificar os sentimentos expressos nos textos de usuários/usuárias nos comentários. Os algoritmos de classificação de texto são utilizados para determinar se um comentário é positivo, negativo ou neutro em relação ao conteúdo do vídeo. Essa análise permitiu compreender a percepção geral dos espectadores em relação ao material apresentado e identificar possíveis pontos de atenção com relação as interações presentes nos vídeos. Além da análise de sentimentos, técnicas de PLN também foram empregadas para a detecção de palavras-chave, mais especificamente adjetivos, presentes nos comentários, para verificação da frequência de ocorrência dos mesmos em cada grupo.

A fim de subsidiar a comparação da relação entre os dados de interação nos vídeos e a média de classificação dos comentários, foi analisada a correlação entre a metodologia utilizada na produção dos vídeos e a quantidade de comentários recebidos, bem como a ocorrência de sentimentos positivos ou negativos nos comentários. Essas análises permitiram uma compreensão mais aprofundada das interações entre os usuários e os vídeos, permitindo identificar fatores que influenciam nos resultados de classificação. Os resultados obtidos ao longo do trabalho foram consolidados e disponibilizados em veículos de comunicação científica, como artigos acadêmicos e conferências da área.

4 RESULTADOS E DISCUSSÕES

A presente pesquisa se propôs a analisar comentários de vídeos do YouTube. Os resultados obtidos são referentes a um recorte que separou vídeos de ensino de programação de canais feitos por homens e canais feitos por mulheres postados na plataforma Youtube.

Os experimentos realizados consistem em: i) uma análise quantitativa das métricas dos comentários e ii) uma análise qualitativa aplicada a base de comentários.

4.1 Experimento 1: Panorama dos vídeos

Com o intuito de se obter um panorama dos dados dos vídeos retirados dos canais analisados, foi construído um mapa de correlação envolvendo os dados dos canais, como número de curtidas, número de comentários, data de postagem do vídeo, número de visualizações e a duração dos vídeos. Para permitir a análise, a data dos vídeos que é retornada com “dia-mês-ano” foi transformada em dias contabilizando a partir da data de publicação do vídeo.

As figuras 2 e 3 mostram os resultados dessa análise feita a partir dos dados retirados dos vídeos de homens e dos dados retirados de vídeos de mulheres respectivamente.

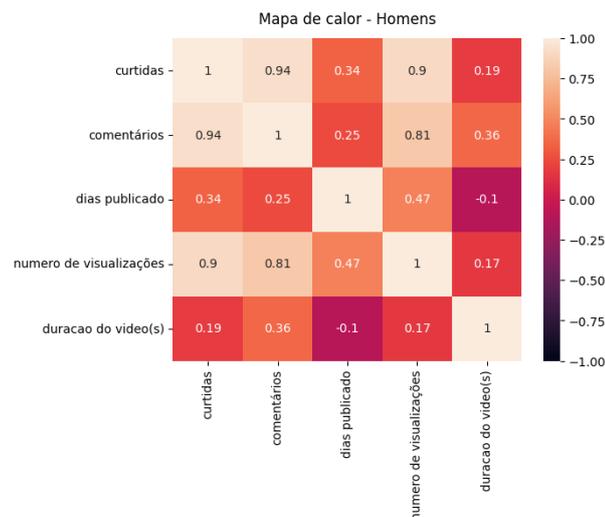


Figura 2 – Mapa de correlação - Dados de vídeos de homens

As figuras mostram organizações a partir do gradiente de cores relativamente parecidos, estipuladas a partir do coeficiente de correlação, onde números acima de 0 mostram uma correlação positiva, ou seja, mostra que o aumento de uma variável está relacionada ao aumento da outra. Números abaixo de 0 mostram uma correlação negativa, onde quando uma variável aumenta a outra diminui, e números iguais a 0 mostram correlação neutra, não há correlação sistemática entre as duas variáveis, fazendo com que poucas ou nenhuma informação a partir desses dados possam ser inferidas relacionadas ao aspecto de gênero analisado.

A partir disso, as análises se voltaram aos comentários postados nos vídeos.

