

# Crianças e Propagandas no TikTok: identificando publicidade infantil na rede social TikTok

Raíssa Gonçalves Lopes Carvalho  
Departamento de Ciência da Computação  
Pontifícia Universidade Católica de Minas Gerais  
Belo Horizonte, MG, Brasil  
raissa.carvalho.1266210@sga.pucminas.br

Humberto Torres Marques-Neto  
Departamento de Ciência da Computação  
Pontifícia Universidade Católica de Minas Gerais  
Belo Horizonte, MG, Brasil  
humberto@pucminas.br

## ABSTRACT

This study addresses the identification of advertising on TikTok and its impact on children's behavior, highlighting the growing presence of this platform in children's routines. Considering the subtle and often disguised nature of advertising aimed at children, the work focuses on identifying and analyzing all advertising done by influencers for children on this social network. Using a machine learning-based methodology, we built a database of influencer posts by automatically classifying advertising and non-advertising content aimed at children. We applied algorithms such as Naive Bayes and Support Vector Machine to classify videos that accurately address child advertising on TikTok. The results showed that the SVM model performed best, with high accuracy and F1-Score. The research also proposes future improvements, including context analysis and the expansion of the keyword base, to enhance the detection of child advertising and protect young users of the platform.

## KEYWORDS

Publicidade Infantil, Criança, TikTok, Rede Social Online, Mídia Social, Caracterização

## 1 INTRODUCTION

No cenário contemporâneo, as redes sociais se tornaram uma parte intrínseca da vida de uma parcela significativa da população, especialmente na vida dos jovens. Nesse cenário, as crianças se veem cada vez mais imersas em uma variedade de plataformas e aplicativos que desempenham um papel fundamental em sua socialização, entretenimento e educação, conforme discutido no estudo realizado por [26]. Atualmente, jogos e redes sociais se destacam como os tipos principais de aplicativos que atraem a atenção das crianças, ocupando uma parcela considerável de seu tempo. O artigo [23] indica que adolescentes passam em média 8½ horas por dia em telas, enquanto pré-adolescentes gastam cerca de 5½ horas diárias. Esse tempo é dedicado principalmente a redes sociais, jogos e mensagens de texto, excluindo atividades escolares. Seja jogando online, assistindo a vídeos no YouTube ou dedicando horas ao Instagram e TikTok, essas atividades se tornaram elementos centrais na vida cotidiana das crianças.

O TikTok, uma plataforma de compartilhamento de vídeos curtos, ganhou notoriedade devido à sua capacidade de envolver milhares de usuários em uma experiência de entretenimento singular,

tornando-se até mesmo uma referência para outros aplicativos. Sua popularidade influenciou as principais redes sociais, como Facebook, Instagram, Twitter, LinkedIn e YouTube, que passaram a introduzir novos recursos para acompanhar as tendências trazidas pelo TikTok, como apresentado por [20]. Essa experiência é marcada pela presença de vídeos criativos, trilhas sonoras envolventes e interações sociais que, quando combinados, mantêm o usuário profundamente imerso ao aplicativo, muitas vezes fazendo-o passar horas diante da tela. Uma pesquisa conduzida em 2022 por uma aplicação de controle parental (*software Qustodio* [24]) revela que jovens com idades entre 4 e 18 anos dedicavam aproximadamente 107 minutos diários ao TikTok, superando o YouTube.

Em 2023, o TikTok foi reconhecido como a marca de rede social mais valiosa globalmente, conforme destacado por uma publicação da *Forbes* [18]. Esta avaliação, baseada no ranking da consultoria *Brand Finance*, indica que o TikTok superou até mesmo o *Facebook*, o anterior líder de mercado. Este ascenso do TikTok reflete uma transformação significativa nos padrões de consumo de mídia, com uma tendência de crescimento contínuo desde o período da Covid-19 até o presente. A *Forbes* enfatiza a contribuição do TikTok para a mudança no cenário de mídia social, citando: "Com sua oferta de conteúdo facilmente digerível e divertido, a popularidade do aplicativo se espalhou pelo mundo. (...) ele também atuou como uma saída criativa e forneceu uma maneira de as pessoas se conectarem durante a pandemia."

Além de diversos conteúdos gerados por usuários comuns, os influenciadores desempenham um papel de considerável relevância. Os influenciadores são usuários que possuem contas consideravelmente populares entre todos, muitos desses têm como alvo principal crianças e adolescentes, utilizando suas plataformas e fama para promover produtos e serviços de maneira frequentemente sutil. Essas promoções abrangem uma ampla gama de itens, todos projetados com apelos especificamente para atrair este público. Essas parcerias possuem o potencial de exercer uma influência significativa nas decisões de compra das crianças, uma vez que essas figuras são frequentemente vistas como autoridades ou modelos a serem seguidos pelos jovens. Este fenômeno é evidenciado pelo crescimento na confiança que os jovens depositam nessas personalidades, os artigos [5] e [16], apresentam uma porcentagem significativa de compras sendo influenciada por suas recomendações, especialmente em mercados emergentes como o Brasil.

Ressalta-se que, no Brasil, a publicidade infantil é considerada abusiva e ilegal. Esta característica regulatória será detalhadamente explorada posteriormente. Nesse contexto, o artigo de [2] apresenta questionamentos e respostas válidas, nos quais mostram a possibilidade de antecipar que a probabilidade de um jovem que

passa grande parte do tempo diante da tela ser influenciado por um vídeo desse tipo é considerável. É nesse contexto que se insere o presente trabalho, buscando identificar e analisar a autoridade que os influenciadores detêm sobre as crianças no TikTok.

Neste trabalho, abordamos a identificação da publicidade no TikTok sobre o comportamento infantil por meio da implementação e aplicação de modelos de aprendizado de máquina, para a identificação de conteúdo publicitário direcionado a crianças. Criamos e pré-processamos um conjunto de dados que inclui vídeos e metadados extraídos da plataforma, complementado por análises de descrições, áudios e *hashtags*. Esse conjunto foi utilizado para treinar e testar a eficácia de diversos modelos de aprendizado de máquina na classificação de conteúdos publicitários. Os modelos aplicados incluem o *Naive Bayes* [19], *Support Vector Machine (SVM)* [10], *Logistic Regression* [27] e *Random Forest* [8].

Os principais resultados indicam a capacidade dos modelos em classificar com precisão os conteúdos publicitários, com o modelo SVM demonstrando superioridade em desempenho nas métricas de precisão e generalização. Essa superioridade indica uma capacidade significativa do modelo em distinguir entre conteúdos publicitários e não publicitários, ressaltando sua utilidade prática em contextos de proteção ao público infantil em plataformas digitais.

