

Desenvolvimento de um conjunto de dados com comentários extraídos da plataforma *Twitch* sobre o jogo *League of Legends*

Lucas D. F. Rodrigues¹, Luiz C. C. L. Junior²,
Antonio F. L. Jacob Junior², Fábio M. F. Lobato^{1,2}

¹Universidade Federal do Oeste do Pará (UFOPA)
Santarém, Pará, Brasil

²Universidade Estadual do Maranhão (UEMA)
São Luís, Maranhão, Brasil

antonio.junior@professor.uema.br, fabio.lobato@ufopa.edu.br

Resumo. *O crescimento das plataformas de transmissão ao vivo como a Twitch, impulsionado pelo aumento no volume de criadores de conteúdo, impactou positivamente em uma indústria economicamente importante, os jogos eletrônicos (e-Sports). O destaque da categoria dentro do tipo Multiplayer Online Battle Arena (MOBA) vai para o League of Legends, que foi um dos responsáveis pela legitimação e profissionalização dos e-Sports. O jogo possui uma vasta gama de criadores e que trazem consigo uma grande quantidade de interações dos internautas que os assistem. Um fenômeno deletério percebido neste cenário é a proliferação de discursos ofensivos, com comentários atacando ou denegrindo pessoas ou grupos, criando uma rede de ódio. Neste ensejo, neste trabalho apresentamos um conjunto de dados construído com comentários extraídos das transmissões dos criadores com maior engajamento na plataforma, visualizando os aspectos característicos e verificando de forma experimental, como o ódio está distribuído. Esta base de dados tem o potencial de auxiliar pesquisas envolvendo a detecção e também na análise desta indústria/domínio de aplicação da temática abordada.*

1. Introdução

A rede mundial de computadores tornou-se um espaço de troca de informação altamente personalizável, na qual os usuários podem adaptar sua forma de exposição [Fuchs 2017]. Para tal, os usuários utilizam plataformas para o compartilhamento de conteúdo como o *Facebook*, *Twitter* e *YouTube* [Perrin 2015]. Dentre os conteúdos dessas plataformas, dos jogos com maior número de espectadores é destacado o *League of Legends*. Este é um jogo da categoria *Multiplayer Online Battle Arena (MOBA)* e a partir do campeonato mundial que ocorreu em 2012, tornou-se um dos responsáveis pela legitimação e profissionalização dos *e-Sports*. Milhões de jogadores participam da categoria regularmente, sendo que os principais torneios oferecem oportunidades com títulos e premiações as quais variam desde itens do jogo até valores em dinheiro [Hinnant 2013]. Como exemplo, o campeonato mundial de *League of Legends* ocorrido na Coreia do Sul em 2018, ofereceu cerca de US\$ 6.450.000,00 em premiações aos participantes.

Devido ao impacto que as transmissões de videogames ao vivo causam na indústria de jogos com os *e-Sports*, a *Twitch* foi considerada um bom ambiente para a

transmissão desse tipo de conteúdo [Johnson and Woodcock 2018]. A plataforma obteve um alto crescimento que vai desde a exibição de jogos mais simples até campeonatos mundiais [Deng et al. 2017]. Além disso, ocorre uma maior conexão entre os participantes e seus visualizadores, pois o ambiente motiva e proporciona uma interação em tempo real entre os usuários e criadores com características e recursos de redes sociais [Hilvert-Bruce et al. 2018].

Contudo, assim como em outras plataformas de mídias sociais, é possível encontrar na *Twitch* casos de disseminação de discurso de ódio no qual é direcionado a atacar pessoas com base em características como raça, etnia, gênero e orientação sexual ou ameaças de violência em relação a outras pessoas [Watanabe et al. 2018]. Esta propagação ocorre tanto por parte dos criadores de conteúdo quanto dos usuários que assistem suas transmissões ao vivo. Mesmo que a plataforma possua diretrizes para coibir tais atividades na comunidade¹, em 2018 foram registrados casos envolvendo discursos de ódio por parte de dois *streamers* populares. Estes foram banidos por um mês devido a declarações homofóbicas, violentas e racistas feitas durante as transmissões de jogos [Moore 2018].

