

Disability-Related News: An Analysis of User-Generated Content on Social Media Posts

Lucas D. F. Rodrigues¹, Ademir B. Junior¹, Fábio M. F. Lobato¹

¹Instituto de Engenharia e Geociências - Universidade Federal do Oeste do Pará
Santarém, Pará, Brasil

{lucas.darlindo, basegiojunior}@gmail.com, fabio.lobato@ufopa.edu.br

Abstract. *Social media plays an essential role for people with disabilities and their family members by providing a channel for denouncement, in addition to inducing support networks. However, hate speech is also found in these media. In this sense, an analysis of social media posts about disability-related news in Brazil is presented in this paper. News media comments were collected and analyzed using topic modeling and sentiment analysis. Moreover, the posts were classified according to its content. The results show that there is a search on the subject, mainly regarding denunciations or for providing mutual support. The findings can guide the formation of a support network for people with disabilities and their families.*

Resumo. *As mídias sociais desempenham um papel essencial para as pessoas com deficiência e seus familiares por proporcionarem um canal de denúncia e induzirem a construção de redes de apoio. No entanto, discursos de ódio também são encontrados nessas mídias. Nesse sentido, uma análise das postagens de mídia social sobre notícias relacionadas à deficiência no Brasil é apresentada neste artigo. Os comentários da mídia foram coletados e analisados usando modelagem de tópicos e análise de sentimentos. Além disso, as postagens foram classificadas de acordo com o seu conteúdo. Os resultados mostram que há uma busca sobre o assunto, principalmente em relação a reclamações ou em busca de apoio mútuo. Os achados da pesquisa podem guiar a formação de uma rede de apoio às pessoas com deficiência e suas famílias.*

1. Introdução

As mídias sociais desempenham um papel importante nos movimentos sociais, proporcionando um canal de denúncia e induzindo à construção de redes de apoio [Caldas et al. 2018]. Tais mídias também contribuem para a propagação de informações relevantes para a população em geral [Figueira and Guimarães 2017]. Essa difusão, proporcionada pela *Web 2.0*, provocou alterações na forma de consumo de informações pelos usuários, substituindo de meios tradicionais de notícias, como jornais e revistas por suas versões digitais [Ramos et al. 2016, Lobato et al. 2017].

No entanto, os discursos de ódio também são veiculados neste tipo de mídia pois elas permitem a interação dos usuários por meio do sistema de comentários e respostas [Djuric et al. 2015]. O anonimato é um importante aspecto que incita esses discursos dentro das plataformas [Mondal et al. 2017], sendo que existem duas principais esferas de direcionamento destes: i) Direcionado a grupos; e ii) Individualmente

[ElSherief et al. 2018]. No que tange ao tema das pessoas com deficiência, é perceptível uma segregação e depreciação destas com impactos negativos para as suas auto-estimas e também a de seus familiares [Lobato et al. 2018].

No Brasil, uma pesquisa conduzida pelo Instituto Brasileiro de Geografia e Estatística (IBGE) em 2018 e apresentado em [Simões et al. 2018] cita algumas porcentagens relacionadas a esses grupos: a quantidade de pessoas com deficiência no país é equivalente à 23,9% (sem linha de corte, dados baseados no Censo Demográfico de 2010) ou 6,7% (com linha de corte do WG¹, dados baseados no Censo Demográfico de 2010), com proporção distribuída entre visual (18,8% ou 3,4%), motora (7% ou 2,3%), auditiva (5,1% ou 1,1%) e mental (1,4% em ambos). [Lobato et al. 2018] destaca a frequência com a qual os discursos de ódio são manifestados nas redes sociais quando relacionada a temática de deficiência. Expressões depreciativas relacionadas as condições de vidas dessas pessoas são comuns, por exemplo, “retardado” é utilizado como uma expressão que evoluiu de um diagnóstico médico para um insulto.

Também percebeu-se a partir da análise exploratória que há uma lacuna na literatura quanto à análise desses discursos nas redes sociais, sendo este o foco do presente trabalho. Com isto, considerou-se diferenciar os padrões utilizados em postagens que contém discurso de ódio e identificar o conteúdo (expressões) e discurso (construção da postagem) que atingem o público-alvo (pessoas com deficiência). A partir das discussões apresentadas nesta seção, foram propostas as seguintes Perguntas de Pesquisa (PP):

1. Qual o principal sentimento manifestado nos comentários de notícias relacionadas à temática da deficiência?
2. Quais são os principais tópicos abordados nos comentários de notícias relacionadas à deficiência?

Nesse panorama, o presente artigo realizou uma análise dos comentários acerca de notícias relacionadas às deficiências físicas, mentais e de aprendizado, para identificar quais os principais tópicos discutidos, qual o grupo principal de sentimento manifestado e o contexto nos quais foram utilizados. Os resultados obtidos dão suporte para a construção de novas estratégias a fim de auxiliar na interlocução entre as pessoas com deficiência e seus familiares, além de outros grupos externos como o terceiro setor.