Este documento está organizado em seções que apresentam os trabalhos relacionados (Seção 2), detalham desde a metodologia (Seção 3), a implementação e o desenvolvimento dos modelos (Seção 4) até os resultados obtidos e discussões (Seção 5), proporcionando uma compreensão aprofundada das complexidades e desafios associados à publicidade infantil no TikTok.

## 2 TRABALHOS RELACIONADOS

O aumento da participação de crianças nas redes sociais tem ganhado destaque na atualidade e vários estudos têm sido conduzidos para compreender os possíveis impactos negativos a longo prazo para essa população. Esta seção apresenta as referências selecionadas para proporcionar uma contextualização mais profunda sobre o assunto.

No Brasil, a legislação classifica a publicidade voltada ao público infantil como abusiva. O Estatuto da Criança e do Adolescente – ECA (Lei nº 8.069/1990) [7], o Código de Defesa do Consumidor (Lei nº 8.078/1990) [6] e a Resolução nº 163 do Conselho Nacional dos Direitos da Criança e do Adolescente – Conanda [12], definem a publicidade que se dirige às crianças como considerada abusiva e ilegal, visto que explora a falta de discernimento e experiência dessa faixa etária. Complementando essa perspectiva, o Instituto Alana [1] apresentou uma discussão detalhada sobre a exploração comercial infantil na plataforma TikTok, destacando uma iniciativa do Instituto através do seu programa Criança e Consumo. Este programa notificou as empresas TikTok Inc. e Bytedance Brasil Tecnologia Ltda., exigindo providências quanto à presença de publicidade dirigida ao público infantil, em contraste com a legislação vigente e os termos de serviço da plataforma. O documento expõe as preocupações relacionadas aos impactos do consumismo na infância, destacando a influência prejudicial da publicidade voltada a esse público, que inclui riscos como a obesidade infantil, o materialismo excessivo e a degradação das relações sociais. A notificação

ênfata a necessidade de proteger as crianças dessas práticas de exploração comercial, realçando a urgência em adaptar as estratégias digitais para resguardar os direitos e o desenvolvimento saudável da criança e do adolescente. O texto analisado revelou conceitos fundamentais sobre a exploração comercial infantil na plataforma TikTok. Além de elucidar as estratégias de publicidade direcionadas ao público infantil, destacando a atuação significativa de alguns influenciadores digitais neste contexto.

Com a expansão das redes sociais, a publicidade vai além dos tradicionais comerciais de televisão, explorando novos territórios e ampliando o destaque para o trabalho dos influenciadores. O artigo intitulado “Social Media Influencers (SMIs) in Context: A Literature Review” [13] aborda de maneira abrangente o impacto dos influenciadores digitais nas redes sociais sobre adolescentes. Este estudo, uma revisão de literatura exaustiva sobre a interação entre jovens e SMIs, destaca como esses influenciadores moldam comportamentos e atitudes consumistas, apresentando as dinâmicas complexas entre adolescentes e a promoção de conteúdo nas mídias sociais. Os autores discutem os métodos de persuasão utilizados e os possíveis impactos positivos e negativos dessa influência, ressaltando a necessidade de conscientização dentro do campo de marketing sobre essas interações.

Ao mesmo tempo em que o impacto dos influenciadores pode ser visto de um ponto de vista negativo, o artigo de [17], por sua vez, apresenta o grande fenômeno TikTok, enfatizando sua rápida ascensão e sua revolução no consumo e compartilhamento de conteúdo. Para influenciadores se torna fácil a conexão com sua audiência de maneira mais orgânica e imersiva. O artigo apresenta que, as empresas aproveitam do potencial da plataforma, adaptando-se para amplificar suas campanhas de marketing e aproveitar sua natureza viral. No entanto, lutam com a rápida evolução da plataforma e as mudanças nas preferências do público também apresentam desafios, exigindo que as marcas estejam continuamente adaptando suas estratégias.

Ambos artigos de [13] e [17] trazem diferentes perspectivas, e, a partir delas, podemos entender o crescimento do TikTok, que, com sua influência crescente, é uma plataforma que não pode ser ignorada no marketing de influenciadores. A busca por um equilíbrio na plataforma deve ser feita, na qual as empresas devem compensar a exploração do potencial da plataforma com uma abordagem ética, especialmente ao engajar com públicos mais jovens e imaturos.

A influência das propagandas infantis não ocorre somente via TikTok. O artigo de [9] aborda a temática da publicidade infantil via Instagram e propõe um classificador baseado em aprendizado de máquina para identificar tais anúncios, aproximando-se do tema apresentado neste estudo. Os autores coletaram postagens de influenciadores brasileiros usando a ferramenta Instaloader, categorizando manualmente quais dessas postagens eram anúncios direcionados a crianças. Sete modelos de aprendizado de máquina foram testados para detectar a publicidade infantil nas redes sociais, sendo que o *Naive Bayes* se destacou como um dos mais eficazes. Concluiu-se que, dada a crescente prevalência de publicidade infantil nas mídias sociais, ferramentas como o classificador proposto são essenciais para monitorar e garantir a proteção dos direitos das crianças em ambientes digitais. Em paralelo a este tema, o trabalho de [14] investiga a eficácia da publicidade infantil presente em vídeos do YouTube. Utilizando um modelo para caracterizar se

um conteúdo é de fato um anúncio infantil no vídeo, os autores utilizaram personas (*bots*) que assistiam a vídeos e anúncios do YouTube. Os autores concluíram que os anúncios em vídeos são cuidadosamente elaborados, usando estratégias para captar e manter o interesse das crianças, impactando diretamente na evolução dessas crianças. Por fim, a pesquisa também enfatiza a necessidade de uma maior compreensão e regulamentação dessas práticas publicitárias, considerando a vulnerabilidade do público infantil e os potenciais impactos éticos e psicológicos.