Neste ensejo, o presente trabalho tem o objetivo de apresentar um conjunto de dados em Inglês, de dez *streamers* com maior popularidade no jogo *League of Legends* extraídos da plataforma de *streaming Twitch*. Além disso, foram conduzidas análises sobre o conjunto de dados construído e que englobam a distribuição geral dos dados (metadados), métodos adjacentes ao Processamento de Linguagem Natural (PLN) e a visualização experimental por meio de ferramentas disponíveis para a detecção de discurso de ódio nos comentários. A motivação deu-se do interesse em fornecer um maior suporte aos pesquisadores na área e no campo das plataformas de transmissão de conteúdo, além de fomentar uma maior busca e desenvolvimento de métodos para detecção e redução de discursos de ódio nas mídias sociais.

O artigo está estruturado como se segue. Na Seção 2, serão apresentados os conceitos relevantes para o desenvolvimento da pesquisa e trabalhos relacionados à área. Na Seção 3, será apresentado o conjunto de dados construído e utilizado, assim como as suas características e características baseadas nas metodologias de PLN. Na Seção 4, serão apresentadas as metodologias das ferramentas de discurso de ódio estudadas e os seus resultados quando executadas na base de dados aqui construída. E na Seção 5, serão discutidas as considerações finais, ameaças a validade do estudo e sugestões de trabalhos futuros.

2. Fundamentação Teórica

Dentre as diversas conceituações utilizadas pela literatura no estado da arte para o discurso de ódio, a de [Fortuna and Nunes 2018] possui um direcionamento maior na taxonomia e nos alvos do discurso, definindo-a como uma linguagem que ataca ou diminui disseminando a violência ou ódio contra grupos, baseando-se em características específicas como aparência física, religião, descendência, nacionalidade ou etnia, orientação sexual, identidade de gênero ou outros. Além disso, ocorrem em diferentes estilos linguísticos, a exemplo da linguagem humorista.

¹<https://www.twitch.tv/p/pt-br/legal/community-guidelines/>

Seguindo um mesmo direcionamento, os conceitos utilizados por [Nascimento et al. 2019, Coutinho and Malheiros 2020, Silva and Roman 2020] definem como uma manifestação hostil, preconceituosa e que visa denegrir pessoas ou grupos baseados em características, como o gênero, religião, nacionalidade, deficiência, orientação sexual e condicionamento físico. Além disso, a manifestação deles ocorrem pela mídia tradicional (*e.g.* portais de notícia) e de forma *online*, como nas mídias sociais [Almeida et al. 2020]. E isso pode evoluir além do meio virtual, tornando-se uma ameaça real e fomentando a execução de crimes de ódio na sociedade.

2.1. Trabalhos Relacionados

É proposto por [Watanabe et al. 2018] a detecção de discursos de ódio em redes sociais e páginas de microblogs. Destaca-se que embora a maioria desses ambientes proíba a propagação desse tipo de discurso, o volume de dados que transita nestes ambientes torna quase impossível controlar todo o conteúdo compartilhado. Além disso, o trabalho demonstra preocupação com a necessidade de detectar e filtrar esses tipos de expressões de forma autônoma. Utilizou-se dados compostos por *tweets* e que apresentaram resultados com precisão de 87,4% para detecção de discurso ofensivo e 78,4% para identificar se o mesmo é caracterizado como discurso de ódio, ofensivo ou neutro.

Quanto a utilização de metodologias para a tarefa de classificação, em [dos Santos and Guedes 2019] é feita a identificação de predadores sexuais *online* por meio do uso de redes neurais. O método aplicado utilizou como núcleo uma CNN para atingir o objetivo abordado. O trabalho mostra que os índices de avaliação do classificador ficam entre 0.99 e 1.0 nos valores de acurácia, precisão, abrangência e F1. Em comparação aos trabalhos supracitados, o presente artigo aproveita-se de uma *API* melhor estruturada fornecida pela *Twitch*, a qual permite que sejam coletados uma maior quantidade e diversidade de dados, aumentando a variedade de conteúdo do conjunto de dados construído. Na área de discurso de ódio, faz a utilização de métodos associados as redes neurais como em [dos Santos and Guedes 2019] que obteve um bom desempenho nas métricas para a realização das tarefas abordadas e também adjacentes com PLN.

3. Conjunto de dados

Nesta seção serão apresentadas algumas informações básicas acerca do conjunto de dados, tais como a *API* e forma de coleta utilizada, a estruturação dos dados na base, informações sobre os metadados, estruturação dos comentários e outros itens relacionados a toda a base de dados.

