

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Pedro Henrique Gonçalves Teixeira

**Análise de sentimentos do *chat* da *Twitch*  
usando redes *Transformers***

**Uberlândia, Brasil**

**2025**

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Pedro Henrique Gonçalves Teixeira

**Análise de sentimentos do *chat* da *Twitch* usando redes  
*Transformers***

Trabalho de conclusão de curso apresentado  
à Faculdade de Computação da Universidade  
Federal de Uberlândia, como parte dos requi-  
sitos exigidos para a obtenção título de Ba-  
charel em Ciência da Computação.

Orientador: Fernanda Maria da Cunha Santos

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Ciência da Computação

Uberlândia, Brasil

2025

Pedro Henrique Gonçalves Teixeira

## **Análise de sentimentos do *chat* da *Twitch* usando redes *Transformers***

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, como parte dos requisitos exigidos para a obtenção título de Bacharel em Ciência da Computação.

Trabalho aprovado. Uberlândia, Brasil, 19 de dezembro de 2025:

---

**Fernanda Maria da Cunha Santos**  
Orientadora

---

**Maria Adriana Vidigal de Lima**  
Professora

---

**Mauricio Cunha Escarpinati**  
Professor

Uberlândia, Brasil  
2025

# Resumo

O crescimento das plataformas de *live streaming*, como a *Twitch*, consolidou um novo formato de comunicação em que criadores de conteúdo interagem em tempo real com grandes audiências por meio de chats de texto. Nesse cenário, compreender o sentimento coletivo do público torna-se uma tarefa desafiadora, tanto pelo alto volume de mensagens quanto pela rapidez com que elas são produzidas, dificultando uma análise manual por parte do *streamer*. Este trabalho tem como objetivo investigar e desenvolver uma ferramenta de apoio para análise de sentimentos no *chat* da *Twitch*, focada na categoria de jogos. Para isso, serão empregadas técnicas de Processamento de Linguagem Natural e modelos de classificação baseados na arquitetura *Transformer*, (i) XLM-RoBERTa, (ii) BERT Multilingual e (iii) um *ensemble* por regra de confiança entre ambos, comparando essas três configurações entre si em um conjunto de mensagens coletadas da plataforma. A ferramenta proposta foi capaz de classificar automaticamente as mensagens em polaridades (positivas, negativas e neutras), fornecendo uma visualização dinâmica do humor da audiência e auxiliando o *streamer* na compreensão das reações do público durante as transmissões. Nos experimentos, a ferramenta foi capaz de classificar as mensagens com acurácia de 46.2%.

**Palavras-chave:** Processamento de Linguagem Natural, Análise de Sentimentos, Redes *Transformers*, *Twitch*, *Live Streaming*, Jogos eletrônicos.

# **Lista de ilustrações**

Figura 1 – Visão geral do modelo XLM-RoBERTa. Fonte: Elaboração própria. . .	15
Figura 2 – Representação esquemática do modelo RoBERTa com o refinamento do BERT. Fonte: Elaboração própria. . . . .	16
Figura 3 – Esquematização dos algoritmos propostos na metodologia para classificação (RoBERTa-only, BERT-only e Ensemble por confiança). Fonte: Elaboração própria. . . . .	21
Figura 4 – Resultados obtidos pelos três classificadores: RoBERTa, BERT Multilingual e Ensemble. Fonte: Elaboração própria. . . . .	25

# Lista de abreviaturas e siglas

PLN	Processamento de Linguagem Natural
NLP	<i>Natural Language Processing</i>
API	<i>Application Programming Interface</i>
ML	<i>Machine Learning</i>
PLM	<i>Pre-trained Language Model</i>
BERT	<i>Bidirectional Encoder Representations for Transformers</i>
GPT	<i>Generative Pretrained Transformer</i>
RoBERTa	<i>Robustly Optimized BERT Pretraining Approach</i>
XLM-RoBERTa	<i>Cross-lingual RoBERTa</i>
CT-BERT	<i>COVID-Twitter BERT</i>
RNN	Rede Neural Recorrente
LSTM	<i>Long Short-Term Memory</i>
GRU	<i>Gated Recurrent Unit</i>
BiGRU	<i>Bidirectional Gated Recurrent Unit</i>
CNN	<i>Convolutional Neural Network</i>
KNN	<i>K-Nearest Neighbors</i>
RF	<i>Random Forest</i>
SVM	<i>Support Vector Machine</i>
SGD	<i>Stochastic Gradient Descent</i>
BOW	<i>Bag of Words</i>
TF-IDF	<i>Term Frequency–Inverse Document Frequency</i>
ROC	<i>Receiver Operating Characteristic</i>
AUC	<i>Area Under the ROC Curve</i>

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>7</b>
<b>1.1</b>	<b>Objetivo Geral</b>	<b>8</b>
<b>1.1.1</b>	<b>Objetivos Específicos</b>	<b>8</b>
<b>1.2</b>	<b>Justificativa</b>	<b>9</b>
<b>1.3</b>	<b>Estrutura do Trabalho</b>	<b>9</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>11</b>
<b>2.1</b>	<b>Processamento de Linguagem Natural</b>	<b>11</b>
<b>2.2</b>	<b>Análise de Sentimentos</b>	<b>12</b>
<b>2.3</b>	<b>Redes Neurais e os Modelos <i>Transformers</i></b>	<b>13</b>
<b>2.4</b>	<b>Modelos de Classificação da Arquitetura <i>Transformer</i></b>	<b>14</b>
<b>2.4.1</b>	<b>Modelo Principal: XLM-RoBERTa</b>	<b>14</b>
<b>2.4.2</b>	<b>Modelo Secundário: BERT Multilingual</b>	<b>15</b>
<b>3</b>	<b>TRABALHOS RELACIONADOS</b>	<b>17</b>
<b>3.1</b>	<b>Análise de Sentimentos em Redes Sociais</b>	<b>17</b>
<b>3.2</b>	<b>Classificação de Texto com Redes Neurais <i>Transformers</i></b>	<b>18</b>
<b>3.3</b>	<b>Síntese e Posicionamento Deste Trabalho</b>	<b>19</b>
<b>4</b>	<b>METODOLOGIA</b>	<b>21</b>
<b>4.1</b>	<b>Etapa 1: Coleta e Organização dos Dados</b>	<b>22</b>
<b>4.2</b>	<b>Etapa 2: Ambiente e Bibliotecas em Python</b>	<b>22</b>
<b>4.3</b>	<b>Etapa 3: Pré-processamento e Limpeza Textual</b>	<b>22</b>
<b>4.4</b>	<b>Etapa 4: As Redes Neurais <i>Transformers</i></b>	<b>23</b>
<b>4.5</b>	<b>Etapa 5: Os Três Pipelines de Inferência</b>	<b>23</b>
<b>4.6</b>	<b>Etapa 6: Visualização dos Resultados</b>	<b>24</b>
<b>5</b>	<b>RESULTADOS</b>	<b>25</b>
<b>6</b>	<b>CONCLUSÃO</b>	<b>28</b>
	<b>REFERÊNCIAS</b>	<b>30</b>

# 1 Introdução

Com o avanço das tecnologias digitais e o crescimento das plataformas de entretenimento online, surgiram novas formas de interação social mediadas pela internet. Entre essas formas, destacam-se as transmissões ao vivo em plataformas como o *YouTube* e a *Twitch*, nas quais criadores de conteúdo, conhecidos como *streamers*, interagem em tempo real com seus espectadores. As plataformas de *live streaming* já movimentam um mercado global estimado em mais de 87 bilhões de dólares em 2023, com previsão de forte crescimento nos próximos anos ([Grand View Research, 2024](#)), e a *Twitch*, especificamente, registra em média cerca de 2,4 milhões de espectadores simultâneos em 2023, consolidando-se como uma das principais plataformas de transmissão ao vivo voltadas para jogos digitais ([StreamScheme, 2023](#)). Esse modelo de comunicação tem transformado o cenário da mídia e do entretenimento, permitindo que o público participe ativamente das transmissões e influencie, por meio do chat, a dinâmica dos programas ao vivo.

