

# DETECTANDO COMPORTAMENTOS TÓXICOS DE USUÁRIOS NA REDE SOCIAL TWITCH EM TRANSMISSÕES DO JOGO ELETRÔNICO LEAGUE OF LEGENDS

## DETECTING TOXIC BEHAVIOR OF USERS ON THE SOCIAL NETWORK TWITCH IN LEAGUE OF LEGENDS ELECTRONIC GAME BROADCASTS

David De Sena Souza\*  
Henrique Viana Oliveira\*\*

### RESUMO

As redes sociais vêm recebendo cada vez mais atenção de usuários que frequentemente compartilham suas opiniões por meio delas. Entretanto, identificar que tipo de opinião é relevante, que tipos de usuários opinam na rede e quais são seus perfis é um trabalho desafiador. Logo, recorrer ao uso de técnicas computacionais que permitam fazer análise de grandes quantidades de dados para descobrir novos conhecimentos é uma tarefa muito valiosa nesse contexto. O presente trabalho propõe um modelo de abordagem léxica da área de Análise de Sentimentos, na rede social de transmissões ao vivo *Twitch*, em transmissões do jogo eletrônico *League of Legends*. A ideia central do trabalho é identificar o comportamento dos usuários que assistem a esse tipo de jogo, classificando-os como positivo, negativo ou neutro. Além disso, buscamos classificar as mensagens dos usuários a fim de descobrir possíveis comportamentos tóxicos.

**Palavras-chave:** 1. Análise de Sentimentos 2. *Twitch* 3. Detecção de Emoções.

### ABSTRACT

Social networks are receiving more and more attention from users who often share their opinions through them. However, identifying what type of opinion is relevant, what types of users opine on the network and what are their profiles is a challenging job. Therefore, resorting to the use of computational techniques that allow the analysis of large amounts of data to discover new knowledge is a very valuable task in this context. The present work proposes a model of lexical approach in the area of Sentiment Analysis, in the social network of live broadcasts Twitch, in broadcasts of the electronic game League of Legends. The central idea of the work is to identify the behavior of users who watch this type of game, classifying them as positive, negative or neutral. In addition, we seek to classify users' messages in order to discover possible toxic behaviors.

**Keywords:** 1. Sentiment Analysis 2. Twitch 3. Emotion Detection.

\* Instituto Federal de Educação, Ciência e Tecnologia do Ceará, david.sena.souza05@aluno.ifce.edu.br

\*\* Instituto Federal de Educação, Ciência e Tecnologia do Ceará, henrique.viana@ifce.edu.br

## 1 INTRODUÇÃO

A rede global de computadores aumentou exponencialmente sua base de usuários ao longo dos anos. Segundo uma pesquisa promovida pelo *Hootsuite* em parceria com a agência *We Are Social*<sup>1</sup>, no ano de 2021, a Internet contava com 4.95 bilhões de usuários ativos, fazendo o compartilhamento de uma grande quantidade de dados. No total, é gerada uma quantidade expressiva de 400 zettabytes de dados, o que corresponde a 40 trilhões de gigabytes por dia. A pesquisa projeta que esse número de dados trafegados na rede se duplique a cada dois anos.

Ainda de acordo com a pesquisa acima, as redes sociais possuem uma grande contribuição nessa quantidade de dados trafegados. Existem aproximadamente 4.62 bilhões usuários acessando-as ativamente, correspondendo a 93.3% dos usuários de toda a Internet. Esses usuários estão usando as redes sociais diariamente para expressar suas emoções, opiniões, pensamentos e comentários sobre seu cotidiano.

A *Twitch*<sup>2</sup>, uma rede social de transmissões ao vivo, possui mais de 140 milhões de usuários, com média de 30 milhões de visualizações diárias e inúmeras mensagens trocadas nos *chats*<sup>3</sup>. Na *Twitch* existem dois tipos de usuários: os criadores de conteúdo e os espectadores. Os criadores de conteúdo, também chamados de *streamers*, geram conteúdos em várias categorias disponíveis como Jogos, Música, Arte, Vida Real, Esportes, etc. Os conteúdos criados pelos *streamers* são popularmente chamadas de *lives*, *streamings* ou *streams*. Já os espectadores, comumente chamados de *viewers*, são os usuários que assistem as *lives*. Eles conseguem interagir com os *streamers* através de textos e *emotes*, na forma de curtos comentários.

Esses dados gerados pelos *viewers*, apesar de parecer irrelevantes, se coletados e bem analisados, podem ter grande potencial de serem extraídas informações importantes. Essa tarefa permite, por exemplo, ajudar empresas a obter conhecimento sobre os usuários, conseguindo traçar e classificar seus perfis. Assim, é possível para as empresas direcionarem produtos ou conteúdos mais assertivos para cada tipo de usuário.

Por outro lado, essa coleta de dados pode também ajudar os *streamers*, permitindo identificar se as suas comunidades tendem a ter comportamentos negativos ou positivos. Ter essas informações mais profundas vão além de saber apenas dados quantitativos gerais, mas também cria a possibilidade de se ter uma análise de dados mais específica de cada usuário. Saber quais usuários contribuem positivamente ou negativamente em sua *lives*, abre a possibilidade dos *streamers* adotarem estratégias para melhorar o desempenho das *lives* ou contornar eventuais problemas que apareçam com determinados usuários.

Porém, uma análise de todos esses dados gerados é uma tarefa inviável para seres humanos, pois é impossível extrair informação significativa em tempo hábil, devido as nossas limitações cognitivas. Logo, se faz necessário o uso de técnicas computacionais que conseguem extrair e analisar informações relevantes.

<sup>1</sup> <https://datareportal.com/reports/digital-2022-global-overview-rep>

<sup>2</sup> <https://www.twitch.tv>

<sup>3</sup> <https://histografias.com/infografia-historia-twitch.html>

Uma dessas técnicas, a Análise de Sentimentos ou Mineração de Opinião, é o ramo da Inteligência Artificial relacionado à análise de opiniões, sentimentos, avaliações, atitudes e emoções (ASGHAR et al., 2015). A Análise de Sentimentos em textos *online* tem sido usada para previsão de ações, resposta a campanhas de *marketing*, monitoramento global da felicidade, etc. Também constata três tipos de abordagens de Análise de Sentimentos: baseadas em regras de dicionário (léxico), em aprendizado de máquina ou em aprendizado profundo (REAGAN et al., 2017). Os sentimentos podem ser extraídos como classificação de polaridade (neutro, negativo, positivo) ou uma pontuação da valência de polaridade (como por exemplo, em uma escala de valores inteiros) (HUTTO; GILBERT, 2014).

Este trabalho tem como objetivo utilizar a Análise de Sentimentos com classificação de polaridade em chats da *Twitch*, através da abordagem léxica. Como a plataforma da *Twitch* possui seu próprio vocabulário, com a utilização frequente de junção entre textos e *emotes*, é necessário então a criação de um dicionário específico para ela. A partir dessa análise léxica, o foco principal do trabalho é identificar comportamentos tóxicos entre os usuários.

O presente trabalho está estruturado em sequência. A Seção 2 apresenta o referencial teórico, onde são explicados alguns conceitos e teorias sobre Análise de Sentimentos, que servem como base para este estudo. A Seção 3 detalha os trabalhos relacionados, envolvendo Mineração de Sentimentos que se aparentam com a ideia proposta para este trabalho. Na Seção 4, é apresentada a proposta do trabalho, onde é explicado todo o processo para conseguir extrair sentimentos, como coleta, análise e visualização final. A Seção 5 apresenta os resultados alcançados do trabalho. Finalmente, a Seção 6 conclui o trabalho com considerações finais e seus trabalhos futuros.

## 2 REFERENCIAL TEÓRICO

### 2.1 Análise de Sentimentos

A Análise de Sentimentos pode ser considerada um processo de classificação que procura identificar e analisar os sentimentos expressos em textos. O processo da Análise de Sentimentos se concentra especialmente em encontrar opiniões, identificar os sentimentos que são expressos e então classificar sua polaridade (por exemplo, em uma escala de valores) (MEDHAT; HASSAN; KORASHY, 2014), conforme as etapas no fluxograma na Figura 1.

Existem três principais níveis de tratamento na classificação em Análise de Sentimentos, são eles: Nível de Documento, Nível de Sentença e Nível de Aspecto.

- **Nível de Documento:** Considera todo o documento uma única informação, falando sobre algum assunto. Classifica um documento como expressando uma opinião de sentimento em positivo ou negativo.
- **Nível de Sentença:** Classifica o sentimento expresso em cada sentença do documento. O primeiro passo é identificar se a sentença é subjetiva ou objetiva. Quando a sentença for

subjetiva, será determinado se ela expressa sentimentos positivos ou negativos. Segundo (LIU, 2012), em geral não há uma diferença relevante entre as classificações de Nível de Documento e Sentença, porque as sentenças são apenas curtos documentos.

- **Nível de Aspecto:** Classifica o sentimento em relação aos aspectos específicos do objeto. Por exemplo, um indivíduo pode dar opiniões diferentes para cada aspecto do mesmo objeto como na frase “A qualidade de áudio deste telefone não é boa, mas a duração da bateria é longa” é dada uma opinião positiva (bateria é longa) e negativa (qualidade do áudio não é boa) sobre o objeto telefone.

Figura 1 – Etapas do processo de Análise de Sentimentos.



Fonte: (MEDHAT; HASSAN; KORASHY, 2014).

As tarefas realizadas na área de Análise de Sentimentos podem ser divididas basicamente em três categorias: Classificação de Sentimentos, Seleção de Características e Detecção de Emoções.