A eficácia dos influenciadores é um tema frequentemente abordado em pesquisas. O trabalho [22] apresenta um estudo realizado pelos autores acerca dos influenciadores do TikTok. Os autores exploram como a credibilidade dos influenciadores afeta a intenção de compra dos usuários do TikTok. Para isso, baseiam-se em dados dos usuários e aplicam o PLS (*Partial Least Squares*) para análise de dados, buscando identificar correlações. O PLS é uma técnica estatística utilizada para modelagem preditiva que busca encontrar relações entre conjuntos de variáveis dependentes e independentes. Com resultados positivos, os autores observaram que fatores como a autenticidade e a expertise do influenciador desempenham um papel crucial na formação da percepção de credibilidade. Muito similar a este estudo, temos o trabalho de [25] que examina os fatores que determinam a atratividade dos influenciadores digitais. Utilizando uma abordagem quantitativa a partir do *framework* AHP, os pesquisadores conduziram um levantamento com consumidores para identificar os principais fatores que determinam a atratividade e confiabilidade em um influenciador. E, conforme levantado pelo trabalho de [22], autenticidade, expertise e a capacidade de estabelecer conexões com a audiência são elementos primordiais. As duas pesquisas enfatizam a necessidade de marcas contarem com influenciadores autênticos e de confiança ao mirar o engajamento e a conversão de sua audiência no TikTok, destacando a importância de escolher colaboradores que exibam esses atributos.

Para além das propagandas, as crianças estão constantemente expostas nas redes sociais. A pesquisa de [4] explora as percepções e os riscos associados a contatos com desconhecidos no ambiente virtual, focando especialmente em jovens e adolescentes. Os autores conduziram entrevistas com jovens procurando entender como ocorre o contato com estranhos online. Os resultados enfatizam que o contato é feito, apesar da consciência de muitos sobre os riscos, colocando-os em posições de vulnerabilidade. As crianças por fim, mostram interesse em aprender maneiras a fim de evitar tais contatos. Em paralelo ao tema deste trabalho, os influenciadores tem o poder de apresentar qualquer tipo de publicidade aos jovens, que muitas vezes não são supervisionadas pelos pais, aumentando a vulnerabilidade e risco de cada uma.

Diante dos desafios que as crianças enfrentam nas redes sociais, o artigo de [21] aborda a crescente preocupação dos pais em relação ao uso da internet por seus filhos e a demanda por sistemas de controle parental. Em sua análise, os autores apresentam um programa em Android, explorando as características essenciais desse sistemas e sua eficácia em monitorar as crianças. O estudo indica que, mesmo com alguns desafios, sua relevância já é reconhecida pelos pais. Ademais, a pesquisa sublinha a união de tecnologia e educação como um meio primordial para garantir a segurança online das crianças.

Os trabalhos relacionados apresentados fundamentam-se na análise compreensiva das dinâmicas de publicidade infantil nas redes sociais, destacando o papel dos influenciadores e a sutileza das estratégias de marketing dirigidas às crianças. Os estudos revisados apontam para uma crescente sofisticação na forma como o conteúdo publicitário é integrado às plataformas favoritas do público infantil, como o TikTok. Evidencia-se, assim, a necessidade de métodos que possam discernir entre entretenimento e publicidade, dada a complexidade e o impacto dessa publicidade no comportamento e desenvolvimento infantil. Neste contexto, o trabalho propõe uma abordagem metodológica utilizando algoritmos de aprendizado de máquina, que identifica conteúdos publicitários dirigidos a crianças. Este estudo, portanto, não só dialoga com as preocupações levantadas pelos autores citados, mas também avança na aplicação prática e técnica, contribuindo para o aprimoramento das estratégias de proteção à infância no ambiente digital.

### 3 METODOLOGIA

Para desenvolver a pesquisa proposta, foi adotada uma metodologia estruturada em etapas. A Figura 1 ilustra de maneira esquemática, cada uma dessas fases do trabalho. Os parágrafos subsequentes apresentam uma descrição detalhada de cada etapa, elucidando as técnicas e os procedimentos empregados ao longo da investigação. Vale destacar que, durante todo o desenvolvimento do trabalho, foi utilizada a ferramenta PyCharm, além de *Application Programming Interface* (APIs) e bibliotecas externas para auxiliar na implementação das análises e classificações necessárias.

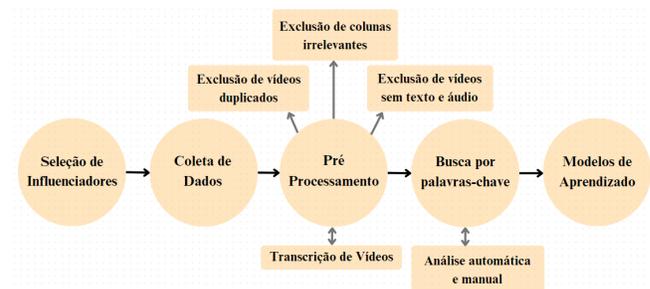


Figura 1: Metodologia Proposta.

#### 3.1 Coleta de Dados

O escopo deste trabalho concentra-se especificamente na análise de vídeos que apresentam anúncios protagonizados por crianças ou adultos, postados no TikTok.

Para a coleta de dados, inicialmente utilizou-se a mesma lista de influenciadores da plataforma Instagram utilizada em um trabalho prévio dos autores [9]. Posteriormente, expandiu-se a lista com influenciadores famosos no TikTok, identificados através de pesquisas na plataforma, focando em contas populares entre o público infantil e que produzem conteúdo voltado para esse público. Com esses influenciadores identificados, foi empregado, de forma complementar, uma ferramenta não oficial disponibilizada pelo site *Apify*, "*TikTok Data Extractor*". Esta plataforma foi desenvolvida com o objetivo de divulgar dados públicos da plataforma, oferecendo dados que podem ser empregados em pesquisas acadêmicas.

Com o apoio dessa API, uma base de dados foi estruturada para este estudo. Nessa base, os influenciadores fornecidos foram consultados com o objetivo de extrair os posts relacionados, permitindo uma análise detalhada do conteúdo anunciado. A base de dados é formada por 24.131 exemplos e inclui 328 características.

### 3.2 Pré-processamento

Com os posts obtidos na etapa anterior, aplicou-se um filtro para identificar apenas o conteúdo relevante para a pesquisa. Nessa etapa, buscou-se identificar posts que estavam relacionados à publicidade infantil. Além disso, foram excluídos vídeos repetidos, sem áudio e outros elementos textuais, bem como colunas desnecessárias para o escopo do trabalho. Ainda nessa etapa, foi necessária a adição das transcrições de cada vídeo.

A identificação de posts não relacionados à publicidade infantil foi realizada através da análise de elementos textuais extraídos dos vídeos publicados. Descrições, áudios e *hashtags* foram utilizados para caracterizar o conteúdo dos vídeos, e uma busca automática foi realizada com base em palavras fornecidas por autores de trabalhos previamente citados [9]. Se os elementos textuais de um vídeo contivessem uma das palavras coletadas, esse vídeo era marcado como publicidade.