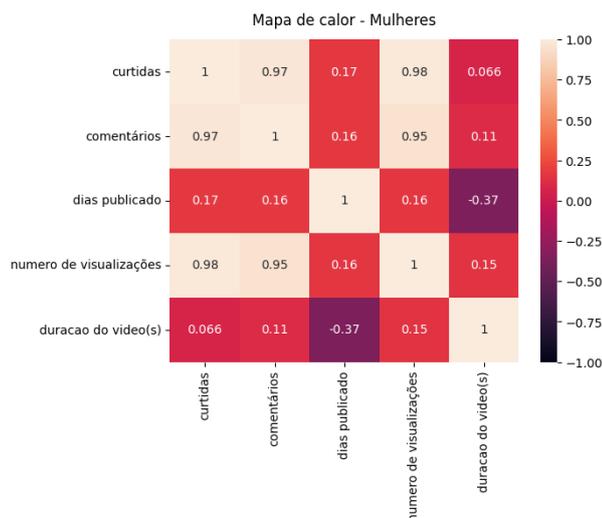


Figura 3 – Mapa de correlação - Dados de vídeos de mulheres

4.2 Experimento 2: Análise quantitativa

Na análise quantitativa, foi inicialmente realizada uma média da quantidade de caracteres dos comentários dos vídeos. A Tabela 1 apresenta valores estatísticos sobre a quantidade de caracteres em comentários pelo gênero do autor do vídeo. São apresentados os valores de média, mediana, desvio padrão, máximo e mínimo de caracteres.

Tabela 1 – Valores estatísticos sobre a quantidade de caracteres em comentários separados pelo gênero do autor do vídeo.

	média	desvio padrão	mediana	máximo	mínimo
Masculino	122,57	275,74	71	9059	1
Feminino	121,42	165,23	73	2083	1

Com os resultados dessa etapa do primeiro experimento, pôde-se observar que os vídeos que eram apresentados por YouTubers mulheres, possuíam nos seus comentários uma média de aproximadamente 121,42 caracteres, sendo uma quantidade menor de caracteres digitados quando comparada aos 122,57 caracteres em média que foram apresentados nos comentários dos vídeos feitos por Youtubers homens.

Conforme o cenário apresentado na Tabela 1, embora a média e a mediana da quantidade de caracteres em comentários para amostras de vídeo de pessoas do gênero masculino e feminino sejam próximas, a quantidade máxima de caracteres e o desvio padrão das amostras para vídeos de pessoas do gênero masculino são substancialmente maiores que os das amostras para o feminino. Dessa forma, os comentários usuais na plataforma não apresentam uma tendência com relação ao tamanho com base no gênero da pessoa apresentadora, porém os comentários com maior quantidade de caracteres estão presentes nos vídeos de canais de pessoas do gênero masculino.

A segunda etapa de análise do primeiro experimento foi baseada na ocorrência de perguntas nos comentários. É importante ressaltar que para essa etapa não foram contabilizadas mais de uma pergunta por comentário, fazendo uma classificação booleana de "contem" ou "não contém", com base na ocorrência do caractere de interrogação existente no comentário. Isso se deve ao fato de que o método de detecção de perguntas feito na base de dados completa utiliza a verificação da ocorrência do ponto de interrogação ("?") para

definir se o comentário é ou não uma pergunta. Deste modo, não foi contabilizado, portanto, comentários com mais de uma pergunta e nem mesmo comentários que não continham o caractere. Nessa análise, pode-se verificar que aproximadamente 23,09% dos comentários de vídeos de homens apresentavam pelo menos uma pergunta, enquanto os vídeos de mulheres apresentavam uma porcentagem ligeiramente menor com 20,74% dos comentários sendo classificados como pergunta, mostrando novamente uma diferença pequena, mas que deve ser considerada, entre os resultados dos dois conjuntos de comentários.

Na terceira etapa do primeiro experimento, foi realizada uma análise quanto aos sentimentos contidos nos comentários. A classificação de sentimentos foi feita com o uso do GPT, utilizando a interface ChatGPT, as diferenças foram pequenas.

Com base na literatura, a princípio foram selecionadas a API Komprehend e a API GPT (a partir da interface com o ChatGPT – por ser uma versão gratuita). Para auxiliar na decisão sobre qual API seria melhor para a análise do domínio português e comentários de vídeos no YouTube, algumas análises estatísticas foram realizadas.

Para verificar a eficácia dos modelos utilizados, foram selecionados os 10 últimos comentários do retorno de cada vídeo e efetuada uma análise manual desta base. A base foi composta por 300 comentários retirados dos vídeos de homens e 300 comentários de vídeos de mulheres. Nessa etapa, duas pessoas classificaram os sentimentos em positivo, negativo ou neutro. Posteriormente, foram utilizadas métricas de verificação estatística de acurácia, *recall* e *F1-score*, dos resultados dos modelos, tomando como base de comparação a classificação das pessoas avaliadoras.

A Tabela 2 mostra o resultado da análise. Com relação ao recorte de base de 600 comentários, notou-se que o GPT, a partir do uso do ChatGPT, realizou uma classificação superior ao Komprehend. Deste modo, optou-se pelo uso do GPT para realizar o percentual dos sentimentos.