3.1. Coleta de dados

A base de dados utilizada foi construída a partir da *API* da plataforma *Twitch* com o auxílio da biblioteca *Twitch-Python*² [Index 2019]. A biblioteca fornece uma abordagem orientada a objetos para acessar a *API* da *Twitch* e permite a extração de uma grande quantidade de atributos de usuários, vídeos e comentários publicados. A Tabela 1 apresenta os dados que foram extraídos para este trabalho.

O processo de construção da base de dados final, foi baseado nos comentários coletados dos usuários que interagem nas transmissões dos criadores na plataforma. Além

²<https://github.com/PetterKraabol/Twitch-Python>

Tabela 1. Dados extraídos para o conjunto de dados.

Atributo	Tipo
<i>user.display_name</i>	<i>String</i>
<i>user.view_count</i>	<i>Integer</i>
<i>video.title</i>	<i>String</i>
<i>video.view_count</i>	<i>Integer</i>
<i>video.created_at</i>	<i>String</i>
<i>comment.created_at</i>	<i>String</i>
<i>comment.commenter.display_name</i>	<i>String</i>
<i>comment.message.body</i>	<i>String</i>

Tabela 2. Descrição da base de dados extraída da Twitch.

<i>Streamer</i>	Arquivo (CSV)	Exemplos
<i>C9Sneaky</i>	<i>C9Sneaky_comments</i>	110.698
<i>DoubleLift</i>	<i>DoubleLift_comments</i>	32.423
<i>foggedftw2</i>	<i>foggedftw2_comments</i>	46.734
<i>lolTyler1</i>	<i>lolTyler1_comments</i>	316.338
<i>Nightblue3</i>	<i>Nightblue3_comments</i>	48.674
<i>Rush</i>	<i>Rush_comments</i>	60.396
<i>TFBlade</i>	<i>TFBlade_comments</i>	191.803
<i>Trick2g</i>	<i>Trick2g_comments</i>	56.933
<i>TSM_Bjergsen</i>	<i>TSM_Bjergsen_comments</i>	5.144
<i>Yassuo</i>	<i>Yassuo_comments</i>	116.623
Combinação	<i>TwitchDataset</i>	985.766

disso, esses internautas foram anonimizados a fim de manter a sua integridade, mesmo que eles façam o uso de *nicknames* para interação na seção de comentários. Ao todo, foram coletados mais de 985 mil comentários, aos quais foi possível identificar em uma inspeção manual a presença de discurso de ódio ou o uso de linguagem ofensiva.

3.2. Conteúdo

O conteúdo extraído e apresentado na Tabela 1 é constituído a partir de dados de dez *streamers* com maior popularidade no jogo *League of Legends*, com todos no idioma Inglês. Seguindo a ordem por atributo na Tabela 1, foram respectivamente coletados o nome do usuário, a quantidade de visualizações do canal, o título do vídeo, número de visualizações do vídeo, período de publicação, data de criação do comentário, nome do usuário que publicou e o texto do comentário. Ao considerar que grande parte das análises tenha como objetivo o conteúdo dos comentários, esses dados foram concatenados em um único arquivo nomeado “*TwitchDataset.csv*”. Este, foi utilizado posteriormente nas etapas de análise e visualização, o qual está disponível de forma aberta na plataforma *Zenodo*³ [LimaJr et al. 2020].

A base de dados é composta originalmente por dez arquivos, todos em formatação *UTF-8*, cada um representando um *streamer* diferente e contém apenas os comentários coletados. O usuário correspondente, nome do arquivo e quantidade de exemplos presentes em cada um destes (atributo “*comment*”), assim como o arquivo concatenado citado anteriormente estão detalhados na Tabela 2.

³<https://zenodo.org/record/3735091>

A distribuição de comentários entre os criadores de conteúdo varia entre aproximadamente 5 mil e 316 mil exemplos, caracterizando uma maior popularidade e nível de interatividade entre o *streamer* e seus espectadores. Seguindo os dados da Tabela 2, a Figura 1 simboliza essa variação entre os usuários.