Após a classificação automática dos vídeos, realizou-se uma análise manual final para identificar e corrigir possíveis classificações errôneas. Ao término da etapa, a tabela continha 24.118 exemplos e 32 características.

### 3.3 Classificação

Finalmente, com os dados pré-processados, foi implementado um modelo de aprendizado de máquina, buscando identificar as postagens com algum tipo de publicidade infantil. Utilizando técnicas de classificação, o modelo foi treinado para discernir entre conteúdos que são e que não são publicidade direcionada às crianças.

Dentre os algoritmos utilizados para essa tarefa, o *Naive Bayes* surgiu como uma opção viável devido à sua eficácia com grandes conjuntos de dados e facilidade de implementação. O SVM também foi um candidato forte, especialmente pela sua capacidade de lidar com a alta dimensionalidade típica dos dados de texto. O trabalho de [11] auxiliou na escolha e entendimento dos modelos.

O sucesso deste modelo dependeu da qualidade e quantidade do conjunto de dados de treinamento, que incluiu exemplos claros e diversificados de publicidade infantil e não-infantil. A escolha final do algoritmo foi influenciada por vários fatores, incluindo a disponibilidade de dados, a necessidade de interpretabilidade e os recursos computacionais disponíveis. Experimentos com vários algoritmos e ajustes de parâmetros foram realizados para identificar a abordagem mais eficaz para esta tarefa específica. Finalmente, a classificação do modelo foi incorporada à tabela criada na etapa anterior.

## 4 IMPLEMENTAÇÃO

A implementação deste trabalho foi feita seguindo os passos detalhados em seguida, buscando uma maneira eficiente para classificar conteúdo publicitário direcionado a crianças no TikTok. Ao longo da seção, foram realizadas descrições das ferramentas e tecnologias utilizadas, incluindo especificações sobre as bibliotecas de software

e plataformas de desenvolvimento. Inicialmente, esta seção começa com uma explanação detalhada do processo de coleta de dados, abordando as metodologias adotadas para extrair e processar os vídeos e metadados da plataforma. Posteriormente, discutimos o pré-processamento dos dados, destacando as técnicas de limpeza e organização de dados para torná-los adequados para análise. A etapa de classificação é explorada, apresentando os modelos testados, a configuração dos experimentos e a justificativa da escolha do modelo. Por fim, a seção conclui com uma avaliação dos resultados obtidos, descrevendo como os modelos performaram e quais informações foram extraídas dessas observações. Cada uma dessas etapas é fundamentada com exemplos e ilustrações, proporcionando uma compreensão das técnicas implementadas.

### 4.1 Coleta de Dados

Esta etapa do estudo envolve a aquisição de dados relevantes para análise. Houveram duas abordagens principais para a coleta desses dados: através da API oficial do TikTok ou utilizando uma ferramenta de extração de dados de terceiros.

**4.1.1 API Oficial do TikTok.** O TikTok oferece sua ferramenta oficial, o "*TikTok for Developers*", desenvolvida com o objetivo de fomentar a transparência. Especificamente neste trabalho, houve a tentativa de utilizar a API "*Research API*", disponibilizada nessa ferramenta, que permite acesso a informações detalhadas de contas específicas e dados de vídeos mediante solicitações específicas. No entanto, devido às restrições de acesso dessa API, limitadas a usuários universitários na América do Norte e Europa, foi feito um pedido de exceção para este projeto. Infelizmente, esse pedido foi negado. Como alternativa, a ferramenta *TikTok Data Extractor* foi identificada como uma opção viável para a coleta de dados.

**4.1.2 TikTok Data Extractor.** O *TikTok Data Extractor* é uma ferramenta de extração de dados não oficial, que extrai mídias e metadados associados a vídeos públicos no TikTok, disponibilizada pelo site *Apify* [3]. Essa ferramenta permite a pesquisa de criadores de conteúdo no TikTok e retorna informações detalhadas de seus vídeos.

Para a finalidade deste trabalho, realizou-se uma seleção criteriosa de influenciadores com base em pesquisas e em duas listas pré-definidas, abrangendo distintos grupos de influenciadores, categorizando-os em:

- **Influenciadores Mirins:** Seleção de contas de jovens, focadas em moda, brinquedos, jogos e entretenimento, incluindo desafios e brincadeiras. Houve a seleção de 17 influenciadores.
- **Criadores de conteúdo:** Contempla adolescentes e adultos com conteúdos que atraem crianças e jovens. O propósito é analisar se essas contas, apesar de não serem voltadas exclusivamente para crianças ou de não produzirem conteúdos puramente infantis, conseguem captar a atenção desse segmento jovem. Totalizando 25 influenciadores.

A aplicação do *TikTok Data Extractor* envolveu a inserção dos nomes dos influenciadores selecionados e a subsequente coleta de postagens. Os dados obtidos incluem o período de 2020 a 2023, cobrindo a era da Covid-19 e pós-pandemia. A escolha desse período específico é importante pois a pandemia da Covid-19 representou

um momento único de mudanças sociais e de comportamento, influenciando significativamente o uso das mídias sociais e a criação de conteúdo, como é apresentado em [15]. Espera-se que, ao analisar esse período, seja possível compreender melhor como essas mudanças afetaram as tendências de influência digital e o engajamento dos usuários, particularmente entre os públicos jovens.

Com base na busca dos influenciadores listados e no período especificado, foram coletados um total de 24.131 exemplos.

## 4.2 Pré-processamento

A partir da etapa anterior, foi gerada uma tabela extensa composta por 328 características, contendo dados altamente relevantes para a pesquisa. No entanto, a análise inicial da tabela revelou a presença de vídeos sem elementos textuais, colunas irrelevantes e a ausência de transcrições dos áudios dos vídeos. Portanto, a etapa de pré-processamento foi crucial neste trabalho.

No processo de pré-processamento dos dados coletados, utilizou-se o *PyCharm*, uma IDE (*Integrated Development Environment*) desenvolvida para *Python*. O *PyCharm* foi escolhido por suas amplas funcionalidades, como análise de código, suporte a várias bibliotecas e frameworks, e ferramentas para desenvolvimento profissional. Sua eficiência e facilidade de uso no tratamento de grandes conjuntos de dados foram decisivas para a escolha.

Inicialmente, a análise neste estudo concentrou-se principalmente nas descrições de texto e *hashtags* presentes nos vídeos fornecidos pela API. Contudo, decidiu-se também realizar a análise dos áudios dos vídeos para enriquecer os dados coletados. Assim, foi implementada uma etapa para a transcrição de cada vídeo. Além disso, foram excluídos da análise colunas irrelevantes e os vídeos que não continham descrição, *hashtags* ou áudio, bem como aqueles que apresentavam duplicidades. Esta seleção garante que apenas dados relevantes sejam considerados, aumentando a qualidade e a confiabilidade da análise.