Tabela 2 – Comparativo de API para análise de sentimento.

	acurácia	precisão	recall	f1-score
GPT (ChatGPT)	73,33%	79,74%	73,33%	71,33%
Komprehend (komprehend.io)	50,00%	55,85%	50,00%	49,28%

A medida de acurácia permite verificar a proporção de acertos de um modelo com relação aos resultados alvo, ou seja, os resultados “corretos” que se espera. Nesse caso em específico, os resultados alvo são as classificações feitas manualmente por um avaliador e uma avaliadora. Contudo, por ser um conjunto de dados desbalanceado, a medida de eficácia dos modelos não é consistente apenas com a acurácia. Em um conjunto de dados desbalanceados, o modelo pode ter uma acurácia alta simplesmente por classificar todo o conjunto com a classe mais recorrente.

A medida de precisão mostra a proporção entre as ocorrências de classificação correta em uma classe e a quantidade de ocorrências total daquela classe, mostrando, por exemplo, quantos comentários classificados como “positivo” pelo modelo era de fato da classe “positivo”. Uma alta precisão indica que o modelo é eficaz na identificação correta das instâncias de uma dada classe, minimizando assim os falsos positivos.

O *recall*, ou sensibilidade, mostra a relação de quantos comentários de uma dada classe foram classificados como tal, por exemplo, quantos comentários que são “positivo” foram classificados como “positivo”. Em termos simples, um recall alto indica que o modelo está minimizando a quantidade de falso-negativos.

O *F1-score* combina a precisão e o recall em uma única pontuação, proporcionando uma visão equilibrada do desempenho do modelo. Essa métrica é especialmente útil quando

há um desequilíbrio nas classes do conjunto de dados, pois leva em conta tanto os falsos positivos quanto os falsos negativos.

Os gráficos da Figura 4 trazem a comparação entre as classificações feitas pelo chatGPT entre os comentários de vídeos de homens e comentários de vídeos de mulheres, esses comentários foram classificados em “positivo”, “negativo” e “neutro”.



Figura 4 – Análise de sentimentos dos comentários (Homens X Mulheres).

As diferenças apresentadas entre as médias foram pequenas com relação aos comentários “positivo” quando comparados os resultados médios dos sentimentos em comentários de vídeos apresentados por homens e apresentados por mulheres, variando apenas 0.7%, com a maior porcentagem sendo dos comentários em vídeos de mulheres. Já no que se refere aos comentários em que o sentimento contido neles foi classificado como negativos houve um aumento de 1,5% no grupo de comentários de vídeos de Homens quando feita uma comparação aos sentimentos da mesma categoria em comentários em vídeos de mulheres.

Uma parcela significativa dos comentários possuía uma classificação de “neutro”. Essa ocorrência, na classificação “neutro” pode ser reflexo do fato de os conteúdos serem educativos e/ou tutoriais, o que resulta em uma porcentagem significativa de comentários expressando dúvidas e/ou fazendo pontuações sobre o conteúdo assim como opiniões das pessoas usuárias. Estes textos não carregam, necessariamente, uma carga sentimental positiva ou negativa em seu conteúdo, podendo ser um motivo que justifique o volume apresentado de comentários que apresentaram a classificação “neutro”. De modo geral as comparações entre os comportamentos expressos nos comentários apresentaram-se bastante consistentes quando analisadas as duas categorias de criadores de vídeos no recorte de gênero.

4.3 Experimento 3: Análise qualitativa

Visando a melhor compreensão dos resultados, também foram realizadas análises qualitativas. Ainda com o apoio do GPT, foi solicitada a extração dos adjetivos dos comentários dos vídeos da base de comentários selecionada para este trabalho. Os adjetivos foram coletados e processados para gerar nuvens de palavras e permitir análises acerca dos diferentes termos utilizados nas bases. Por fim, manualmente, foram selecionados alguns comentários que permitem uma discussão acerca de gênero nos vídeos analisados. O uso das ferramentas para a execução da pesquisa foram o Google Collaboratory¹ como

¹ <https://colab.google>

Tabela 3 – Comentários que envolvem características da apresentadora ou indicam interesse em relacionamento.

Comentários em vídeos de mulheres
Ela é tao linda q tenho q ver o video 2x pra prestar atenção.
Além de linda e simpática ensina muito!
Nem sei o que vc ta falandokkkk mais fico te assistindo por que te acho linda d+
tão linda apaixonado por vc kakaka ganhou 1 escrito kk
Voz meiga, inteligente e linda, tô apaixonado !????
LINDA E INTELIGENTE
Meu Deus, que mulher linda..... Infelizmente o meu modo Gado foi ativado
n tem como prestar atenção na aula vei, ela é linda demais ;3
Linda e programadora...Ta usando cheater no jogo da vida UHAUHAUAHUAHAU
Nossa que gata é essa
Sempre vejo varios videos seus, muito boa a didatica.. ps: vc tá mt gata nesse video mds
Cabelos lindos
Playlist maravilhosa!!! Adorei sua voz!!
Voz meiga, inteligente e linda, tô apaixonado !????
Muito bom, sua voz é muito boa de se ouvir, ela não vai me expressar
Parabéns você ensina muitos bem com uma voz bem netidez que faz com que a pessoa entenda bem a tonalidade da sua voz
Além de bonita muito inteligente
num entendi nada..... professora bonita demais, tive que deixar só o audio.....aí entendi tudo..... p.s. Brincadeira! nasceu pra ser professora!!!!
Ta bonita la profe

peessoas que interagem nestes vídeos tem, geralmente, o intuito de tirar dúvidas ou fazer ponderações positivas acerca daquele conhecimento.