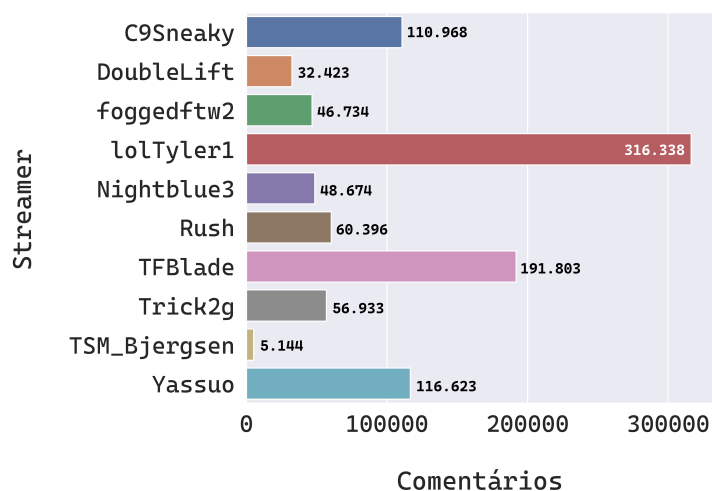


Figura 1. Quantidade de comentários por *streamer*.

Percebe-se que o usuário com maior popularidade em número de comentários publicados é o “*lolTyler1*”, com 316.338 exemplos, seguido por “*TFBlade*” com 191.803 e “*Yassuo*” com 116.623. A menor quantidade registrada é para o criador “*TSM_Bjergsen*”, com apenas 5.144. Em contrapartida, a quantidade de visualizações para cada um deles difere na proporção da publicação de comentários em seus conteúdos e é apresentada na Tabela 3.

Tabela 3. Visualizações e proporção de interações por canal.

<i>Streamer</i>	Visualizações	Comentários	Proporção
<i>Nightblue3</i>	218.398.725	48.674	$2,22 \times 10^{-4}$
<i>Trick2g</i>	144.710.525	56.933	$3,93 \times 10^{-4}$
<i>loltyler1</i>	134.955.884	316.338	$2,34 \times 10^{-3}$
<i>C9Sneaky</i>	104.128.583	110.698	$1,06 \times 10^{-3}$
<i>TSM_Bjergsen</i>	92.608.310	5.144	$5,55 \times 10^{-5}$
<i>Doublelift</i>	71.162.230	32.423	$4,56 \times 10^{-4}$
<i>Yassuo</i>	54.649.565	116.623	$2,13 \times 10^{-3}$
<i>TFBlade</i>	40.392.663	191.803	$4,75 \times 10^{-3}$
<i>Rush</i>	19.790.153	60.396	$3,05 \times 10^{-3}$
<i>foggedftw2</i>	4.909.203	46.734	$9,52 \times 10^{-3}$

Quando visualizada a proporção entre a quantidade de comentários pela de visualizações, torna-se perceptível a relação entre os dois itens. Quanto mais visualizações um determinado canal tem, a taxa de publicação de comentários decai, sendo essa uma variável inversamente proporcional à primeira. Por exemplo, o *streamer* “*foggedftw2*” contabilizou 4.909.203 e quando relacionado com os valores da Tabela 1, é o que possui a maior proporção da Tabela 3, demonstrando o fator de divergência entre as visualizações e interações.

3.3. Estrutura dos comentários

Ao considerar que os dados coletados são de uma plataforma de *streaming online*, onde os usuários podem interagir mediante as regras do canal e da plataforma *Twitch*, grande parte deles fazem o uso de figuras que representam uma ideia ou palavra (*emoticons*) disponibilizados de forma gratuita ou paga, ou então postam textos curtos. A distribuição por grupos de tamanho dos comentários publicados é apresentada na Figura 2.

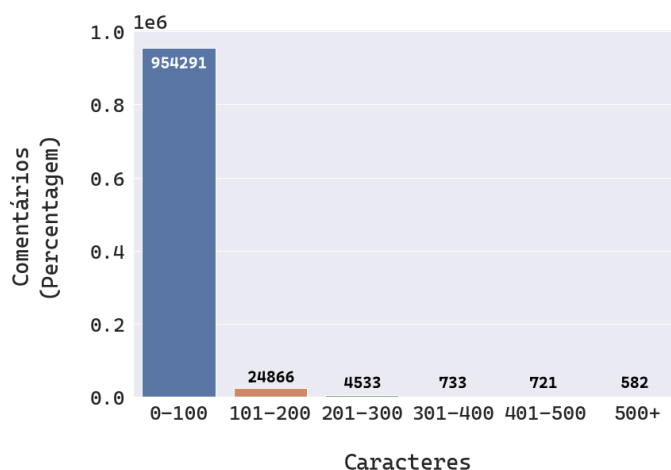


Figura 2. Quantidade de caracteres nos comentários.