A tabela resultante, composta por 32 características distintas e 24.118 exemplos, incluiu uma ampla gama de informações, detalhando aspectos como:

- **Vídeo:** Detalhes gerais sobre o vídeo, incluindo um identificador único, descrição ou texto associado, data e hora de criação, links diretos para o vídeo na plataforma, contagens de engajamento (curtidas, compartilhamentos, visualizações, comentários, e se o vídeo foi baixado), e menções a outros usuários e *hashtags* associadas;
- **Criador do vídeo:** Informações detalhadas sobre o autor do conteúdo, como identificador único, nome de usuário, apelido, status de verificação, descrição pessoal, link para o avatar do autor, e estatísticas de engajamento social (número de seguidores, curtidas recebidas, número total de vídeos postados, e total de aprovações);
- **Áudio do vídeo:** Informações sobre a música utilizada no conteúdo, incluindo nome da música, autor, se é uma música original ou não, nome do álbum, e um link direto para onde a música pode ser reproduzida;

Ainda na fase de pré-processamento, uma etapa crucial foi a distinção entre vídeos publicitários e não publicitários. Essa identificação foi conduzida por dois métodos complementares: uma busca automatizada, empregando palavras-chave específicas, e uma

investigação manual. Essa abordagem dupla assegurou uma seleção mais abrangente e precisa, resultando em um modelo preliminar de treinamento essencial para o processo subsequente de classificação.

**4.2.1 Transcrição de Vídeo.** Antes de iniciar a análise das palavras-chave nos elementos textuais dos vídeos, foi necessário converter os áudios de cada vídeo em transcrições textuais. Espera-se que a combinação das descrições dos vídeos com suas respectivas transcrições resulte em uma base de dados robusta e adequada para a pesquisa de palavras-chave. O processo de obtenção das transcrições envolveu três etapas principais: download do vídeo, transcrição do áudio e o armazenamento do resultado.

Primeiramente, os vídeos foram baixados utilizando uma biblioteca disponível no *Python*, que permite baixar vídeos e extrair informações de diversas plataformas, como YouTube, TikTok e Vimeo. Em seguida, cada arquivo de áudio foi submetido à transcrição utilizando a API *Speech-to-Text* do Google, que oferece reconhecimento de voz de alta precisão, suporta vários idiomas e dialetos, e possui recursos de filtragem de ruídos, o que ajuda a ignorar a música de fundo nos vídeos do TikTok.

Dado o volume substancial da base de dados, aproximadamente 24.000 entradas, foi essencial implementar um processo automatizado e eficiente em termos de recursos. Todo o procedimento foi executado através de comandos no *PyCharm*, utilizando também as ferramentas de *Cloud* do Google. Assim, para lidar com os arquivos de áudio gerados a partir dos vídeos, o *Google Cloud Storage* (GCS) foi implementado, devido à sua escalabilidade e integração perfeita com outras soluções do *Google Cloud*, facilitando o manejo de grandes volumes de dados de forma segura e acessível. O GCS foi empregado para armazenar temporariamente os arquivos de áudio antes da transcrição, o que possibilitou uma gestão eficiente dos dados e uma integração robusta com a API *Speech-to-Text*. *Buckets* no GCS foram utilizados para organizar e armazenar os arquivos de áudio extraídos, sendo cada arquivo acessado pela API para transcrição. As transcrições obtidas são então recuperadas e armazenadas no sistema para análises subsequentes.

Para o armazenamento das transcrições, foi desenvolvida uma tabela complementar dedicada exclusivamente a esse fim. Com o objetivo de otimizar o processo de manejo de dados, as transcrições foram processadas e inseridas na tabela em blocos de 1.000 registros. Esse método assegura que a tabela complementar seja continuamente atualizada de forma eficiente, mantendo a integridade e a organização dos dados.

O processo resultou na transcrição correta dos vídeos de diversos influenciadores, fornecendo uma base sólida para a pesquisa subsequente de palavras-chave nos textos transcritos.

**4.2.2 Busca automática.** A publicidade em conteúdos voltados para o público infantil tende a apresentar duas características principais: o emprego de palavras-chave que atraem a atenção das crianças e a utilização de certos elementos visuais e temáticos ao longo do vídeo. Neste estudo, a análise concentrou-se exclusivamente na identificação de termos que denotam apelo infantil nas descrições, nas transcrições e *hashtags* dos vídeos.

Para efetuar esta análise, empregou-se um conjunto pré-definido de palavras-chave, incluindo termos como 'brinquedo', 'parceria', 'Mattel', 'Estrela', 'l.o.l.', 'surpresa', 'presente', 'diversão', entre outros. As palavras foram fornecidas por trabalhos complementares

[9]. Além disso, para garantir a abrangência da pesquisa, foram utilizadas também as palavras traduzidas para o inglês, resultando na inclusão de 66 termos no processo de busca, aplicado às colunas de descrição e *hashtags* de cada vídeo.

Após a análise de 24.118 *posts*, constatou-se que 5.131 deles (aproximadamente 20%) apresentaram indicativos de conteúdo publicitário, conforme identificado pela metodologia empregada.

**4.2.3 Busca manual.** Durante o processo manual de verificação, foi analisada uma amostra de 1.200 vídeos previamente etiquetados como publicitários, representando aproximadamente 20% dessa base. O objetivo foi identificar e corrigir possíveis equívocos na categorização inicial. Para tanto, os vídeos foram assistidos individualmente e analisados com base nas seguintes questões:

- O influenciador promove a própria marca ou marca exterior no vídeo?
- O conteúdo publicitário possui linguagem ou tom infantil?
- Se há produtos infantis com marcas identificáveis, eles são o foco principal da publicação?

Ao revisar a amostra, verificou-se que 70% dos vídeos foram categorizados adequadamente como conteúdo publicitário. Observou-se que a hashtag #publi não era comumente empregada, o que serviria como indicativo de conteúdo promocional. No entanto, alguns influenciadores recorrem a termos estratégicos como "slime" e "Netflix" para promover produtos próprios ou conteúdos exclusivos. Em relação aos vídeos indevidamente marcados, muitos consistiam apenas em desafios ou entretenimento casual promovidos pelos criadores de conteúdo. Esta análise ressalta a complexidade na identificação de publicidade em plataformas digitais, onde as linhas entre o conteúdo promocional e o entretenimento podem ser tênues e frequentemente se sobrepõem.