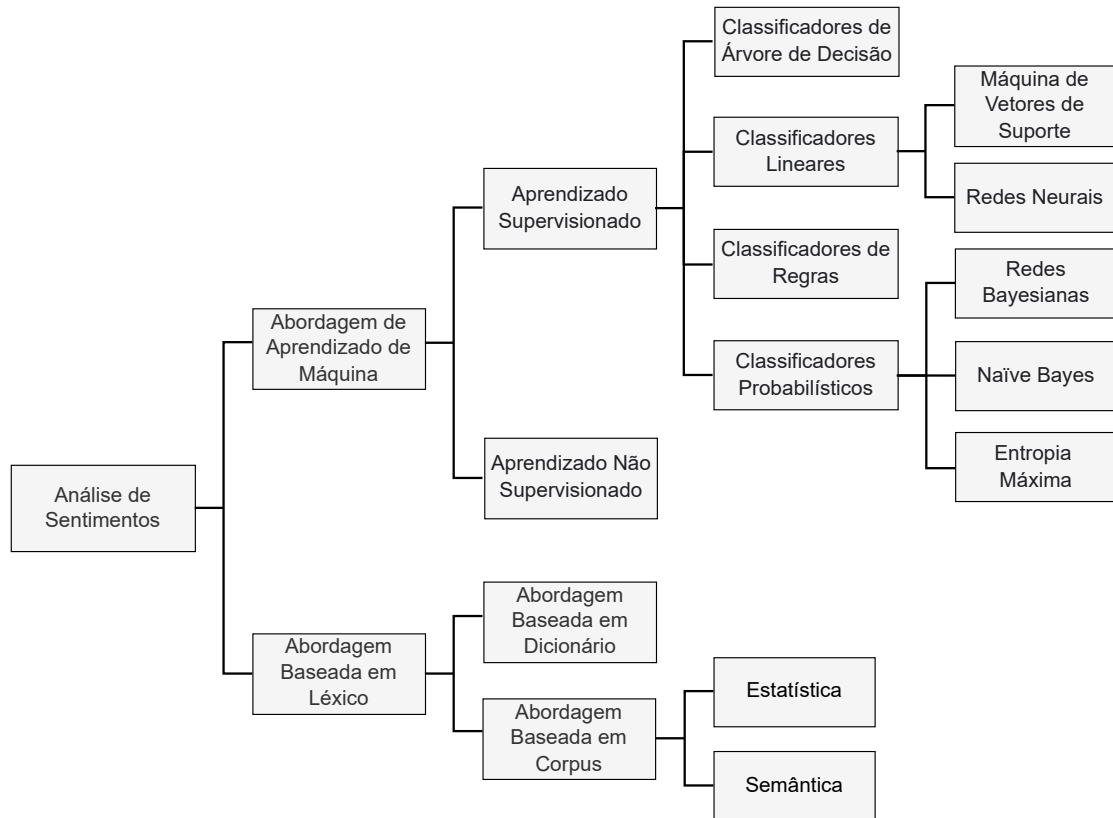
- **Classificação de Sentimentos:** O primeiro passo no problema de Classificação de Sentimentos é extrair e selecionar recursos do texto. Algumas das principais características para classificar são (BHONDE; PRASAD, 2015):
  - **Presença de termos e Frequência:** Essas características encontradas como palavras individuais ou n-gramas de palavras e suas contagens de frequência. São dadas às palavras uma ponderação binária (zero se a palavra aparecer ou um se caso contrário), ou pesos são usados como frequência de termo para indicar a importância relativa de características (MEJOVA; SRINIVASAN, 2011).
  - **Partes do discurso:** Encontrar adjetivos, pois são importantes indicadores de opiniões.

- **Palavras e frases de opinião:** São palavras geralmente usadas para expressar opiniões, incluindo boas ou ruins, gostar ou odiar. De outra forma, algumas frases expressam opiniões sem usar palavras de opinião. Por exemplo: "*me custou os olhos da cara*".
- **Negações:** A presença de palavras negativas pode mudar a sentindo da opinião, como *não bom* é equivalente a *ruim*.
- **Seleção de Características:** Os métodos de Seleção de Características podem ser divididos em métodos baseados em léxico, que precisam de anotação humana e métodos estatísticos que são processos automáticos usados com mais frequência. As abordagens baseadas em léxico geralmente começam com um pequeno conjunto de palavras chamadas de “*sementes*”. Em seguida, se inicializa esse conjunto por meio da detecção de sinônimos para obter um léxico maior. As abordagens estatísticas, por outro lado, são totalmente automáticas (WHITELAW; GARG; ARGAMON, 2005).
- **Detecção de Emoções:** A Detecção de Emoções busca extrair e analisar emoções, enquanto elas podem estar explícitas ou implícitas nas frases. Existem oito emoções básicas e prototípicas que são *alegria, tristeza, raiva, medo, confiança, desgosto, surpresa e antecipação* (PLUTCHIK, 1980). É implementado com mais frequência usando a abordagem Baseada em Léxico.

Para realizar as tarefas descritas acima, pode-se citar duas das mais relevantes abordagens na literatura: abordagem baseada em Léxico e Aprendizado de Máquina, como mostradas na Figura 2. Ambas as abordagens geralmente classificam o sentimento de qualquer texto em sentimento de polaridade (positivo, negativo ou neutro). Como cada método tem suas vantagens e limitações, pesquisadores precisam verificar prudentemente a precisão da classificação (BROWN et al., 1990) para evitar agir em resultados imprecisos da análise de dados (CANHOTO; PADMANABHAN, 2015).

A abordagem de Aprendizado de Máquina requer uma fase de treinamento que é conduzida pelos próprios pesquisadores. Uma parte dos dados classificados manualmente é usada como um conjunto de dados de treinamento e a outra parte dos dados classificadores é utilizada para aprender. Essa abordagem frequentemente treina os sentimentos utilizando recursos como unigramas ou bigramas (PANG; LEE; VAITHYANATHAN, 2002) aplicando diferentes técnicas de aprendizado Supervisionado, como Máquina de Vetores de Suporte, Naïve Bayes, Entropia Máxima, etc. Embora os métodos que usam conjuntos de dados de treinamento para automatizar a classificação de dados sejam vantajosos, eles ainda exigem rotulagem manual de exemplos de treinamento, cujo tamanho e qualidade afetam o desempenho do modelo treinado. A rotulagem de grande qualidade de um conjunto de dados de treinamento pode ser demorada, enquanto limitar o tamanho do conjunto de dados de treinamento leva a uma precisão de classificação mais baixa. Além disso, a amostragem do conjunto de dados de treinamento pode ter um impacto significativo no desempenho do modelo treinado.

Figura 2 – Abordagens de Análise de Sentimentos.



Fonte: (MEDHAT; HASSAN; KORASHY, 2014).

A abordagem baseada em Léxico não requer pré-processamento ou treinamento de um classificador. Normalmente se baseia em um dicionário de palavras de opinião, também conhecido como dicionário de sentimentos ou léxico de sentimentos para identificar e determinar a orientação de polaridade do sentimento como positivo ou negativo. A construção de um Léxico de Sentimentos é necessária ser feita manualmente, demanda esforço e tempo consideráveis. A maioria dos Léxicos de Sentimentos não são específicos para um determinado tópico (GODBOLE; SRINIVASIAH; SKIENA, 2007), mas existe a possibilidade de criação e modificação de diferentes Léxicos de Sentimentos para temas específicos, podendo ser considerando uma vantagem. Métodos baseados em Léxico normalmente analisam palavras comuns, erros de ortografia, coloquialismos, palavras estrangeiras e abreviações, que geralmente não estão incluídas nos dicionários. Além disso, a desvantagem de usar a abordagem baseada em Léxico para Análise de Sentimentos é que a classificação de polaridade pode variar em diferentes contextos. Por exemplo, o adjetivo “imprevisível” pode ter uma orientação positiva em uma crítica de filme (o final do filme foi imprevisível), mas uma orientação negativa para as habilidades condutivas de direção de uma pessoa (o motorista a frente tem movimentos imprevisíveis) (TURNERY, 2002).

As aplicações mais usuais da Análise de Sentimentos são nas áreas de avaliações de produtos e serviços de consumo. Existem vários sites que dispõem de comentários de usuários contando sua experiência sobre um produto adquirido, que podem ser processados por técnicas

de Análise de Sentimentos. Um exemplo perceptível disso é o *Google Shopping*<sup>4</sup>, um serviço do Google, que permite aos usuários fazer busca de produtos *onlines* e comparação de preços.

A Análise de Sentimentos inclusive pode produzir valor informativo aos candidatos que concorrem a cargos. Ela possibilita que os responsáveis de campanha observem como os eleitores se relacionam com os discursos e ações dos candidatos, como eles se sentem sobre diferentes questões dos candidatos. Uma Análise de Sentimentos de *tweets* referente as eleições presidências do Brasil no ano de 2018 pode ser vista na pesquisa feita por (CRISTIANI; LIEIRA; CAMARGO, 2020).

Também pode ser feita a aplicação de Análise de Sentimentos na área de mercado financeiro. Há numerosas informações em artigos, blogs, notícias etc. sobre as empresas. Um sistema de Análise de Sentimentos pode usar essas diversas fontes para encontrar informações das empresas e adicionar sentimento sobre elas, com uma pontuação de polaridade única, que pode ser usada por um sistema de negociação automático, que avalia o risco de uma ação, beneficiando o investidor. Um desses sistemas é o *The Stock Sonar*<sup>5</sup>. Outra aplicação que tem como foco supervisionar a reputação de uma marca. É o *Tweetfeel*<sup>6</sup>, uma aplicação que realiza a Análise de Sentimento nos *tweets* (conteúdo compartilhado pelos usuários do *Twitter*<sup>7</sup>) que contêm uma determinada palavra.

## 2.2 Twitch

A *Twitch* foi lançada em junho de 2011 derivando-se de outra plataforma de *streaming* chamada Justin.tv e é, atualmente uma das maiores redes sociais de *streaming* (GANDOLFI, 2016). Sua popularidade cresceu rapidamente e em 2013, 45 milhões de usuários já à utilizava, cada um em média assistindo 100 minutos de transmissões por dia (EWALT, 2013). Em 2014, a empresa foi comprada por US\$ 970 milhões, e passou a ser de propriedade da Amazon (WEINBERGER, 2016) e em 2022, segundo os relatórios da *Twitchtracker*<sup>8</sup> (site que registra dados sobre a plataforma), estima-se que até o presente momento, a *Twitch* tem 1000 bilhões de minutos de *streams* assistidos, com média de 7.9 milhões de canais fazendo transmissões por mês, sendo cerca de 94.800 mil de canais simultâneos por dia, com uma média de 2.6 milhões de espectadores simultâneos mensalmente.

