



Trabalho de Conclusão de Curso

Análises e modelos de predição de *tweets* de jornais e de seus comentaristas sobre as Vacinas de Covid-19

Alvino Lessa de Lima Júnior
alvino@laccan.ufal.br

Orientador:
Prof. Dr. André Luiz Lins de Aquino

Maceió, 15 de Setembro de 2021

Alvino Lessa de Lima Júnior

Análises e modelos de predição de *tweets* de jornais e de seus comentaristas sobre as Vacinas de Covid-19

Monografia apresentada como requisito parcial para obtenção do grau de Bacharel em Engenharia de Computação do Instituto de Computação da Universidade Federal de Alagoas.

Orientador:

Prof. Dr. André Luiz Lins de Aquino

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecário: Valter dos Santos Andrade – CRB4-1251

L732a Lima Junior, Alvino Lessa de.
Análises de modelos de predição de tweets de jornais e de seus comentaristas sobre as vacinas de Covid-19 / Alvino Lessa de Lima Junior. - 2021.
23 f.: il.

Orientador: André Luiz Lins de Aquino.
Monografia (Trabalho de Conclusão de Curso em Engenharia de Computação) – Universidade Federal de Alagoas, Centro de Tecnologia, Maceió, 2021.

Bibliografia: f. 22-23.

1. Modelo de predição. 2. Análise de sentimento. 3. Tweets. 4. Redes sociais. 5. Twitter. 6. Vacinas – Covid-19. I. Título.

CDU: 004

Monografia apresentada como requisito parcial para obtenção do grau de Bacharel em Engenharia de Computação do Instituto de Computação da Universidade Federal de Alagoas, aprovada pela comissão examinadora que abaixo assina.

Prof. Dr. André Luiz Lins de Aquino - Orientador
Instituto de Computação
Universidade Federal de Alagoas

Prof. Dr. Leonardo Viana Pereira - Examinador
Instituto de Computação
Universidade Federal de Alagoas

Prof. Dr. Jorge Artur Peçanha de Miranda Coelho - Examinador
Faculdade de Medicina
Universidade Federal de Alagoas

Maceió, 15 de Setembro de 2021

Agradecimentos

Primeiramente, agradeço a minha família, e em especial aos meus pais que sempre se esforçaram para me proporcionar a melhor educação que eles puderam. Meu pai (Lessa) sempre foi um grande exemplo, amigo e admirador. Minha mãe (Luciene) sempre foi e é minha grande incentivadora e entusiasta em minha vida de estudos. Ela sempre sonhou que eu estudasse na UFAL. Sempre que visitávamos meus avós ela dizia ao passar em frente a UFAL: "Meu Júnior ainda vai estudar aí". E finalmente estou próximo de concluir essa etapa de graduação.

Ao meu irmão (Erik) muito obrigado pelas palavras de carinho e incentivo, sei que sempre poderei contar com meu irmão, temos uma grande amizade e admiração recíproca.

Agradeço as minhas tias e tios, em especial minha Tia Betânia que sempre foi uma segunda mãe. Agradeço a minha Tia Nilda sempre carinhosa. Meus tios amigos Valdir e Antônio. Agradeço a todos os meus primos e primas, em especial Laura e Letícia que serão as próximas a ingressar na UFAL. E ao priminho Joaquim Caleb que se mostra sempre estudioso e aplicado para os estudos.

Agradeço a todos os meus mestres professores, em especial Leandro Sales que sempre me incentivou e foi meu primeiro orientador. Agradeço ao professor Alla por me acolher no LaCCAN e me proporcionar oportunidades incríveis de aprendizado e que diversas vezes acreditou em mim, mesmo quando não acreditei. Agradeço o professor Heitor que foi meu orientador de monitoria pelos conselhos e que será orientador de Mestrado. Agradeço ao professor Jorge Artur pelas lições de pesquisa.

Agradeço a todos os amigos do LaCCAN, em especial ao Pedro Henrique que sempre foi um grande amigo e entusiasta da carreira acadêmico científica, agradeço ao Matheus que sempre incentivou e foi um grande amigo, agradeço ao Demétrios com quem compartilhei várias dificuldades nas disciplinas da UFAL. Agradeço ao Christopher, Geymerson e Randy pelo apoio em momentos difíceis da jornada de graduação. Agradeço a Eduarda Chagas, Pedro Henrique e Christopher Freitas por sempre compartilharem de suas experiências de pesquisa.

Agradeço a todos os amigos do Conversa Mole que começou na ufal. Agradeço aos amigos da época de colégio que estão comigo até hoje, João Lucas e Ícaro.

Quero agradecer ao meu psicoterapeuta Silvino que me ajudou a encarar meus medos, me aceitar e superar diversos momentos difíceis ao longo de anos de psicoterapia.

E por fim quero me agradecer por não ter desistido nos momentos difíceis, noites em claro com ansiedade e dias de derrotas.

Resumo

Este trabalho propõe uma análise de *tweets* feitos por dois jornais (The Guardian e The Economist), e dos comentários recebidos (*replies*), sobre o termo *vaccine* no ano de 2020. A amostra de dados coletadas do *Twitter* possui 1343 *tweets* feitos pelos dois jornais e 11784 comentários. Para as análises foram utilizadas as seguintes técnicas: análise de sentimentos (*Vader*), análise de toxicidade (*Perspective API*), tópicos não-supervisionados (LDA) e modelos de predição baseados em um classificador de regressão logística. As análises indicam diferenças entre os Jornais, com o jornal The Economist sendo cerca de 40% mais positivo que The Guardian em suas publicações no *Twitter*, o mesmo não foi observado para os comentários. Porém, percebemos uma alta de comentários tóxicos nas semanas de aprovação das vacinas da *Pfizer* e *Moderna*, e também quando a primeira pessoa foi vacinada com a primeira dose no Reino Unido. Os modelos de predição propostos possuem 92% de acurácia para o modelo baseado nos *tweets* feitos pelos jornais e 68,5% de acurácia para o modelo baseado nos comentários.

Palavras-chave: covid-19, vacina, análise em redes sociais, análise de sentimentos, toxicidade, modelo de predição textual, twitter, comentários

Abstract

This work proposes an analysis of *tweets* written by two newspapers (The Guardian and The Economist), and the comments received (*replies*), about the term *vaccine* in the year 2020. A sample data collected from *Twitter* has 1343 *tweets* written by two newspapers and 11784 comments. For the analysis, the following techniques were used: sentiment analysis (*Vader*), toxicity analysis (*Perspective API*), unsupervised topics (LDA) and prediction models based on a logistic regression classifier . The analyzes indicate differences between the newspapers and The Economist is 40% more positive than The Guardian in their publications on *Twitter*, this difference was not observed for the comments. However, we saw a spike in toxic comments in the weeks of approval for the *Pfizer* and *Moderna* vaccines, and also when the first person was vaccinated with the first dose in the UK. The proposed prediction models have 92% accuracy for the newspapers *tweets* based model and 68,5% accuracy for the comments based model.