A maior parte está contida entre 0 e 100 caracteres (*0-100*), seguida por *101-200* e *201-300*. Os demais três grupos (*301-400*, *401-500* e *500+*) possuem uma quantidade muito baixa quando comparado aos demais e tornam-se imperceptíveis na escala de visualização dos dados. Essas características estão normalmente presentes em publicações de mídias sociais, onde os traços de informalidade e textos curtos são predominantes. Além desses itens, os tópicos tendem a estar mais direcionados ao conteúdo abordado ou a temas adjacentes. Nesse caso, convergentes à plataforma, aos criadores de conteúdo ou ao vídeo/jogo em questão. Diante disso, procurou-se executar a modelagem de tópicos para visualizar se as hipóteses supracitadas são satisfeitas.

O método utilizado foi o *Non-Negative Matrix Factorization (NMF)* por ter um melhor desempenho para tarefas de classificação textual em textos curtos quando comparado ao *Latent Dirichlet Allocation (LDA)*. Como entrada, foi gerada uma matriz multi-dimensional de termos, a *Bag-of-Words (BoW)* com o peso estruturado no *Term Frequency-Inverse Document Frequency (TF-IDF)*, como utilizado em outros trabalhos próximos à área [Rodrigues et al. 2020, Junior et al. 2020]. A saída para cinco tópicos e palavras de modo geral e para cada *streamer* é listada nas Tabelas 4 e 5, respectivamente.

Tabela 4. Tópicos e palavras para todos os comentários.

Nº	Tópico	Palavras
1	<i>Emoticons</i>	kekw cs lul lulw head
2		tridance trikool pepends hyoon idfc
3	<i>League of Legends</i>	league weirdcmp normies weird lo
4	<i>Emoticons</i>	pog lul pogu clap trird
5		sosnowy icecold trird santat reindeer

Tabela 5. Tópicos e palavras dos comentários por streamers.

<i>Streamer</i>	Nº Tópico	Palavras
<i>C9Sneaky</i>	1	kek w monkaw ez clap ayaya
	2	peped ricardoflick pepegaske go sneakyclaus
	3	csneaky tt sneaky wt play
	4	pepodance pepegapls pepedance delet last
	5	subscribed months theyve twitch prime
<i>DoubleLift</i>	1	doublelift play high pogu sneaky
	2	powerupr powerupl mydog unleashed sdowdogs
	3	kek w tt lul pog play
	4	kek w loser pov pjsalt na
	5	doublelift live minutes seconds fans
<i>foggedftw2</i>	1	foggedyoda coachme foggedrng rigged must
	2	fogged kek w tt pog foggedftw
	3	foggedult foggedrip foggedw foggedfiesta trkbad
	4	foggedfrog spam foggedwizard pogcmp frog
	5	foggedftw wt think like food
<i>lolTyler1</i>	1	kek w pog pogu lulw lul
	2	league normies weirdcmp weird lo
	3	pog pepelaugh lul pogu dansgame
	4	head pepega trird heimer peepoleave
	5	loltyler play tyler omegalul variety
<i>Nightblue3</i>	1	nightblue kek w pepega lul kappa
	2	rust trust trustin rustrust want
	3	ttours pjsalt ttoursttours residentsleeper kappapride
	4	play dog cane jhin nightblue
	5	heal gun water drink eat
<i>Rush</i>	1	hwaiting greece grandmaster graphic grasp
	2	hwaiting greece grandmaster graphic grasp
	3	hwaiting greece grandmaster graphic grasp
	4	hwaiting greece grandmaster graphic grasp
	5	hwaiting greece grandmaster graphic grasp
<i>TFBlade</i>	1	kek w lul na spam pog
	2	forsencd yes clap trynd tfblade
	3	trird clap santat icecold cmonbr
	4	sosnowy icecold reindeer santat candycane
	5	peepofinger tfblade cheater head cmonbr
<i>Trick2g</i>	1	kek w icecold topt mana team
	2	trkcmp trklegend trkoverlay rezzy mute
	3	sosnowy peped headbang pepejam icecold
	4	trklol trkpotato trkmald bike trksmile
	5	pog hyperclap boys mcmac topt
<i>TSM_Bjergsen</i>	1	tsm bjergsen wt worlds play
	2	lozky pog pepends elixir use
	3	subscribed months theyve tier twitch
	4	urf play zed yuumi please
	5	bjerg pog stream lul love
<i>Yassuo</i>	1	tridance trikool pepends hyoon kek w
	2	kek w lulw time moe lol
	3	yassuo moe play wt tt
	4	yasspit thu clap yasokay ratirlspit
	5	na peped weirdcmp moe right