Entretanto, diferentemente das interações presenciais, nas quais expressões faciais, gestos e entonação de voz auxiliam na interpretação emocional, os chats dessas transmissões se baseiam predominantemente em mensagens de texto. Essa limitação torna mais difícil compreender o sentimento coletivo da audiência, uma vez que o texto, isolado dos sinais não verbais, não transmite com clareza o contexto emocional que os participantes desejam expressar. Tal dificuldade é ainda mais evidente na categoria de **jogos**, em que emoções como empolgação, frustração, tensão e euforia surgem de maneira espontânea e intensa no chat, especialmente em momentos decisivos de partidas competitivas ou eventos sazonais.

Nesse contexto, o uso de técnicas de PLN e Análise de Sentimentos mostra-se uma estratégia promissora para interpretar automaticamente as mensagens enviadas pelos espectadores. O PLN, enquanto subárea da Inteligência Artificial, busca permitir que sistemas computacionais processem, representem e interpretem a linguagem humana em suas diversas formas ([GUPTA, 2014](#)). Já a Análise de Sentimentos é uma aplicação específica de PLN voltada à identificação automática de opiniões e emoções em textos, frequentemente por meio da classificação da polaridade das mensagens como **positivas, neutras ou negativas** ([HASAN; MALIHA; ARIFUZZAMAN, 2019](#)). A aplicação conjunta dessas técnicas possibilita tratar grandes volumes de mensagens em tempo real, extraíndo indicadores que auxiliam na compreensão do humor coletivo do público.

Dessa forma, este trabalho propõe o desenvolvimento de uma ferramenta computacional de apoio para *streamers* da categoria de jogos na plataforma *Twitch*, capaz de classificar automaticamente os sentimentos expressos nas mensagens do chat. A ferra-

menta busca oferecer uma análise dinâmica e visual do humor da audiência, permitindo que o criador de conteúdo acompanhe o engajamento e a percepção emocional do público em tempo quase real e, a partir dessas informações, ajuste sua postura, estratégias de interação e decisões durante a transmissão.

## 1.1 Objetivo Geral

O objetivo geral deste estudo é desenvolver um modelo computacional capaz de identificar e classificar automaticamente os sentimentos predominantes nas mensagens enviadas pelos participantes durante as transmissões ao vivo da plataforma *Twitch*, na categoria de jogos. Assim, o modelo utiliza técnicas de Processamento de Linguagem Natural (PLN) para realizar a seleção e formatação dos textos das mensagens, e as redes neurais *Transformers* farão a classificação destas, com o intuito de gerar indicadores que auxiliem os *streamers* na compreensão das emoções do público.

### 1.1.1 Objetivos Específicos

Os objetivos específicos do trabalho podem ser descritos como:

- **Coleta de Dados:** extrair e armazenar mensagens provenientes de transmissões ao vivo na plataforma *Twitch*, com foco na categoria de jogos;
- **Pré-Processamento:** aplicar técnicas de limpeza e normalização textual para remover ruídos, emojis, links e símbolos, preparando os dados para análise;
- **Classificação de Sentimentos:** empregar modelos pré-treinados de redes neurais *Transformers* (XLM-RoBERTa e BERT Multilingual) para classificar as mensagens em categorias positivas, neutras ou negativas;
- **Refinamento dos Resultados:** realizar uma segunda etapa de classificação com um modelo complementar, a fim de aumentar a confiabilidade das previsões;
- **Visualização dos Sentimentos:** gerar gráficos que representem a distribuição e a variação dos sentimentos do público ao longo da transmissão, possibilitando a interpretação visual e intuitiva dos resultados;
- **Proposição de Ferramenta:** apresentar uma ferramenta funcional capaz de auxiliar o *streamer* na leitura emocional do *chat* em tempo real, servindo como apoio ao gerenciamento de interações durante transmissões.

As ferramentas computacionais utilizadas para o desenvolvimento do trabalho incluem:

- Linguagem de programação **Python**;
- Bibliotecas Python voltadas para **Processamento de Linguagem Natural** (*Transformers, Torch, Pandas*);
- Bibliotecas Python para **visualização de dados** (*Matplotlib e Seaborn*);
- Base de dados composta por ***mensagens de chats da Twitch***.

## 1.2 Justificativa

As interações em chats de transmissões ao vivo representam uma forma moderna e dinâmica de comunicação digital, mas carecem de mecanismos que permitam interpretar o tom emocional das mensagens. A categoria de jogos, em especial, apresenta alto volume de mensagens curtas e informais, frequentemente marcadas por gírias e expressões emocionais típicas de comunidades online. Isso torna inviável a interpretação manual do sentimento coletivo em tempo real.

Nesse cenário, o uso de técnicas de PLN juntamente com as redes neurais *Transformers* treinadas para realizar a análise de sentimentos, se mostra como uma alternativa promissora. Além de ampliar o entendimento sobre o comportamento do público, tais ferramentas podem contribuir para a melhoria da interação entre *streamer* e audiência, oferecendo suporte para decisões imediatas durante a transmissão, como ajustar o tom da conversa, moderar o *chat* ou promover engajamento positivo.

Assim, este trabalho busca aliar os avanços tecnológicos da inteligência artificial com as necessidades práticas do ambiente de *streaming*, oferecendo uma aplicação concreta que conecta a pesquisa acadêmica à realidade das plataformas digitais.

## 1.3 Estrutura do Trabalho

Além deste capítulo, este trabalho está organizado em mais cinco capítulos. O Capítulo 2 apresenta a fundamentação teórica necessária para o desenvolvimento do projeto, abordando conceitos de Processamento de Linguagem Natural, Análise de Sentimentos e a arquitetura neural *Transformer*.

O Capítulo 3 descreve os principais trabalhos relacionados ao tema, destacando pesquisas que aplicam técnicas de PLN e modelos baseados em *Transformers* para análise de sentimentos em redes sociais, bem como estudos voltados especificamente para ambientes de *streaming* e interação em tempo real.

No Capítulo 4 é apresentada a metodologia adotada no desenvolvimento da ferramenta computacional proposto neste trabalho, incluindo desde o processo de coleta das

mensagens do *chat*, os procedimentos de pré-processamento dos dados, a construção dos conjuntos de treino e teste, a configuração dos modelos e as métricas empregadas para avaliação do desempenho.

## 2 Fundamentação Teórica

### 2.1 Processamento de Linguagem Natural

O Processamento de Linguagem Natural (PLN) é uma área de pesquisa que estuda como permitir que computadores compreendam, representem e manipulem a linguagem humana de forma semelhante à maneira como as pessoas se comunicam. De forma geral, o PLN busca desenvolver sistemas capazes de tratar textos ou fala em linguagem natural e realizar tarefas como classificação, extração de informação, sumarização e geração de texto (GUPTA, 2014). Entre suas aplicações práticas podem ser citadas a tradução automática, a análise de sentimentos, sistemas de perguntas e respostas, *chatbots* e filtros de conteúdo em redes sociais (HASAN; MALIHA; ARIFUZZAMAN, 2019; TUNSTALL; WERRA; WOLF, 2022).

Historicamente, o PLN surgiu nos anos 1950 em um contexto de convergência entre linguística e inteligência artificial, tendo como uma de suas primeiras motivações a tradução automática entre idiomas. As abordagens iniciais eram fortemente baseadas em regras e gramáticas formais, exigindo um grande esforço manual de especialistas. Com o passar do tempo, técnicas estatísticas e, posteriormente, métodos de aprendizado de máquina passaram a dominar a área, culminando no uso de redes neurais profundas e modelos de linguagem pré-treinados na última década (GUPTA, 2014; TUNSTALL; WERRA; WOLF, 2022).

Em geral, projetos de PLN seguem um fluxo composto por etapas como coleta de dados, pré-processamento, representação dos textos em uma forma numérica e aplicação de um modelo de aprendizado de máquina ou rede neural para realizar a tarefa desejada.