Keywords: covid-19, vaccine, social media analysis, sentiment analysis, toxicity, text prediction models, twitter, comments

Sumário

Lista de Figuras	v
Lista de Tabelas	vi
1 Introdução	1
1.1 Motivação	1
1.2 Trabalhos Relacionados	2
1.3 Objetivo	3
1.4 Contribuições	3
1.5 Estrutura do texto	4
2 Fundamentação Teórica	5
2.1 Análise de Sentimento	5
2.2 Campos Emocionais	5
2.3 Toxicidade	6
2.4 Latent Dirichlet Allocation	6
2.5 Regressão Logística	8
3 Nossa proposta	9
3.1 Coleta de dados	9
3.2 Pré-Processamento	9
3.3 Indicadores de Polarização	10
3.4 Indicadores de Toxicidade	10
3.5 Principais tópicos	10
3.6 Predição de <i>tweets</i>	11
4 Resultados e Discussões	12
4.1 Análises	12
4.1.1 <i>Tweets</i> dos Jornais	12
4.1.2 Comentários	14
4.2 Predizendo <i>tweets</i> de cada Jornal	17
4.3 Predizendo comentários para cada jornal	17
5 Considerações Finais	21
Referências bibliográficas	22

Lista de Figuras

4.1	Histogramas de comprimentos de textos	13
4.2	Campos Emocionais dos Jornais	13
4.3	Principais tópicos citados pelos Jornais.	14
4.4	Campos Emocionais dos Comentaristas	15
4.5	Heatmap Toxicity para comentários feitos ao The Guardian.	16
4.6	Heatmap Toxicity para comentários feitos ao The Economist.	16
4.7	Principais Tópicos citados pelos comentaristas.	16
4.8	Características que mais contribuem para o modelo de predição dos Jornais. . .	18
4.9	Características que mais contribuem para o modelo de predição dos Comentários.	20

Lista de Tabelas

2.1	Atributos padrão do Perspective API.	7
2.2	Atributos do Perspective API que foram treinados com comentários do The New York Times.	7
4.1	Métricas de avaliação para o modelo de predição dos Jornais	17
4.2	Métricas de avaliação para o modelo de predição dos Comentários	18

1

Introdução

1.1 Motivação

A pandemia de Covid-19 no ano de 2020 trouxe à tona o debate sobre medidas preventivas, medicamentos de combate e prevenção da doença. A prevenção a longo prazo só será possível por meio da vacinação (Chou and Budenz, 2020), e por isso estamos em meio a uma corrida para produção de diferentes vacinas feitas por organizações competentes, como universidades e empresas de nações diferentes, com a esperança de que alguma delas ou todas elas possam resolver a epidemia global que vivemos.

Porém, tal medida só será efetiva se adotada pela maior parte das pessoas em todos os países. O que nos leva à problemática de pessoas que se opõem ao método de vacinação como forma eficaz de combate a doenças. Conhecidos como anti-vacinas, essas pessoas se baseiam em desinformação sobre o tema, as quais compartilham em massa principalmente nas mídias sociais, como o Twitter. Apesar dessas comunidades online serem menores em relação aos pró-vacinas e indecisos, existe uma tendência de aumento ao convencer pessoas indecisas em prol das ideias anti-vacinas (Johnson et al., 2020a). Tal aumento da comunidade anti-vacina no Facebook já foi confirmado recentemente (Johnson et al., 2020b).

Com a pandemia afetando mundialmente a vida das pessoas e conseqüentemente a economia e a política, o tema passou a ser retratado em manchetes jornalísticas, mesmo em jornais que não costumavam noticiar sobre saúde, e a corrida vacinal contra covid-19 passou a ser falada semanalmente e até diariamente. Pois além da pandemia, o ano de 2020 foi marcado pelas eleições presidenciais nos EUA, o que tornou a discussão de saúde pública ainda mais polarizada.

Alguns jornais, como The Guardian ¹ e The Economist ², não possuem em suas plataformas online de notícias um ambiente destinado para os comentários de leitores, o que torna suas

¹<https://www.theguardian.com/international>

²<https://www.economist.com/>

contas no Twitter um dos principais ambientes para comentar sobre suas notícias. Portanto, o presente estudo resolveu analisar a emoção nos esforços de comunicação de mídias tradicionais no Twitter sobre o termo em inglês “vaccine” ao longo do tempo, bem como os comentários nas respectivas publicações feitos por seus leitores.

O fato de a pandemia ter surgido na China, o debate de forma politicamente polarizada sobre o tema e as incertezas do momento podem gerar comportamentos racistas e xenófobos (Lwin et al. (2020); Ruiz et al. (2020)). Então o presente estudo analisou também as notícias e comentários com o intuito de saber se existe xenofobia ou preferencialismo por determinada vacina de acordo com sua origem, ou se a hesitação da vacina de covid-19 é algo mais geral.

A metodologia desse trabalho é a seguinte. Primeiro foram coletados tweets feitos sobre “vaccine” pelos jornais escolhidos The Guardian e The Economist, e os comentários que receberam em suas postagens entre 1 de janeiro de 2020 até 31 de dezembro de 2020. Em seguida utilizamos histogramas e nuvens de palavras para perceber o padrão textual de cada jornal e também dos comentaristas.

Foram medidos os sentimentos expressados pelos jornais e comentaristas em suas publicações através do Vader (Ribeiro et al., 2016). Utilizando a métrica de campos emocionais para entender melhor os resultados da análise de sentimento percebemos que a média dos tweets do jornal The Economist são 40% mais positivos que a média dos tweets do The Guardian.

Porém, o mesmo não ocorreu para os comentários, então o Perspective API foi utilizado para mensurar a toxicidade dos comentários coletados e quais os momentos de maior pico de tais comentários (Guimarães et al., 2020). As análises foram finalizadas com o LDA que permitiu perceber os principais tópicos de interesse textual dos jornais e dos comentários (Mutanga and Abayomi, 2020).

Por fim, as análises motivaram a investigar se as diferenças textuais no discurso dos Jornais podem ser identificadas por um classificador de regressão logística. Para os comentários, além de investigar as diferenças entre os comentaristas de cada jornal, também consideremos os interesses textuais comuns aos comentaristas de cada jornal, o mecanismo social de homofilia (Massachs et al., 2020).

1.2 Trabalhos Relacionados

Na literatura já foram analisados tweets e comentários em diversos outros contextos. Partindo da análise de sentimentos, Ribeiro et al. (2016) analisaram as principais técnicas em diversos cenários de texto e concluíram que o melhor desempenho médio nos experimentos realizados foi do Vader.

No cenário de notícias de saúde geral, Lee and McElroy (2019) evidenciam um perfil mais neutro e positivo dos comentários em plataformas de notícias e sugere que o tema de vacinação pode ser um ambiente de comentários mais polarizados emocionalmente que o tema geral de

saúde, e assim o presente estudo decide investigar o tema de vacinação e seus comentários em jornais no twitter.