O restante deste artigo está organizado como segue. Os trabalhos relacionados são apresentados na Seção 2. A metodologia adotada nos experimentos é descrita na Seção 3. Os resultados obtidos são discutidos na Seção 4. Por fim, as considerações finais, ameaças a validade do estudo e projeções para trabalhos futuros são apresentadas na Seção 5.

2. Trabalhos Relacionados

[Buchholz et al. 2018] desenvolveram um trabalho com a finalidade de verificar como parentes e cuidadores fornecem suporte comunicativo para pessoas com deficiência cognitiva ou comunicativa. Para tal, os autores conduziram um estudo moldado a partir de questionários com cinco diferentes grupos e vinte e um indivíduos definidos para fornecer assistência a esses. Concluiu-se que pessoas com deficiência necessitam de um maior acesso a tecnologia e informação, assim como meios adaptados para as suas necessidades específicas.

¹*Group on Disability Statistics*

Redes de apoio para pessoas com deficiência impactam diretamente no bem-estar destas e de seus familiares. A partir disso, [Kilburn and Shapiro 2018] efetuaram um estudo com 88 mães de crianças com deficiência de até 2 anos de idade. O foco da pesquisa foi no relacionamento e no suporte fornecido a partir das redes sociais. Os autores encontraram sinais de estresse, depressão e transtorno de ansiedade nas mães que participavam de redes com baixo suporte, concluindo-se que essas plataformas são relevantes para esses indivíduos.

[Correa et al. 2018] efetuaram uma análise de sentimentos em postagens publicadas no *Twitter* sobre os filmes indicados ao Oscar de 2017 e se há uma correlação entre essas postagens e os vencedores. Sua abordagem deu-se com três diferentes algoritmos: i) *Naïve Bayes*; ii) *Distant Supervision Learning*; e iii) *Polarity Function*. Notou-se que há mais preferência a comentários positivos sobre os filmes. Na tarefa de classificação, o *Multinomial Naïve Bayes* foi o que apresentou o melhor resultado. Visando encontrar a polaridade de *tweets* relacionados ao trânsito, [Xavier 2018] avaliou a performance de três métodos de aprendizado de máquina: i) *Support Vector Machine*; ii) *Naïve Bayes*; iii) *Maximum Entropy*. Para a tarefa abordada no estudo, o *Support Vector Machine* apresentou o melhor desempenho do que os demais algoritmos, com 99% de acurácia.

3. Metodologia

Nesta seção será apresentada a metodologia utilizada para o presente estudo, desde a coleta dos dados até a interpretação dos resultados, descrevendo as etapas de pré-processamento e métodos utilizados nas análises. O fluxo utilizado é semelhante ao descrito em [Silva et al. 2017].

3.1. Coleta de Dados

A principal fonte para coleta dos dados foi selecionada a partir do *ranking* da *Alexa*², sendo este atualizado diariamente com os dados de tráfego gerado pelos usuários dentro de um período de três meses [Yesbeck 2018]. O Portal G1³ foi escolhido com base em sua posição e por englobar notícias de todas as regiões do país. Foram coletadas apenas as informações básicas das matérias, além do conteúdo e itens presentes nos comentários. Os termos de busca utilizados estão listados na Tabela 1 e foram adaptados do trabalho de [Lobato et al. 2018], onde os termos ali dispostos foram validados por um analista em rede social, um profissional da comunicação e um psicólogo pertencentes ao grupo de pesquisa no qual o projeto está inserido.

Tabela 1. Palavras-chave utilizadas para coleta.

Termos			
Deficiência	Deficiência Mental	Cego(s)	Cegueira
Surdo(s)	Surdez	Autismo	Deficiência Auditiva
Deficiência Física	Deficiência Intelectual	Síndrome de Down	Deficiência Visual
Pessoa com Deficiência	PcD	Microcefalia	Acessibilidade

Para a coleta, criou-se um *web crawler* desenvolvido na linguagem *Python* com o uso da biblioteca do *Scrapy Framework*⁴. Os dados foram armazenados em arqui-

²<https://www.alex.com/topsites/countries/BR>

³<https://g1.globo.com/>

⁴<https://scrapy.org/>

vos no formato *Comma-Separated Values (CSV)*, tal como em [Almeida et al. 2017, Teixeira et al. 2018]. Para a interação com essas páginas, utilizou-se a *Application Programming Interface (API) Selenium*⁵ para permitir o controle autônomo do navegador (*Geckodriver*⁶). Esta integração se fez necessária para o correto carregamento de todos os itens presentes na plataforma. Os dados extraídos estão presentes na Tabela 2.