Na etapa de pré-processamento, o objetivo é reduzir ruídos do texto e padronizar a entrada para facilitar o trabalho do modelo. Entre as técnicas tradicionalmente empregadas estão a remoção de pontuações e *stopwords*, a normalização de caixa (conversão para minúsculas), a derivação (*stemming*) e a lematização, que busca agrupar variações morfológicas de uma palavra em um mesmo lema (KHYANI; S, 2021). Em seguida, é necessário transformar o texto em uma representação numérica adequada. Técnicas como *Bag-of-Words*, *Term Frequency–Inverse Document Frequency* (TF-IDF) e, mais recentemente, *word embeddings* contínuos, como *Word2Vec*, *GloVe* e representações baseadas em modelos *Transformers*, permitem representar palavras e sentenças em espaços vetoriais de forma a capturar relações semânticas e contextuais (MIKOLOV et al., 2013; TUNSTALL; WERRA; WOLF, 2022; GILLIOZ et al., 2020).

No contexto deste trabalho, o PLN é empregado para tratar mensagens de texto

curtas e informais presentes no *chat* da *Twitch*. Esse tipo de dado apresenta particularidades, como uso intenso de gírias, abreviações, emojis, onomatopeias e mistura de idiomas, que tornam a tarefa mais desafiadora. Assim, foram consideradas técnicas de pré-processamento específicas, como remoção de URLs, menções a usuários, normalização de repetições de caracteres e substituição de certas sequências de emojis ou emotes por marcadores textuais simples, além de uma etapa de tokenização adequada aos modelos baseados em *Transformers*. As etapas detalhadas de pré-processamento adotadas neste trabalho são descritas no Capítulo 4.

## 2.2 Análise de Sentimentos

A Análise de Sentimentos é uma aplicação específica de PLN voltada à identificação automática de opiniões, emoções ou atitudes expressas em textos, buscando, em geral, classificar o conteúdo em categorias como positivo, negativo ou neutro (HASAN; MALIHA; ARIFUZZAMAN, 2019). Em redes sociais, fóruns *online* e plataformas de comunicação em tempo real, permitem compreender de forma agregada a percepção do público sobre produtos, eventos, marcas ou, como no caso deste trabalho, transmissões de jogos digitais.

Segundo Hasan, Maliha e Arifuzzaman (2019), tarefas de análise de sentimentos podem ser formuladas em diferentes níveis de granularidade, como nível de documento, sentença ou até mesmo aspecto (focada em partes específicas de um produto ou serviço). Independentemente do nível, a abordagem tradicional envolve um conjunto de etapas que vão desde a coleta e preparação dos dados até a classificação supervisionada por meio de modelos de aprendizado de máquina ou redes neurais (TUNSTALL; WERRA; WOLF, 2022).

Em linhas gerais, o processo inicia-se com a coleta e organização das mensagens que compõem o conjunto de dados. Em seguida, são aplicadas técnicas de pré-processamento, como remoção de ruídos (URLs, menções, *hashtags* irrelevantes), normalização de texto, tratamento de emojis e *emoticons*, remoção de *stopwords* e, em alguns casos, lematização ou derivação (KHYANI; S, 2021; TUNSTALL; WERRA; WOLF, 2022). Após essa etapa, os textos são transformados em representações numéricas por meio de técnicas de vetorialização, que podem variar desde representações esparsas (como *Bag-of-Words* e TF-IDF) até *embeddings* densos produzidos por modelos pré-treinados (MIKOLOV et al., 2013; GILLIOZ et al., 2020).

A etapa seguinte consiste na aplicação de um classificador supervisionado, que pode envolver desde algoritmos tradicionais de aprendizado de máquina, como Máquinas de Vetores de Suporte, *Random Forest* e regressão logística, até redes neurais profundas e modelos baseados em *Transformers* (GILLIOZ et al., 2020; TUNSTALL; WERRA;

WOLF, 2022). Os resultados obtidos são, então, avaliados por meio de métricas como acurácia, precisão, revocação e *F1-score*, além de eventualmente serem agregados para fornecer uma visão temporal ou global do humor do público.

Quando aplicada a contextos dinâmicos, como chats de transmissões ao vivo, a análise de sentimentos enfrenta desafios adicionais. Além da informalidade da linguagem e da alta presença de sarcasmo, ironia e gírias, a rapidez com que as mensagens são produzidas exige métodos capazes de processar grandes volumes de dados em tempo quase real. Trabalhos anteriores já exploraram o uso de técnicas de análise de sentimentos em redes sociais como o *Twitter*, mostrando que modelos modernos de PLN conseguem capturar, com boa precisão, a polaridade emocional de textos curtos (HASAN; MALIHA; ARIFUZZAMAN, 2019; TUNSTALL; WERRA; WOLF, 2022). Este trabalho se apoia nessas abordagens para investigar a viabilidade de aplicar métodos semelhantes ao *chat* da *Twitch*.

## 2.3 Redes Neurais e os Modelos *Transformers*

Redes neurais profundas têm sido amplamente utilizadas em tarefas de PLN, em especial arquiteturas recorrentes, como LSTMs e redes convolucionais (TUNSTALL; WERRA; WOLF, 2022). No entanto, essas arquiteturas apresentam limitações para lidar com dependências de longo alcance em sequências textuais e dificultam o paralelismo durante o treinamento.

Em 2017, Vaswani et al. (2017) propuseram a arquitetura *Transformer*, baseada em mecanismos de atenção e autoatenção (*self-attention*), dispensando o uso de recorrência. Nessa arquitetura, cada *token* de uma sentença pode “prestar atenção” a todos os demais, aprendendo quais posições do texto são mais relevantes para a representação atual. Isso permite capturar o contexto global da frase de maneira mais eficiente e paralelizável (VASWANI et al., 2017; GILLIOZ et al., 2020).

Com base nessa arquitetura, diversos modelos de linguagem pré-treinados foram desenvolvidos, como BERT, GPT e suas variações, que se tornaram padrão de fato em tarefas modernas de PLN (GILLIOZ et al., 2020; TUNSTALL; WERRA; WOLF, 2022). Em ambientes práticos, esses modelos são disponibilizados em bibliotecas especializadas, como a *Transformers*, da *HuggingFace*, que fornece implementações prontas de modelos para diferentes tarefas, incluindo classificação de sentenças, análise de sentimentos e geração de texto (Hugging Face, 2024). Nesta biblioteca encontram-se, por exemplo, os modelos XLM-RoBERTa e BERT Multilingual utilizados neste trabalho, bem como ferramentas de tokenização e *pipelines* de inferência que facilitam sua integração em aplicações reais.

A partir da arquitetura *Transformer*, diversos modelos de linguagem pré-treinados (*Pre-trained Language Models* — PLMs) foram desenvolvidos, como BERT (*Bidirectional*

*Encoder Representations from Transformers*), GPT (*Generative Pre-trained Transformer*) e suas variações (GILLIOZ et al., 2020; RADFORD et al., 2018; TUNSTALL; WERRA; WOLF, 2022). Esses modelos são inicialmente treinados em grandes coleções de textos não rotulados, aprendendo representações ricas da linguagem, e posteriormente podem ser ajustados (*fine-tuned*) para tarefas específicas, como classificação de sentimentos, resposta a perguntas e sumarização (GILLIOZ et al., 2020; TUNSTALL; WERRA; WOLF, 2022).

O uso de PLMs trouxe ganhos significativos de desempenho em tarefas de PLN, especialmente em cenários com dados rotulados limitados. Em vez de treinar um modelo do zero, aproveita-se o conhecimento linguístico adquirido no pré-treinamento, ajustando apenas as últimas camadas para a tarefa de interesse. Essa abordagem é particularmente útil em contextos como o deste trabalho, em que o conjunto de mensagens de *chat* rotuladas é relativamente pequeno se comparado a grandes corpora de domínio geral.

## 2.4 Modelos de Classificação da Arquitetura *Transformer*

Nesta seção são apresentados os principais modelos utilizados neste trabalho para a tarefa de classificação de sentimentos no *chat* da *Twitch*. Ambos se baseiam na arquitetura *Transformer* e são disponibilizados como modelos pré-treinados para que, posteriormente, sejam ajustados para a tarefa específica.