Por outro lado, na análise qualitativa, notam-se diferenças que são omitidas a partir dos resultados quantitativos. Com relação às palavras, o termo “Didática” é repetidamente colocado em vídeos de pessoas do gênero feminino. Não é possível afirmar, mas talvez cause uma certa surpresa de ter mulheres explicando sobre assuntos classificados como *hard skills* e vistos mais frequentemente ministrados por homens. De fato, as pesquisas apresentam que os canais de homens são mais populares (AMARASEKARA; GRANT, 2019), porém não há dados que comprovem serem superiores em qualidade que os canais de mulheres.

Não obstante, termos como “simpática”, “linda”, “bonita”, “gata” também figuram entre a nuvem de palavras dos vídeos de mulheres. Tais adjetivos são frequentemente empregados ao se referir a imagem feminina idealizada socialmente. Assim, é perceptível a existência de vieses implícitos nos papéis de gênero relacionados às figuras femininas e masculinas, i.e, os adjetivos expressam a forma com que as pessoas vislumbram o que é feminino e o que é masculino, de modo que para o público masculino, por exemplo, o adjetivo “lindo” não foi pontuado no aspecto de beleza física.

Considerando as manifestações de pessoas usuárias por meio dos comentários atribuídos às apresentadoras do gênero feminino (como apresentado na Tabela 3), o teor destes corroboram para evidenciar a existência de violências simbólicas e sexistas nos canais de comunicação, mesmo que aconteçam de maneira velada. Nesse sentido, apesar de adjetivos relacionados à didática terem sido evidenciados nos vídeos femininos, há comentários que sinalizam uma possível deslegitimação e desqualificação do conhecimento disseminado em vídeo, a exemplo do comentário “*Nem sei o que vc ta falandokkkk mais fico te assistindo por que te acho linda d+*”. Portanto, isso suscita discussões quanto à objetificação e vulnerabilidade de mulheres, mesmo em mídias sociais, em que adjetivos

como “linda”, “gata”, “bonita” são usuais para denotar pontos positivos e/ou neutros relacionados a um vídeo. Conforme Costa e Carvalho (2020), esta posição de vulnerabilidade no YouTube está atrelada a preconceitos e estereótipos decorrentes de processos socioculturais, econômicos e políticos que também possuem efeito na esfera da Internet.

Deste modo, ainda que o assunto de um vídeo seja algo de interesse da comunidade, discutindo, muitas das vezes, aspectos técnicos relacionados ao ensino de programação, parte das pessoas comentadoras atribuem pouca seriedade ao tema. São enfatizados aspectos como voz, cabelo e/ou outros adjetivos empregados na sociedade para representar características tidas como “femininas” no imaginário social, conseqüentemente, resultando em discussões mais rasas sobre o tópico abordado nos vídeos.

Os resultados também permitem trazer como discussão que assuntos que envolvem gênero nem sempre é possível de observar se há ou não diferenças somente a partir de dados quantitativos. Tal análise deve servir como reflexão para que trabalhos futuros se atentem às possíveis violências simbólicas e/ou de mensagens de cunho sexistas e discursos de ódio que podem ser interpretadas por classificadores como mensagens positivas.

Os resultados da pesquisa indicam que existem semelhanças e diferenças nos comentários em vídeos de ensino de computação produzidos por homens e por mulheres. Em termos de análise quantitativa, não houve diferenças significativas entre os gêneros em relação à média de caracteres nos comentários, presença de perguntas nos comentários e sentimento geral expresso nos comentários. Tanto homens quanto mulheres receberam predominantemente comentários positivos e neutros.

No entanto, a análise qualitativa revelou alguns padrões distintos. A análise da nuvem de palavras mostrou que as palavras em destaque nos vídeos produzidos por homens e mulheres se assemelham, porém as palavras que possuem menos frequência trazem questões de gênero. A análise qualitativa também revelou comentários que focavam na aparência e características físicas das apresentadoras, indicando a presença de comentários objetificadores e sexistas. Esses resultados sugerem que, embora haja semelhanças na análise de sentimento e no engajamento nos comentários de vídeos de ensino de programação em geral, também existem dinâmicas específicas de gênero. As mulheres podem enfrentar desafios adicionais relacionados à objetificação e estereótipos de gênero, que podem se manifestar na forma de comentários que focam em sua aparência no lugar de suas habilidades com o conhecimento. Os resultados destacam a importância de refletir acerca de preconceitos de gênero e promover ambientes inclusivos no campo da educação em programação.