Tabela 2. Dados extraídos das notícias.

Fonte	Item	Tipo de Dado
<i>Notícias</i>	Título	<i>String</i>
	Data de Publicação	Data
	Palavra(s)-chave	<i>String</i>
	<i>Link</i>	URL
<i>Comentários</i>	Autor	<i>String</i>
	Data de Publicação	Data
	Comentário	<i>String</i>
	Interações	"Curti", "Não curtir"

3.2. Pré-processamento

As técnicas de pré-processamento utilizadas foram escolhidas com o intuito de remover inconsistências do conjunto de dados a fim de melhorar a confiabilidade do resultado final [Cirqueira et al. 2018]. Os passos de pré-processamento foram implementados em *Python*, usando a biblioteca *Natural Language Toolkit (NLTK)*⁷ por possuir suporte para o idioma Português. Foi efetuada a remoção de *stopwords*, acentuação, caracteres numéricos, pontuações e especiais (*e.g.* !, ?), saudações ou cumprimentos (*e.g.* "olá"), *emojis* (*i.e.* figura que transmite sentimento ou ideia) e transformação para caixa baixa.

3.3. Análise de Sentimentos

A análise de sentimentos classifica os dados textuais conforme a sua polaridade expressa no *corpus* do texto. Usualmente as polaridades são divididas em três categorias principais, a saber: positivo, negativo e neutro [Cirqueira et al. 2017]. Para se realizar a análise de sentimentos utilizou-se a biblioteca *Polyglot*⁸ pois a mesma disponibiliza um suporte nativo para o idioma Português. Como a ferramenta produz como saída um *score* numérico variando de -1 à 1, as polaridades foram discretizadas de acordo com os intervalos apresentados na Tabela 3.

Tabela 3. Intervalos para cada tipo de sentimento.

Pontuação (P)	Sentimento
$-1.0 \leq P < 0.0$	Negativo
$P = 0.0$	Neutro
$0.0 < P \leq 1.0$	Positivo

⁵<https://www.seleniumhq.org/>

⁶<https://firefox-source-docs.mozilla.org/testing/geckodriver/>

⁷<https://www.nltk.org/>

⁸<https://polyglot.readthedocs.io/en/latest/>

3.4. Modelagem de Tópicos

Para a visualização dos tópicos mais utilizados dentro do conjunto de comentários, utilizou-se o mesmo fluxo seguido por [Lobato et al. 2018] para a análise de *tweets*. A partir disto, o algoritmo utilizado foi o *Latent Semantic Analysis (LSA)* por permitir uma verificação da semântica proveniente dos comentários a fim de uma melhor identificação do conteúdo abordado [Williams and Betak 2018]. Além disso, o *LSA* está presente em diversos trabalhos, sendo bem aceito dentro do campo das análises de redes sociais. O peso utilizado na matriz de termos foi o *Term Frequency-Inverse Document Frequency (TF-IDF)*.

O *TF-IDF* permite uma verificação estatística do quão importante uma palavra é dentro de um ou mais documentos. O *TF* mede a frequência de um termo e resulta em um valor normalizado a partir da divisão pelo tamanho do documento (*i.e.* número de termos). O *IDF* proporciona um ajuste ao *TF*, haja vista a sua consideração de todos os vocábulos como igualmente relevantes. Os tópicos são definidos em um modelo conhecido como *Bag of Words (BoW)* e representa as informações textuais como uma matriz documento-termo, com assimilação de pesos para cada termo ou palavra encontrada no conjunto. Contudo, não é gerada uma rotulação ou categorização desses tópicos a partir dos algoritmos e com isso é necessário efetuar essa tarefa manualmente a partir de uma análise subjetiva do autor ou anotadores [Chen et al. 2013].

Foram adotadas diferentes composições para testes nos tópicos mais utilizados nas postagens. O objetivo destas composições é rotular, comparar e selecionar a melhor configuração, dada que a escolha desses parâmetros é feita arbitrariamente [Brookes and McEnery 2019]. As principais modificações deram-se sobre a quantidade de tópicos (T) e de palavras (P). Os parâmetros utilizados foram T(5) com P(5), P(10) e T(10) com P(5) e P(10). A rotulação dos tópicos obtidos foi feita colaborativamente pelos autores, até chegar a um consenso. Os resultados são apresentados na Seção 4.

3.5. Classificação Manual

Para que houvesse uma melhor caracterização dos comentários coletados, adotou-se a mesma estratégia de [Lobato et al. 2018]. Com isso, efetuou-se a classificação dos exemplos baseadas em cinco diferentes classes, todas listadas na Tabela 4 com seus respectivos significados. O processo foi feito de forma manual por três avaliadores, onde um comentário era de uma dada classe apenas se o mesmo fosse classificado por dois dos três anotadores, estratégia similar a adotada por [Magalhães et al. 2017].