A *Twitch* oferece uma plataforma de transmissão ao vivo onde um usuário pode criar o seu canal e transmitir conteúdos (*video streaming*) relacionados a jogos, estilo de vida ou conteúdo criativo a partir do seu computador ou celular. Esses usuários, chamados de *streamers*, podem criar um canal e transmitir nas categorias disponíveis pela *Twitch* diretamente e em tempo real para seu público, que são chamados de espectadores. Cada canal tem uma sala de *chat* ao

<sup>4</sup> <https://shopping.google.com.br/>

<sup>5</sup> <http://www.thestocksonar.com>

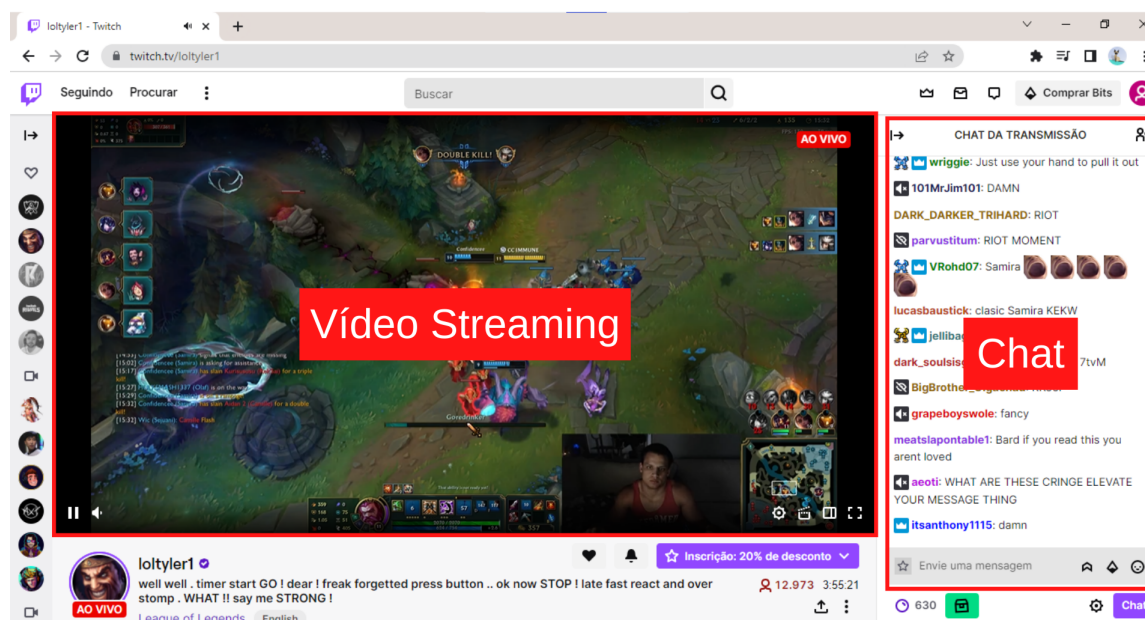
<sup>6</sup> <http://www.tweetfeel.com/>

<sup>7</sup> <https://twitter.com/>

<sup>8</sup> <https://twitchtracker.com/statistics>

vivo, na qual os usuários podem publicar comentários, interagindo com os outros espectadores e com o *streamer*, através do *chat*, falando sobre os eventos da transmissão (ver Figura 3).

Figura 3 – Plataforma da *Twitch*.



Fonte: Elaborado pelos autores.

Também é notável nesta plataforma, a presença de *emotes*: pequenas imagens incorporadas em mensagens de texto que são exibidas quando certas palavras são inseridas. Cada um desses *emotes* tem um significado específico, que pode variar dependendo do contexto de uso. Alguns *emotes* representam opiniões ou sentimentos, como aprovação ou alegria, enquanto outros têm significados bem específicos, referindo-se a jogadores famosos ou figuras da comunidade. Essas expressões desempenham um papel importante na maneira como o significado é transmitido por meio de mensagens de bate-papo.

Os *emotes* usados na comunicação do usuário estão sujeitos a mudanças constantes, seja pela adição de novos *emotes*, ou às vezes, porque eles simplesmente perdem popularidade. Esse ambiente de constante mudança, significa que é necessário que a análise atenda uma ampla variedade de *emotes*. Existem três fontes principais de *emotes* da *Twitch*:

























- **Emotes globais:** Nativos da *Twitch*, podem ser usados em todo e qualquer canal.
- **Emotes de assinante:** Limitados apenas para assinantes no canal do *streamer*.
- **Emotes de terceiros:** Podem ser utilizados globalmente, mas o *streamer* tem que adicioná-los de ferramentas externas, para que fiquem disponíveis para uso no seu canal. Por exemplo, *Better Twitch TV*<sup>9</sup> e *FrankerFaceZ*<sup>10</sup>.

<sup>9</sup> <https://betterttv.com/>

<sup>10</sup> <https://www.frankerfacez.com/>



Tabela 1 – Alguns *emotes* globais da *Twitch*.











	KEKW		BigBrother		BatChest
	PogChamp		LUL		AYAYA
	GigaChad		OmegaLUL		WutFace
	Sadge		BabyRage		Ez
	Kappa		ArsonNoSexy		BegWan
	CoolStoryBob		HeyGuys		4Head
	TakeNRG		SMOrc		PotFriend
	CurseLit		EntropyWins		PJSalt

Fonte: <https://twitchemotes.com>.

Os *emotes* estão muito presentes na comunicação da *Twitch*. Eles podem ser usados de forma mais rápida para expressar a opinião do usuário do que um comentário convencional de texto, principalmente em um ambiente que oferece pouco tempo de tela para que outras pessoas consigam ler suas mensagens. Partindo da ideia que *emotes* podem expressar ideias e opiniões, para classificar a expressão de *emotes* pelos usuários, em (GUIBON; OCHS; BELLOT, 2016) são identificados três casos de uso para a Análise de Sentimentos de *emojis* (ver Tabela 2): Expressão de Sentimento, Modificação de Sentimento e Melhoria de Sentimento.

- **Expressão de Sentimento:** Um *emoji* pode adicionar um sentimento a uma mensagem que, de outra forma, seria considerado neutro. Por exemplo, considere as mensagens (1) “Certo, vou hoje à noite” e (2) “Certo, vou hoje à noite 😭”. A primeira mensagem apenas descreve uma ação enquanto que a última emite a ideia de reclamação.
- **Melhoria de Sentimento:** *Emojis* podem ser usados para melhorar um sentimento, considere os exemplos (1) “Isso não foi legal” e (2) “Isso não foi legal 😡”. Enquanto a primeira mensagem transmite um sentimento de descontentamento em relação a algo, a segunda aprimora o sentimento, deixando-o muito mais claro.
- **Modificação de Sentimento:** Um *emoji* também pode ser usado para modificar um sentimento presente. O autor fornece a mensagem como exemplo “Estou tão triste que ele está morto 🤡”. O *emoji* rindo modifica o sentimento de tristeza do restante da frase e revela os sentimentos real da pessoa.


Tabela 2 – Alguns *emoji*.

	Rosto sorridente		Rosto de cabeça para baixo
	Rosto pensante		Careta
	Rosto mandando um beijo		Rosto sorridente com lágrima
	Cabeça explodindo		Rosto gritando de medo
	Rosto de palhaço		Rosto sorridente com corações

Fonte: <https://unicode.org/emoji/charts/emoji-list.html>.

Embora os *emotes* da *Twitch* sejam diferentes dos *emojis*, esses três casos de uso identificados pelo autor também se aplicam aos *emotes* do chat da *Twitch* e, portanto, será usado em nosso modelo de Análise de Sentimentos. A Tabela 3 fornece algumas mensagens de exemplos extraídas dos *chats* da *Twitch* que estão inclusas nesses três casos de uso.

Tabela 3 – Análise de Sentimentos de *emotes* da *Twitch*.

Expressão de Sentimento	“o que é isso  ”
	“eu sou muito burro para pensar  ”
	“Os irmãos estão de volta  ”
Melhoria de Sentimento	“isso não parece nada bom  ”
	“caramba, esses gráficos parecem bons  ”
	“Ele perdeu isso  ”
Modificação de Sentimento	“Boa produção  ”
	“tão animado  ”
	“tão interessante  ”

Fonte: (REIS, 2020).

### 2.3 *League of Legends*

O jogo *League of Legends*<sup>11</sup> (mais conhecido como LoL) é desenvolvido pela Riot Games e foi lançado em outubro de 2009. É um dos mais populares jogos de *e-sports*, contando com mais de 32 milhões de contas registradas e com média 11.5 milhões de jogadores mensais. A final do torneio de *e-sports* em novembro de 2021 atraiu um pico de mais de 73 milhões de espectadores simultâneos<sup>12</sup>.

<sup>11</sup> <https://www.leagueoflegends.com/>

<sup>12</sup> <https://www.sportsbusinessjournal.com/Esports/Sections/Media/2021/11/Worlds-2021-Finals-AMA.aspx>

Neste jogo, duas equipes de cinco jogadores cada competem em partidas que duram entre 20 e 40 minutos em média, durante as quais os participantes devem empregar estratégia e técnica para obter vantagens e, finalmente, destruir a base inimiga.

Muitos jogadores consideram *League of Legends* ser competitivo e viciante. Muitos passam muito tempo jogando para melhorar e subir de *rank* e esperam as performances ótimas de todos os outros jogadores. Algumas pessoas parecem amigáveis no começo, mas se tornam bastante tóxicas se você não se sair bem no jogo. Na cultura de jogo altamente competitiva de LoL, os jogadores aspiram a posições mais altas e usam as posições para mediar suas interações. Não é surpreendente observar como o *rank* dos jogadores também ajuda a induzir a toxicidade.

Esse comportamento é recorrente na comunidade em geral e acaba se estendendo para fora do jogo. Muitos jogadores acabam tendo esse tipo de atitude com os *streamers* do jogo. Piadas, insultos ou até mesmo ameaças são algumas formas de toxicidade que são aplicadas contra os *streamers* nas plataformas de transmissões ao vivo.

Figura 4 – Final do torneio mundial de *League of Legends* de 2021.



Fonte: Elaborado pelos autores.

### 3 TRABALHOS RELACIONADOS

Nesta seção são apresentados, de forma breve, algumas aplicações que possuem relação com o assunto a ser abordado neste artigo, no caso a Análise de Sentimentos.

No trabalho *Sentiment Analysis of Social Media Usernames and Titles on YouTube and Twitch* (NTUNGILA, 2021), é analisada a influência do sentimento do nome de usuário e do título do vídeo em relação à audiência. A pesquisa é realizada nas redes sociais *YouTube* e *Twitch*. Foram achadas evidências de que o sentimento e as emoções do título afetam a

audiência. Resultaram na criação de uma aplicação *web* que implantou os modelos de Análise de Sentimentos, como objetivo um ambiente de consultoria para sugerir mudança no título do vídeo para aumentar a audiência. A recomendação seria otimizar consistentemente a emoção “confiança” nos títulos. Este aplicativo *web* recebe como entrada um nome de usuário ou título de vídeo e retorna a classificação de sentimento. A análise dos dados nesse trabalho de pesquisa foi realizada usando *Python* como sua principal linguagem de programação.

*Twitter Sentiment Analysis* (SARLAN; NADAM; BASRI, 2014) é um trabalho, na área de Análise de Sentimento, realizado na rede social *Twitter* e que busca classificar sentimento a partir da extração de *tweets*. Aos *tweets* foram atribuídos valores positivos ou negativos a cada palavra, de acordo com o dicionário léxico. Conseqüentemente, os usuários e seus pensamentos podem ser classificados em positivo e negativo. Para isso, foi criado um programa que faz uso da abordagem de Aprendizado em Máquina, juntamente com técnicas de Processamento de Linguagem Natural.