Durante a busca manual, observou-se um fenômeno recorrente. Muitos influenciadores não indicam explicitamente em suas descrições e *hashtags* que um vídeo possa ser publicitário, deixando essa informação implícita no decorrer do vídeo. Nesse contexto, a etapa de transcrições se mostrou essencial. Um exemplo foi observado em um vídeo postado por um influenciador analisado. O vídeo apresenta uma criança muito feliz por ter "recebido" um brinquedo, sem deixar explícita a parceria. É necessário também observar os comentários da postagem, onde muitos jovens demonstram interesse no brinquedo, conforme ilustrado na Figura 2.

### 4.3 Modelo de Aprendizado Supervisionado

Com a base de dados final formada pelos elementos textuais necessários e a classificação inicial, foram implementados os modelos de classificação automática. No nosso estudo, investigamos o desempenho de quatro modelos de aprendizagem de máquina, sendo eles o *Naive Bayes*, *SVM*, *Logistic Regression* e *Random Forest*, para identificar vídeos com publicidade infantil na plataforma TikTok, marcando cada vídeo como publicidade ou não publicidade. A escolha dos modelos foi feita a partir de uma análise de suas vantagens, priorizando aqueles que se sobressaem na classificação de elementos textuais e no manejo de grandes volumes de dados.

Utilizamos o vetorizador TF-IDF para extrair características dos textos, uma abordagem padrão para transformar dados textuais em um formato que os modelos possam processar eficientemente.

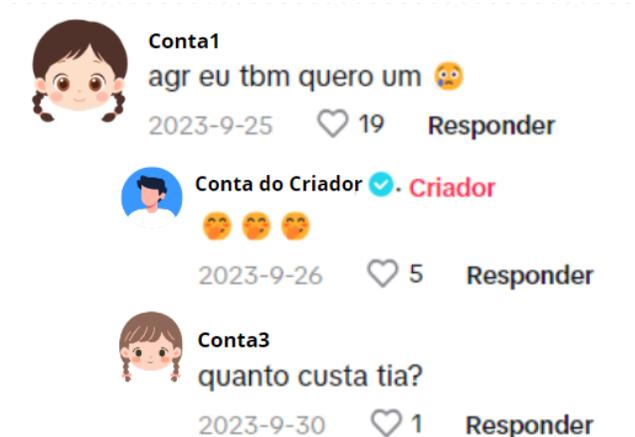


Figura 2: Comentário em uma possível publicidade infantil.

Adaptamos a implementação dos modelos baseados nas recomendações e práticas comuns na literatura de aprendizado de máquina, garantindo que cada modelo fosse configurado para maximizar sua eficácia. Para avaliar e comparar os modelos, empregamos uma técnica de validação cruzada, particionando os dados coletados em um conjunto de 80% dos dados para treinamento e 20% para teste, garantindo que o desempenho do modelo seja verificado contra dados não vistos durante o treinamento. O SVM foi o modelo que se destacou durante os testes. Durante o treinamento, utilizou-se o método *GridSearchCV* especificamente no modelo SVM para otimizar seus parâmetros, resultando em uma regularização com valor 100, um kernel linear e o coeficiente gamma definido como *scale*. A escolha desses parâmetros foi guiada por métricas de desempenho, visando maximizar a precisão e a generalidade do modelo. A partir dessa otimização final, o modelo se destacou em relação aos outros, sendo capaz de receber os elementos textuais de um vídeo e retornar sua classificação, sendo publicidade ou não.

As métricas escolhidas para avaliar o desempenho dos modelos foram Acurácia, Macro-F1 e Área sob a Curva ROC (AUC). Cada uma dessas métricas fornece uma visão importante sobre diferentes aspectos do desempenho do modelo:

- Acurácia mede quão frequentemente o modelo prediz corretamente, seja como publicidade ou não-publicidade.
- Macro-F1 é uma média do F1-Score calculado para cada classe e fornece uma medida de precisão e *recall* equilibrada, tratando igualmente as classes minoritárias e majoritárias.
- AUC proporciona uma visão abrangente da capacidade do modelo de distinguir entre as classes em diferentes limiares, o que é crucial para ajustar a sensibilidade do modelo a casos de publicidade infantil.

Essas métricas nos ajudam a entender não apenas a eficácia geral dos modelos, mas também como eles performam na identificação precisa e equilibrada de publicidade direcionada ao público infantil.

## 5 RESULTADOS

As métricas apresentadas na Tabela 1 forneceram uma visão compreensiva do desempenho de cada modelo. O modelo SVM, após

**Tabela 1: Desempenho dos Modelos de Classificação.**

Modelo	Acurácia	Macro-F1	AUC
Naive Bayes	0.84	0.68	0.83
SVM	0.97	0.95	0.99
Logistic Regression	0.92	0.87	0.98
Random Forest	0.95	0.93	0.99

o método *GridSearchCV*, se mostrou superior, apresentando uma acurácia e F1-Score particularmente altos, o que indica uma boa capacidade de generalização e balanceamento entre precisão e sensibilidade. Os outros modelos apresentaram uma boa performance, mas, não superior ao SVM, como pode ser observado nessa tabela.

Durante uma validação do modelo, a partir de uma base não classificada, observou-se que o modelo SVM apresentou precisão de 92%. Uma análise detalhada desses casos sugeriu que certas palavras no texto desses vídeos podem ter semelhanças com termos frequentemente encontrados em publicidades, levando a classificações errôneas. Um exemplo notável da limitação do classificador ocorre quando influenciadores narram histórias em seus vídeos. Frequentemente, eles mencionam eventos ou objetos sem intenção publicitária, no entanto, o classificador, guiado pela presença de palavras-chave específicas, pode erroneamente identificar tais menções como publicidade. Essa tendência de captar palavras-chave isoladas, independentemente do contexto narrativo em que são usadas, destaca uma área significativa para aprimoramento na sensibilidade e precisão do modelo.

A análise das palavras-chave utilizadas em conteúdos de publicidade infantil revelou uma frequência significativa de termos específicos. A Figura 3 apresenta uma *Word Cloud* que destaca as palavras mais frequentemente mencionadas nos dados analisados. Termos como 'desafio', 'brinquedo', 'publicidade', 'sbt', e 'ad' aparecem com destaque, indicando sua prevalência nas campanhas publicitárias direcionadas ao público infantil. Essa visualização ajuda a identificar os principais temas e tendências nas estratégias de marketing voltadas para crianças.