### 2.4.1 Modelo Principal: XLM-RoBERTa

O modelo principal empregado é uma variação do XLM-RoBERTa, modelo multilíngue baseado em BERT projetado para lidar com mais de 100 idiomas. Em particular, foi utilizado o modelo *cardiffnlp/twitter-xlm-roberta-base-sentiment*, treinado originalmente em textos de redes sociais, como o *Twitter*, e adaptado para análise de sentimentos. Esse modelo utiliza tokenização em subpalavras e representações contextuais, o que o torna adequado para lidar com textos curtos, informais e com mistura de idiomas, características presentes nas mensagens do *chat* da *Twitch* (GILLIOZ et al., 2020; TUNSTALL; WERRA; WOLF, 2022).

No *pipeline* deste trabalho, o XLM-RoBERTa recebe como entrada as mensagens já pré-processadas e tokenizadas, produzindo como saída três escores de probabilidade associados às classes **positivo**, **neutro** e **negativo**. Após a aplicação da função *softmax*, a classe com maior probabilidade é escolhida como rótulo previsto, enquanto o valor dessa probabilidade é interpretado como uma medida de confiança da predição. O fluxo de inferência descrito pode ser observado na Figura 1.

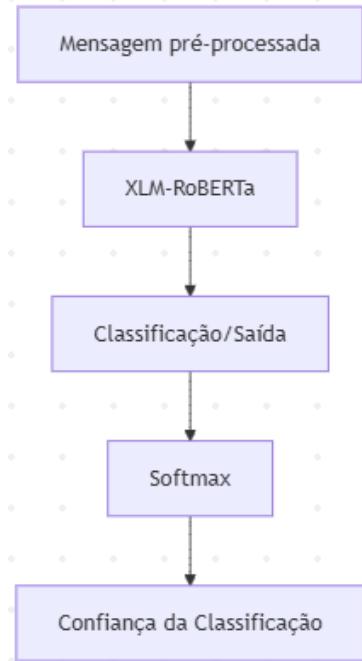


Figura 1 – Visão geral do modelo XLM-RoBERTa. Fonte: Elaboração própria.

#### 2.4.2 Modelo Secundário: BERT Multilingual

Como modelo secundário, foi adotado o *nlptown/bert-base-multilingual-uncased-sentiment*, uma variação multilíngue de BERT treinada originalmente para classificação de sentimentos em escala ordinal (por exemplo, de 1 a 5 estrelas). Neste trabalho, suas saídas foram reagrupadas de forma a representar as três categorias de sentimento de interesse: positivo, neutro e negativo.

O modelo secundário atua como um refinador das previsões do modelo principal. Sempre que a confiança do XLM-RoBERTa for inferior a um limiar pré-definido (por exemplo, 70%), a mensagem é reenviada ao BERT Multilingual para obtenção de uma segunda opinião. O esquema descrito pode ser observado na Figura 2. Essa esquematização mostra a interação entre os dois modelos.

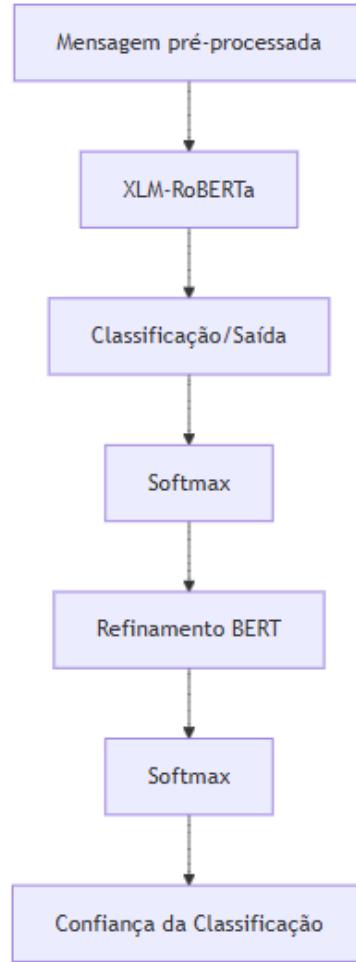


Figura 2 – Representação esquemática do modelo RoBERTa com o refinamento do BERT.  
Fonte: Elaboração própria.

Essa abordagem com duas redes *Transformers* aproxima-se de um sistema de *ensemble*, no qual a decisão final se beneficia de mais de um modelo. A utilização de um modelo secundário treinado em um contexto ligeiramente diferente fornece maior robustez às previsões, especialmente em casos ambíguos ou em mensagens cujo conteúdo se afasta do domínio original de treinamento do modelo principal.

### 3 Trabalhos Relacionados

Este capítulo apresenta trabalhos que dialogam com a proposta deste estudo, tanto no que diz respeito à análise de sentimentos em textos de redes sociais quanto ao uso de modelos de aprendizado de máquina para tarefas de classificação de texto. São destacadas, em especial, pesquisas que tratam em sua base de dados de textos curtos, informais e produzidos em ambientes *online*, cenário próximo ao encontrado nos *chats* de transmissões ao vivo na *Twitch*.

#### 3.1 Análise de Sentimentos em Redes Sociais

Diversos trabalhos na literatura exploram a aplicação de técnicas de PLN para análise de sentimentos em redes sociais. Em [Hasan, Maliha e Arifuzzaman \(2019\)](#), por exemplo, os autores investigam a análise de sentimentos em dados do *Twitter* utilizando técnicas de PLN e algoritmos de aprendizado de máquina supervisionado. O estudo segue um *pipeline* típico que inclui pré-processamento dos textos, extração de características e treinamento de classificadores para estimar a polaridade das mensagens (positiva, negativa ou neutra), evidenciando o potencial dessa abordagem para compreender opiniões em plataformas de grande circulação.

No contexto brasileiro, [Rosa \(2015\)](#) apresenta uma tese voltada à análise de sentimentos e afetividade em textos extraídos de redes sociais. A autora discute as particularidades da linguagem informal utilizada nesses ambientes, incluindo o uso de gírias, abreviações, emojis e outros recursos gráficos, e propõe métricas específicas para mensurar polaridade e afetividade em português, como o *Sentimeter-Br* e o *AFM-Br*. Esses trabalhos evidenciam tanto a relevância quanto a complexidade de se trabalhar com dados textuais provenientes de redes sociais abertas, cenário que guarda semelhanças importantes com o *chat* da *Twitch*.

Já [Fraga \(2019\)](#) desenvolve um estudo aplicado de análise de sentimentos em redes sociais, detalhando as principais etapas do processo: coleta de dados, pré-processamento, representação vetorial dos textos e classificação supervisionada. O trabalho enfatiza o uso de técnicas como remoção de *stopwords*, normalização e tokenização, além da comparação entre diferentes algoritmos de classificação, reforçando a importância das etapas de tratamento dos dados para o desempenho final dos modelos.

De forma complementar, [Silva \(2022\)](#) apresenta uma revisão dos principais conceitos relacionados à análise de sentimentos, discutindo níveis de análise (documento, sentença, aspecto), abordagens baseadas em léxicos e em aprendizado de máquina, bem

como aplicações práticas em diferentes domínios. Embora o foco do trabalho esteja em textos de resenhas, a sistematização conceitual oferecida é útil para embasar a formulação da tarefa de análise de sentimentos adotada neste estudo.

Os trabalhos mencionados demonstram que a análise de sentimentos em redes sociais é um tema consolidado, porém a maior parte dos estudos concentra-se em plataformas como *Twitter* ou em bases de dados estáticas de comentários e resenhas. A aplicação dessa técnica a *chats* de transmissões ao vivo, como os da *Twitch*, ainda é menos explorada, especialmente quando se consideram características como alto volume, linguagem informal extrema e necessidade de análise em tempo quase real.