Em *Text Mining of Audience Opinion in eSports Events* (MAINIERI et al., 2017), são analisados os *chats* do campeonato *League Championship Series (LCS)*, de *League of Legends*, um jogo online desenvolvido pela empresa *Riot Games*. A Análise de Sentimento é feita na plataforma *Twitch* com a abordagem baseada em léxico e com o objetivo de entender a expressão opinativa dos usuários. Na análise são considerados tanto texto como os *emotes*. São buscados detectar os times e os jogadores mais populares com base nas mensagens de texto, aos que podem se tornar opções mais interessantes para potenciais patrocinadores. Nessa abordagem são utilizadas técnicas das áreas de *Big Data* e *Text Mining*. As operações de contagem de menções e medição de opinião são aplicadas a um conjunto de mais de meio milhão de mensagens de *chat* coletadas em três finais de semana da competição. Os dados processados são representados graficamente com polaridade de sentimento positivo e negativo, expondo a existência de times favoritos dos torcedores e jogadores mais citados.

Na mesma linha de pesquisa, o trabalho *Sentiment Analysis: The Case of Twitch Mining user feedback from livestream chats* (REIS, 2020) é concentrado em desenvolver modelo de Análise de Sentimentos na *Twitch*. O estudo é feito utilizando abordagens baseadas em léxico e em regras, capazes de interpretar mensagens de bate-papo do usuário e considerando textos e *emotes*, a fim de identificar se as mensagens são negativas, positivas ou neutras. O modelo foi testado nos eventos de lançamento de produtos da *Blizzcon*, *Nike* e *E3*, todos no ano de 2018. Também, foi desenvolvida uma aplicação para melhorar a visualização e análise dos resultados do modelo.

No artigo *Towards the Understanding of Gaming Audiences by Modeling Twitch Emotes* (BARBIERI et al., 2017), é apresentando uma abordagem para o entendimento da mensagem implícita dos *emotes* da *Twitch*. O artigo tem como alvo o problema de predição de *emotes*, onde dada uma coleção de mensagens, prever qual *emote* o usuário é mais provável usar. A base de dados contém mensagens do bate-papo dos 300 canais mais populares da época, coletados no período de cinco meses. Nos experimentos foram avaliados a eficiência de três modelos de Aprendizado de Máquina.

Por fim, *Toxic Comment Detection in Online Discussions* (RISCH; KRESTEL, 2020), é um estudo que tem como objetivo a identificação de comentários tóxicos, respectivamente com as subclasses: linguagem obscena, insultos, ameaças e discurso de ódio. São colocadas em estudo várias abordagens de Aprendizado Profundo e Redes Neurais, adaptadas à Análise de Sentimentos em discussões *online*.

#### 4 PROPOSTA

Esse trabalho tem como metodologia fazer uso da Abordagem Baseado em Léxico, para realizar a Análise de Sentimentos das mensagens da *Twitch* utilizando como referência textos e *emotes* enviados pelos usuários, com o uso de regras que permitam ter uma classificação dessas mensagens. Em especial, o trabalho tem como objetivo realizar a tarefa de Detecção de Emoções dos *chats* da *Twitch*. Tendo como proposta fazer a análise individualmente de mensagens e *emotes* e também fazer a análise conjunta, pois a união de textos com *emotes* podem mudar o significado geral do sentimento da mensagem.

O projeto foi conduzido seguindo as etapas mostradas na Figura 5. Elas consistem (i) na coleta de dados dos *chats* das *lives* da *Twitch*; (ii) na análise das mensagens dos usuários através de um modelo de classificação; e (iii) na visualização dos dados resultantes da classificação. Essas etapas são descritas na sequência.

Figura 5 – Fluxo da Metodologia.



Fonte: Elaborado pelos autores.

A ideia principal dessa análise é identificar comportamentos tóxicos dos usuários da *Twitch*. Para isso, primeiramente classificamos as mensagens como emoções positivas (por exemplo, confiança, admiração, antecipação, surpresa, amor e alegria), negativas (por exemplo, medo, raiva, tristeza e nojo) ou neutras. Dentre as emoções negativas, classificamos quais delas caracterizam-se como tóxicas. Por exemplo, tristeza é uma emoção negativa, mas nem sempre podemos classificá-la como sendo tóxica. Já a emoção de raiva podemos caracterizar como sendo certamente tóxica. Assim, esperamos conseguir identificar quem são esses usuários tóxicos e que tipo de mensagem eles enviam, também verificando se é um comportamento recorrente. Além disso, pretendemos identificar quais os tipos de mensagens tóxicas são mais comuns e em que determinado contexto ou frequência elas aparecem.

Com a aplicação dessa técnica pretende-se ter um *feedback* automático mais profundo sobre a comunidade da *Twitch*, em especial aos espectadores do jogo *League of Legends*, o que seria inviável fazer de modo manual pelo grande fluxo de mensagens compartilhada por segundo. Assim, o *streamer* conseguirá entender em geral o sentimento que sua comunidade expressa e possível elaborar uma estratégia para evitar alguns tipos de comportamento, por exemplo, punindo usuários que tenham um comportamento tóxico.

#### 4.1 Coleta de Dados

Para a coleta de dados foi utilizada a ferramenta *RechatTool*<sup>13</sup>. Executada por linha de comando, a ferramenta usa o *link* de uma *live* salva no canal de um *streamer* para baixar o *log* de *chat*. Por exemplo, o comando

```
.\RechatTool.exe -d https://www.twitch.tv/videos/1311229744
```

Faz o *download* da reprodução do *chat* do ID de vídeo 1311229744 da *Twitch* e salva com saída um arquivo no formato JSON processado no diretório atual. O arquivo contém uma série de dados referentes aos estados do *chat* (ver Figura 6).

Figura 6 – Formato do arquivo JSON retornado pela ferramenta *RechatTool*.

```

1  [{
2    "id": "9eedb15-0ee3-4216-a728-24a89a4ffcba",
3    "created_at": "2022-05-28T09:25:12.128Z",
4    "updated_at": "2022-05-28T09:25:12.128Z",
5    "channel_id": "51496027",
6    "content_type": "video",
7    "content_id": "1311229744",
8    "content_offset_seconds": 2.428,
9    "commenter": {
10     "display_name": "manicx90",
11     "id": "26706468",
12     "name": "manicx90",
13     "type": "user",
14     "bio": "Turkish Master toss living in GermanyPlaying 1v1's and teamgames :)",
15     "created_at": "2011-12-10T22:58:40.990177Z",
16     "updated_at": "2022-06-07T17:34:29.538103Z",
17     "logo": "https://static-cdn.jtvnw.net/jtv_user_pictures/manicx90-profile_image-d210c2eb0de86379-30
18   },
19   "source": "chat",
20   "state": "published",
21   "message": {
22     "body": "@metotoro thats what he said too",
23     "fragments": [{"text": "@metotoro thats what he said too"}],
24     "is_action": false,
25     "user_badges": [{"_id": "premium", "version": "1"}],
26     "user_color": "#8A2BE2", "user_notice_params": {}
27   }
28 }
29 ...]

```

Fonte: Elaborado pelos autores.

Os dados utilizados do arquivo JSON para criar a base de dados do problema são: *name* (nome do usuário), *body* (mensagem do usuário) e *created\_at* (hora que a mensagem foi enviada), conforme mostrado na Figura 7. A base de dados e o desenvolvimento do restante do trabalho

<sup>13</sup> <https://github.com/jdpurcell/RechatTool>

foi realizado com a linguagem R (TEAM, 2013), que se trata de um ambiente de *software* livre para computação estatística e gráficos.

Figura 7 – Arquivo com os dados organizados na Linguagem R.

	name	body	created_at
1	manicx90	@metototo thats what he said too	2022-05-28T09:25:12.128Z
2	eric4211	Pog	2022-05-28T09:25:12.275Z
3	id_noided	Pog	2022-05-28T09:25:12.509Z
4	snugglestheteddy	Pog	2022-05-28T09:25:12.608Z
5	jusherx	huh	2022-05-28T09:25:13.682Z
6	swoopityswoop_	SUP SUP	2022-05-28T09:25:14.716Z
7	greeneggsandslam	Pog	2022-05-28T09:25:14.757Z
8	thebostonmarket	live fr holy shit	2022-05-28T09:25:14.953Z
9	slymuffins	Pog	2022-05-28T09:25:15.099Z
10	manicx90	nvm he went live	2022-05-28T09:25:15.339Z
11	crayo0	!BigBroher	2022-05-28T09:25:15.642Z
12	swoopityswoop_	Pog	2022-05-28T09:25:15.794Z

Showing 1 to 13 of 102,627 entries, 3 total columns

Fonte: Elaborado pelos autores.

Como objeto de estudo, os dados para análise foram coletados do canal do *streamer* loltyler1<sup>14</sup>, um dos mais famosos *streamers* americano de *League of Legends*. Ao todo, foram coletados dados de cinco *lives*, ocorridas durante o período de 28 de maio de 2022 até 01 de junho do mesmo ano. A Tabela 4 mostra o total de mensagens de cada *live*, totalizando 505.614 mensagens ao longo dos cinco dias de *live*.

Tabela 4 – Quantidade de mensagens de cada live durante os cinco dias.

Live	Data	Quantidade de Horas	Mensagens
1	28/05/2022	12h 53m 48s	102.627
2	29/05/2022	16h 12m 18s	103.165
3	30/05/2022	9h 59m 24s	73.834
4	31/05/2022	12h 35m 38s	92.828
5	01/06/2022	18h 40m 26s	133.160

Fonte: Elaborado pelos autores.

## 4.2 Análise de Dados

### 4.2.1 Preparação de Dados

O presente trabalho foi desenvolvido na versão 4.1.2 da linguagem de programação R, com a utilização principal das bibliotecas *dplyr*, *tidyr* e *tidytext*. Estas bibliotecas oferecem ferramentas para operar sobre entradas de texto, dividindo-as em palavras únicas, removendo palavras de parada, colocando-as em tabelas e, em seguida, executando operações tradicionais de

<sup>14</sup> <https://www.twitch.tv/loltyler1>

tabelas, como junções entre tabelas (SILGE; ROBINSON, 2017). As operações citadas consistem em técnicas de mineração de texto, contagem de frequência e avaliação de opinião.

Primeiramente, o arquivo JSON é transformado em uma tabela, com as instâncias já mencionadas. É feita a tokenização das tabelas, que é o processo de dividir os textos em *tokens*. Um *token* é uma parte do texto completo, como uma palavra, que estamos interessados em usar para a análise. Assim, o formato dos nossos dados organizados fica definido como um *token* por linha na tabela. O *token* armazenado em cada linha geralmente é uma única palavra, mas também pode ser um *n*-grama (uma junção de *n tokens*), que será discutido mais a frente.

Em seguida, é realizada uma limpeza nos dados, retirando palavras que não tem valor semântico de sentimento, são elas: palavras de parada, números, menções de usuários, robôs de assistente de *chat*, nomes de personagens e de termos específicos do jogo *League of Legends*. A Figura 10 mostra um exemplo de tokenização e limpeza de dados de uma *live*. Por exemplo, a mensagem “*thats what he said too*” do usuário *manicx90* foi retirada pela limpeza, pois cada token criado dessa mensagem (‘*thats*’, ‘*what*’, ‘*he*’, ‘*said*’, ‘*too*’) se trata de uma palavra de parada e não apresenta sentido ou emoção. Em geral, palavras como substantivos, pronomes, conjunções, advérbios são retiradas nessa etapa.