É importante ressaltar que essa pesquisa representa uma análise específica com base em uma amostra limitada de vídeos e comentários. Os resultados devem ser interpretados no contexto do estudo e podem não ser generalizáveis para todos os vídeos de ensino de programação ou canais do YouTube. Outra limitação é que a classificação de sentimentos foi feita com a interface ChatGPT, sendo que a técnica que alimenta a interface é classificada como não determinística. Pesquisas futuras com conjuntos de dados maiores e mais diversos podem fornecer uma compreensão mais abrangente das dinâmicas entre gênero, comentários e educação em programação no YouTube.

5 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Com o aumento da quantidade de dados analisados, fica cada vez mais evidente a necessidade de se debruçar sobre esses dados a fim de entender um pouco melhor os processos de troca interpessoal nos meios de comunicação digital. A partir disso, busca-se através desse trabalho inicialmente uma revisão da literatura com o intuito de entender o panorama atual do uso de ferramentas na classificação de textos. Assim como também proporcionar uma base para a execução de próximos trabalhos, bem como para servir de arcabouço para outras pesquisas no ramo.

Juntamente com isso, fazer a criação de uma base de dados contendo informações sobre vídeos de ensino de programação na plataforma YouTube. Dados como duração dos vídeos, índices de interação, didática utilizada, etc. farão parte dessa base de dados. Além disso, refinar os parâmetros de avaliação e criação dessa base de dados para obter um conjunto de dados com maior confiabilidade, aumentando a representatividade de resultados mais condizente com a realidade.

Com a criação de uma base de dados com comentários dos vídeos, espera-se aplicar diversas técnicas de classificação de texto, buscando identificar características com relação às percepções dos usuários, assim como tendências que possam estar presentes nos mesmos. A partir disso ter um panorama que possa direcionar conclusões a respeito do cenário atual em se tratando de vídeos de ensino de programação.

Ao fim espera-se ter um conjunto de informações a respeito das relações entre os formatos e características dos vídeos de ensino de programação existentes na plataforma YouTube e as métricas de interatividade e feedback da plataforma. Ao término, espera-se obter uma melhor compreensão do que os dados disponibilizados podem nos providenciar de informações relevantes e que podem servir para direcionar o ensino de programação através de vídeos.

Referências

- AHMAD, A. A.; OBEIDALLAH, R. The impact of social networks on students' academic achievement in practical programming labs. *International journal of advanced computer science and applications*, Science and Information (SAI) Organization Limited, v. 10, n. 11, 2019. Citado na página 9.
- AHMAD, A. A.; OBEIDALLAH, R. Studying the effectiveness of a proposed methodology for teaching programming labs online and students' perspectives toward it during covid-19: A case study of hashemite university. *iJIM*, v. 16, n. 05, p. 53, 2022. Citado na página 10.
- ALHUJAILI, R. F.; YAFOOZ, W. M. Sentiment analysis for youtube videos with user comments. In: IEEE. *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*. [S.l.], 2021. p. 814–820. Citado na página 6.
- ALSHAMRANI, S. et al. Hate, obscenity, and insults: Measuring the exposure of children to inappropriate comments in youtube. In: *Companion Proceedings of the Web Conference 2021*. [S.l.: s.n.], 2021. p. 508–515. Citado na página 8.
- AMARASEKARA, I.; GRANT, W. J. Exploring the youtube science communication gender gap: A sentiment analysis. *Public Understanding of Science*, SAGE Publications Sage UK: London, England, v. 28, n. 1, p. 68–84, 2019. Citado 2 vezes nas páginas 2 e 19.
- BARROSO, F.; ANTUNES, M. Tecnologia na educação: ferramentas digitais facilitadoras da prática docente. *Pesquisa e Debate em Educação*, v. 5, n. 1, p. 124–131, jul. 2020. Disponível em: <<https://periodicos.ufjf.br/index.php/RPDE/article/view/31969>>. Citado na página 2.
- BORDIGON, A. C. de A. et al. Processamento de linguagem natural na identificação e modelagem de processos de negócio: Uma revisão sistemática da literatura. *Anais do XIV Simpósio Brasileiro de Sistemas de Informação*, SBC, p. 191–198, 2018. Citado na página 6.
- BRINGULA, R. P. et al. "why computing students are not using e-resources?" evidence from the university of the east. In: *Proceedings of the Western Canadian Conference on Computing Education*. [S.l.: s.n.], 2014. p. 1–5. Citado na página 10.
- BRINGULA, R. P. et al. Cluster and sentiment analyses of youtube textual feedback of programming language learners to enhance learning in programming. In: SPRINGER. *Proceedings of the Future Technologies Conference (FTC) 2019: Volume 2*. [S.l.], 2020. p. 913–924. Citado na página 10.
- CARDOSO, É. E. C.; DAVID, T. de. A falta de profissionais de tecnologia de informação no mercado de trabalho. *Uma Nova Pedagogia para a Sociedade Futura*, p. 697–700, 2016. Citado na página 1.
- CARLISLE, M. C. Using you tube to enhance student class preparation in an introductory java course. In: *Proceedings of the 41st ACM Technical Symposium on Computer Science Education*. New York, NY, USA: Association for Computing Machinery, 2010. (SIGCSE '10), p. 470–474. ISBN 9781450300063. Disponível em: <<https://doi.org/10.1145/1734263.1734419>>. Citado na página 1.