Tabela 4. Classes e seus significados.

Classe	Significado da classificação para o comentário
Informativo	Apresentar ou explicar algo sobre algum conteúdo referente a temática da notícia.
Ofensivo e Pejorativo	Denegrir ou utilizar os termos com caráter ofensivo.
Relatos de Experiência	Relato acerca de um acontecimento ou experiência do autor.
Indignação e Denúncia	Expressa uma denúncia ou indignação acerca de uma notícia ou outro comentário.
Outros	Não se enquadra nas demais categorias listadas.

Para medir o índice de concordância entre os anotadores, o coeficiente *Kappa* de *Cohen (K)* é adotado entre os três avaliadores. Com o intuito de automatizar o processo, desenvolveu-se também um classificador automático baseado em aprendizado de máquina, tal como descrito na subseção a seguir.

3.6. Classificação Automática

Utilizando-se dos dados anotados na etapa anterior, investigou-se alguns algoritmos de classificação para automatizar este processo. Neste ensejo, foram testados dois algoritmos também utilizados em [Xavier 2018] e [Lima and Maia 2018] com o intuito de selecionar o qual desempenhava melhor na tarefa supracitada, a saber: i) *Support Vector Machine (SVM)*, com *kernel Linear*; e ii) *Multinomial Naïve Bayes (MNB)*. Ambos estão disponíveis na biblioteca *Scikit-learn*⁹ [Pedregosa et al. 2011], utilizados com os seus parâmetros padrões. Para avaliar os resultados, adotou-se a acurácia e a precisão das classes como principal medida de desempenho e também a avaliação da matriz de confusão.

4. Resultados Obtidos

As notícias analisadas foram publicadas entre o dia 11 de outubro de 2011 até 13 de junho de 2019. Ao todo, foram coletadas 1.065 notícias e 16.350 comentários. A Figura 1 apresenta um exemplo de comentário.

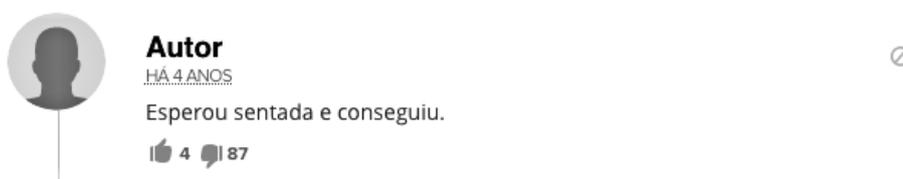


Figura 1. Exemplo de comentário anonimizado.

A Figura 1 demonstra em forma de captura de tela algumas características presentes nos comentários da plataforma, além de indicar certa ironia no exemplo utilizado quando o contexto da notícia ao qual ele foi publicado¹⁰ é considerado. Todos os comentários tiveram os seus respectivos autores e fotos anonimizados para preservar a sua privacidade e segurança. Por meio de análise exploratória, identificou-se também a frequência de comentários por temas de notícias, de acordo com as palavras-chave adotadas. A Figura 2 sumariza essa distribuição.

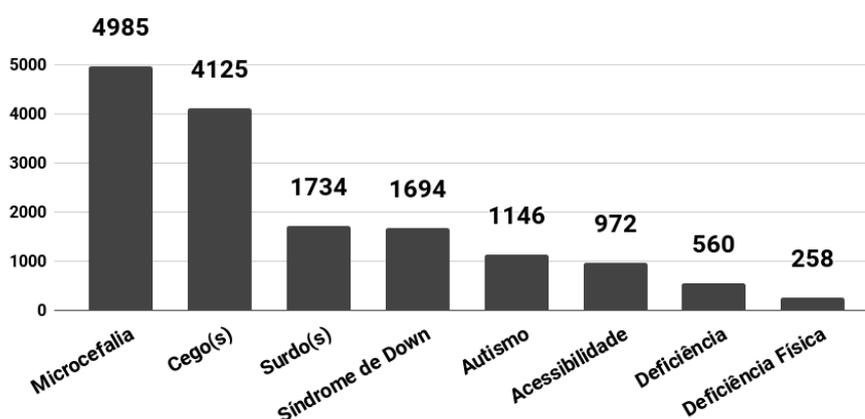


Figura 2. Distribuição dos comentários por palavra-chave.

⁹<https://scikit-learn.org/stable/index.html>

¹⁰<http://glo.bo/1KHwPXa>

Por meio da análise da Figura 2 é possível perceber que notícias sobre cegueira e microcefalia foram as que tiveram maior interação por parte dos usuários. Após a etapa de pré-processamento, os comentários foram submetidos à análise de sentimentos. A Figura 3 apresenta a distribuição dos comentários de acordo com a polaridade.