### 3.2 Classificação de Texto com Redes Neurais *Transformers*

Além da análise de sentimentos, há um corpo significativo de pesquisas voltadas à classificação de texto em geral, muitas delas relacionadas à detecção de notícias falsas, discurso de ódio ou outros tipos de conteúdo problemático. Embora a tarefa-alvo não seja necessariamente a mesma deste trabalho, esses estudos compartilham um *pipeline* semelhante: coleta de textos, pré-processamento, representação vetorial e classificação, e, principalmente, exploram arquiteturas de modelos que também podem ser aplicadas à análise de sentimentos.

Em [Oshikawa, Qian e Wang \(2018\)](#), por exemplo, é apresentado um levantamento abrangente sobre o uso de técnicas de PLN para detecção de notícias falsas, discutindo tanto abordagens baseadas em características linguísticas e léxicas quanto modelos de aprendizado de máquina e redes neurais. No contexto da língua portuguesa, [Silva et al. \(2020\)](#) e [Monteiro et al. \(2018\)](#) propõem corpora e modelos para filtragem automática de notícias falsas, explorando diferentes estratégias de representação textual e classificadores tradicionais. Posteriormente, [Garcia, Afonso e Papa \(2022\)](#) introduzem o *FakeRecogna*, um novo corpus em português brasileiro com maior quantidade de amostras e categorias, utilizado para avaliar algoritmos tradicionais e redes neurais na tarefa de detecção de *fake news*.

A partir da popularização dos modelos baseados em *Transformers*, diversos trabalhos passaram a explorá-los como classificadores para tarefas de PLN. Em [Praseed, Rodrigues e Thilagam \(2023\)](#), os autores propõem um conjunto de modelos de linguagem pré-treinados baseados em *Transformers* para detecção de notícias falsas em Hindi, idioma com recursos limitados. A arquitetura combina modelos como BERT, RoBERTa e ELECTRA em um esquema de votação, avaliando diferentes estratégias de agregação, como votação majoritária e votação com voto. Os resultados obtidos mostram que a combinação de múltiplos modelos pré-treinados pode superar o desempenho de cada modelo isoladamente, destacando o potencial de abordagens em *ensemble* para tarefas de

classificação de texto.

De forma semelhante, [Dev et al. \(2024\)](#) apresentam um modelo híbrido que combina uma rede LSTM bidirecional com uma CNN para detecção de notícias falsas. O estudo evidencia que arquiteturas híbridas podem explorar de maneira complementar as capacidades de redes recorrentes e convolucionais na captura de dependências locais e de longo prazo em textos. Embora o domínio seja o de notícias jornalísticas, os resultados reforçam a importância da escolha de arquiteturas capazes de lidar com nuances da linguagem natural em tarefas de classificação.

No contexto específico da pandemia de COVID-19, [Alghamdi, Lin e Luo \(2023\)](#) investigam o uso de modelos baseados em *Transformers*, como BERT e CT-BERT, combinados com redes recorrentes, como GRU e BiGRU, para detecção de notícias falsas relacionadas à doença. Os autores comparam diferentes configurações de modelos e técnicas de ajuste fino, concluindo que modelos pré-treinados especializados no domínio (como CT-BERT) alcançam desempenho de estado da arte quando combinados com arquiteturas recorrentes adequadamente configuradas.

Ainda no escopo de comparação entre modelos de classificação, [Souza \(2023\)](#) e [Garrido-Merchan, Gozalo-Brizuela e Gonzalez-Carvajal \(2023\)](#) analisam o desempenho de redes neurais recorrentes e modelos baseados em *Transformers* em tarefas de classificação de texto, incluindo detecção de notícias falsas e outras aplicações. De modo geral, os resultados apontam que modelos pré-treinados como BERT tendem a superar métodos tradicionais e redes neurais treinadas do zero, especialmente quando os dados de treinamento rotulados são relativamente limitados.

Embora esses trabalhos tenham como foco principal a detecção de notícias falsas ou outras formas de classificação textual, eles são relevantes para este TCC por demonstrarem, de maneira consistente, a eficácia de modelos pré-treinados baseados em *Transformers* e de arquiteturas híbridas em tarefas de classificação de texto. A partir dessas evidências, justifica-se a escolha de modelos como XLM-RoBERTa e BERT Multilingual, bem como a adoção de uma estratégia em duas camadas para aumentar a robustez das predições no contexto do *chat* da *Twitch*.

### 3.3 Síntese e Posicionamento Deste Trabalho

A literatura revisada indica que:

- a análise de sentimentos em redes sociais é um campo amplamente estudado, com diversos trabalhos tratando de textos curtos e informais, especialmente em plataformas como *Twitter* ([HASAN; MALIHA; ARIFUZZAMAN, 2019; ROSA, 2015; FRAGA, 2019; SILVA, 2022](#));

- modelos de aprendizado de máquina e, mais recentemente, modelos baseados em *Transformers* têm apresentado desempenho superior em tarefas de classificação de texto, incluindo detecção de notícias falsas e outras aplicações de PLN (OSHI-KAWA; QIAN; WANG, 2018; PRASEED; RODRIGUES; THILAGAM, 2023; DEV et al., 2024; ALGHAMDI; LIN; LUO, 2023; SOUZA, 2023; GARRIDO-MERCHAN; GOZALO-BRIZUELA; GONZALEZ-CARVAJAL, 2023);
- abordagens que combinam diferentes modelos ou arquiteturas, como *ensembles* de *Transformers* ou modelos híbridos, costumam alcançar resultados melhores que modelos isolados em cenários complexos.

Entretanto, observa-se uma lacuna específica na aplicação de técnicas de análise de sentimentos ao contexto de *chats* de transmissões ao vivo, como os da *Twitch*, em que mensagens são curtas, altamente informais e produzidas em grande volume e velocidade. A maior parte dos trabalhos existentes concentra-se em dados de redes sociais assíncronas (como *Twitter* e *Facebook*) com notícias e resenhas estáticas.

Dessa forma, o presente trabalho se diferencia por investigar a viabilidade de utilizar modelos pré-treinados baseados em *Transformers*, em particular XLM-RoBERTa e BERT Multilingual, para realizar análise de sentimentos em mensagens do *chat* da *Twitch* na categoria de jogos, propondo ainda uma arquitetura em duas camadas que busca aumentar a confiabilidade das predições. Além disso, o trabalho visa integrar essa classificação a uma visualização dinâmica do humor da audiência, aproximando a análise de sentimentos de um cenário em tempo quase real para os *streamers*.

## 4 Metodologia

A metodologia deste trabalho foi estruturada em etapas sequenciais para desenvolver um sistema na linguagem de programação *Python*, capaz de identificar os sentimentos predominantes nas mensagens enviadas em chats de transmissões ao vivo na plataforma *Twitch*, com foco na categoria de jogos. O modelo implementa três configurações: (i) **RoBERTa-only**, (ii) **BERT-only** e (iii) **Ensemble por confiança** (RoBERTa com *fallback* para BERT em baixa confiança). Esses três *pipelines* estão representados na Figura 3. Nesta Figura a primeira etapa consiste em coletar os dados para montar a base de dados, a segunda etapa é a aplicação dos modelos, sendo o 1º *pipeline* a classificação usando RoBERTa, o 2º usando BERT e, por fim, o 3º a junção dos dois modelos, que é chamado de *ensemble*.