Figura 8 – Tabelas de antes e depois do processo de tokenização e limpeza.

name	body	created_at	name	created_at	word
1 manicx90	thats what he said too	2022-05-28T09:25:12.128Z	1 eric4211	2022-05-28T09:25:12.275Z	pog
2 eric4211	Pog	2022-05-28T09:25:12.275Z	2 id_noided	2022-05-28T09:25:12.509Z	pog
3 id_noided	Pog	2022-05-28T09:25:12.509Z	3 snugglestheteddy	2022-05-28T09:25:12.608Z	pog
4 snugglestheteddy	Pog	2022-05-28T09:25:12.608Z	4 jusherx	2022-05-28T09:25:13.682Z	huh
5 jusherx	huh	2022-05-28T09:25:13.682Z	5 greeneggsandslam	2022-05-28T09:25:14.757Z	pog
6 swoopityswoop_	sup sup	2022-05-28T09:25:14.716Z	6 thebostonmarket	2022-05-28T09:25:14.953Z	live
7 greeneggsandslam	Pog	2022-05-28T09:25:14.757Z	7 thebostonmarket	2022-05-28T09:25:14.953Z	fr
8 thebostonmarket	live fr holy shit	2022-05-28T09:25:14.953Z	8 thebostonmarket	2022-05-28T09:25:14.953Z	holy
9 slymuffins	Pog	2022-05-28T09:25:15.099Z	9 thebostonmarket	2022-05-28T09:25:14.953Z	shit
10 manicx90	nvm he went live	2022-05-28T09:25:15.339Z	10 slymuffins	2022-05-28T09:25:15.099Z	pog
11 crayo0		2022-05-28T09:25:15.642Z	11 manicx90	2022-05-28T09:25:15.339Z	nvm

Showing 1 to 12 of 96,903 entries, 3 total columns

Showing 1 to 12 of 173,261 entries, 3 total columns

Fonte: Elaborado pelos autores.

Na Tabela 5 pode ser observada a quantidade de *tokens* brutos e quantos restaram após a limpeza. É notório afirma que mais da metade dos *tokens* não apresentam valor semântico.



Tabela 5 – Quantidade de *tokens* antes e depois do processo de limpeza.

Live	Data	Mensagens	<i>Tokens</i> Brutos	<i>Tokens</i> Tratados
1	28/05/2022	102.627	380.926	173.281
2	29/05/2022	103.165	406.239	178.142
3	30/05/2022	73.834	311.781	141.826
4	31/05/2022	92.828	422.357	190.317
5	01/06/2022	133.160	553.078	246.243

Fonte: Elaborado pelos autores.

Para realizar a Análise de Sentimentos, foi utilizado o dicionário léxico NRC *Emotion Lexicon*<sup>15</sup>. Ele contém um lista de palavras em inglês, totalizando 14.182 unigramas (*tokens*) e suas associações com dois sentimentos (negativo e positivo) e oito emoções básicas (raiva, medo, antecipação, confiança, surpresa, tristeza, alegria e nojo). Todos os sentimentos e emoções estão associados a cada unigrama presente no dicionário léxico. Os unigramas que não têm sentimento ou emoção associado a eles recebem valor zero para aquele sentimento ou emoção. Por exemplo, na Figura 9, a palavra “*abandoned*” está associada aos sentimentos e emoções: negativo, raiva, medo e tristeza.

Figura 9 – Dicionário léxico de sentimentos NRC.

	A	B	C	D		A	B	C	D		A	B	C	D
1	word	emotion	value	type	23	abandon	anticipation	0	text	45	abandonment	fear	1	text
2	aback	anger	0	text	24	abandon	disgust	0	text	46	abandonment	joy	0	text
3	aback	anticipation	0	text	25	abandon	fear	1	text	47	abandonment	negative	1	text
4	aback	disgust	0	text	26	abandon	joy	0	text	48	abandonment	positive	0	text
5	aback	fear	0	text	27	abandon	negative	1	text	49	abandonment	sadness	1	text
6	aback	joy	0	text	28	abandon	positive	0	text	50	abandonment	surprise	1	text
7	aback	negative	0	text	29	abandon	sadness	1	text	51	abandonment	trust	0	text
8	aback	positive	0	text	30	abandon	surprise	0	text	52	abate	anger	0	text
9	aback	sadness	0	text	31	abandon	trust	0	text	53	abate	anticipation	0	text
10	aback	surprise	0	text	32	abandoned	anger	1	text	54	abate	disgust	0	text
11	aback	trust	0	text	33	abandoned	anticipation	0	text	55	abate	fear	0	text
12	abacus	anger	0	text	34	abandoned	disgust	0	text	56	abate	joy	0	text
13	abacus	anticipation	0	text	35	abandoned	fear	1	text	57	abate	negative	0	text
14	abacus	disgust	0	text	36	abandoned	joy	0	text	58	abate	positive	0	text
15	abacus	fear	0	text	37	abandoned	negative	1	text	59	abate	sadness	0	text
16	abacus	joy	0	text	38	abandoned	positive	0	text	60	abate	surprise	0	text
17	abacus	negative	0	text	39	abandoned	sadness	1	text	61	abate	trust	0	text
18	abacus	positive	0	text	40	abandoned	surprise	0	text	62	abatement	anger	0	text
19	abacus	sadness	0	text	41	abandoned	trust	0	text	63	abatement	anticipation	0	text
20	abacus	surprise	0	text	42	abandonment	anger	1	text	64	abatement	disgust	0	text
21	abacus	trust	1	text	43	abandonment	anticipation	0	text	65	abatement	fear	0	text
22	abandon	anger	0	text	44	abandonment	disgust	0	text	66	abatement	joy	0	text

Fonte: Elaborado pelos autores.

Além disso, é proposta uma abordagem semelhante usando *emotes* para o nosso modelo de Análise de Sentimentos, que é uma extensão de sentimentos e emoções com *emoticons* e *emojis*. Foram utilizados apenas os *emotes* globais da *Twitch*, pois os *emotes* de assinantes e *emotes* de terceiros são essencialmente inviáveis de extrair, pela grande quantidade, assim como também pela sua mudança constante.

<sup>15</sup> <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>

Para isso, foi criado um dicionário léxico com 281 *emotes* globais nativos da *Twitch*, com as mesmas polaridades de sentimentos e emoções presentes no dicionário léxico NRC, fazendo apenas a adição de apenas um terceiro sentimento, chamado “*posneg*”. Os *emotes* que foram definidos com esse sentimento são caracterizados por terem um duplo sentido de sentimento, positivo e/ou negativo. O sentimento expresso em uma mensagem pode sofrer modificação de sentido dependendo da presença desses tipos de *emotes*. Esse ponto será discutido mais adiante quando for abordada a classificação de sentimentos tóxicos. Por fim, esse novo dicionário léxico de *emotes* criado foi agregado ao dicionário léxico NRC existente. Alguns exemplos do dicionário léxico de *emotes* podem ser visto na Figura 10.

Figura 10 – Dicionário Léxico de emoção de *emotes* criado a partir da *Twitch*.

	A	B	C	D		A	B	C	D		A	B	C	D
144083	omegalul	anger	0	emote	144105	sadge	anger	0	emote	144127	copeler1	anger	0	emote
144084	omegalul	anticipation	0	emote	144106	sadge	anticipation	0	emote	144128	copeler1	anticipation	0	emote
144085	omegalul	disgust	0	emote	144107	sadge	disgust	0	emote	144129	copeler1	disgust	0	emote
144086	omegalul	fear	0	emote	144108	sadge	fear	0	emote	144130	copeler1	fear	0	emote
144087	omegalul	joy	0	emote	144109	sadge	joy	0	emote	144131	copeler1	joy	0	emote
144088	omegalul	negative	0	emote	144110	sadge	negative	1	emote	144132	copeler1	negative	1	emote
144089	omegalul	positive	0	emote	144111	sadge	positive	0	emote	144133	copeler1	positive	0	emote
144090	omegalul	sadness	0	emote	144112	sadge	sadness	1	emote	144134	copeler1	sadness	1	emote
144091	omegalul	surprise	0	emote	144113	sadge	surprise	0	emote	144135	copeler1	surprise	0	emote
144092	omegalul	trust	0	emote	144114	sadge	trust	0	emote	144136	copeler1	trust	0	emote
144093	omegalul	posneg	1	emote	144115	sadge	posneg	0	emote	144137	copeler1	posneg	0	emote
144094	ez	anger	0	emote	144116	lol	anger	0	emote	144138	copium	anger	0	emote
144095	ez	anticipation	0	emote	144117	lol	anticipation	0	emote	144139	copium	anticipation	0	emote
144096	ez	disgust	0	emote	144118	lol	disgust	0	emote	144140	copium	disgust	0	emote
144097	ez	fear	0	emote	144119	lol	fear	0	emote	144141	copium	fear	0	emote
144098	ez	joy	0	emote	144120	lol	joy	0	emote	144142	copium	joy	0	emote
144099	ez	negative	1	emote	144121	lol	negative	0	emote	144143	copium	negative	1	emote
144100	ez	positive	0	emote	144122	lol	positive	1	emote	144144	copium	positive	0	emote
144101	ez	sadness	0	emote	144123	lol	sadness	0	emote	144145	copium	sadness	1	emote
144102	ez	surprise	0	emote	144124	lol	surprise	0	emote	144146	copium	surprise	0	emote
144103	ez	trust	0	emote	144125	lol	trust	0	emote	144147	copium	trust	0	emote
144104	ez	posneg	0	emote	144126	lol	posneg	0	emote	144148	copium	posneg	0	emote

Fonte: Elaborado pelos autores.

#### 4.2.2 Análise de Sentimentos

A etapa de Análise de Sentimentos tem como objetivo principal identificar as palavras em três casos de uso diferentes: Expressão de Sentimento, Modificação de Sentimento e Melhoria de Sentimentos. Esses casos de uso serão aplicados através da análise de unigramas e bigramas. Vale ressaltar que pode haver um abuso de notação em relação às palavras sentimento e emoções, mas em geral destacamos sentimentos como as polaridades positivas, negativas e *posneg* e emoções destacamos como as oito emoções básicas (raiva, medo, antecipação, confiança, surpresa, tristeza, alegria e nojo).

Após o processo de tokenização, é feita a classificação do sentimento e emoção de cada palavra (também chamada de unigrama), que compreendem textos ou *emotes*, presentes nas nossas tabelas usando o dicionário léxico NRC. Cada unigrama classifica uma palavra através da sua Expressão de Sentimento. O exemplo da Figura 11 ilustra o resultado dessa etapa. Pode-se ver que nas linhas 1 e 2 da tabela mostrada nessa figura o usuário “eric4211” enviou o *emote* “pog” que está classificado como positivo e surpresa.