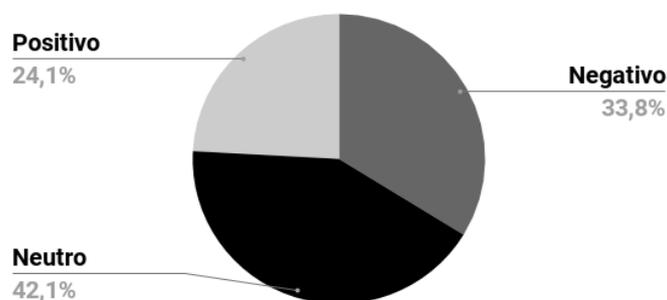


Figura 3. Análise de Sentimentos via Polyglot.

Diante da 1ª PP, a análise gerada pode ser interpretada de forma a correlacionar o grupo de sentimentos neutros e positivos com 66.2% de prevalência. Tais comentários incluem desde discursos que visam expressar apoio até relatos de experiências ou dúvidas. Ambos os exemplos estão inclusos na Tabela 5, respectivamente nos Itens 2-3 e 4-5. O grupo negativo (33.8%) pode estar relacionado aos comentários de cunho depreciativo ou mesmo em discursos de ódio, como o Item 1 listado na Tabela 5.

Tabela 5. Exemplos de comentários extraídos das notícias.

Item	Comentário
1	<i>“Não quero ser chato mas... É um mais estranho do que o outro... Nem sei pra que essas porcarias vão a escola se não aprendem nada”</i>
2	<i>“Que triste!!! Que Deus abençoe a família e esta criança, para que possa ter uma vida mais digna e feliz!!!”</i>
3	<i>“Parabéns a essa mãe. Que “investiu” em seu filho. Essa inclusão que ela promoveu não foi só para ele, foi para todas as crianças portadoras do Down. Tive um menino com essa síndrome e fiquei imaginando o meu filho nessa inclusão. Pena que ele se foi aos seis meses de vida.”</i>
4	<i>“Boa noite, tenho uma filha cadeirante com 10 anos de idade. Como faço para adquirir esse carro ou até outros mesmo...”</i>
5	<i>“Minha esposa está grávida de 10 semanas, além do repelente quais as outras medidas de precaução ela deve tomar?”</i>
6	<i>“É vergonhoso como o ser humano trata, nos dias de hoje o seu semelhante, principalmente quando esse outro tem ligações genealógicas.”</i>
7	<i>“Coloquem um surdo pra ser jurado de reality musical”</i>

Para a anotação manual, foi selecionada uma amostra de 376 publicações do total de 16.350 comentários. Esta amostra foi calculada utilizando como parâmetros 95% de intervalo de confiança e 5% para a taxa de erro. A seleção da amostra foi feita de forma completamente aleatória. Após a definição e consenso da rotulação pelos três anotadores, analisou-se a distribuição do conteúdo entre elas. A Figura 4 exibe a distribuição das classes anotadas.

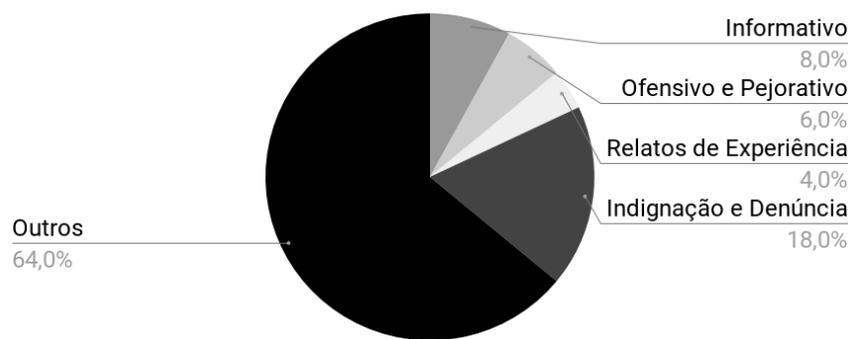


Figura 4. Distribuição das classes dos comentários.

Por meio da análise da Figura 4, é possível perceber que a quantidade majoritária dos comentários (64%) está caracterizada dentro da classe “Outros”. Essa grande quantidade é justificada pelo nível de ruído presente nos exemplos anotados que contém termos relacionados a deficiência que foram utilizados para outros acontecimentos em temáticas diferentes. Ademais, também englobam discursos positivos ou que manifestam apoio ou suporte, a exemplo do Item 2 e 3 da Tabela 5. Em sequência, a classe “Indignação e Denúncia” (18%) representa a parcela relacionada a notícias que relatavam falta de recursos ou acontecimentos com um grau de impacto elevado, como o Item 6 da Tabela 5. A classe “Informativo” contabilizou 8%, seguida da “Ofensivo e Pejorativo” com 6% e “Relato de Experiência” com 4%.