**Figura 3: Word Clouds**

A palavra "desafio" apresenta diferentes implicações no contexto do TikTok. Ela é amplamente utilizada tanto em vídeos classificados como publicidade quanto em vídeos não são. Esta alta frequência

reflete a popularidade dos desafios na plataforma, que frequentemente incluem danças e brincadeiras entre amigos. No entanto, muitos influenciadores também utilizam o termo para promover suas próprias marcas, criando tendências e engajamento entre seus seguidores. Essa dualidade no uso da palavra evidencia como um mesmo termo pode servir a propósitos diferentes, dependendo do contexto em que é empregado.

## 6 CONCLUSÃO

O presente estudo criou uma base de dados e implementou um modelo capaz de classificar publicidade infantil no TikTok, a partir dos seus elementos textuais, evidenciando o papel crescente desta plataforma nas atividades diárias do público infantil. O trabalho enfocou principalmente em identificar e classificar conteúdos publicitários direcionados a crianças, destacando a necessidade crítica de discernir e entender como esses anúncios se integram às rotinas delas. Utilizando uma base de dados, coletada e processada, foram aplicados modelos de aprendizado de máquina, incluindo *Naive Bayes*, *SVM*, *Logistic Regression* e *Random Forest*, para alcançar uma classificação precisa de vídeos no TikTok, distinguindo de maneira eficaz, entre conteúdos publicitários dirigidos ao público infantil e conteúdos não publicitários.

Além disso, a base de dados construída é de extrema importância, pois fornece um recurso para futuras pesquisas sobre publicidade infantil nas redes sociais. Ela não só permite a análise de padrões atuais, mas também oferece um ponto de partida para a implementação de modelos mais sofisticados e precisos. A base de dados pré-processada está disponível para uso acadêmico a todos os interessados. Sua distribuição será realizada sob demanda, mediante contato com os autores deste trabalho.

Os resultados mostraram que o modelo SVM, ajustado com otimização de parâmetros, obteve o melhor desempenho em termos de acurácia, F1-Score e AUC, mostrando-se eficaz na distinção entre conteúdos publicitários e não publicitários. No entanto, observamos limitações no classificador, especialmente em contextos onde influenciadores utilizam termos de maneira ambígua ou não explícita.

Além disso, vale ressaltar algumas limitações encontradas, como:

- **Acesso aos Dados:** A coleta de dados foi realizada por meio de uma ferramenta não oficial devido às restrições de acesso à API oficial do TikTok. Esta limitação pode ter impactado a abrangência e a representatividade dos dados coletados, uma vez que a ferramenta não oficial pode não capturar todos os dados disponíveis na plataforma.
- **Volume de Dados:** Embora a amostra utilizada seja significativa, com aproximadamente 24.000 exemplos analisados, o volume de dados ainda pode ser considerado limitado quando comparado ao total de conteúdos disponíveis no TikTok. Isso pode influenciar a variabilidade e a diversidade dos dados, afetando a robustez dos resultados.

Este trabalho visa oferecer uma metodologia validada com resultados que sirva de base para futuras pesquisas relacionadas ao impacto da publicidade digital nas crianças, em especial no TikTok. Além disso, o estudo pode fornecer a base para a criação de ferramentas de identificação automática de publicidade infantil, que podem ser utilizadas tanto pelos pais quanto pelo próprio TikTok para monitorar e proteger os jovens usuários da plataforma.

Em resumo, este estudo estabelece uma base para a identificação de publicidade infantil no TikTok, mas reconhece a necessidade de melhorias contínuas e inovações metodológicas. Durante a elaboração da base de dados, foram realizadas pesquisas sobre a ferramenta adequada e o conteúdo relevante do TikTok a ser analisado. Posteriormente, a identificação de palavras-chave facilitou a compreensão dos conceitos associados ao tema. Através das classificações automáticas e manuais, foi possível estabelecer uma conexão mais profunda com os vídeos analisados, proporcionando uma perspectiva mais clara sobre como a publicidade pode estar velada. Finalmente, a análise demonstrou que esta temática segue determinados padrões textuais, os quais podem ser efetivamente identificados e categorizados. A combinação de todas essas etapas resultaram em passos para desenvolver uma ferramenta de detecção de publicidade infantil, protegendo assim os jovens usuários de conteúdos potencialmente manipuladores.

Para aprimorar a precisão e a sensibilidade da classificação, trabalhos futuros podem focar em várias direções:

- **Análise de Contexto:** Desenvolver modelos que não se limitem a analisar palavras isoladas, mas que considerem todo o contexto em que as palavras são usadas. Isso pode incluir o uso de técnicas de processamento de linguagem natural (NLP) mais avançadas, como redes neurais recorrentes (RNNs) ou transformadores, que são capazes de capturar a sequência e o contexto das palavras.
- **Integração de Nomes de Empresas:** Expandir a base de dados de palavras-chave para incluir nomes de empresas e marcas conhecidas, o que pode ajudar a identificar conteúdos publicitários de maneira mais eficaz. Esta inclusão pode melhorar significativamente a detecção de publicidades implícitas e parcerias de marca.
- **Monitoramento Contínuo:** Implementar um sistema de monitoramento contínuo que atualize a base de dados com novos conteúdos e ajuste os modelos de classificação com base em padrões e tendências emergentes na plataforma.
- **Aprofundamento na Análise dos Resultados:** Concentrar esforços para aprimorar a análise dos resultados obtidos, incorporando uma abordagem detalhada e buscando padrões presentes nos resultados. Além disso, realizar uma análise das características com o objetivo de identificar sua relevância para o resultado da classificação, bem como possíveis correlações e padrões entre elas.