Alguns trabalhos avaliaram o ódio nos comentários de mídias sociais, [Guimarães et al. \(2020\)](#) utiliza o Perspective API para realizar uma caracterização de comentários brasileiros no *Facebook* e conclui que apesar dos comentários tóxicos serem minoria, estes estão concentrados nas páginas dos principais responsáveis por 60% dos comentários tóxicos. [Lima et al. \(2020\)](#) também utilizou o Perspective API para realizar uma caracterização de uma mídia social recente (*Gab*) que não possui moderação de comentários. E ao realizar um comparativo com o *Twitter* conclui que uma rede social sem moderação possui muito mais comentários tóxicos.

[Mutanga and Abayomi \(2020\)](#) analisou *tweets* sobre a pandemia de covid-19 na África do Sul utilizando a modelagem de tópicos não-supervisionados com LDA. E identificou termos como: *lockdown, gates, president, africa e vaccine*.

[Massachs et al. \(2020\)](#) coletou e analisou publicações feitas na página oficial de apoiadores do Trump no *Reddit*. Além de analisar o sentimento expressado nas publicações, tem como principal proposta prever se uma publicação foi realizada na comunidade a favor do Trump ou em uma comunidade qualquer do *Reddit*. Com isso examina homofilia dos adeptos do grupo Trump através das características aprendidas pelo modelo de predição.

Então o presente trabalho propõe utilizar as diferentes técnicas dos trabalhos correlatos para analisar os dados que coletamos do *Twitter* e assim contribuir para a literatura.

1.3 Objetivo

Analisar tweets dos jornais The Guardian e The Economist. Analisar comentários no *Twitter* sobre matérias de jornais acerca de temas relacionados à vacina para Covid-19. Através das análises propostas pretende-se identificar os padrões de sentimento no discurso de cada um dos dois jornais, identificar picos na quantidade de comentários tóxicos ao longo das semanas de 2020, identificar os tópicos de interesse mais desenvolvidos pelos jornais e os tópicos mais escritos pelos comentaristas. Além disso, propomos um modelo de predição para o discurso dos jornais no *Twitter* e um modelo de predição para o discurso dos comentaristas.

1.4 Contribuições

As contribuições deste trabalho são:

- Percepção de padrões de sentimento no discurso dos jornais;
- Identificação de picos de toxicidade nos comentários ao longo das semanas de 2020;
- Um modelo de predição para o discurso dos jornais no *Twitter*;

- Um modelo de predição para o discurso dos comentaristas.

1.5 Estrutura do texto

Este trabalho se divide como segue: Capítulo 2 define os conceitos das principais técnicas utilizadas para compor o presente trabalho; O Capítulo 3 apresenta a metodologia da abordagem proposta por este trabalho; O Capítulo 4 apresenta os resultados obtidos nas etapas de análises dos tweets e comentários, como também para as predições propostas; e, por último, o Capítulo 5 apresenta as conclusões e considerações deste trabalho.

2

Fundamentação Teórica

Esta sessão enuncia as principais técnicas utilizadas ao longo do presente trabalho.

2.1 Análise de Sentimento

Existem diversas ferramentas para análise de sentimentos na literatura, neste trabalho foi utilizada a técnica *Vader* que se baseia em um dicionário léxico para gerar uma nota de acordo com as palavras que compõem uma frase, essa nota evidencia um sentimento em uma escala de $v \in [-1, 1]$ e pode ter 3 classes, sendo elas positivo, negativo ou neutro. [Ribeiro et al. \(2016\)](#) demonstrou que o *Vader* foi o melhor algoritmo na média dos experimentos de 3 classes, no entanto não apareceu em primeiro lugar em nenhum experimento específico e na terceira posição em dados de mídias sociais. Apesar dos algoritmos concorrentes se saírem melhor nos experimentos, alguns deles possuem algum tipo de viés para uma determinada classe ou se ajustam bem apenas para temas específicos. Sendo assim, o *Vader* permite analisar como está o sentimento dos jornais e dos comentaristas sobre a corrida de vacinação.

2.2 Campos Emocionais

Uma maneira de analisar os sentimentos obtidos com o *Vader* é calcular uma métrica que representa a polarização do discurso tal qual uma normalização semanal. Logo, não nos interessa a intensidade do sentimento, mas se ele é positivo ou negativo. Então, se $v \geq 0.05$, o sentimento é considerado positivo; no entanto, se $v \leq -0.05$, o sentimento será negativo. Com essas faixas de representação definidas, podemos então calcular os quantificadores $h_+(t)$ e $h_-(t)$. Sendo a equação de $h_+(t)$:

$$h_+(t) = \frac{\sum_i^{n_+} \mathbb{1}(v_i(t) \geq \epsilon_+)}{n(t)}. \quad (2.1)$$

De maneira análoga, temos $h_-(t)$ dado por:

$$h_-(t) = \frac{\sum_i^{n_-} -\mathbb{1}(v_i(t) \leq \epsilon_-)}{n(t)}. \quad (2.2)$$

Onde $n(t)$ é o número total de *tweets* na semana t , e sendo considerado *tweets* positivos, negativos e neutros para compor $n(t)$. E $v_i(t)$ é a valência de sentimento de cada *tweet* avaliado pelo *Vader* e $+= 0.05$ e $-= -0.05$. Assim, $h_+(t)$ e $h_-(t)$ representam a proporção de expressões positivas ou negativas, respectivamente, para a semana de interesse t .

2.3 Toxicidade

*Perspective API*¹ é uma iniciativa do *Google* através do *Jigsaw* para monitorar e reduzir a toxicidade nos comentários feitos em plataformas online. A *API* disponibiliza uma série de modelos de aprendizado de máquina treinados para detectar diferentes tipos de comentários tóxicos. Assim, cada comentário em formato de texto pode ser requisitado a *API* que retorna um valor de probabilidade de $[0,1]$ para cada modelo solicitado. Essa ferramenta tem sido utilizada para moderação de comentários em plataformas de notícias online, como *The New York Times*, *El País*, *Le Monde*, *The Wall Street Journal* e *Financial Times*, bem como em mídias sociais e fóruns de jogos, como *reddit*, *Taringa!* e *Faceit*. Com isso as plataformas têm aumentado seu engajamento com diálogos construtivos e colaborado para melhorias da *API*.

Os modelos disponibilizados pela *Perspective API* e utilizados para nossa análise estão listados nas tabelas 2.1 e 2.2 que especificam o título do atributo e uma breve descrição. A tabela 2.1 lista os atributos definidos como padrão para a *Perspective API* e que foram treinados com diversos modelos de predição e inúmeros comentários de diversas fontes. Enquanto na tabela 2.2 estão listados os atributos que foram treinados utilizando exclusivamente comentários feitos no site de notícias *The New York Times*.

2.4 Latent Dirichlet Allocation

Uma técnica amplamente utilizada para extrair os principais tópicos de uma discussão textual é o algoritmo de *Latent Dirichlet Allocation*. Também conhecido como LDA, é uma técnica que consiste em um modelo probabilístico generativo de um corpus textual. A premissa do algoritmo é que cada documento é uma mistura de tópicos latentes e um tópico pode ser definido como uma distribuição de palavras que geralmente permitem compreender qual a ideia de cada tópico (Jelodar et al., 2019).

