CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA PAULA SOUZA
Faculdade de Tecnologia de Jundiaí – "Deputado Ary Fossen"
Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas

Mel Iza Costa Augusto

CLASSIFICAÇÃO DE FAKE NEWS A PARTIR DE DADOS TEXTUAIS COM MACHINE LEARNING

Jundiaí
2023

Mel Iza Costa Augusto

CLASSIFICAÇÃO DE FAKE NEWS A PARTIR DE DADOS TEXTUAIS COM MACHINE LEARNING

Trabalho de Graduação apresentado à Faculdade de Tecnologia de Jundiaí - “Deputado Ary Fossen” como requisito parcial para a obtenção do título de Tecnólogo em Análise e Desenvolvimento de sistemas, sob a orientação do Professor Mestre Peter Jandl Junior.

Jundiaí
2023

Dedico este trabalho à ciência brasileira e à classe trabalhadora.

AGRADECIMENTOS

Agradeço a todas as pessoas que me apoiaram durante o desenvolvimento desse trabalho e ao longo do curso. Gostaria de agradecer ao meu orientador professor Peter Jandl por sempre me ensinar, apoiar, incentivar a buscar conhecimento, além de ser uma fonte de inspiração profissional. Agradeço à minha família pelo suporte, amor e apoio incondicional durante essa jornada.

À medida que a inteligência artificial (IA) expande o escopo dos agentes autônomos, o desafio de como projetar esses agentes para que honrem o conjunto mais amplo de valores e leis que os humanos exigem dos agentes morais humanos torna-se cada vez mais urgente. A humanidade realmente quer que os computadores tomem decisões moralmente importantes?[1]

(WALLACH; ALLEN, 2009)

[1] Tradução livre nossa: “As artificial intelligence (AI) expands the scope os autonomous agents, the challenge of how to design these agents so that they honor the broader set of values and laws humans demand of human moral agents becomes increasingly urgent. Does humanity really want computers making morally importante decisions?” Moral Machines: Teaching robots right from wrong (WALLACH; ALLEN, 2009)

AUGUSTO, Mel Iza Costa. **Classificação de *fake news* a partir de dados textuais com Machine Learning.** 75 páginas. Trabalho de Conclusão de Curso de Tecnólogo em Análise e Desenvolvimento de Sistemas. Faculdade de Tecnologia de Jundiaí - “Deputado Ary Fossen”. Centro Estadual de Educação Tecnológica Paula Souza. Jundiaí. 2023.

RESUMO

A pandemia da COVID-19, causada pelo coronavírus SARS-CoV-2, impôs desafios significativos para a saúde e sociedade desde a declaração de Emergência de Saúde Pública de Importância Internacional pela OMS[2]. Além das mudanças na rotina e no trabalho, a forma de comunicação foi transformada, resultando em um aumento expressivo no compartilhamento de informações. No entanto, esse cenário também propiciou a disseminação em massa de informações falsas e desinformação, principalmente por meio de redes sociais e aplicativos de mensagens. Por meio da implementação de modelos de machine learning treinados em conjuntos de dados sobre vacinação, o objetivo do presente trabalho foi classificar e categorizar precisamente notícias e/ou mensagens como falsas ou verdadeiras com base nos padrões e estudos sobre vacinação. Os conjuntos de dados foram coletados em diversas fontes, variando entre grupos de redes sociais e agência verificadoras de fatos. Estima-se que o presente trabalho auxilie a identificação de notícias falsas ou *fake news* sobre o tema selecionado.

Palavras-chave: Inteligência Artificial (IA), aprendizagem de máquina, classificação de dados, notícias falsas.

[2] Organização mundial da Saúde

AUGUSTO, Mel Iza Costa. **Classificação de *fake news* a partir de dados textuais com machine learning.** 75 páginas. Trabalho de Conclusão de Curso de Tecnólogo em Análise e Desenvolvimento de Sistemas. Faculdade de Tecnologia de Jundiaí - “Deputado Ary Fossen”. Centro Estadual de Educação Tecnológica Paula Souza. Jundiaí. 2023.