*Short-Term Memory (LSTM)*⁶ são implementações da plataforma *Kaggle*, com uma base de dados aberta⁷ de 159.571 exemplos para o treinamento dos algoritmos.

A escolha de classificadores baseados em redes neurais foi baseada em trabalhos recentes na detecção de discurso de ódio e análise de sentimentos [Pitsilis et al. 2018, Gambäck and Sikdar 2017, Wei et al. 2017, Wang 2018]. Os métodos implementados em [Erlapally 2020], [Sakhiya 2020] e [Atef 2020] foram utilizados para a condução dos testes experimentais para visualização da distribuição desses discursos no conjunto de dados. Os valores obtidos para todos os comentários da base estão visíveis na Figura 4.

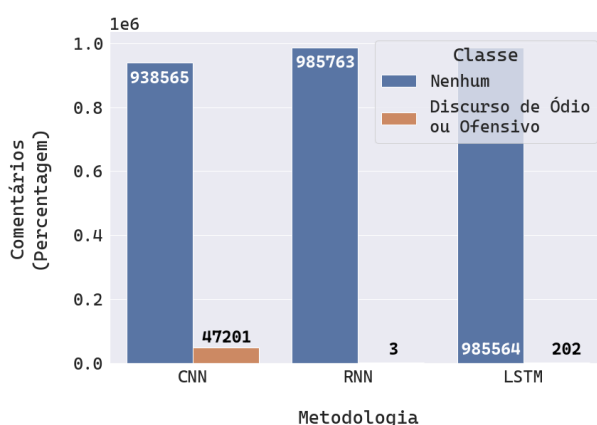


Figura 4. Quantidade de comentários classificados por metodologia.

Para a *RNN* e a *LSTM*, os dados foram quase todos para a classe “*Nenhum*”, respectivamente com 985.763 e 985.564 comentários. Apenas 3 e 202 para cada uma dessas metodologias supracitadas foram classificadas como “*Discurso de Ódio ou Ofensivo*”. A divergência mais significativa foi para a *CNN*, com 938.565 para “*Nenhum*” e 47.201 para “*Discurso de Ódio ou Ofensivo*”. Mesmo com a baixa quantidade de exemplos classificados como discurso de ódio, com maioria para a *CNN*, pode-se inferir que há a presença desses dentre os comentários coletados e além disso, que se manifestam na plataforma estudada. A execução dessas metodologias para testar o conjunto é feita de forma demonstrativa e com o principal objetivo de dar um panorama generalizado do desempenho em suas configurações padrões. Alguns dos textos classificados como “*Discurso de Ódio ou Ofensivo*” nas três metodologias estão listados na Tabela 7, com algumas das letras de seus termos sendo substituídas por asteriscos.

A partir dos comentários listados na Tabela 7, é possível visualizar algumas segmentações dentre os tipos de discurso de ódio. Por exemplo, identifica-se a presença de textos com caráter homofóbico, sexista ou xenofóbico, e utilizando termos ofensivos à pessoas com deficiência como “*retard*” (*retardado*), além de direcionamentos diretos à ofensas ao alvo no qual o texto foi direcionado no momento de sua publicação. Em todos eles, a *CNN* classificou dentro de “*Discurso de Ódio ou Ofensivo*” (Classe 1), enquanto as demais por terem uma proporção menor, classificaram em sua maioria como “*Nenhum*” (Classe 0).

⁶<https://www.kaggle.com/nadergo/classifying-hate-speech-with-a-pytorch-transformer>

⁷<https://www.kaggle.com/mrinaal007/hate-speech-detection>

Tabela 7. Comentários e classes conforme metodologia.