Dado um corpus textual C contendo N documentos e com o documento d contendo M_d

¹<https://www.perspectiveapi.com/>

TOXICITY	Um comentário rude, desrespeitoso ou irracional que provavelmente fará as pessoas abandonarem uma discussão.
IDENTITY_ATTACK	Comentários negativos ou odiosos dirigidos a alguém por causa de sua identidade.
INSULT	Comentário insultuoso, inflamatório ou negativo dirigido a uma pessoa ou grupo de pessoas.
PROFANITY	Palavrões ou outra linguagem obscena ou profana.
THREAT	Descreve a intenção de infligir dor, lesão ou violência contra um indivíduo ou grupo.
SEXUALLY_EXPLICIT	Contém referências a atos sexuais, partes do corpo ou outro conteúdo obsceno.
FLIRTATION	Frases de efeito, aparência elogiosa, insinuações sexuais sutis, etc.

Tabela 2.1: Atributos padrão do Perspective API.

ATTACK_ON_AUTHOR	Ataque ao autor de um artigo ou postagem.
ATTACK_ON_COMMENTER	Ataque a outro comentarista.
INCOHERENT	Difícil de entender, absurdo.
INFLAMMATORY	Com a intenção de provocar ou inflamar.
LIKELY_TO_REJECT	Medida geral da probabilidade de o comentário ser rejeitado de acordo com a moderação do NYT.
OBSCENE	Linguagem obscena ou vulgar, como palavrões.
SPAM	Conteúdo comercial irrelevante e não solicitado.
UNSUBSTANTIAL	Comentários triviais ou curtos.

Tabela 2.2: Atributos do Perspective API que foram treinados com comentários do The New York Times.

palavras ($d \in 1, \dots, N$), um modelo LDA de D atua conforme o seguinte procedimento de geração (Jelodar et al., 2019):

- (a) Escolher uma distribuição multinomial σ_t para o tópico $t (t \in 1, \dots, T)$ de uma distribuição Dirichlet com parâmetro α ;
- (b) Escolher uma distribuição multinomial ω_d para o documento $d (d \in 1, \dots, N)$ de uma distribuição Dirichlet com parâmetro β ;
- (c) Para uma palavra $p_m (m \in 1, \dots, M)$ em um documento d ;
 - 1. (a) i. Selecione um tópico z_m de ω_d .
 - ii. Selecione uma palavra p_m de σ_{z_m}

Sendo palavras e documentos variáveis observadas, ω e σ são variáveis latentes, α e β são hiperparâmetros. Com o intuito de inferir os hiperparâmetros e as variáveis latentes, é feito um cálculo de probabilidade que é maximizado para os dados observados D . Em seguida, realizando a iteração pelos itens definidos acima os documentos e suas palavras serão alocados de acordo com os tópicos predominantes do corpus (Jelodar et al., 2019).

2.5 Regressão Logística

Na regressão logística utilizada como modelo de predição, o resultado de cada tentativa de predição é uma probabilidade de que o resultado pertença a uma classe numérica acima de um valor de corte ajustado pelo modelo. Caso contrário, visando prever se algo pertence a classe contrária, o valor de probabilidade deve estar abaixo do valor de corte. Dada a seguinte equação, a regressão logística retornará um valor entre um e zero:

$$F(t) = \frac{1}{1 + e^{-t}} \quad (2.3)$$

3

Nossa proposta

3.1 Coleta de dados

Para tal análise, tweets foram coletados do ano de 2020 para as páginas no Twitter: “@guardian” e “@TheEconomist” com a palavra-chave: vaccine. Ao longo do ano de 2020, “@guardian” produziu 727 tweets, enquanto “@TheEconomist” produziu 616 tweets. Para esses tweets, foi utilizado a ferramenta Twint para obtenção dos dados.

Além dos tweets, foram coletados 7821 comentários feitos ao “@guardian” e 3963 comentários feitos ao “@TheEconomist”, para isso foi desenvolvida uma ferramenta de coleta de comentários, já que não conseguimos obter as informações diretamente. Partindo do link de um tweet coletado na primeira fase, conseguimos coletar no máximo 100 comentários de texto feitos em inglês.

3.2 Pré-Processamento

O processamento de texto pelo computador ainda não alcançou o mesmo nível de compreensão humana, necessitando um pré-processamento dos dados para formatá-los em um padrão estruturado compreensível para o algoritmo. No caso do Twitter, é necessário remover termos contendo hashtags (#), menções (@) e links (<https://www...>). Além disso, processos como tokenização para separar e identificar as palavras relevantes; lematização, para remover formas flexionadas das palavras e analisá-las como equivalentes; e remoção de stopwords, que são palavras retiradas sem alterar o sentido do texto de origem, e em geral são palavras comuns na língua, como preposições e artigos.

3.3 Indicadores de Polarização

Após o pré-processamento, os tweets e comentários foram submetidos ao algoritmo de análise de sentimentos *Vader*. Em seguida, os dados absolutos foram utilizados para compor a dinâmica de sentimentos ao longo de todo o ano de 2020. Porém, mesmo para intervalos de tempo semanais é difícil de interpretar os números absolutos. E por isso é utilizada a métrica de campos emocionais definida no capítulo anterior.

Então são definidos dois campos emocionais, um positivo e outro negativo. Por exemplo, para compor um campo emocional positivo, suponha que em um intervalo de tempo semanal foram contabilizados 10 tweets, desses 5 são positivos, 2 são negativos e 3 são neutros. Logo, o campo emocional positivo será uma razão de $5/10$, o que gera um valor de 0,5. Enquanto que no campo emocional negativo, cada tweet possui uma unidade negativa, e será uma razão de $-2/10$, obtendo um valor de -0,2.

3.4 Indicadores de Toxicidade

Com o intuito de compreender os comentários além do sentimento expressado na escrita, foi utilizada a ferramenta Perspective API para avaliar os comentários quanto a toxicidade e outros atributos. A Interface de Programação de Aplicações (API) retorna para cada comentário uma probabilidade que varia entre $[0,1]$ que avalia se o comentário pertence ao atributo requisitado.

Lima et al. (2020) define que para toxicidade os valores de probabilidade devem ser maiores que 0,8. Enquanto que para os demais atributos a probabilidade deve ser maior que 0,5, porém, o presente estudo utiliza valores maiores que 0,6 para garantir um melhor resultado. Além de toxicidade, foram avaliados também os seguintes atributos: *identity_attack*, *insult*, *inflammatory*, *attack_on_commenter* e *attack_on_author*.

3.5 Principais tópicos

The Guardian e *The Economist* somaram 1343 *tweets* feitos durante o ano de 2020 sobre o termo *vaccine*. Esses *tweets* pré-processados foram utilizados para compreender os tópicos mais citados pelos jornais. Para isso foi utilizado o LDA e após examinar diversos resultados com diferentes números de tópicos foi escolhido empiricamente um número de 5 tópicos como sendo o que melhor representou os dados de maneira interpretável.

Os 11784 comentários coletados nas postagens dos dois jornais passaram por um procedimento análogo e o resultado que melhor representou os dados de maneira interpretável foi utilizando 10 tópicos.