## REFERÊNCIAS

- [1] Instituto Alana. 2022. Notificação enviada ao TikTok sobre exploração comercial infantil na plataforma. <https://criancaconsumo.org.br/nossa-atuacao/atuacao-juridica/acoes-juridicas/tiktok-exploracao-comercial-infantil-no-aplicativo-junho-2022/>
- [2] Mohamed Amara. 2021. Navigating the Influencer Era: Unraveling the Impact on Children's Lifestyles and Well-being. <https://medium.com/@simohamed.amara/navigating-the-influencer-era-unraveling-the-impact-on-childrens-lifestyles-and-well-being-ed0a97dc34f2>
- [3] Apify. 2024. *Apify*. <https://apify.com/> Acessado em 20 de março de 2024.
- [4] Karla Badillo-Urquiola, Diva Smriti, Brenna McNally, Evan Golub, Elizabeth Bon-signore, and Pamela Jean Wisniewski. 2019. Stranger Danger! Social Media App Features Co-Designed with Children to Keep Them Safe Online. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (Boise, ID, USA) (IDC '19). Association for Computing Machinery, New York, NY, USA, 394–406. <https://doi.org/10.1145/3311927.3323133>
- [5] Lucas Berti. 2022. How social media influencers shape Brazilian consumption. <https://brazilian.report/liveblog/2022/06/06/social-media-influencers-consumption/>
- [6] Brasil. 1990. Código de Defesa do Consumidor. Lei nº 8.078, de 11 de setembro de 1990. Dispõe sobre a proteção do consumidor e dá outras providências..
- [7] Brasil. 1990. Estatuto da Criança e do Adolescente. Lei nº 8.069, de 13 de julho de 1990. Dispõe sobre o Estatuto da Criança e do Adolescente e dá outras providências..
- [8] Leo Breiman. 2001. Random Forests. *Machine Learning* 45 (2001), 5–32. <https://doi.org/10.1023/A:1010933404324>
- [9] Josemar Alves Caetano, Cecília Capurro Bouchardet, and Humberto Torres Marques-Neto. 2023. Identifying and Characterizing Child Advertising on Instagram. In *Proceedings of the 3rd International Workshop on Open Challenges in Online Social Networks* (Rome, Italy) (OASIS '23). Association for Computing Machinery, New York, NY, USA, 9–13. <https://doi.org/10.1145/3599696.3612896>
- [10] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 3, Article 27 (may 2011), 27 pages. <https://doi.org/10.1145/1961189.1961199>
- [11] KaiSiang Chong and Nathar Shah. 2022. Comparison of Naive Bayes and SVM Classification in Grid-Search Hyperparameter Tuned and Non-Hyperparameter Tuned Healthcare Stock Market Sentiment Analysis. *International Journal of Advanced Computer Science and Applications* 13, 12 (2022). <https://doi.org/10.14569/IJACSA.2022.0131213>
- [12] CONANDA. 2014. Resolução nº 163 do Conselho Nacional dos Direitos da Criança e do Adolescente. Dispõe sobre a abusividade do direcionamento de publicidade e de comunicação mercadológica à criança e ao adolescente..
- [13] Charles Alves de Castro, Isobel O'Reilly, and Aiden Carthy. 2021. Social Media Influencers (SMIs) in Context: A Literature Review. *Journal of Marketing Management* 9, 2 (Dec 2021), 59–71. <https://doi.org/10.15640/jmm.v9n2a9>
- [14] Flavio Figueiredo, Felipe Giori, Guilherme Soares, Mariana Arantes, Jussara Marques Almeida, and Fabricio Benevenuto. 2020. Understanding Targeted Video-Ads in Children's Content. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media* (Virtual Event, USA) (HT '20). Association for Computing Machinery, New York, NY, USA, 151–160. <https://doi.org/10.1145/3372923.3404787>
- [15] Caroline Forsey. 2022. How the Pandemic Changed Social Media: 6 Quarantine Trends Here to Stay [Data + Expert Insights]. <https://blog.hubspot.com/marketing/quarantine-trends-social-media>
- [16] Karla Garcia Gil. 2022. The power of influencers in purchasing decisions. <https://www.bbva.ch/en/news/the-power-of-influencers-in-purchasing-decisions/>
- [17] Teresa Guarda, Maria Fernanda Augusto, Jose Avelino Victor, Luis Miguel Mazon, Isabel Lopes, and Pedro Oliveira. 2021. The Impact of TikTok on Digital Marketing. In *Marketing and Smart Technologies*. Springer Singapore, Singapore, 35–44. [https://link.springer.com/chapter/10.1007/978-981-33-4183-8\\_4#citeas](https://link.springer.com/chapter/10.1007/978-981-33-4183-8_4#citeas)
- [18] Monique Lima. 2023. TikTok é a marca de rede social mais valiosa do mundo em 2023. <https://forbes.com.br/forbes-money/2023/01/tiktok-e-a-marca-de-rede-social-mais-valiosas-do-mundo-em-2023/>
- [19] Christopher David Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. Cambridge University Press. <https://nlp.stanford.edu/IR-book/html/htmledition/naive-bayes-text-classification-1.html>
- [20] Iliia Markov. 2022. The Influence of TikTok on the Big 5 Social Media Platforms. <https://www.brandwatch.com/blog/tiktok-influence-on-social-networks/>
- [21] Rafidah Md Noor, Sharifah Noor Sahila Syed Jamal, and Khairil Hafizze Zakaria. 2012. Parental mobile control system for children's internet use. In *International Conference on Information Society (i-Society 2012)*. 511–513. <https://ieeexplore.ieee.org/document/6285036>
- [22] Jasmine Andhini Rivai, Aimee Gian Haryanto, Kezia Tehila, and Adilla Anggraeni. 2023. The Impact of Influencer Credibility on Purchase Intention of TikTok Users. In *2023 11th International Conference on Information and Communication Technology (ICoICT)*. IEEE, 57–62. <https://doi.org/10.1109/ICoICT58202.2023.10262423>
- [23] David Rosenberg and Natalia Szura. 2023. How much time do kids spend on devices – playing games, watching videos, texting and using the phone? <https://theconversation.com/how-much-time-do-kids-spend-on-devices-playing-games-watching-videos-texting-and-using-the-phone-210118>
- [24] William Schendes. 2023. TikTok é a rede social que as crianças mais passam tempo diariamente. <https://olhardigital.com.br/2023/02/17/internet-redes-sociais/tiktok-e-a-rede-social-que-as-criancas-mais-passam-tempo-diariamente/#:~:text=Os%20dados%20v%C3%AAm%20do%20levantamento,de%20107%20minutos%20por%20dia.> Online; Acesso em 15-09-2023.
- [25] Ruhie Sethi and Shahzia Khan. 2023. Prioritization of Digital Influencers Attractive Influencing Consumer Purchase Intention Using AHP. In *2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*. 484–488. <https://doi.org/10.1109/ICCIKE58312.2023.10131125>
- [26] Arunkumar Sivakumar, Sudarsan Jayasingh, and Shahanaz Shaik. 2023. Social Media Influence on Students' Knowledge Sharing and Learning: An Empirical Study. *Education Sciences* 13, 7 (2023). <https://doi.org/10.3390/educ13070745>
- [27] Hsiang-Fu Yu, Fang-Lan Huang, and Chih-Jen Lin. 2011. Dual coordinate descent methods for logistic regression and maximum entropy models. *Mach Learn* 85 (2011), 41–75. <https://doi.org/10.1007/s10994-010-5221-8>