CASTILLA, A. C. *Instrumento de investigação clínico-epidemiológica em Cardiologia fundamentado no processamento de linguagem natural*. Tese (Doutorado) — Universidade de São Paulo, 2007. Citado na página 6.

CHATTOPADHYAY, S.; FORD, D.; ZIMMERMANN, T. Developers who vlog: dismantling stereotypes through community and identity. *Proceedings of the ACM on Human-Computer Interaction*, ACM New York, NY, USA, v. 5, n. CSCW2, p. 1–33, 2021. Citado na página 10.

COSTA, V. S. d.; CARVALHO, C. A. Mulheres não podem falar de ciência? análise de comentários sexistas em vídeo do canal nerdologia. *Em Questão*, v. 26, n. 1, p. 42–64, 2020. Citado na página 20.

COUTINHO, V. M. d. M. S.; MALHEIROS, Y. Detecção de mensagens homofóbicas em português no twitter usando análise de sentimentos. In: SBC. *Anais do IX Brazilian Workshop on Social Network Analysis and Mining*. [S.l.], 2020. p. 1–12. Citado na página 3.

EBIED, M. M. A.; KAHOUF, S.; RAHMAN, S. A. A. Effectiveness of using youtube in enhance the learning of computer in education skills in najran university. *International Interdisciplinary Journal of Education*, v. 5, n. 3, p. 619–625, 2016. Citado na página 1.

FALCÃO, L. C. de J.; LOPES, B.; SOUZA, R. R. Absorção das tarefas de processamento de linguagem natural (nlp) pela ciência da informação (ci): uma revisão da literatura para tangibilização do uso de nlp pela ci. *Em Questão*, p. 13–34, 2022. Citado na página 6.

FERREIRA, J. P. B. et al. Análise de vídeos como recurso educacional em plataforma não formal de aprendizagem. In: SBC. *Anais do XXXI Simpósio Brasileiro de Informática na Educação*. [S.l.], 2020. p. 1733–1742. Citado na página 3.

GARCIA, M. B.; JUANATAS, I. C.; JUANATAS, R. A. Tiktok as a knowledge source for programming learners: a new form of nanolearning? In: IEEE. *2022 10th International Conference on Information and Education Technology (ICIET)*. [S.l.], 2022. p. 219–223. Citado na página 1.

HOSSEINI, R.; BRUSILOVSKY, P. A study of concept-based similarity approaches for recommending program examples. *New Review of Hypermedia and Multimedia*, Taylor & Francis, v. 23, n. 3, p. 161–188, 2017. Citado na página 8.

HOU, M. Social media celebrity and the institutionalization of youtube. *Convergence*, SAGE Publications Sage UK: London, England, v. 25, n. 3, p. 534–553, 2019. Citado na página 6.

ISOTANI, S.; BITTENCOURT, I. I. *Dados abertos conectados: em busca da web do conhecimento*. [S.l.]: Novatec Editora, 2015. Citado na página 6.

JANNAH, H. A.; HERMAWAN, D. Analysis of indonesian society's perceptions of the covid-19 vaccine in youtube comments using machine learning algorithms. In: IEEE. *2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*. [S.l.], 2022. p. 72–77. Citado na página 9.

JUNGES, D. de L. V.; GATTI, A. Estudando por vídeos: o youtube como ferramenta de aprendizagem. *Informática na educação: teoria & prática*, v. 22, n. 2, 2019. Disponível em: <<https://seer.ufrgs.br/index.php/InfEducTeoriaPratica/article/view/88586>>. Citado na página 2.

KADRIU, A. et al. Investigating trends in learning programming using youtube tutorials. *International Journal of Learning and Change*, Inderscience Publishers (IEL), v. 12, n. 2, p. 190–208, 2020. Citado na página 1.