3.6 Predição de *tweets*

Para entender as principais diferenças entre os *tweets* feitos por cada um dos jornais e as diferenças entre o discurso dos comentaristas de cada jornal, foi proposto um classificador de regressão logística (Feng et al., 2014) que se mostrou melhor nas métricas de avaliação que outros algoritmos experimentados. Para implementação utilizamos a biblioteca scikit-learn ¹ que disponibiliza tanto o modelo, como as métricas de avaliação necessárias para o modelo.

Os *tweets* dos dois jornais devidamente pré-processados, inclusive retirando termos que identificam os jornais de origem, rotulados e vetorizados por meio do *TF-IDF* (Yun-tao et al., 2005) foram submetidos a fase de treino do modelo que utilizou 70% dos dados. Para a fase de teste os demais *tweets* que representam 30% foram utilizados.

Com o intuito de entender as características textuais que mais contribuem para definir cada classe do modelo treinado, foi utilizada a técnica de visualização e interpretação de modelos *SHAP* (Lundberg et al., 2020).

Então, utilizamos as ferramentas listadas na presente subseção para compor um modelo de predição para os *tweets* feitos pelos jornais e um modelo de predição para o discurso dos comentaristas.

¹https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html



Resultados e Discussões

Nesta sessão apresentamos os resultados e discussões sobre o trabalho. Primeiramente teremos os resultados das análises feitas utilizando as técnicas descritas no capítulo 2. Em seguida, os resultados obtidos através de um modelo baseado em um classificador de regressão logística para os tweets e depois para os comentários. Por fim, será apresentado os resultados de interpretabilidade das características de cada classificador.

4.1 Análises

Roadmap: Tweets dos Jornais

- Comprimento dos Textos
- Polarização da Discussão
- Modelagem de Tópicos com LDA

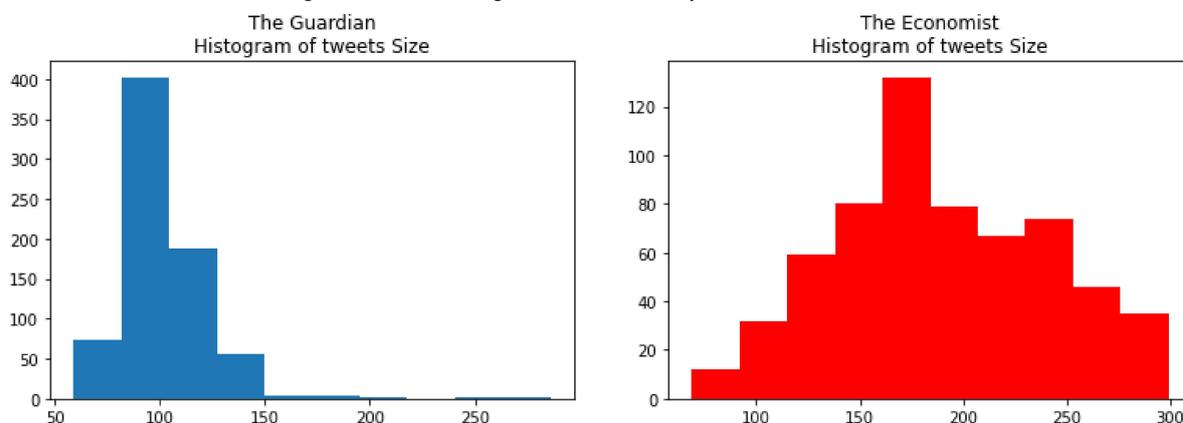
4.1.1 *Tweets dos Jornais*

Comprimento dos Textos

Os jornais *The Guardian* e *The Economist* podem apresentar um discurso contrastativo ou equivalente. Sendo assim, os resultados desta sessão podem ajudar a entender melhor como se posicionam os jornais sobre o tema de vacinação.

Uma das características observadas foi a frequência de *tweets* com tamanhos específicos. Observando a figura 4.1, notamos que o *The Guardian* possui uma variedade menor que o *The Economist*, em geral o *The Guardian* escreve com uma média de 100 caracteres, enquanto o *The Economist* escreve com uma média de 188 caracteres. Tal característica observada foi o primeiro indício para propormos um classificador para os *tweets* de jornais.

Figura 4.1: Histogramas de comprimentos de textos

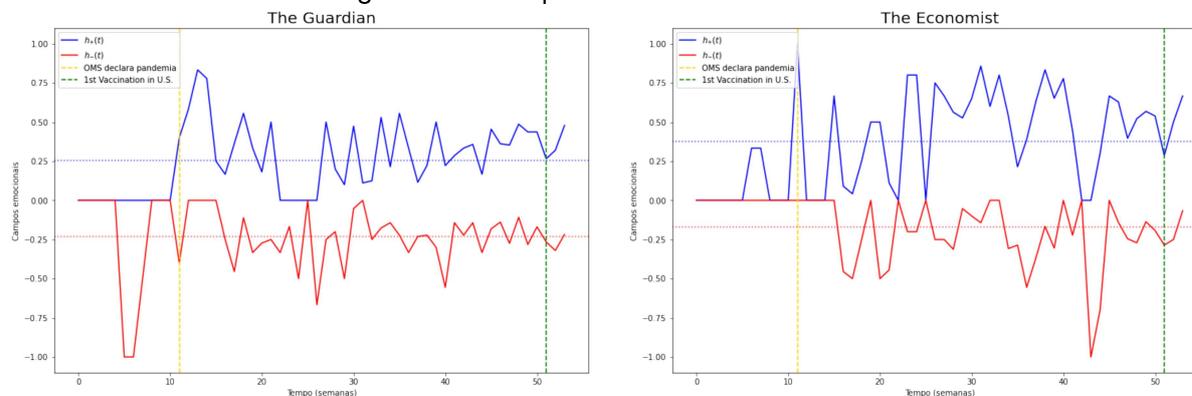


Polarização da Discussão

Além do Vader, decidimos testar outras ferramentas de análise de sentimentos, como foi o caso do SentiStrength que apesar de bem conhecido e utilizado em eventos como as Olimpíadas(2012) e Super Bowl(2014) possui um grande viés neutro. Por fim o Vader prevaleceu para analisar os *tweets* após pré-processados (Ribeiro et al., 2016).

Para analisar a dinâmica de sentimentos ao longo das semanas de 2020 utilizamos o conceito de campos emocionais para aplicar uma média semanal de sentimento como podemos observar na figura 4.2. Percebe-se que a média do The Guardian para o campo emocional positivo é de 0,25, enquanto que o The Economist possui uma média de 0,35. Para o campo emocional negativo o The Guardian apresenta uma média de $-0,25$ e o The Economist apresenta $-0,17$. Sendo assim, percebe-se um padrão mais positivo do The Economist em seus *tweets* sobre a vacina de covid-19 no ano de 2020.

Figura 4.2: Campos Emocionais dos Jornais



Modelagem de Tópicos com LDA

Em seguida, o LDA permitiu evidenciar os 5 principais tópicos abordados pelos jornais em seus *tweets*, como visto na figura 4.3 uma palavra comum a todos os tópicos é *vaccine*, o que era esperado por ser o termo utilizado para coletar os *tweets*.