Destaca-se também que houve um baixo nível de concordância entre os anotadores. A seguir estão os coeficientes obtidos para os anotadores: i) Anotador A e B: 28%; ii) Anotador A e C: 33%; e iii) Anotador B e C: 17,23%. A baixa percentagem obtida a partir do *Kappa* de *Cohen* dá-se por duas principais causas: i) Alto grau de ruído no conjunto de dados coletados, especialmente quando relacionado à classe “Outros”; e ii) A análise subjetiva dos comentários está sujeita a diferentes interpretações para cada um dos anotadores e ocasiona em divergências entre as concordâncias.

Em continuidade ao *framework* experimental apresentado na seção anterior, efetuou-se o treino do classificador automático. Para tal, utilizou-se o conjunto rotulado manualmente. Para fins de validação, avaliou-se a distribuição obtida entre a classificação automática e a manual. A Figura 5 exhibe a distribuição resultante.

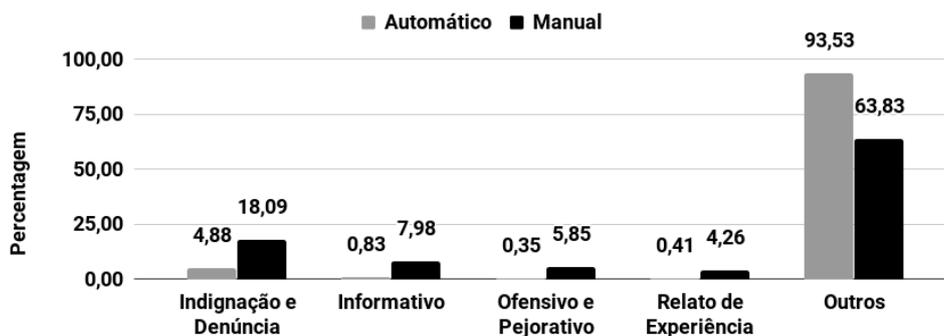


Figura 5. Relação entre a classificação automática e manual.

Analisando a Figura 5 é possível verificar uma diferença considerável na distribuição das classes. Tal diferença é justificada pelo desbalanceamento do conjunto de dados e pelas características do problema em questão. A subjetividade e dificuldade na rotulação dos dados contribuiu para a prevalência da classe “Outros”.

A acurácia obtida pelo *SVM* e *MNB* foi de 0.97 e 1.00, respectivamente. No entanto, notou-se que ao avaliar a precisão para cada classe o *SVM* apresentou uma performance superior aos resultados obtidos com o *MNB*, dado que a precisão das classes neste modelo foi considerada insatisfatória para identificar as múltiplas classes deste conjunto.

A modelagem de tópicos foi conduzida tanto para todo o conjunto de dados quanto para o conjunto de dados estratificado (por classe), tal como mostrado nas Tabelas 6 e 7, respectivamente.

Tabela 6. Tópicos gerais mais frequentes.

Tópico	Palavras
Família	sao pessoas pais fazer vida
Terapia com Animais	sao pessoas paulo policiais animais
Epidemia de Zika (2015/2016)	microcefalia zika virus sao pais
Projetos Educacionais da Polícia	pais povo policia fazer educacao
Epidemia de Zika (2015/2016)	deus sao vc crianca microcefalia

Tabela 7. Tópicos mais frequentes por classe.

Informativo	Discussões sobre Aborto	vida microcefalia gravidez inicio aborto
	Epidemia de Zika (2015/2016)	crianca casos sao virus microcefalia
	Campanhas sobre Autismo	crianca sao autismo olhar autista
	Suporte Materno	sucesso tera mae estao preciso
	Pesquisa & Informação	casos meses bebe ciencia proximos
Ofensivo e Pejorativo	Condição Socioeconômica	pobre coisa mulheres tema exemplo
	Surdez	acho surdo passar cara fazer
	Não Rotulado	pais sei deficiencia coisa hein
	Resiliência e Sororidade	tema exemplo mulheres coisa passar
Indignação e Denúncia	Não Rotulado	acho povo casos sei deficiencia
	Assistência Pública	faz sao fazer atendimento publico
	Saúde Pública	faz estado tipo rua medicos
	Transporte Público	deveria crianca mae onibus tipo
	Discussão sobre Aborto	aborto pessoas sao defender estao
Relatos de Experiência	Discussão sobre Aborto	estar deveriam estado tipo aborto
	Suporte Familiar	peessoa sindrome vida familia sao
	Inclusão Social	sao crianas sociedade filho maior
	Deficiência na Infância	criancas anos idade sindrome vida
	Suporte Familiar	mae familia crianas filho entrar
Outros	Realizações	fazer grande boa grandes vida
	Políticas Públicas	sao pessoas pode governo estao
	Políticas Públicas	peessoas pode estado matar povo
	Políticas Públicas	sao governo manifestacoes tema povo
	Empoderamento	ver vida pessoas poder sao
Religiosidade	deus sao pode estado feliz	