Figura 11 – Cada unigrama com sentimento e/ou emoção.

	name	created_at	word	emotion		name	created_at	word	emotion
1	eric4211	2022-05-28T09:25:12.275Z	pog	positive	12	thebostonmarket	2022-05-28T09:25:14.953Z	shit	negative
2	eric4211	2022-05-28T09:25:12.275Z	pog	surprise	13	slymuffins	2022-05-28T09:25:15.099Z	pog	positive
3	ld_noided	2022-05-28T09:25:12.509Z	pog	positive	14	slymuffins	2022-05-28T09:25:15.099Z	pog	surprise
4	ld_noided	2022-05-28T09:25:12.509Z	pog	surprise	15	swoopityswoop_	2022-05-28T09:25:15.794Z	pog	positive
5	snugglestheteddy	2022-05-28T09:25:12.608Z	pog	positive	16	swoopityswoop_	2022-05-28T09:25:15.794Z	pog	surprise
6	snugglestheteddy	2022-05-28T09:25:12.608Z	pog	surprise	17	liverism	2022-05-28T09:25:16.616Z	pog	positive
7	greeneggsandslam	2022-05-28T09:25:14.757Z	pog	positive	18	liverism	2022-05-28T09:25:16.616Z	pog	surprise
8	greeneggsandslam	2022-05-28T09:25:14.757Z	pog	surprise	19	slymuffins	2022-05-28T09:25:17.402Z	pog	positive
9	thebostonmarket	2022-05-28T09:25:14.953Z	holly	positive	20	slymuffins	2022-05-28T09:25:17.402Z	pog	surprise
10	thebostonmarket	2022-05-28T09:25:14.953Z	shit	anger	21	acesomeone	2022-05-28T09:25:17.41Z	pog	positive
11	thebostonmarket	2022-05-28T09:25:14.953Z	shit	disgust	22	acesomeone	2022-05-28T09:25:17.41Z	pog	surprise

Fonte: Elaborado pelos autores.

A Figura 12 mostra a proporção de categorias dos dados classificados e não classificados durante as cinco *lives*. É possível observar que os *emotes* são bastante utilizados pelos usuários e em média compreende que é proporcional a quantidade de palavras de todo o conjunto de dados classificados, confirmando que são relevantes para serem analisados.

Figura 12 – Categorias dos dados classificados (*emote* e *texto*) e não classificados (NA) de cada *live*.



Fonte: Elaborado pelos autores.

Apesar da classificação do unigrama relacionar cada texto ou *emote* a um ou mais sentimentos e emoções, não necessariamente eles vão expressar seus sentimentos reais, pois vão depender do contexto onde estão inseridas, ou seja, de que palavras ou *emotes* elas estão acompanhadas. Logo, pode ser que venha ocorrer alguns casos de unigramas falsos-positivos ou falsos-negativos no contexto geral da frase. Nesses casos há a necessidade de criar um modelo baseado em regras, para a classificação correta das mensagens. São estabelecidas cinco regras para corrigir possíveis casos de falsos positivos e negativos. As regras são aplicadas em bigramas (uma sequência de dois *tokens*).

Essas regras são usadas para encontrar padrões combinando palavras e *emotes*, a partir de sequências de palavras específicas que podem ser consideradas indicador se a mensagem é positiva ou negativa (sendo a mensagem tóxica incluída nessa categoria). Essas sequências funcionam em uma junção de uma palavra da esquerda para a direita. Algumas das regras abaixo tiveram como inspiração o trabalho apresentado em (REIS, 2020).

Foram pensadas duas categorias de regras para os bigramas: Modificação de Sentimento (regras 1 e 2) e Melhoria de Sentimento (regras 3 e 4), que serão descritas na sequência.

**REGRA 1** (palavra **not** + palavra **positiva**  $\Rightarrow$  emoção **negativa**) (1)

(palavra **not** + palavra **negativa**  $\Rightarrow$  emoção **positiva**) (2)

Nessa regra são tratadas palavras que tem como antecessor palavras do tipo *not*. No inglês as palavras que contém alguma espécie de *not* são: *not, cannot, ain't, aren't, can't, couldn't, didn't, doesn't, don't, hadn't, hasn't, haven't, isnt, mustn't, shan't, shouldn't, wasn't, weren't, won't e wouldn't*. Essas palavras foram adicionas no dicionário léxico NRC com a emoção do tipo “*negation*”. Como esses termos têm a função linguística de negar palavras que apareçam na sequência, é feita a correção com a emoção inversa que primeiramente elas foram classificadas (ver equações (1) e (2)). Por exemplo, se palavra sucessora do *not* foi classificada como positiva, com a aplicação dessa regra ela vai passar a ser negativa. Alguns resultados de sentimentos e emoções inversos após da aplicação da **REGRA 1** podem ser vistos na Figura 13.

Figura 13 – Resultados da aplicação da **REGRA 1**.

	name	created_at	word1	word2	emotion2	emotion
1	mudkeepmudkeep	2022-05-28T09:43:49.337Z	not	including	positive	negative
2	blueblooper	2022-05-28T09:44:40.664Z	wont	ban	negative	positive
3	aashas_	2022-05-28T09:44:46.656Z	not	durability	positive	negative
4	aashas_	2022-05-28T09:44:46.656Z	not	durability	trust	fear
5	blueblooper	2022-05-28T09:47:24.778Z	arent	stupid	negative	positive
6	blueblooper	2022-05-28T09:47:51.268Z	dont	ban	negative	positive
7	aashas_	2022-05-28T09:48:06.166Z	not	worthless	anger	love
8	aashas_	2022-05-28T09:48:06.166Z	not	worthless	disgust	appreciation
9	aashas_	2022-05-28T09:48:06.166Z	not	worthless	negative	positive
10	aashas_	2022-05-28T09:48:06.166Z	not	worthless	sadness	joy
11	tomm0_____	2022-05-28T09:57:33.734Z	not	ff	negative	positive
12	fygonjinnnn	2022-05-28T09:59:05.62Z	dont	worry	anticipation	surprise

Showing 1 to 13 of 1,035 entries, 6 total columns

Vale ressaltar que foram adicionadas duas novas emoções nos resultados da inversão da **REGRA 1**, as quais não existiam previamente no dicionário NRC, são elas amor e apreciação. O resultado da inversão das emoções pode ser visto na Tabela 6.

Tabela 6 – Sentimentos e Emoções inversos.

Emoção	Inverso
<i>positive</i>	<i>negative</i>
<i>negative</i>	<i>positive</i>
<i>trust</i>	<i>fear</i>
<i>anticipation</i>	<i>surprise</i>
<i>joy</i>	<i>sadness</i>
<i>surprise</i>	<i>anticipation</i>
<i>fear</i>	<i>trust</i>
<i>anger</i>	<i>love</i>
<i>sadness</i>	<i>joy</i>
<i>disgust</i>	<i>appreciation</i>

Fonte: Elaborado pelos autores.

**REGRA 2** (palavra **positiva** + *emote* **posneg** ⇒ emoção **ironia**) (3)

(palavra **negativa** + *emote* **posneg** ⇒ emoção **humor**) (4)

Diferente da **REGRA 1** que analisa duas palavras, na **REGRA 2** são analisadas uma palavra e um *emote*. Como dito anteriormente, nesse trabalho existem três categorias de sentimentos referentes os *emotes* (*positive*, *negative* e *posneg*). Os *emotes* que tem o sentimento *posneg* podem ter duplo sentido, possivelmente modificando o sentimento da palavra anterior. Na equação (3) são tratados os casos em que se tem palavra com sentimento ou emoção positivos seguidos por um *emote* com o sentimento *posneg*. É feita a correção desses casos classificando como a emoção *irony*. Já na equação (4) são tratados os casos em que se tem palavra com sentimento ou emoção negativos e com o sucessor *emote* com o sentimento *posneg*, no qual é feita a classificação da emoção como *humor*. Alguns resultados podem ser vistos na Figura 14.

Figura 14 – Resultados da aplicação da **REGRA 2**. À esquerda tem-se um resultado da equação (3) e à direita um resultado da equação (4).

name	word1	word2	emotion1	emotion2	emotion
1 orchidshrub	baby	bigbrother	joy	negative	irony
2 jimcarry32467	diamond	bigbrother	joy	negative	irony
3 si4yerkid	diamond	copeler1	joy	negative	irony
4 greeneggsandsiam	nervous	kappa	anticipation	posneg	irony
5 cavalo_preto	winning	3head	anticipation	negative	irony
6 akali_xxx_vayne_rul...	hug	omegalul	joy	posneg	irony
7 marqer1	diamond	omegalul	joy	posneg	irony
8 horsefightherr	diamond	4head	joy	negative	irony
9 horsefightherr	confirmed	4head	positive	negative	irony
10 flygonjinnnn	diamond	bigbrother	joy	negative	irony
11 mafuyui	clap	bigbrother	anticipation	negative	irony
12 for_demacia	lose	copeler1	surprise	negative	irony

Showing 1 to 12 of 440 entries, 6 total columns

name	word1	word2	emotion1	emotion2	emotion
1 areatangens	trip	kekww	surprise	positive	humor
2 itstoxicmascullinity	pressure	kekww	negative	positive	humor
3 fenex5	god	pepelaugh	fear	positive	humor
4 papiben	joke	kekww	negative	positive	humor
5 greeneggsandsiam	nervous	kappa	fear	posneg	humor
6 johntan0815	quiet	kekww	sadness	positive	humor
7 jackehoce	broken	lol	anger	positive	humor
8 daddy_christmas	ban	pepelaugh	negative	positive	humor
9 tiltedbrocollil	trash	kekww	disgust	positive	humor
10 paingzayn	gap	batchest	negative	positive	humor
11 kekww_spammer_	anxiety	kekww	anger	positive	humor
12 imaboomer	anxiety	kekww	anger	positive	humor

Showing 1 to 12 of 778 entries, 6 total columns

Fonte: Elaborado pelos autores.

**REGRA 3** (palavra **positiva** + *emote* **positivo**  $\Rightarrow$  emoção adicional **positiva**) (5)

(palavra **negativa** + *emote* **negativo**  $\Rightarrow$  emoção adicional **negativa**) (6)

Diferente das duas últimas regras citadas, onde é feita a modificação de emoção, na **REGRA 3** é feita uma adição da emoção já presente. Na equação (5) são tratados os casos em que se tem palavra com sentimento ou emoção positivos seguidos por um *emote* com emoção *positive*. É feita a adição de desses casos classificando como acréscimo sentimento *positive*. Logo na equação (6) são tratados os casos em que se tem palavra com emoção negativa seguida por um *emote* com a emoção *negative*. É feita a adição desses casos classificando com acréscimo de sentimento *negative*. Alguns resultados de sentimentos e emoções após da aplicação da **REGRA 3** podem ser vistos na Figura 15.