KOKINDA, E.; RODEGHERO, P. Streaming software development: Accountability, community, and learning. *Journal of Systems and Software*, Elsevier, v. 199, p. 111630, 2023. Citado na página 10.

LAUSCH, A.; SCHMIDT, A.; TISCHENDORF, L. Data mining and linked open data—new perspectives for data analysis in environmental research. *Ecological Modelling*, Elsevier, v. 295, p. 5–17, 2015. Citado na página 6.

LAUSCH, A.; SCHMIDT, A.; TISCHENDORF, L. Data mining and linked open data – new perspectives for data analysis in environmental research. *Ecological Modelling*, v. 295, p. 5–17, 2015. ISSN 0304-3800. Use of ecological indicators in models. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0304380014004335>>. Citado na página 7.

LIN, P. Developing an intelligent tool for computer-assisted formulaic language learning from youtube videos. *ReCALL*, Cambridge University Press, v. 34, n. 2, p. 185–200, 2022. Citado na página 1.

LOUZADA, C. S. et al. Um mapeamento das publicações sobre o ingresso das mulheres na computação. In: *CLEI 2014: Conferência Latino-americana em Informática-VI Congresso da Mulher Latino-americana na Computação*. Montevideu. [S.l.: s.n.], 2014. p. 16. Citado na página 2.

LUCHANINOV, D. V. et al. It-based e-learning environment design in web engineering for students: on the experience of regional universities. In: IEEE. *2021 International Conference on Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS)*. [S.l.], 2021. p. 686–690. Citado na página 10.

MACLEOD, L.; STOREY, M.-A.; BERGEN, A. Code, camera, action: How software developers document and share program knowledge using youtube. In: *2015 IEEE 23rd International Conference on Program Comprehension*. [S.l.: s.n.], 2015. p. 104–114. Citado na página 2.

MARAZA-QUISPE, B. et al. Youtube assessment as a means of documentary research students. In: *Proceedings of the 11th International Conference on Education Technology and Computers*. [S.l.: s.n.], 2019. p. 316–321. Citado na página 8.

MAZIRIRI, E. T.; GAPA, P.; CHUCHU, T. Student perceptions towards the use of youtube as an educational tool for learning and tutorials. *International Journal of Instruction*, ERIC, v. 13, n. 2, p. 119–138, 2020. Citado na página 1.

MEDHAT, W.; HASSAN, A.; KORASHY, H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, v. 5, n. 4, p. 1093–1113, 2014.

- ISSN 2090-4479. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2090447914000550>>. Citado na página 7.
- MOURA, M. A. *O discurso do ódio em redes sociais*. [S.l.]: Lura Editorial (Lura Editoração Eletrônica LTDA-ME), 2016. Citado na página 3.
- MULLEN, T.; COLLIER, N. Sentiment analysis using support vector machines with diverse information sources. In: *Proceedings of the 2004 conference on empirical methods in natural language processing*. [S.l.: s.n.], 2004. p. 412–418. Citado na página 7.
- NADKARNI, P. M.; OHNO-MACHADO, L.; CHAPMAN, W. W. Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, BMJ Group BMA House, Tavistock Square, London, WC1H 9JR, v. 18, n. 5, p. 544–551, 2011. Citado na página 6.
- NISSENSON, P. M. et al. Creation of an online video tutorial library at a state university. In: *2019 ASEE Annual Conference & Exposition*. [S.l.: s.n.], 2019. Citado na página 9.
- OCHOA-HERNÁNDEZ, J. L.; ALEMÁN, Y. Teammx at politices 2022: Analysis of feature sets in spanish author profiling for political ideology. In: CEUR-WS. *CEUR Workshop Proceedings*. [S.l.], 2022. v. 3202. Citado na página 9.
- PAL, S. et al. A semi-automatic metadata extraction model and method for video-based e-learning contents. *Education and Information Technologies*, Springer, v. 24, n. 6, p. 3243–3268, 2019. Citado na página 6.
- PARK, S.; BIER, L. M.; PARK, H. W. The effects of infotainment on public reaction to north korea using hybrid text mining: Content analysis, machine learning-based sentiment analysis, and co-word analysis. *Profesional de la Información*, v. 30, n. 3, 2021. Citado na página 9.
- PEZZINI, A. Mineração de textos: conceito, processo e aplicações. *Revista Brasileira De Contabilidade E Gestão*, v. 5, n. 10, p. 58–61, 2017. Citado na página 6.
- PLATH, H. O. et al. Detecção de discurso de ódio contra mulheres em textos em português brasileiro: Construção da base mina-br e modelo de classificação. *Revista Eletrônica de Iniciação Científica em Computação*, v. 20, n. 3, 2022. Citado na página 7.
- POCHÉ, E. et al. Analyzing user comments on youtube coding tutorial videos. In: IEEE. *2017 IEEE/ACM 25th International Conference on Program Comprehension (ICPC)*. [S.l.], 2017. p. 196–206. Citado 2 vezes nas páginas 1 e 3.
- RAHMAN, S. A. E.; ALOTAIBI, F. A.; ALSHEHRI, W. A. Sentiment analysis of twitter data. In: IEEE. *2019 international conference on computer and information sciences (ICCIS)*. [S.l.], 2019. p. 1–4. Citado na página 7.
- REGUEIRA, U.; ALONSO-FERREIRO, A.; DA-VILA, S. Women on youtube: Representation and participation through the web scraping technique. *Comunicar, Comunicar*, v. 28, n. 63, p. 31–40, 2020. Citado na página 2.
- RIBEIRO, A.; SILVA, N. Inf-hateval at semeval-2019 task 5: Convolutional neural networks for hate speech detection against women and immigrants on twitter. In: *Proceedings of the 13th International Workshop on Semantic Evaluation*. [S.l.: s.n.], 2019. p. 420–425. Citado na página 8.