Por meio da análise das Tabelas 6 e 7 é possível responder a 2ª PP. Neste sentido, percebe-se que os conteúdos mais abordados são acerca temas como o surto de microcefalia ocorrido no Brasil em 2015 e 2016 e representado como a palavra-chave mais

frequente. Esse fato é corroborado quando correlaciona-se a distribuição dos comentários por termos (Figura 2). Além disso, temáticas relacionadas a denúncias ao atendimento público ou problemas de acessibilidade, relatos acerca da Síndrome de Down e usos depreciativos ou irônicos de termos relacionados (Tabela 5, Item 7) também são visualizados.

Em análise final, os comentários das notícias são muito variados, perpassando sobre políticas públicas, discussões sobre aborto, campanhas sobre deficiência, projetos educacionais e de terapia com animais, por exemplo. Sendo a família, governo e assistencialismo, tópicos centrais quando se fala sobre deficiências física, mental ou de aprendizado.

5. Considerações Finais

Neste artigo uma análise de comentários de notícias sobre a temática de deficiência no Brasil foi apresentada. Para a condução do estudo, 16.350 comentários de 1.065 notícias foram coletados e analisados usando modelagem de tópicos e análise de sentimentos. Um conjunto de dados rotulado manualmente também foi construído e posteriormente utilizado para a implementação de um classificador automático baseado em aprendizado de máquina. Dois algoritmos foram testados, o *Multinomial Naïve Bayes* e o *Support Vector Machine*, sendo que o último foi adotado por ter apresentado melhor desempenho. Os resultados obtidos permitiram a identificação dos: i) Assuntos mais comentados em comentários de notícia sobre deficiência (microcefalia e deficiência visual); ii) As polaridades das postagens; iii) Também a classificação dos comentários em relação ao seu conteúdo; e iv) Os tópicos mais frequentemente discutidos (aborto, políticas públicas e epidemia de Zika).

Como contribuições técnico-científicas decorrentes do estudo destacam-se a construção de um conjunto de dados sobre o tema e o relatório de análises. O primeiro permitiu a construção de um classificador automático de postagens. Em relação ao segundo, com as informações obtidas é possível fornecer um maior suporte a fim de fomentar a criação de redes de apoio e melhorar a forma na qual o conteúdo é abordado pela mídia.

Algumas ameaças a validade da presente pesquisa foram detectadas, por exemplo, os *emojis* dos comentários foram removidos, podendo causar variação nos resultados. E ainda, os ruídos encontrados na classificação podem gerar uma maior incidência da classe “Outros”. Os autores pretendem como trabalhos futuros reestruturar o *framework* para uma melhor classificação dos comentários e reduzir as ameaças a validade do estudo supraditas.

Agradecimentos

Este trabalho foi financiado por meio do Programa Institucional de Bolsas de Iniciação Científica (PIBIC) da Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação Tecnológica da Universidade Federal do Oeste do Pará (PROPPIT/UFOPA). Agradecemos também aos revisores pelos comentários e sugestões que muito ajudaram na construção e melhoria do manuscrito.