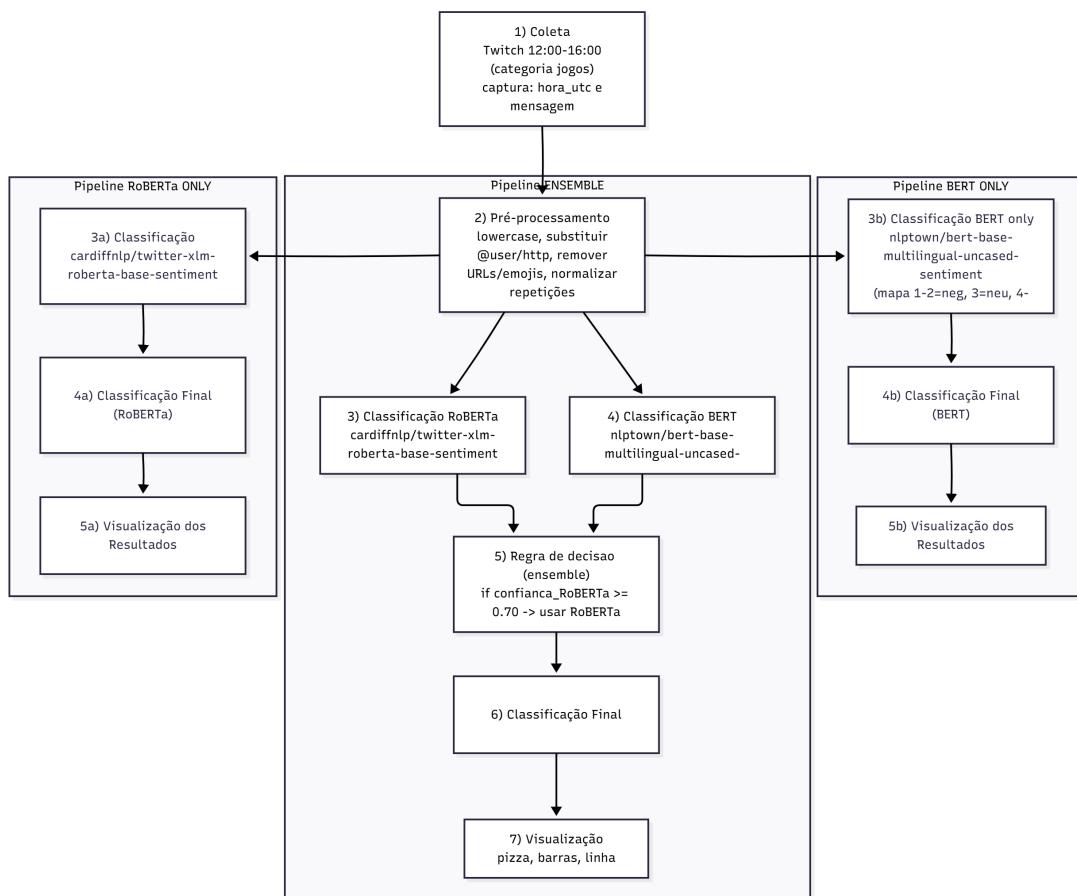


Figura 3 – Esquematização dos algoritmos propostos na metodologia para classificação (RoBERTa-only, BERT-only e Ensemble por confiança). Fonte: Elaboração própria.

## 4.1 Etapa 1: Coleta e Organização dos Dados

As mensagens foram coletadas diretamente do *chat* da *Twitch* na categoria de jogos, em uma janela contínua entre **12h e 16h** do dia 3 de Junho de 2025, totalizando 3.512 mensagens/linhas (uma mensagem por linha).<sup>1</sup> Para cada entrada, foram capturados: (i) *hora\_utc* (marcação temporal), (ii) *usuario* (identificador textual) e (iii) *mensagem* (conteúdo textual).

Na análise, foi considerada a linha completa do conteúdo textual da mensagem; o identificador do usuário foi desconsiderado, por não contribuir para a tarefa de análise de sentimento e para preservar privacidade do participante. O tratamento posterior assumiu apenas o par  $\{hora\_utc, mensagem\}$  como insumo para a modelagem.

## 4.2 Etapa 2: Ambiente e Bibliotecas em Python

Na implementação do modelo proposto utilizou as seguintes bibliotecas da linguagem Python:

- **Transformers** (*Hugging Face*) para carregamento de modelos e *tokenizers*;
- **PyTorch** (`torch`) para execução dos modelos (CPU/GPU), com inferência sob `torch.no_grad()`;
- **NumPy** e **SciPy** (`softmax`) para operações numéricas sobre os *logits* (*scores* retornados por cada modelo);
- **pandas** para manipulação tabular das mensagens e integração com o fluxo de inferência;
- **tqdm** para barra de progresso em processamento sequencial;
- **Matplotlib** e **Seaborn** para geração das visualizações.

## 4.3 Etapa 3: Pré-processamento e Limpeza Textual

Adotou-se um **pré-processamento mínimo**, devido as características do texto em estudo, que são: textos curtos, ruidosos e ricos em gírias/emotes. Os processamentos realizados foram:

1. **Padronização:** menções convertidas em `@user` e URLs em `http`;

<sup>1</sup> Horário conforme a janela definida para a captura do estudo.

2. **Preservação do texto:** não se removeu gírias, risadas, alongamentos e outros sinais afetivos, por carregarem pistas de polaridade;
3. **Tokenização** feita pelos *tokenizers* nativos de cada modelo.

Essa estratégia mantém o sinal semântico/afetivo enquanto reduz variabilidade inútil em identificadores e links.

## 4.4 Etapa 4: As Redes Neurais Transformers

As classificações foram implementadas com:

- **Bibliotecas:** *transformers* (tokenização e carregamento dos modelos), **torch** (inferência), **numpy/scipy** (pós-processamento com *softmax*) e **pandas** (manuseio em memória das mensagens).
- **Modelos:**
  1. RoBERTa (saída nativa em 3 classes: *negative/neutral/positive*);
  2. BERT (saída nativa em 5 estrelas; mapeamento para 3 classes: 1–2→negativo; 3→neutro; 4–5→positivo).
- **Estratégia do ensemble:** inferir com RoBERTa; se a maior probabilidade (*confidence*)  $< 0,70$ , reclassificar a mesma mensagem com BERT e usar o rótulo do BERT.

## 4.5 Etapa 5: Os Três Pipelines de Inferência

Foram definidos três fluxos consistentes:

### (a) RoBERTa-only

1. Pré-processar a mensagem;
2. Tokenizar e classificar com RoBERTa;
3. Aplicar **softmax** aos *logits* do modelo, convertendo-os em uma distribuição de probabilidades (valores entre 0 e 1, cuja soma é 1), permitindo selecionar o rótulo pela maior probabilidade e definir a confiança como esse valor máximo.

### (b) BERT-only

1. Pré-processar a mensagem;
2. Tokenizar e classificar com BERT (cabeça 1–5 estrelas);
3. Aplicar *softmax* aos *logits* do BERT (saídas brutas para 1–5 estrelas), convertendo-os em probabilidades normalizadas (valores entre 0 e 1 cuja soma é 1). Em seguida, selecionar a classe com maior probabilidade como predição e usar essa probabilidade máxima como medida de confiança.
4. Mapear a predição de 5 classes (estrelas) para 3 classes: 1–2 → negativo; 3 → neutro; 4–5 → positivo.

### (c) Ensemble por confiança (RoBERTa → BERT)

1. Executar o RoBERTa e obter o rótulo previsto e sua **confiança** (isto é, a probabilidade do rótulo escolhido).
2. Se a **confiança** for maior ou igual a 0,70, aceitar o rótulo do RoBERTa como resultado final.
3. Caso contrário, submeter a mesma mensagem ao BERT e substituir pelo rótulo do BERT (após o mapeamento de 5 para 3 classes).

O limiar de 0,70 foi definido de forma pragmática para reduzir ambiguidades em mensagens curtas e/ou ruidosas.

## 4.6 Etapa 6: Visualização dos Resultados

Para interpretação dos resultados e apoio ao *streamer*, foram utilizadas as funções de visualizações em *Python*:

- **Distribuição de classes** (barras/pizza) por configuração: RoBERTa-only, BERT-only, Ensemble;
- **Tendência temporal**: as mensagens são mapeadas para  $-1, 0, +1$  (negativo, neutro, positivo) e suavizadas via média móvel;
- **Análises complementares**: dispersão por confiança e comparação lado a lado dos três fluxos.

## 5 Resultados

Este capítulo apresenta e discute os resultados obtidos a partir dos três cenários de classificação implementados em *Python* com a biblioteca *transformers*: (i) **RoBERTa** (XLM-RoBERTa, *cardiffnlp/twitter-xlm-roberta-base-sentiment*); (ii) **BERT Multilingual** (*nlptown/bert-base-multilingual-uncased-sentiment*, convertido de 5 para 3 classes); e (iii) **Ensemble** (RoBERTa como classificador principal e BERT como refinador quando a confiança do primeiro  $< 0,70$ ). As inferências foram executadas com PyTorch em modo *no-grad*, aplicando *softmax* aos *logits* para obter probabilidades por classe.