Podemos perceber que no tópico 1 que as palavras *Oxford* e *University* se destacam sugerindo um tópico onde predomina o assunto sobre a vacina produzida pela universidade de Oxford em parceria com a empresa *Astrazeneca*. Para o tópico 2 algumas palavras em destaque são *Pfizer*, *first* e *dose* sugerindo um tema sobre a primeira dose da vacina produzida pelas empresas *Pfizer* e *Biontech*. O tópico 3 possui as palavras *Russia*, *approve* e *approval*, o que remete a época em que a Rússia aprovou sua vacina *Sputnik-V* nos respectivos testes.

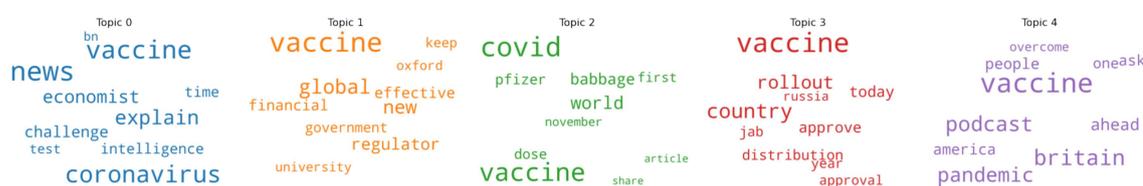


Figura 4.3: Principais tópicos citados pelos Jornais.

Roadmap: Comentários

- Comprimento dos Textos
- Polarização da Discussão
- Toxicidade da discussão
- Modelagem de Tópicos com LDA

4.1.2 Comentários

Comprimento dos Textos

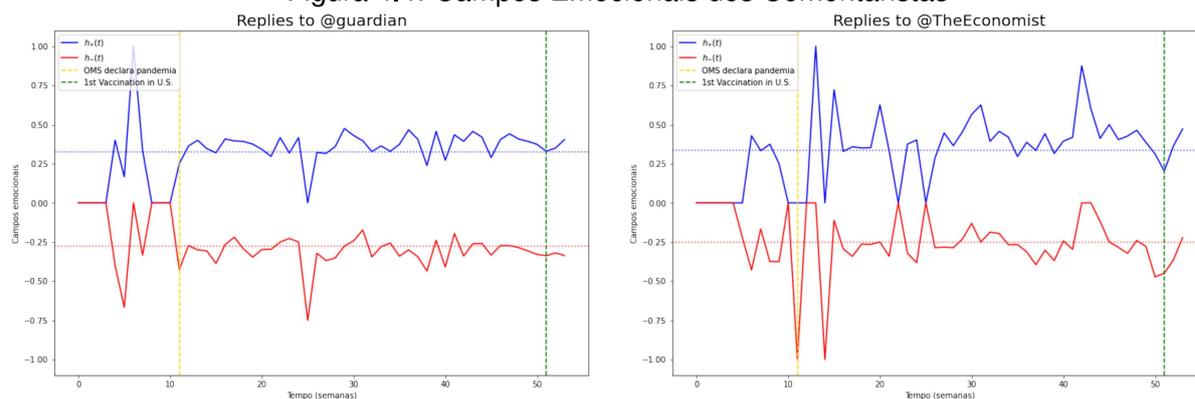
Para os comentários também foram analisadas as frequências dos comentários com comprimentos específicos. Porém, não obtivemos um padrão textual que diferisse os comentaristas do *The Guardian* e os comentaristas do *The Economist*. O que sugere uma possível homofilia entre os usuários do *Twitter* que realizaram os comentários, o que será abordado em outra subseção desse capítulo quando propomos um classificador para os comentaristas.

Polarização da Discussão

De maneira análoga aos *tweets* feitos pelos jornais, os comentários também foram analisados pelo *Vader*. Em seguida, foi possível estabelecer a métrica de campos emocionais, que vemos

na 4.4. Os comentários feitos ao The Guardian apresentaram uma média de 0,30 para o campo emocional positivo e o mesmo valor foi observado para a média do The Economist, o mesmo ocorreu para o campo emocional negativo que em ambos teve média de $-0,25$. Assim como os comprimentos de texto dos comentários, também não identificamos diferenças nas emoções dos comentaristas.

Figura 4.4: Campos Emocionais dos Comentaristas



Toxicidade da discussão

Com intuito de analisar melhor os comentários acerca da toxicidade utilizamos o Perspective API. Como podemos perceber através da escala das figuras 4.5 e 4.6, o The Guardian tem uma maior concentração de comentários por semanas em suas publicações. Apesar disso, em ambos percebemos um pico de toxicidade a partir da semana 49 e que se estende até a semana 51.

O intervalo de semanas mencionado inicia com a aprovação da vacina da Pfizer em todas as fases de teste. Em seguida, a Moderna também teve sua vacina finalizada em fases de teste. E por fim desse intervalo a primeira pessoa foi imunizada com a primeira dose da Pfizer no Reino Unido. Sendo assim, a aprovação de vacina e início de seu uso provocou muito mais comentários tóxicos que em qualquer outro momento do ano.

Modelagem de Tópicos com LDA

Foram evidenciados pelo LDA os 10 principais tópicos abordados pelos comentaristas. Na figura 4.7 temos alguns termos como *side_effect*, *china*, *chinese* e *trump* que são destacados em alguns tópicos. O termo *side_effect* sugere que os comentaristas estão preocupados com possíveis efeitos colaterais das vacinas. Os termos *china* e *chinese* sugerem críticas ao lugar que foi o primeiro epicentro da doença e possivelmente xenofobia sobre a população chinesa. Enquanto *trump* pode ser uma referência para possíveis críticas ao governo Trump.

4.2 Predizendo *tweets* de cada Jornal

Regressão Logística

Como resultado do classificador proposto, obtivemos uma acurácia de predição de 92% para identificar os *tweets* de cada jornal, como pode ser visto na tabela 4.1. As demais métricas também obtiveram valores elevados, o que constata a excelente acurácia obtida. Portanto, de fato existem diferenças no discurso dos jornais em seus *tweets* sobre o tema da vacina de covid-19.

Tabela 4.1: Métricas de avaliação para o modelo de predição dos Jornais

Método	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.921	0.889	0.972	0.929

Interpretabilidade de Características

Utilizando a ferramenta de visualização de dados *SHAP*, podemos constatar quais características textuais fornecem mais influência para o modelo na tomada de decisão para classificar se um *tweet* foi feito pelo *The Guardian* ou pelo *The Economist*.

Na figura 4.8 os valores *SHAP* que podem ser mensurados no eixo-x determinam a contribuição da feature para cada classe do modelo. Sendo assim, foi observado que os valores *SHAP* a esquerda fazem contribuição para o modelo classificar como *The Economist*. Enquanto que os valores a direita do valor zero no eixo-x contribuem para classificar um *tweet* como escrito pelo *The Guardian*.