Referências

- Almeida, G. R. d., Cirqueira, D. R., and Lobato, F. M. (2017). Improving Social CRM through electronic word-of-mouth: a case study of ReclameAqui. *XIV Workshop de Trabalhos de Iniciação Científica (WTIC 2017)*.
- Brookes, G. and McEnery, T. (2019). The utility of topic modelling for discourse studies: A critical evaluation. *Discourse Studies*, 21(1):3–21.
- Buchholz, M., Ferm, U., and Holmgren, K. (2018). Support persons' views on remote communication and social media for people with communicative and cognitive disabilities. *Disability and Rehabilitation*, 0(0):1–9.
- Caldas, L. V. A., Jacob, A. F. L., Silva, S. S. C., Pontes, F. A. R., and Lobato, F. M. F. (2018). Development of a social network for research support and individual well-being improvement. In *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 383–386.
- Chen, Z., Mukherjee, A., Liu, B., Hsu, M., Castellanos, M., and Ghosh, R. (2013). Leveraging multi-domain prior knowledge in topic models. In *Twenty-Third International Joint Conference on Artificial Intelligence*.
- Cirqueira, D., Fontes Pinheiro, M., Jacob, A., Lobato, F., and Santana, A. (2018). A Literature Review in Preprocessing for Sentiment Analysis for Brazilian Portuguese Social Media. In *Proceedings of the 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 746–749. IEEE.
- Cirqueira, D., Jacob, A., Lobato, F., de Santana, A. L., and Pinheiro, M. (2017). Performance Evaluation of Sentiment Analysis Methods for Brazilian Portuguese. In Abramowicz, W., Alt, R., and Franczyk, B., editors, *Business Information Systems Workshops: BIS 2016 International Workshops, Leipzig, Germany, July 6-8, 2016, Revised Papers*, pages 245–251. Springer International Publishing, Cham.
- Correa, I. T., Abdala, D. D., Miani, R. S., and Faria, E. R. (2018). Sentiment Analysis of Twitter Posts About the 2017 Academy Awards. In *Anais do XV Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2018)*, pages 320–331. Sociedade Brasileira de Computação - SBC.
- Djuric, N., Zhou, J., Morris, R., Grbovic, M., Radosavljevic, V., and Bhamidipati, N. (2015). Hate Speech Detection with Comment Embeddings. In *Proceedings of the 24th International Conference on World Wide Web - WWW '15 Companion*, pages 29–30, New York, New York, USA. ACM Press.
- ElSherief, M., Kulkarni, V., Nguyen, D., Wang, W. Y., and Belding, E. (2018). Hate Lingo: A Target-based Linguistic Analysis of Hate Speech in Social Media. In *Proceedings of the 12th International AAAI Conference on Web and Social Media*, pages 42–51.
- Figueira, Á. and Guimarães, N. (2017). Detecting Journalistic Relevance on Social Media. *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 - ASONAM '17*, pages 1136–1139.
- Kilburn, J. E. and Shapiro, C. J. (2018). The structure and function of social networks of mothers of young children with disabilities. *Topics in Early Childhood Special Education*, page 0271121418767306.

- Lima, J. M. C. and Maia, J. E. B. (2018). A Topical Word Embeddings for Text Classification. In *Anais do XV Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2018)*, pages 25–35. Sociedade Brasileira de Computação - SBC.
- Lobato, F., Pinheiro, M., Jacob, A., Reinhold, O., and Santana, Á. (2017). Social CRM: Biggest challenges to make it work in the real world. In Abramowicz, W., Alt, R., and Franczyk, B., editors, *Business Information Systems Workshops: BIS 2016 International Workshops, Leipzig, Germany, July 6-8, 2016, Revised Papers*, pages 221–232. Springer International Publishing, Cham.
- Lobato, F. M. F., da Silva, M., Coelho, K., da Costa Silva, S., and Pontes, F. (2018). Vamos falar sobre deficiência? Uma análise dos Tweets sobre este tema no Brasil. *Proceedings of the 7th Brazilian Workshop on Social Network Analysis and Mining*, pages 274–279.
- Magalhães, G. G. M. S. d., Lima, F., Santos, E. F., Junior, P., and Rosa, L. (2017). Seleção de Técnicas de Mineração de Dados para Segmentação de Mercado. In *Proceedings of the 6th Brazilian Workshop on Social Network Analysis and Mining*, pages 658–669.
- Mondal, M., Silva, L. A., and Benevenuto, F. (2017). A Measurement Study of Hate Speech in Social Media. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media - HT '17*, pages 85–94, New York, New York, USA. ACM Press.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Ramos, P., Reis, J., and Benevenuto, F. (2016). Uma Análise da Polaridade Expressa nas Manchetes de Notícias Brasileiras. In *Proceedings of the 5th Brazilian Workshop on Social Network Analysis and Mining*, pages 187–198.
- Silva, W., Santana, Á., Lobato, F., and Pinheiro, M. (2017). A Methodology for Community Detection in Twitter. In *Proceedings of the International Conference on Web Intelligence*, pages 1006–1009.
- Simões, A., Athias, L., and Botelho, L. (2018). *Panorama Nacional e Internacional da Produção de Indicadores Sociais*. Instituto Brasileiro de Geografia e Estatística (IBGE), Rio de Janeiro, 6th edition.
- Teixeira, M. A. M., Lobato, F. M. F., Chagas, B. N. R., and Jacob Junior, A. F. L. (2018). A System of Acquisition and Analysis of Data for Extraction of Knowledge of the Ebit Platform. In *Proceedings of the 15th International Conference on Information Systems & Technology Management*, pages 4195–4206.
- Williams, T. and Betak, J. (2018). A Comparison of LSA and LDA for the Analysis of Railroad Accident Text. In *Proceedings of the 9th International Conference on Ambient Systems, Networks and Technologies (ANT 2018)*, pages 98–102.
- Xavier, C. C. (2018). Polarity Classification of Traffic Related Tweets. In *Anais do XV Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2018)*, pages 206–215. Sociedade Brasileira de Computação - SBC.
- Yesbeck, J. (2018). How are Alexa’s traffic rankings determined?