- ROSA, G. A. M. e.; SANTOS, B. R. d.; FALEIROS, V. d. P. Opacidade das fronteiras entre real e virtual na perspectiva dos usuários do facebook. *Psicologia USP*, Instituto de Psicologia da Universidade de São Paulo, v. 27, n. 2, p. 263–272, May 2016. ISSN 0103-6564. Disponível em: <<https://doi.org/10.1590/0103-656420130026>>. Citado na página 1.
- SABLIĆ, M.; MIROSAVLJEVIĆ, A.; ŠKUGOR, A. Video-based learning (vbl)—past, present and future: An overview of the research published from 2008 to 2019. *Technology, Knowledge and Learning*, Springer, v. 26, n. 4, p. 1061–1077, 2021. Citado na página 1.
- SANTANA, F. C. et al. Mulheres gamers: uma análise da hostilidade enfrentada pelas mulheres em jogos virtuais. *Cadernos de Gênero e Tecnologia*, v. 15, n. 46, p. 120–138, 2022. Citado na página 3.
- SANTOS, L. d. C.; SILVA, S. S. da; MEIRELES, S. P. de. Utilizando o youtube para ensino da linguagem c em cursos de engenharia. In: SBC. *Anais do XXVIII Workshop sobre Educação em Computação*. [S.l.], 2020. p. 156–160. Citado na página 8.
- SANTOS, R. E. et al. Técnicas de processamento de linguagem natural aplicadas ao processo de mineração de textos: resultados preliminares de um mapeamento sistemático. *Revista de Sistemas e Computação-RSC*, v. 4, n. 2, 2015. Citado na página 6.
- SANTOS, V. L.; CARVALHO, T. F.; BARRETO, M. do S. Mulheres na tecnologia da informação: Histórico e cenário atual nos cursos superiores. In: *Anais do XV Women in Information Technology*. [S.l.: s.n.], 2021. Citado na página 2.
- SARKAR, A. et al. End-user encounters with lambda abstraction in spreadsheets: Apollo’s bow or achilles’ heel? In: IEEE. *2022 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. [S.l.], 2022. p. 1–11. Citado na página 10.
- SBC. *Educação Superior em Computação: Estatísticas 2020*. 2023. <https://www.sbc.org.br/documentos-da-sbc/send/133-estatisticas/1420-educacao-superior-em-computacao-estatisticas-2020>. Citado na página 2.
- SCARPA, A. D. *Técnicas de processamento de linguagem natural aplicadas às Ciências Sociais*. Tese (Doutorado) — ”, 2017. Citado na página 6.
- SOUZA, A.; NAKAMURA, E.; NAKAMURA, F. Detecção de discurso de Ódio: Homofobia. In: *Anais do XVI Brazilian e-Science Workshop*. Porto Alegre, RS, Brasil: SBC, 2022. p. 73–80. ISSN 2763-8774. Disponível em: <<https://sol.sbc.org.br/index.php/bresci/article/view/20477>>. Citado na página 7.
- STEINBECK, H.; ZOBEL, T.; MEINEL, C. Using the youtube video style in a mooc:(re-) testing the effect of visual experience in a field-experiment. In: *Proceedings of the Ninth ACM Conference on Learning@ Scale*. [S.l.: s.n.], 2022. p. 142–150. Citado na página 9.
- TONTODIMAMMA, A. et al. Thirty years of research into hate speech: topics of interest and their evolution. *Scientometrics*, Springer, v. 126, p. 157–179, 2021. Citado na página 6.
- YIN, W. et al. Comparative study of cnn and rnn for natural language processing. *arXiv preprint arXiv:1702.01923*, 2017. Citado na página 7.