Comentário	Classe		
	<i>CNN</i>	<i>RNN</i>	<i>LSTM</i>
<i>ST*U HORSEY F**K A*S B**CH</i>	1	0	1
<i>GAY GAY GAY GAY GAY GAY GAY GAY GAY</i>	1	0	1
<i>What's up fu**er</i>	1	0	0
<i>shut the f**k up dutch</i>	1	0	0
<i>Knock this retard out</i>	1	0	0
<i>f**k the s**t out of her</i>	1	0	0

5. Considerações Finais

Com os dados apresentados neste trabalho, é possível visualizar um panorama de como os usuários e os *streamers* interagem na plataforma *Twitch*, com a temática principal sendo o jogo *League of Legends*. O uso de *emoticons* é bastante aplicado, por normalmente expressarem uma ideia associada à sua figura e mantém características de mídias sociais, a informalidade e mensagens curtas. Além disso, os tópicos mais abordados são em sua maioria associados ao ambiente em que o espectador está. Isto é, ao jogo, ao *host* da transmissão, ou assuntos adjacentes ao que está sendo exibido no momento em que o comentário é feito e também, combinado com a interação de outros usuários.

O conjunto de dados apresentado neste artigo permite fomentar a anotação manual para o desenvolvimento de um detector para discurso de ódio, por exemplo. A execução de metodologias como as baseadas em aprendizado supervisionado, não supervisionado ou semi-supervisionado podem gerar resultados diferentes em suas métricas e no desempenho final do mesmo. Os testes baseados em *RNN* ou *CNN*, como apresentado na Seção 4, servem como uma visualização experimental de desempenho no *dataset* apresentado e também podem auxiliar com a criação ou aperfeiçoamento de uma ferramenta associada à área. Dessa forma, a base está disponível na plataforma *Zenodo* [Lima Jr et al. 2020] para que outros pesquisadores possam replicar e melhorar os resultados obtidos, além de implementarem novos métodos.

5.1. Ameaças à validade do estudo

É importante mencionar algumas ameaças a validade do estudo. A primeira preocupação é relacionada à ausência de *fine tuning* dos métodos de detecção empregados. Nos experimentos descritos, a parametrização-padrão dos métodos implementados foi utilizada. Visando mitigar esta ameaça, os códigos dos experimentos encontram-se disponíveis no repositório https://github.com/fabiolobato/hate_speech_twitch.

Outro item é em relação a ausência de anotação manual dos dados. Apesar desta tarefa ter sido iniciada, houve uma divergência grande entre os anotadores. Uma das hipóteses levantadas foi relacionada ao idioma, pois todos os anotadores voluntários têm o Inglês como segunda língua. Por este motivo, os resultados dessa ação não foram incluídos no presente estudo e entra para a sugestão de trabalhos futuros, tal como descrito a seguir.

5.2. Trabalhos Futuros

Como trabalhos futuros, sugere-se a atualização do conjunto a partir da *API Twitch-Python*⁸ [Index 2019], mantendo o formato dos dados e fazendo alterações necessárias apenas nos alvos da coleta, expandindo a lista de *streamers*. A anotação da base de dados por falantes nativos de inglês também se faz interessante. Assim, será possível prosseguir para a fase de desenvolvimento de um detector e classificador de discurso de ódio e linguagem ofensiva, capaz de ser incorporado a sistemas existentes e aprimorar a capacidade de filtragem desses conteúdos nas mais diversas mídias sociais.

Comparativos entre cada metodologia adotada para os classificadores podem ser considerados a partir de métricas de avaliação que envolvem a acurácia, pontuação F1, precisão, dentre outras. Adicionalmente, a verificação dos índices de concordância entre os anotadores por meio do *Kappa* de *Cohen* podem ser feitas, por exemplo.

Agradecimentos

O trabalho apresentado neste artigo foi parcialmente financiado pela FAPEMA e pelo CNPq (processos 308334/2020-5 e 147336/2020-1).