Pode ser observado, na Figura 4, a distribuição de polaridades (positiva, negativa e neutra) produzida pelos três classificadores utilizando as redes *Transformers*.

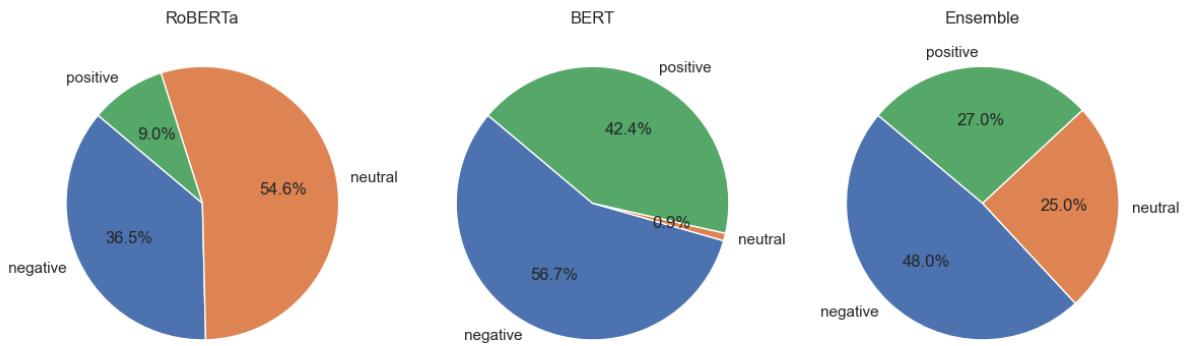


Figura 4 – Resultados obtidos pelos três classificadores: RoBERTa, BERT Multilingual e Ensemble. Fonte: Elaboração própria.

Para facilitar a leitura dos dados que constam na Figura 4, a Tabela 1 apresenta os valores em percentuais de cada polaridade, positiva, negativa e neutra, resultantes dos três *pipelines*.

Cenário	Negativo	Neutro	Positivo
RoBERTa	36,5%	54,6%	9,0%
BERT	56,7%	0,9%	42,4%
Ensemble	48,0%	25,0%	27,0%

Tabela 1 – Distribuição percentual de sentimentos por cenário de classificação.

O RoBERTa, treinado originalmente em dados de redes sociais, mostrou **predomínio de neutros** (54,6%), seguido de negativos (36,5%) e uma fração menor de positivos (9,0%). Esse perfil é compatível com o comportamento esperado para conversas rápidas de

*chat*: muitas mensagens curtas, reativas e contextuais, que o modelo tende a interpretar como neutras quando sinais semânticos de polaridade são fracos ou ambíguos.

O BERT Multilingual (mapeado de 5 → 3 classes) produziu **baixíssima taxa de neutros (0,9%)**, com forte concentração em negativo (56,7%) e positivo (42,4%). Esse comportamento é típico quando:

1. O modelo de 5 níveis foi treinado em domínio de *reviews*, onde o *meio-termo* (3 estrelas) costuma ser relativamente raro, e
2. O mapeamento para 3 classes contrai a região de decisão do “neutro” (3 estrelas) a um intervalo muito estreito, empurrando mensagens levemente ambíguas para as classes *positiva* ou *negativa*.

Na prática, o BERT acentua a polarização e capta melhor elogios/reclamações explícitos, mas não representa de forma objetiva os estados neutros, que são característicos de chats com mensagens telegráficas ou repletas de gírias/emotes.

A combinação por limiar (0,70) produziu uma distribuição mais **equilibrada**: 48,0% negativos, 25,0% neutros e 27,0% positivos. Em linhas gerais:

- O RoBERTa fornece uma base conservadora (tende ao neutro quando pouco confiante).
- O BERT intervém apenas nos casos incertos, deslocando parte do neutro para polaridades quando há evidências suficientes.

Portanto, diante dos resultados apresentados, sugere a redução dos extremos de cada modelo isolado: o *overshoot* de neutros do RoBERTa-only e a quase ausência de neutros do BERT-only. Para contextualizar essa escolha, os três cenários testados seguiram pipelines distintos. No RoBERTa-only, cada mensagem foi pré-processada (substituição de menções por `@user` e links por `http`), tokenizada pelo *tokenizer* do modelo RoBERTa; em seguida, os *logits* foram convertidos em probabilidades via *softmax*, e o rótulo final foi definido pela classe de maior probabilidade, usando essa mesma probabilidade como medida de confiança. No BERT-only, a mensagem foi tokenizada pelo modelo BERT, a saída (1–5 estrelas) foi obtida a partir da maior probabilidade após *softmax*, e então mapeada para três polaridades (1–2 → negativo; 3 → neutro; 4–5 → positivo). Por fim, no ensemble por confiança, adotou-se o RoBERTa como classificador principal: quando a confiança do RoBERTa foi maior ou igual a 0,70, sua predição foi mantida; caso contrário, a mesma mensagem foi submetida ao BERT e o rótulo final passou a ser o do BERT (já após o mapeamento de 5 para 3 classes). Dessa forma, o pipeline proposto reduz decisões frágeis do RoBERTa em mensagens curtas/ambíguas, ao mesmo tempo em que evita a

polarização excessiva do BERT em casos onde o RoBERTa já apresenta alta confiança, tornando o resultado final mais consistente para uso contínuo durante transmissões.

## 6 Conclusão

Este trabalho investigou a criação de modelos computacionais baseados em redes *Transformers* para análise de sentimentos em mensagens retiradas do *chat* da *Twitch* na categoria de jogos, para apoiar *streamers* no acompanhamento do humor dos telespectadores. Foram avaliados três cenários de inferência da rede *Transformer*: RoBERTa, BERT Multilingual e um *ensemble*, quando executa um BERT apenas quando a confiança do RoBERTa fica abaixo de um valor pré-definido.

Os resultados dos três modelos mostraram comportamentos complementares. O RoBERTa produziu maior proporção de rótulos neutros, indicando postura mais conservadora frente a textos curtos e ruidosos típicos de *chat*, na prática, isso reduz decisões precipitadas entre positivo/negativo quando a mensagem é ambígua. O BERT Multilingual apresentou baixa incidência de neutros e maior polarização entre positivo e negativo, coerente com seu treinamento original. Essa característica aumenta a capacidade de destacar reações, mas diminui a presença da classe “neutra”. O *ensemble* combinou as duas abordagens por meio de uma regra de confiança: quando a confiança do RoBERTa foi maior ou igual a 0,70, manteve-se o rótulo do RoBERTa, caso contrário, a mesma mensagem foi reclassificada pelo BERT e o rótulo final passou a ser o do BERT. Dessa forma, o resultado final preserva decisões mais estáveis quando há alta confiança e aciona uma segunda opinião apenas nos casos de maior incerteza.

Sobre as limitações, destacam-se a ausência de ajuste específico ao domínio de *chat* da *Twitch*, a dependência de modelos pré-treinados em contextos diferentes do *chat* da *Twitch* (por exemplo, textos de redes sociais como o *Twitter* e avaliações em lojas online, como a *Amazon*) e a falta de um conjunto rotulado manualmente para mensurar acurácia. Em síntese, os experimentos confirmam a viabilidade de *Transformers* pelo modelo *ensemble* para prover dados úteis em quase tempo real sobre o humor dos telespectadores do *chat*, sendo a alternativa mais promissora para uso prático em transmissões ao vivo.

Para trabalhos futuros, listam-se as seguintes sugestões:

1. **Ajuste fino** (*fine-tuning*) de modelos em um conjunto rotulado de mensagens reais da *Twitch* (com emotes/abreviações), para reduzir vieses de domínio.
2. **Limiar adaptativo** de confiança no *ensemble*, calibrado por canal e horário, e **suavização temporal** (médias móveis, EWMA) para dashboards mais estáveis.
3. **Modelagem de emotes e emojis**: dicionários específicos e mapeamentos semânticos (p.ex., Kappa, PogChamp) para sinalizar ironia, hype e apoio.