Portanto, os termos que melhor definem o discurso do *The Economist* tem como principais palavras *natashaloder*, *world*, *country*, *britain*, *effective*. Natasha Loder é editora de políticas de saúde no *The Economist*, as demais palavras podem representar notícias sobre diferentes localidades e a eficácia de vacinas ou medidas preventivas.

Para o *The Guardian*, temos as palavras *covid*, *coronavirus*, *uk*, *hope*, *oxford* e *trump*. As primeiras quatro palavras sugerem notícias sobre a pandemia, e notícias da vacina no Reino Unido (United Kingdom, U.K.). Além disso, prevalecem notícias sobre a vacina pesquisada pela Universidade de Oxford e por fim, possíveis críticas ou tomadas de decisões do Governo Trump.

4.3 Predizendo comentários para cada jornal

Regressão Logística

Um classificador de regressão logística também foi proposto para identificar os comentaristas do *The Guardian* e do *The Economist*. Como resultado obtivemos uma acurácia de 68,5%, um valor

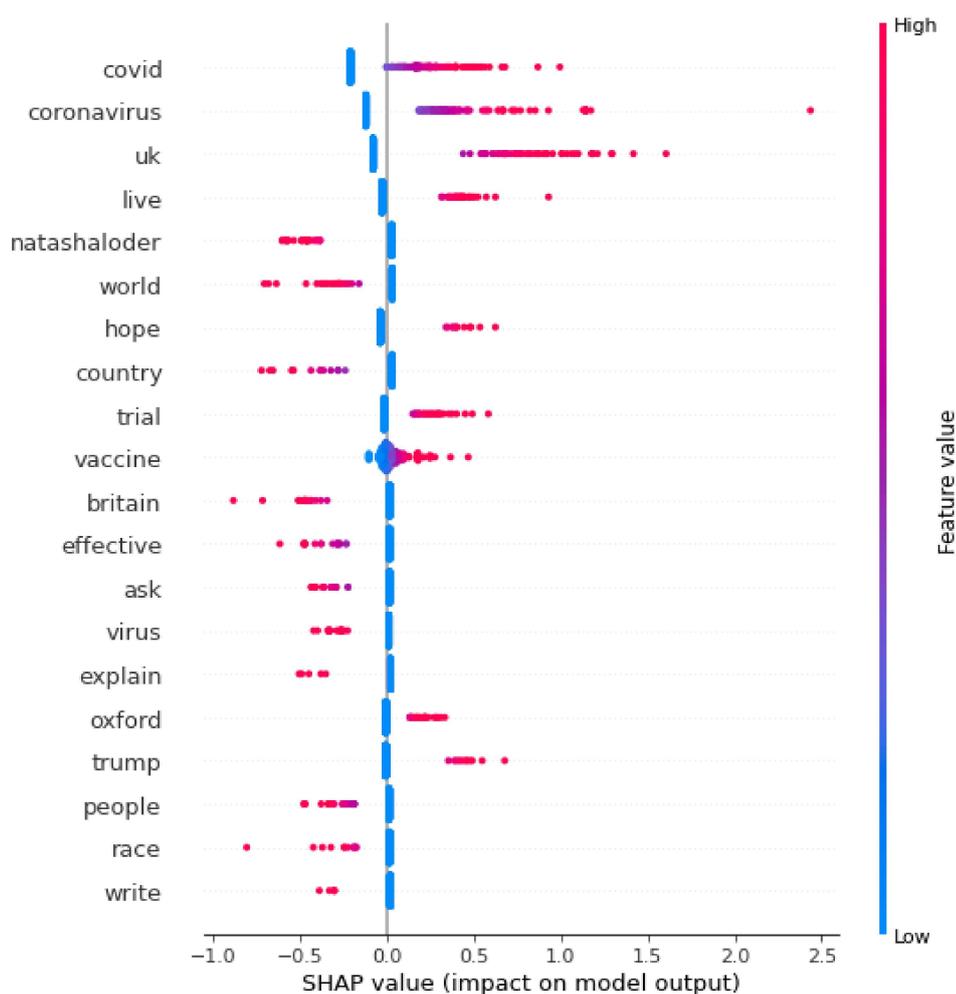


Figura 4.8: Características que mais contribuem para o modelo de predição dos Jornais.

que condiz com a tarefa de encontrar padrões textuais semelhantes entre os comentaristas de determinado jornal. Os valores das demais métricas na tabela 4.2 também foram condizentes com a acurácia obtida. Sendo assim, podemos diferenciar comentaristas do *The Guardian* e *The Economist* com uma acurácia de 68,5%, o que sugere um resultado promissor para definir homofilia sobre os indivíduos que comentam nas publicações de um mesmo jornal.

Tabela 4.2: Métricas de avaliação para o modelo de predição dos Comentários

Método	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.685	0.694	0.928	0.794

Interpretabilidade de Características

A ferramenta *SHAP* também foi utilizada para visualizar e entender quais características textuais influenciam mais na tomada de decisão do modelo para classificar se um comentário foi feito para responder o *The Guardian* ou o *The Economist*. Na figura 4.9 os valores do eixo-x,

assim como para os *tweets* de jornais, nos comentários os valores a esquerda do valor zero contribuíram mais para o modelo escolher um comentário como tendo sido feito para responder o jornal *The Economist*. Enquanto do lado direito, os valores contribuíram mais para classificar como um comentarista do *The Guardian*.

Os termos que definem os temas de interesse de comentaristas do *The Economist* geralmente são *gates*, *china*, *oxford*, *conspiracy*. A combinação dos termos *gates* e *conspiracy* sugerem que os comentaristas podem se referir a teorias conspiratórias envolvendo Bill Gates. A combinação dos termos *china* e *conspiracy* podem sugerir conspirações envolvendo o primeiro país a ser epicentro da doença covid-19. Por fim, o termo *oxford* que sugere comentários sobre a vacina desenvolvida pela universidade de Oxford, apesar desse termo ter aparecido na sessão anterior do estudo como uma característica que define o jornal *The Guardian*.

Para o *The Guardian* alguns dos termos que mais contribuíram foram *uk*, *brexit*, *woman*, *government*. Onde *uk* pode se referir as notícias feitas pelo *The Guardian* e que também foram fator decisivo para o modelo do respectivo jornal. Também podem ser percebidos comentários que fazem menções ao *Brexit*, possíveis críticas ao governo e menções feitas a mulheres.

Os resultados sugerem que apesar de o *The Economist* receber menos comentários possui temas que sugerem conspirações, o que realmente é um forte indício de grupos que possuem um padrão de texto semelhante e características que sugerem homofilia.