Figura 15 – Resultados da aplicação da **REGRA 3**. À esquerda tem-se um resultado da equação (5) e à direita um resultado da equação (6).

name	word1	word2	emotion1	emotion2	emotion
1	gues201	time pog	anticipation	positive	positive
2	areatangens	trip kekw	surprise	positive	positive
3	herosparky	time kekw	anticipation	positive	positive
4	ilverism	clap pepelaugh	anticipation	positive	positive
5	fenex5	god pepelaugh	anticipation	positive	positive
6	angeltheplayer	rising pog	anticipation	positive	positive
7	jshwan	safe kekw	joy	positive	positive
8	johnatan0815	quiet kekw	positive	positive	positive
9	kekvw_spammer_	airport kekw	anticipation	positive	positive
10	daddy_christmas	draft kekw	anticipation	positive	positive
11	paingzayn	child batchest	anticipation	positive	positive
12	kekvw_spammer_	anxiety kekw	anticipation	positive	positive

Showing 1 to 13 of 616 entries, 6 total columns

name	word1	word2	emotion1	emotion2	emotion
1	420promemer	gang ez	anger	negative	negative
2	itchyratt	lost sadge	negative	negative	negative
3	eil_timefly	blower bigbrother	negative	negative	negative
4	andrezinho_	shit bigbrother	anger	negative	negative
5	sevechenco	losing bigbrother	anger	negative	negative
6	mafuyui	loss bigbrother	anger	negative	negative
7	cavalo_preto	winning 3head	disgust	negative	negative
8	vivatroma	sin sadge	anger	negative	negative
9	for_demacia	lose copeler1	anger	negative	negative
10	seanp605	cannon sadge	anger	negative	negative
11	mafuyui	shit bigbrother	anger	negative	negative
12	flygonjinnn	doomed sadge	fear	negative	negative

Showing 1 to 13 of 339 entries, 6 total columns

Fonte: Elaborado pelos autores.

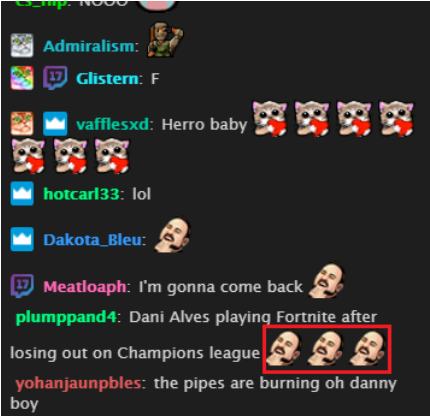
**REGRA 4** (*emote* **posneg** + *emote* **posneg**  $\Rightarrow$  emoção adicional **ironia**) (7)

Na **REGRA 4** são tratados os casos que possuem *emotes* com emoções *posneg* seguidos por *emotes* também do tipo *posneg*. Aos bigramas que caem nesse caso são feitas a adição da emoção *irony*. Nesse trabalho consideramos que quando os usuários da *Twitch* querem expressar sarcasmo, eles enviam uma sequência de *emotes* do tipo *posneg*, o que definido nesse trabalho com a emoção *ironia*. Alguns resultados de sentimentos e emoções após da aplicação da **REGRA 4** podem ser vistos na Figura 16, bem como um exemplo desse comportamento.

Figura 16 – Resultado da aplicação da **REGRA 4** à esquerda e à direita um exemplo de ironia no *chat* da *Twitch*.

name	word1	word2	emotion1	emotion2	emotion
1 swoopityswoop_	smorc	smorc	posneg	posneg	irony
2 akashi69245	lul	lul	posneg	posneg	irony
3 bloomstowerdefense1	smorc	smorc	posneg	posneg	irony
4 anonymouse_blaze	smorc	smorc	posneg	posneg	irony
5 tylerd12344	smorc	smorc	posneg	posneg	irony
6 bloomstowerdefense1	smorc	smorc	posneg	posneg	irony
7 ddickraida66	smorc	smorc	posneg	posneg	irony
8 rtsthatsreal	whysoserious	whysoserious	posneg	posneg	irony
9 iloveagua	kappa	kappa	posneg	posneg	irony
10 charmainediyoza	kappawealth	kappawealth	posneg	posneg	irony
11 rog_408	lul	lul	posneg	posneg	irony
12 herpasierp	lul	lul	posneg	posneg	irony

Showing 1 to 13 of 368 entries, 6 total columns

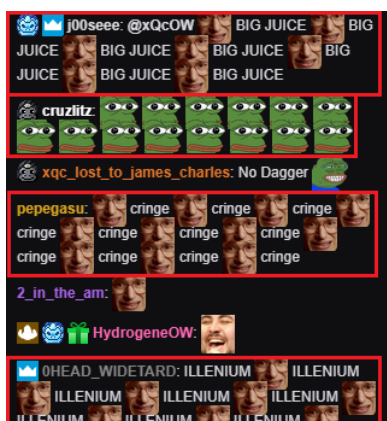


Fonte: Elaborado pelos autores.

**REGRA 5** ( $SPAM \geq 50 \text{ emotes} \Rightarrow$  remoção de sentimento e emoção.) (8)

Finalmente, na **REGRA 5** são tratados os casos em quem se tem uma sequência de cinquenta ou mais *emotes*. O comportamento de publicar *spams* (mensagens em massa) podem inflar o sentimento, que pode deixar modelo de Análise de Sentimento com irregularidades de classificação. Assim, é feita a remoção de todos os *emotes* classificados que se enquadram a **REGRA 5**. Na Figura 17 pode ser visto um exemplo desse comportamento.

Figura 17 – Exemplo de *spam* no *chat* da *Twitch*.



Fonte: Elaborado pelos autores.

### 4.3 Visualização de Dados

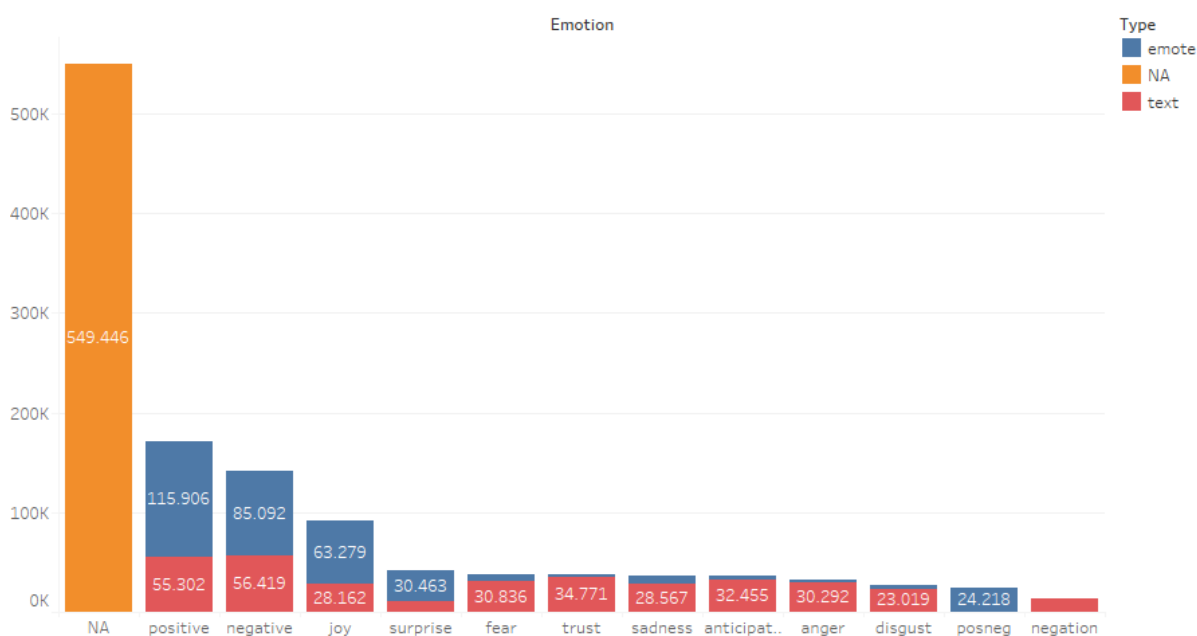
Para fazer a geração de todos os gráficos, foi utilizado a ferramenta *Tableau*<sup>16</sup>. Uma plataforma de análise visual com consultas de dados por meio de uma interface intuitiva. Os dados são apresentados visualmente traduzidos e serão mostrados na seção seguinte.

<sup>16</sup> <https://www.tableau.com/>

## 5 RESULTADOS

Nesta seção são apresentados os resultados obtidos de acordo com a metodologia criada para a obtenção dos sentimentos dos usuários dos *chats* da *Twitch*. Na Figura 18 são mostradas as quantidades de unigramas totais (929.809), divididos entre os categorizados (380.363) e os não categorizados (549.446) de todas as *lives* em cada tipo de emoção. Nela serão aplicadas as etapas de limpeza de dados e classificação de sentimento. Apesar da quantidade de palavras não classificadas serem maior que as classificadas, elas não têm relevância na análise, pois apenas aproximadamente 8.3% dessas palavras não classificadas possuem mais de 10 ocorrências, e classificá-las seria um trabalho exaustivo e não impactante no resultado final.

Figura 18 – Quantidade dos dados classificados e não classificados.

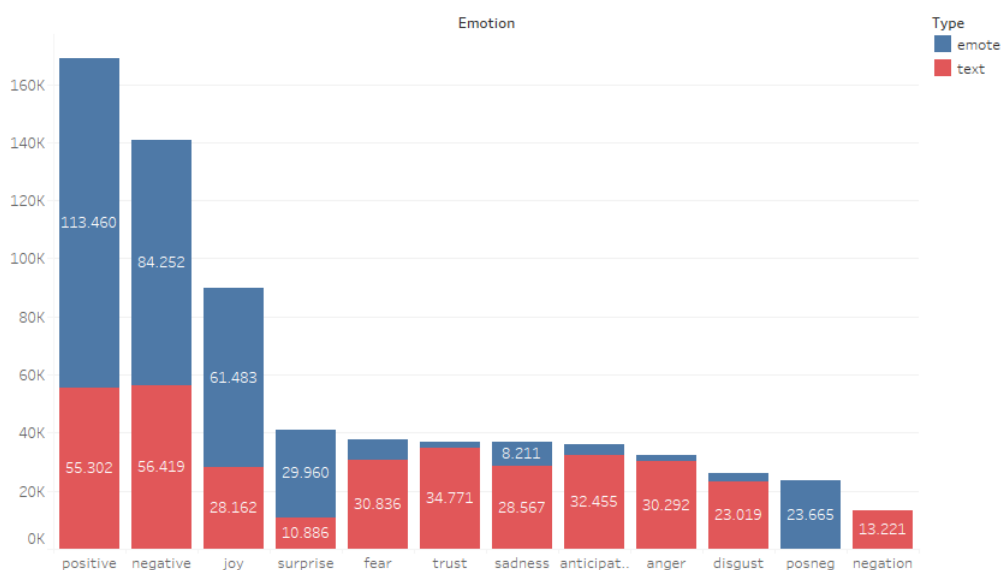


Fonte: Elaborado pelos autores.