4. **Métricas adicionais**: além de contagens, incorporar *confidence-weighted sentiment*, *sentiment velocity* e *burst detection* para detectar viradas rápidas de humor.
5. **Avaliação humana** (*human-in-the-loop*) para auditoria periódica, alinhamento ao tom do canal e coleta contínua de *feedback* para re-treinamento.
6. **Integração de tempo real** com buffers assíncronos e lotes pequenos (*micro-batching*) para latência baixa e uso eficiente de GPU/CPU.

## Referências

- ALGHAMDI, J.; LIN, Y.; LUO, S. Towards covid-19 fake news detection using transformer-based models. **Knowledge-Based Systems**, v. 274, p. 110642, 2023. ISSN 0950-7051. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0950705123003921>>. Citado 2 vezes nas páginas 19 e 20.
- DEV, D. G.; BHATNAGAR, V.; BHATI, B. S.; GUPTA, M.; NANTHAAMORNPHONG, A. LstmCNN: A hybrid machine learning model to unmask fake news. **Helion**, v. 10, n. 3, p. e25244, 2024. ISSN 2405-8440. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2405844024012751>>. Citado 2 vezes nas páginas 19 e 20.
- FRAGA, B. E. d. A. A. **Análise de Sentimentos nas Redes Sociais**. Dissertação (Trabalho de Conclusão de Curso (Bacharelado em Engenharia de Computação)) — Centro Federal de Educação Tecnológica de Minas Gerais, Timóteo, MG, 2019. Citado 2 vezes nas páginas 17 e 19.
- GARCIA, G. L.; AFONSO, L. C. S.; PAPA, J. P. Fakerecogna: A new brazilian corpus for fake news detection. In: PINHEIRO, V.; GAMALLO, P.; AMARO, R.; SCARTON, C.; BATISTA, F.; SILVA, D.; MAGRO, C.; PINTO, H. (Ed.). **Computational Processing of the Portuguese Language**. Cham: Springer International Publishing, 2022. p. 57–67. ISBN 978-3-030-98305-5. Citado na página 18.
- GARRIDO-MERCHAN, E. C.; GOZALO-BRIZUELA, R.; GONZALEZ-CARVAJAL, S. Comparing bert against traditional machine learning models in text classification. **Journal of Computational and Cognitive Engineering**, BON VIEW PUBLISHING PTE, v. 2, n. 4, p. 352–356, 2023. Disponível em: <<http://dx.doi.org/10.47852/bonviewJCCE3202838>>. Citado 2 vezes nas páginas 19 e 20.
- GILLIOZ, A.; CASAS, J.; MUGELLINI, E.; KHALED, O. A. Overview of the transformer-based models for nlp tasks. In: **2020 15th Conference on Computer Science and Information Systems (FedCSIS)**. [S.l.: s.n.], 2020. p. 179–183. Citado 4 vezes nas páginas 11, 12, 13 e 14.
- Grand View Research. **Live Streaming Market Size, Share & Trend Analysis Report, 2024–2030**. 2024. <<https://www.grandviewresearch.com/industry-analysis/live-streaming-market-report>>. Citado na página 7.
- GUPTA, V. A survey of natural language processing techniques. **International Journal of Computer Science & Engineering Technology**, Citeseer, v. 5, n. 1, p. 14–16, 2014. Disponível em: <<https://arxiv.org/abs/2212.05773>>. Citado 2 vezes nas páginas 7 e 11.
- HASAN, M. R.; MALIHA, M.; ARIFUZZAMAN, M. Sentiment analysis with nlp on twitter data. In: IEEE. **2019 international conference on computer, communication, chemical, materials and electronic engineering (IC4ME2)**. [S.l.], 2019. p. 1–4. Citado 6 vezes nas páginas 7, 11, 12, 13, 17 e 19.
- Hugging Face. **Transformers Documentation**. 2024. <<https://huggingface.co/docs/transformers/index>>. Acessado em 10 nov. 2025. Citado na página 13.

- KHYANI, D.; S, S. B. An interpretation of lemmatization and stemming in natural language processing. **Journal of University of Shanghai for Science and Technology**, v. 22, p. 350–357, 2021. Disponível em: <[https://www.researchgate.net/publication/348306833\\_An\\_Interpretation\\_of\\_Lemmatization\\_and\\_Stemming\\_in\\_Natural\\_Language\\_Processing](https://www.researchgate.net/publication/348306833_An_Interpretation_of_Lemmatization_and_Stemming_in_Natural_Language_Processing)>. Citado 2 vezes nas páginas 11 e 12.
- MIKOLOV, T.; CHEN, K.; CORRADO, G.; DEAN, J. **Efficient Estimation of Word Representations in Vector Space**. 2013. Disponível em: <<https://arxiv.org/abs/1301.3781>>. Citado 2 vezes nas páginas 11 e 12.
- MONTEIRO, R. A.; SANTOS, R. L. S.; PARDO, T. A. S.; ALMEIDA, T. A. de; RUIZ, E. E. S.; VALE, O. A. Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In: **Computational Processing of the Portuguese Language**. [S.l.]: Springer International Publishing, 2018. p. 324–334. ISBN 978-3-319-99722-3. Citado na página 18.
- OSHIKAWA, R.; QIAN, J.; WANG, W. Y. A survey on natural language processing for fake news detection. **arXiv preprint arXiv:1811.00770**, 2018. Citado 2 vezes nas páginas 18 e 20.
- PRASEED, A.; RODRIGUES, J.; THILAGAM, P. S. Hindi fake news detection using transformer ensembles. **Engineering Applications of Artificial Intelligence**, v. 119, p. 105731, 2023. ISSN 0952-1976. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0952197622007217>>. Citado 2 vezes nas páginas 18 e 20.
- RADFORD, A.; NARASIMHAN, K.; SALIMANS, T.; SUTSKEVER, I. Improving language understanding by generative pre-training. 2018. Disponível em: <[https://cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf)>. Citado na página 14.
- ROSA, R. L. **Análise de sentimentos e afetividade de textos extraídos das redes sociais**. Tese (Tese (Doutorado em Engenharia Elétrica)) — Escola Politécnica da Universidade de São Paulo, São Paulo, 2015. Citado 2 vezes nas páginas 17 e 19.
- SILVA, O. B. **Análise de sentimento: principais conceitos e aplicação**. Dissertação (Trabalho de Conclusão de Curso (Bacharelado em Engenharia da Computação)) — Universidade Federal do Maranhão, São Luís, MA, 2022. Citado 2 vezes nas páginas 17 e 19.
- SILVA, R. M.; SANTOS, R. L.; ALMEIDA, T. A.; PARDO, T. A. Towards automatically filtering fake news in portuguese. **Expert Systems with Applications**, v. 146, p. 113199, 2020. ISSN 0957-4174. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0957417420300257>>. Citado na página 18.
- SOUZA, C. E. d. Detecção de fake news em redes sociais com o uso de redes neurais recorrentes, redes neurais gráficas e transformers. **REPOSITORIO PUCSP: Página inicial**, Pontifícia Universidade Católica de São Paulo, 2023. Disponível em: <<https://repositorio.pucsp.br/jspui/handle/handle/39999>>. Citado 2 vezes nas páginas 19 e 20.
- StreamScheme. **Twitch Demographic & Growth Statistics [2023 Updated]**. 2023. <<https://streamscheme.com/twitch-statistics/>>. Citado na página 7.

TUNSTALL, L.; WERRA, L. V.; WOLF, T. **Natural language processing with transformers**. [S.l.]: "O'Reilly Media, Inc.", 2022. Citado 4 vezes nas páginas 11, 12, 13 e 14.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, Ł.; POLOSUKHIN, I. Attention is all you need. **Advances in Neural Information Processing Systems**, v. 30, 2017. Disponível em: <<https://arxiv.org/abs/1706.03762>>. Citado na página 13.