ABSTRACT

The COVID-19 pandemic, caused by the SARS-CoV-2 coronavirus, has posed significant challenges for health and society since the declaration of a Public Health Emergency of International Concern by the WHO . In addition to changes in routine and work, the form of communication was transformed, resulting in a significant increase in information sharing. However, this scenario also led to the mass dissemination of false information and misinformation, mainly through social networks and messaging applications. Through the implementation of machine learning models trained on vaccination datasets, the objective of the present work is to precisely classify and categorize news and/or messages as false or true based on patterns and studies on vaccination. Datasets were collected from a variety of sources, ranging from social media groups to fact-checking agencies. It is estimated that the present work will help to identify false news or fake news on the selected topic.

Keywords: Artificial intelligence (AI), machine learning, data classification, fake news

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

Existem diversas situações cotidianas em que precisamos saber se algo pertence a uma categoria ou não, por exemplo, se vai chover ou não, se a comida está pronta ou não, ou se o limite do cartão foi ultrapassado, entre muitos outros exemplos. Em muitos casos não temos dados suficientes para saber questões como essas; e até nos casos que temos, acabamos adivinhando mal, ainda que tenhamos clareza do contexto. Tanto nossos raciocínios indutivos como dedutivos muitas vezes são falhos em cenário que precisamos fazer predições ou suposições, muito embora tenhamos a melhor situação com as melhores informações sobre algo (KAHNEMAN, 2012).

As inteligências artificiais, por outro lado, conseguem fazer isso de maneira eficiente e muito mais precisa - levando em conta a melhor situação e os melhores dados disponíveis. Por isso elas são empregadas em diferentes contextos para resolver problemas. De acordo com a concepção de agente inteligente sob a abordagem da ação humana, as inteligências artificiais seriam capazes de possuir as capacidades de 1) processar a linguagem natural, 2) representar o conhecimento, 3) automatizar o raciocínio e 4) aprender para se adaptar a novas circunstâncias (RUSSEL; NORVIG, 2013) Isso inclui processar dados para representação do conhecimento e aprendizado, e consequentemente resolver diversos dos tipos de problemas como os exemplos mencionados acima.

O risco de cometer erros ao fazer suposições sobre objetos e o mundo é inerente ao aprendizado humano, pois é impossível – dada a limitação da nossa cognição - que se conheça tudo o que há para conhecer sobre todas as coisas. Essa é uma característica presente no aprendizado de máquina, em que o objetivo final é diminuir o erro. Sendo assim, o ato de classificar ou fazer predições numéricas por uma máquina é uma tarefa em princípio, matemática, que envolve ciclos iterativos para diminuição de erros.

Agentes inteligentes racionais tomam a melhor decisão possível dada uma situação. Isso compreende ter à disposição um input de coleção de percepções, que serão

transformadas em informações a partir de padrões e assim transformar isso em uma medida. Essa medida é o aprendizado (NORVIG, RUSSEL, 2013). Algoritmos de aprendizado de máquina funcionam a partir de um conjunto de dados, que serão analisados e terão padrões extraídos para obter previsões de valores, categorias, agrupamentos ou mesmo prever anomalias.

Casos de aplicações variam segundo o tipo de problema abordado, recursos disponíveis e tipos de dados e características que eles possuem. A depender do tipo de combinação entre essas variáveis, pode-se optar por alguns caminhos dentro dos tipos principais de machine learning. De acordo com (ZUBAREV, 2019) os tipos principais se dividem em 1) aprendizado de máquina clássico, 2) aprendizado por reforço, 3) ensembles (ou conjuntos de modelos) e 4) redes neurais e aprendizado profundo.