A Figura 19 mostra os unigramas categorizados com a aplicação das REGRA 5, apresentada na Seção 4.2.2, que consiste na eliminação de *spams* dos *emotes*. É possível observar que todos sentimentos e emoções dos *emotes* tiveram redução. Vale ressaltar a emoção *posneg* criada, que contribuem para modificação e adição da emoção ironia na próxima tabela.



Figura 19 – Quantidade de unigramas classificados com aplicação da Regra 5.



Fonte: Elaborado pelos autores.

Já na Figura 20 temos duas tabelas: a primeira tabela representa o gráfico da Figura 18 e a tabela à direita representa a aplicação das Regras 1, 2, 3, 4 e 5, apresentadas na Seção 4.2.2. Nela é possível observar que todas as emoções existentes tiveram modificação em sua quantidade, bem como a aparição das emoções inversas adicionadas.

Figura 20 – Quantidade de unigramas classificados com aplicação das Regras 1,2,3,4 e 5.

Emotion	Quantity
anger	32.128
anticipation	36.099
disgust	26.220
fear	37.517
joy	89.645
negation	13.221
negative	140.671
positive	168.762
posneg	23.665
sadness	36.778
surprise	40.846
trust	37.027

➔

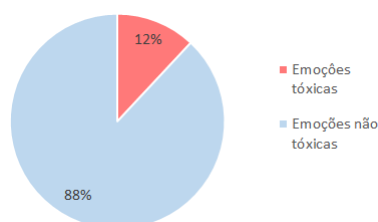
Emotion	Quantity
anger	29.320
anticipation	33.774
appreciation	415
disgust	24.129
fear	34.900
humor	3.887
irony	4.422
joy	88.427
love	631
negative	135.960
positive	166.103
sadness	34.059
surprise	39.813
trust	35.289

Fonte: Elaborado pelos autores.

Após a aplicação das regras, é possível observar a proporção de emoções tóxicas e não tóxicas compartilhadas no *chat* das cinco *lives* (ver Figura 21). Para este trabalho, são

consideradas emoções tóxicas as emoções do tipo *anger*, *disgust*, *irony* e *fear*. Da quantidade total de emoções classificadas (682.579), temos que 88% (600.669) correspondem a emoções não tóxicas e 12% (81.910) correspondem a emoções tóxicas.

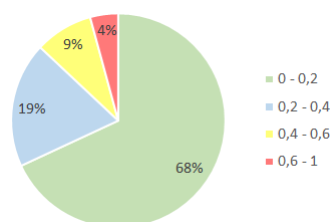
Figura 21 – Proporção de emoções tóxicas no *chat* das cinco *lives*.



Fonte: Elaborado pelos autores.

Por fim, foram detectados uma quantidade total de 31.837 usuários presentes durante as cinco *lives*, os quais tiveram alguma mensagem classificada em algum sentimento. Para cada usuário foi medida a proporção de emoções tóxicas de suas mensagens pela sua quantidade total de mensagens. Foi então gerada uma escala (ver Figura 22), que configura com as seguintes denominações: não tóxico (0-0.2), levemente tóxico (0.2-0.4), moderadamente tóxico (0.4-0.6) e tóxico (0.6-1). Os usuários que tiveram um comportamento tóxico, ou seja, maior que 60% das suas mensagens compartilhadas, correspondem a 4% do total de usuários (cerca de 1.273 usuários).

Figura 22 – Nível de comportamento tóxico dos usuários.



Fonte: Elaborado pelos autores.

## 6 CONCLUSÕES

Este trabalho apresentou um modelo de Análise de Sentimentos para detectar os sentimentos e emoções básicas e comportamentos tóxicos dos usuários nos *chats* da *Twitch* em cenários do jogo *League of Legends*. O modelo pretende funcionar como uma ferramenta auxiliar para os *streamers*, onde eles consigam saber o sentimento geral da sua comunidade e contornar possíveis usuários tóxicos dando penalidade a esses.

No decorrer deste estudo foram coletados dados dos *chats* compartilhados nas *lives* pela comunidade do *streamer* loltyler1, e desenvolvido um *script* na linguagem R (disponível no link: <https://github.com/dsenna/analise-de-sentimento-r>) capaz de fazer análise dessas mensagens

no formato de unigramas e bigramas. A análise foi cumprida com sucesso, confirmando a proposta inicial. Os resultados apresentados neste trabalho são promissores. O modelo é capaz de identificar as emoções da comunidade bem como detectar usuários com comportamentos tóxicos. Ele pode ser utilizado por todos e é funcional em qualquer cenário proposto no trabalho.

No decorrer desse trabalho foram encontradas algumas dificuldades, uma delas é a grande diversidade de *emotes* presentes na plataforma e a adição constantemente de novos por terceiros. Contornamos essa dificuldade tratando apenas os *emotes* globais permanentes nesse trabalho, mas não é descartado que a adição dos *emotes* de terceiros. É possível acrescentar ainda mais ao resultado do modelo. Outro ponto é a abordagem léxica estar limitada a apenas ao sentimento específico da palavra, o que implica no sentimento geral da frase.

Como trabalhos futuros, além de criação de novas regras e melhorias nas existentes, pretende-se disponibilizar o modelo de forma *online*, criando uma página *web* ou uma extensão de navegador, onde é recebido o *link* do vídeo salvo na *Twitch* e retornando informações dos sentimentos presentes no *chat* e os usuários com maior taxa de comportamento tóxico referente ao vídeo pesquisado. Também é vislumbrado utilizar outras abordagens, como Aprendizado de Máquina, para melhorar a classificação das emoções dos usuários.

## REFERÊNCIAS

- ASGHAR, M. Z. et al. Sentiment analysis on youtube: A brief survey. **arXiv preprint arXiv:1511.09142**, 2015.
- BARBIERI, F. et al. Towards the understanding of gaming audiences by modeling twitch emotes. In: ACL (ASSOCIATION FOR COMPUTATIONAL LINGUISTICS). **Third Workshop on Noisy User-generated Text (W-NUT 2017); 2017 Sep 7; Copenhagen, Denmark. Stroudsburg (PA): ACL; 2017. p. 11-20.** [S.l.], 2017.
- BHONDE, S. B.; PRASAD, J. R. Sentiment analysis-methods, applications & challenges. **International Journal of Electronics Communication and Computer Engineering**, International Journal of Electronics Communication and Computer Engineering , v. 6, n. 6, p. 634, 2015.
- BROWN, D. et al. Computers and qda—can they help it? a report on a qualitative data analysis programme. **The Sociological Review**, SAGE Publications Sage UK: London, England, v. 38, n. 1, p. 134–150, 1990.
- CANHOTO, A. I.; PADMANABHAN, Y. we (dont) know how you feel—a comparative study of automated vs. manual analysis of social media conversations. **Journal of Marketing Management**, Taylor & Francis, v. 31, n. 9-10, p. 1141–1157, 2015.
- CRISTIANI, A.; LIEIRA, D.; CAMARGO, H. A sentiment analysis of brazilian elections tweets. In: SBC. **Anais do VIII Symposium on Knowledge Discovery, Mining and Learning.** [S.l.], 2020. p. 153–160.
- EWALT, D. M. **The ESPN of videogames.** [S.l.]: FORBES INC 60 FIFTH AVE, NEW YORK, NY 10011 USA, 2013. 36–36 p.

GANDOLFI, E. To watch or to play, it is in the game: The game culture on twitch. tv among performers, plays and audiences. **Journal of Gaming & Virtual Worlds**, Intellect, v. 8, n. 1, p. 63–82, 2016.

GODBOLE, N.; SRINIVASIAH, M.; SKIENA, S. Large-scale sentiment analysis for news and blogs. **Icwsn**, Citeseer, v. 7, n. 21, p. 219–222, 2007.

GUIBON, G.; OCHS, M.; BELLOT, P. From Emojis to Sentiment Analysis. In: Lab-STICC and ENIB and LITIS. **WACAI 2016**. Brest, France, 2016. Disponível em: <<https://hal-amu.archives-ouvertes.fr/hal-01529708>>.

HUTTO, C.; GILBERT, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: **Proceedings of the international AAAI conference on web and social media**. [S.l.: s.n.], 2014. v. 8, n. 1, p. 216–225.

LIU, B. Sentiment analysis and opinion mining. **Synthesis lectures on human language technologies**, Morgan & Claypool Publishers, v. 5, n. 1, p. 1–167, 2012.

MAINIERI, B. O. et al. Text mining of audience opinion in esports events. **SBC–Proceedings of SBGames 2017**, p. 1224–1227, 2017.

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

MEJOVA, Y.; SRINIVASAN, P. Exploring feature definition and selection for sentiment classifiers. In: **Proceedings of the International AAAI Conference on Web and Social Media**. [S.l.: s.n.], 2011. v. 5, n. 1, p. 546–549.

NTUNGILA, J. Sentiment analysis of social media usernames and titles on youtube and twitch. 09 2021.

PANG, B.; LEE, L.; VAITHYANATHAN, S. Thumbs up? sentiment classification using machine learning techniques. **arXiv preprint cs/0205070**, 2002.

PLUTCHIK, R. A general psychoevolutionary theory of emotion. In: **Theories of emotion**. [S.l.]: Elsevier, 1980. p. 3–33.

REAGAN, A. J. et al. Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs. **EPJ Data Science**, Springer, v. 6, p. 1–21, 2017.

REIS, J. M. G. B. **Sentiment analysis: the case of twitch chat-mining user feedback from livestream chats**. Tese (Doutorado), 2020.

RISCH, J.; KRESTEL, R. Toxic comment detection in online discussions. In: **Deep learning-based approaches for sentiment analysis**. [S.l.]: Springer, 2020. p. 85–109.

SARLAN, A.; NADAM, C.; BASRI, S. Twitter sentiment analysis. In: IEEE. **Proceedings of the 6th International conference on Information Technology and Multimedia**. [S.l.], 2014. p. 212–216.

SILGE, J.; ROBINSON, D. **Text mining with R: A tidy approach**. [S.l.]: "O'Reilly Media, Inc.", 2017.

TEAM, R. C. R: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria. [http://www. R-project. org/](http://www.R-project.org/), 2013.

TURNEY, P. D. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. **arXiv preprint cs/0212032**, 2002.

WEINBERGER, M. Amazon's 970 million purchase of twitch makes so much sense now: It's all about the cloud. **Business Insider**, 2016.

WHITELAW, C.; GARG, N.; ARGAMON, S. Using appraisal groups for sentiment analysis. In: **Proceedings of the 14th ACM international conference on Information and knowledge management**. [S.l.: s.n.], 2005. p. 625–631.