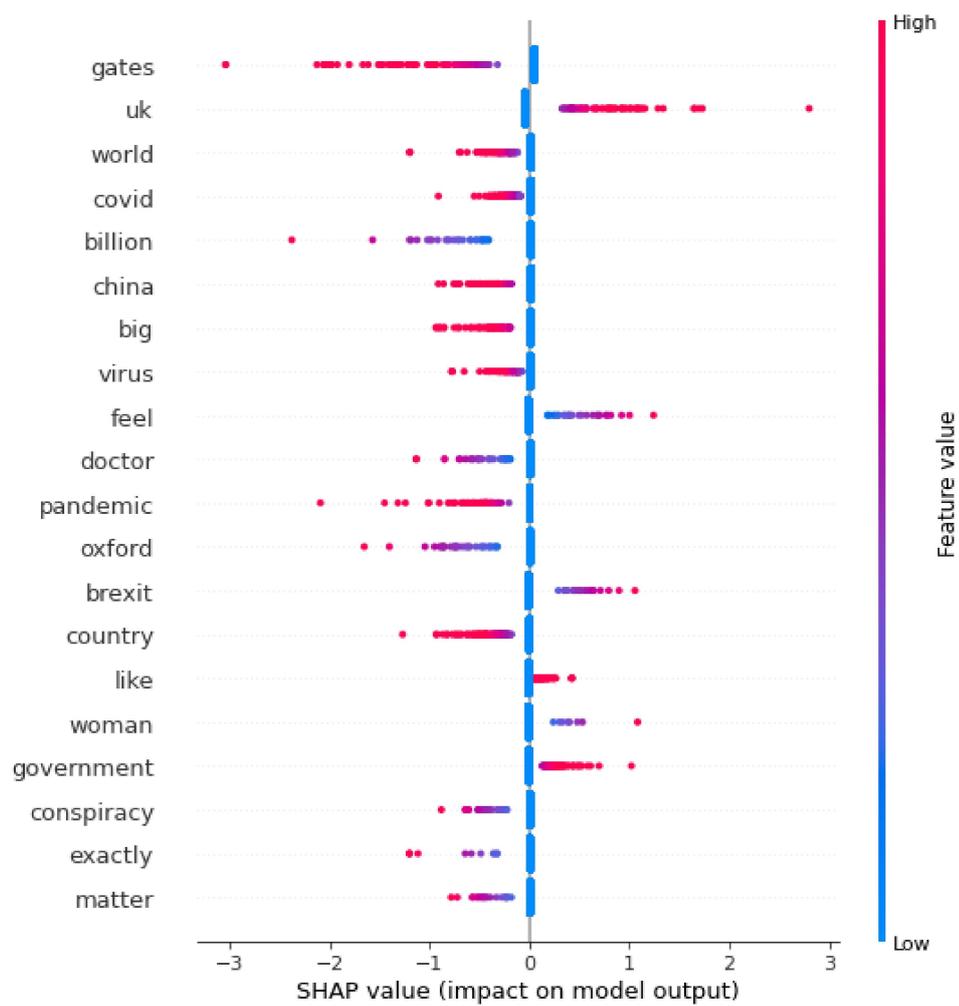


Figura 4.9: Características que mais contribuem para o modelo de predição dos Comentários.

5

Considerações Finais

Nesse trabalho, nós monitoramos a discussão sobre a vacinação da COVID-19 no contexto dos jornais The Guardian e The Economist, além dos comentários relacionados. Para isso, recorreremos à análise de sentimentos, indicadores de polarização e toxicidade para caracterizar os textos, e observamos que os tweets dos feitos pelo The Economist possuem média 40% mais positiva que o valor do The Guardian. Enquanto que os comentários de cada um dos jornais possuem médias iguais.

Além disso, observou-se um pico de comentários tóxicos no intervalo da semana 49 até a semana 51, sendo esse intervalo referente a ocasião em que as vacinas da *Pfizer* e da *Moderna* terminaram suas fases de teste, e a primeira pessoa foi vacinada no Reino Unido.

O modelo de predição proposto para os tweets de jornais possui acurácia de 92% e comprova que há diferenças nos discursos dos jornais sobre o tema de vacinação. Enquanto que o modelo de predição para os comentaristas de cada jornal conseguiu acurácia de 68,5% e conseguiu evidenciar que os termos textuais que melhor generalizam os comentaristas do The Economist são sobre conspirações, Bill Gates e a China.

Como direcionamentos futuros, podemos investigar e realizar uma caracterização coletando tweets de diversos jornais e seus comentários sobre o tema de vacinação da covid-19, inclusive no ano de 2021.

Referências bibliográficas

- Wen-Ying Sylvia Chou and Alexandra Budenz. Considering emotion in covid-19 vaccine communication: addressing vaccine hesitancy and fostering vaccine confidence. *Health communication*, 35(14):1718–1722, 2020.
- Jiashi Feng, Huan Xu, Shie Mannor, and Shuicheng Yan. Robust logistic regression and classification. *Advances in neural information processing systems*, 27:253–261, 2014.
- Samuel S Guimarães, Julio CS Reis, Filipe N Ribeiro, and Fabrício Benevenuto. Characterizing toxicity on facebook comments in brazil. In *Proceedings of the Brazilian Symposium on Multimedia and the Web*, pages 253–260, 2020.
- Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11):15169–15211, 2019.
- Neil F Johnson, Nicolas Velásquez, Nicholas Johnson Restrepo, Rhys Leahy, Nicholas Gabriel, Sara El Oud, Minzhang Zheng, Pedro Manrique, Stefan Wuchty, and Yonatan Lupu. The online competition between pro-and anti-vaccination views. *Nature*, 582(7811):230–233, 2020a.
- NF Johnson, N Velasquez, R Leahy, N Johnson Restrepo, O Jha, and Y Lupu. Not sure? handling hesitancy of covid-19 vaccines. *arXiv preprint arXiv:2009.08413*, 2020b.
- Na Yeon Lee and Kathleen McElroy. Online comments: The nature of comments on health journalism. *Computers in Human Behavior*, 92:282–287, 2019.
- Lucas Lima, Julio CS Reis, Philipe Melo, Fabricio Murai, and Fabrício Benevenuto. Characterizing (un) moderated textual data in social systems. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 430–434. IEEE, 2020.
- Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, 2(1):56–67, 2020.

- May Oo Lwin, Jiahui Lu, Anita Sheldenkar, Peter Johannes Schulz, Wonsun Shin, Raj Gupta, and Yinping Yang. Global sentiments surrounding the covid-19 pandemic on twitter: analysis of twitter trends. *JMIR public health and surveillance*, 6(2):e19447, 2020.
- Joan Massachs, Corrado Monti, Gianmarco De Francisci Morales, and Francesco Bonchi. Roots of trumpism: Homophily and social feedback in donald trump support on reddit. In *12th ACM Conference on Web Science*, pages 49–58, 2020.
- Murimo Bethel Mutanga and Abdultaofeek Abayomi. Tweeting on covid-19 pandemic in south africa: Lda-based topic modelling approach. *African Journal of Science, Technology, Innovation and Development*, pages 1–10, 2020.
- Filipe N Ribeiro, Matheus Araújo, Pollyanna Gonçalves, Marcos André Gonçalves, and Fabrício Benevenuto. Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, 5(1):1–29, 2016.
- Neil G Ruiz, Juliana Menasce Horowitz, and Christine Tamir. Many black and asian americans say they have experienced discrimination amid the covid-19 outbreak. *Pew Research Center*, 1, 2020.
- Zhang Yun-tao, Gong Ling, and Wang Yong-cheng. An improved tf-idf approach for text classification. *Journal of Zhejiang University-Science A*, 6(1):49–55, 2005.