Referências

- Almeida, G., Guimarães, I., Jr., A. J., and Lobato, F. (2020). Fontes de dados gerados por usuários: quais plataformas considerar? In *Anais do IX Brazilian Workshop on Social Network Analysis and Mining*, pages 25–36, Porto Alegre, RS, Brasil. SBC.
- Atef, N. (2020). Classifying Hate Speech with a pyTorch Transformer.
- Coutinho, V. M. and Malheiros, Y. (2020). Detecção de mensagens homofóbicas em português no twitter usando análise de sentimentos. In *Anais do IX Brazilian Workshop on Social Network Analysis and Mining*, pages 1–12, Porto Alegre, RS, Brasil. SBC.
- Deng, J., Tyson, G., Cuadrado, F., and Uhlig, S. (2017). Internet scale user-generated live video streaming: The twitch case. In *International Conference on Passive and Active Network Measurement*, pages 60–71. Springer.
- dos Santos, L. and Guedes, G. (2019). Identificação de predadores sexuais brasileiros por meio de análise de conversas realizadas na internet. In *Anais do VIII Brazilian Workshop on Social Network Analysis and Mining*, pages 143–154, Porto Alegre, RS, Brasil. SBC.
- Erlapally, D. (2020). Hate Speech Detection — 1D CNN — Glove Embedding.
- Fortuna, P. and Nunes, S. (2018). A survey on automatic detection of hate speech in text. *ACM Comput. Surv.*, 51(4).
- Fuchs, C. (2017). *Social media: A critical introduction*. Sage.
- Gambäck, B. and Sikdar, U. K. (2017). Using convolutional neural networks to classify hate-speech. In *Proceedings of the first workshop on abusive language online*, pages 85–90.

⁸<https://github.com/PetterKraabol/Twitch-Python>

- Hilvert-Bruce, Z., Neill, J. T., Sjöblom, M., and Hamari, J. (2018). Social motivations of live-streaming viewer engagement on twitch. *Computers in Human Behavior*, 84:58–67.
- Hinnant, N. C. (2013). Practicing work, perfecting play: League of legends and the sentimental education of e-sports. *Thesis, Georgia State University*.
- Index, P. P. (2019). Twitch python.
- Johnson, M. R. and Woodcock, J. (2018). The impacts of live streaming and twitch.tv on the video game industry. *Media, Culture & Society*, page 10–28.
- Junior, L. F., Junior, J. S., and Lobato, F. (2020). Um olhar sobre turismo gastronômico: Um caso no tripadvisor. In *Anais do XVII Encontro Nacional de Inteligência Artificial e Computacional*, pages 519–530. SBC.
- LimaJr, L. C. C., Rodrigues, L. D. F., JacobJr, A. F. L., and Lobato, F. M. F. (2020). League of Legends and hate speech: a corpus for comments in Twitch.tv. Disponível em: <https://doi.org/10.5281/zenodo.3735091>.
- Moore, M. (2018). 2 Popular Twitch Streamers Banned for a Month Over Hate Speech.
- Nascimento, G., Carvalho, F., Cunha, A. M. d., Viana, C. R., and Guedes, G. P. (2019). Hate speech detection using brazilian imageboards. In *Proceedings of the 25th Brazilian Symposium on Multimedia and the Web, WebMedia '19*, page 325–328, New York, NY, USA. Association for Computing Machinery.
- Perrin, A. (2015). Social media usage. *Pew research center*, pages 52–68.
- Pitsilis, G. K., Ramampiaro, H., and Langseth, H. (2018). Effective hate-speech detection in twitter data using recurrent neural networks. *Applied Intelligence*, 48(12):4730–4742.
- Rodrigues, L., Junior, A., and Lobato, F. (2020). Notícias relacionadas a pessoas com deficiência: uma análise do conteúdo gerado pelos usuários em postagens de mídias sociais. In *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional*, pages 811–822, Porto Alegre, RS, Brasil. SBC.
- Sakhiya, N. (2020). Hate Speech Detection: RNN.
- Silva, A. and Roman, N. (2020). Hate speech detection in portuguese with naïve bayes, svm, mlp and logistic regression. In *Anais do XVII Encontro Nacional de Inteligência Artificial e Computacional*, pages 1–12, Porto Alegre, RS, Brasil. SBC.
- Wang, C. (2018). Interpreting neural network hate speech classifiers. In *Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)*, pages 86–92.
- Watanabe, H., Bouazizi, M., and Ohtsuki, T. (2018). Hate speech on twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection. *IEEE Access*, 6:13825–13835.
- Wei, X., Lin, H., Yang, L., and Yu, Y. (2017). A convolution-lstm-based deep neural network for cross-domain mooc forum post classification. *Information*, 8(3):92.