Dentro do escopo de machine learning clássico, a divisão dos algoritmos se baseia no critério de categoria do dado: se o dado ou conjunto de dados possuir uma categoria (seja ela numérica ou não), ele é considerado como parte do aprendizado supervisionado. Por exemplo: predizer a temperatura (valor numérico) ou predizer se fará calor ou frio (valor categórico) na cidade de Jundiaí na primeira semana de Junho baseado nos dados dos últimos meses é um exemplo de problema de aprendizado supervisionado – no primeiro caso, um exemplo de problema de regressão linear e no segundo caso um exemplo de regressão logística. Algumas aplicações do aprendizado supervisionado envolvem predizer valores do mercado financeiro, fazer filtragem de spam, fazer diagnósticos médicos, análise de sentimentos, detecção de fraudes (ZUBAREV 2019) e saber se um nome é de um pokémon ou framework de machine learning[3].

[3] Is it Pokémon or #bigdata technology?2' e 'Valohai: Pokémon or MLOps' é um projeto da área do entretenimento em que o público acessa uma aplicação ou formulário e seleciona ou tenta adivinhar se o nome que está aparecendo na tela é de um Pokémon ou de um framework de Machine Learning.Disponível em: https://valohai.com/mlops-or-pokemon/ . Esse inicialmente seria o tema deste trabalho, mas por meio da discussão de alguns critérios sobre o alcance desse conjunto de técnicas, o tema sobre o impacto social foi um fator altamente decisivo.

Caso os dados não possuam nenhum tipo de informações prévia, esse tipo de problema se encaixa na categoria de aprendizado não supervisionado. Nesse tipo de aprendizado, os dados são segmentados e/ou divididos por similaridade entre si (clustering). Geralmente exemplos de uso de algoritmos de clustering estão associados com segmentação do mercado (tipos de clientes, fidelização). Mesclar pontos próximos em um mapa. Compressão de imagens e analisar e rotular dados novos, além de detectar comportamento anormal[4] (ZUBAREV 2019).

O presente trabalho se encaixa dentro do escopo do aprendizado supervisionado, tendo como principal questionamento base a pergunta: “Como saber se uma mensagem ou informação sobre vacinação é falsa ou verdadeira?”. Desde a experiência com a pandemia da COVID-19 a relação com a forma de nos comunicarmos foi ressignificada, tendo em vista períodos de isolamento social em que as pessoas não podiam se encontrar pessoalmente como em momentos anteriores à pandemia. Muitos aplicativos de mensageria e redes sociais se tornaram uma das maiores fontes de comunicação, transmitindo milhares de mensagens. Em um estudo sobre a utilização de mídias sociais durante a pandemia da COVID-19 e análise comportamental dos usuários:

> Particularidades importantes da pandemia do COVID-19 alteraram a forma ao qual utilizamos as mídias sociais, podendo evidenciar uma busca ampliada por entretenimento e positividade como forma de aliviar o estresse cotidiano, a socialização passando a ser em grande parte digital o que forçou a maior parte das pessoas a uma adaptação veloz e aspectos do cotidiano que foram introduzidos nas plataformas de mídia social trazendo uma comunicação mais informal e intimista. (SANDRINI BEZERRA, L.; GIBERTONI, D, 2021, p. 156)

Com a utilização das redes e maior circulação de informações durante esse período, também abriu espaço para a veiculação de informações falsas e errôneas sobre diversos assuntos, principalmente sobre saúde pública. As notícias falsas ou *fake*

[4] Tradução livre nossa: “For market segmentation (types of customers, loyalty).To merge close points on a map. For image compression. To analyze and label new data. To detect abnormal behavior”

news sobre saúde, vacinação e uso de medicamentos se tornou extremante preocupante e de difícil manejo para aplicação de medidas sanitárias durante a pandemia:

> a disseminação de teorias da conspiração em redes sociais e a politização de questões de saúde pública – como a conveniência do uso de determinados medicamentos para tratar a doença – são fatores que tornam o combate do vírus pela sociedade muito mais difícil (BAICKER; BOGGIO 2020 APUD HARTMANN, I. A., & IUNES, J 2020, P. 390)

Dado o desafio de lidar com mensagens desse tipo, é de importância social que se reconheça em primeiro lugar a identificar tais notícias. E como agentes artificiais são capazes de reconhecer padrões de forma eficiente, serão a base técnica do trabalho na detecção de *fake news* sobre vacinação. Neste trabalho foram coletados dados de redes sociais e agências verificadoras de fatos, bem como dados de projetos já publicados sobre o tópico. A partir de uma filtragem por assunto que envolvem os termos sinônimos a vacinação, os conjuntos de dados foram pré-processados, treinados e testados em modelos de classificação.

Ao longo do desenvolvimento do trabalho os seguintes tópicos estão organizados da seguinte forma: No capítulo 2 será abordado a especificação do programa, escopo e principais referências e clientes de software. No capítulo 3 o tópico discutido será sobre os requisitos do sistema, tanto requisitos funcionais como não funcionais. No capítulo 4 será abordado a definição do projeto, que envolve: inteligência artificial e aprendizagem, processamento de linguagem natural e descrição, desenvolvimento dos processos e diagramas e documentos de caso de uso. No capítulo 5 será apresentado as análises e resultados. No capítulo 6 a escolha da arquitetura da solução será detalhada e por último, no capítulo 7, serão apresentadas as considerações finais sobre o desenvolvimento da produção aqui desenvolvida.

2 ESPECIFICAÇÃO DO PROGRAMA

A especificação de software é uma das principais atividades que compõe o desenvolvimento profissional de software e é prosseguida pelo desenvolvimento do software, validação do software e evolução (SOMMERVILE, 2011). A especificação como parte integrante do desenvolvimento de software é importante pois ajuda a esclarecer tudo aquilo que o sistema ou aplicação deve possuir, quais comportamentos deve ter diante de determinados ambientes e alinhar as expectativas quanto ao que o cliente quer e precisa que seja implementado.

Sendo assim, o trabalho se destina a produzir a aplicação de algoritmos de classificação para identificar notícias falsas sobre vacinação a partir da coleta e processamento de dados. Para tal, será estruturado um pipeline para processar os dados, aplicar a classificação e avaliar seu desempenho, bem como suas implicações e análises.

Algumas atividades chave para esse processo incluem: 1) delimitar o assunto abordado para um campo de pesquisa específico, 2) coletar e processar dados sobre o tema selecionado, 3) executar uma breve análise sobre os dados adquiridos, 4) aplicar um modelo de classificação e 4) avaliar o desempenho do modelo por meio de métricas.

Por se tratar de um trabalho de software não convencional de acordo com o desenvolvimento de um sistema, apenas alguns tipos de diagramas foram selecionados. Nas sessões seguintes será descrito o escopo do software e todas as funcionalidades previstas, bem como os clientes de software.

2.1 ESCOPO

O escopo do sistema possui como objetivo descrever e listar todas as características que devem estar presentes no funcionamento do software bem como discutir e levantar suas funcionalidades e requisitos. De acordo com Sommervile (2011, p. 57): “Os requisitos de um sistema são as descrições do que o sistema deve fazer, os serviços oferecem e as restrições a seu funcionamento”.

Dentre os tópicos abordados, será importante explicitar o problema base a partir da premissa sobre a necessidade de identificar notícias falsas sobre vacinação; e elucidar os assuntos pertencentes aos clientes de software.

2.2 CLIENTES DE SOFTWARE

O público-alvo reconhecido como principal cliente de software da aplicação envolve dois perfis específicos. A primeira categoria inclui profissionais da área: cientistas de dados e analistas de dados e a segunda categoria se deriva de pessoas interessadas no assunto que porventura tenham curiosidade de testar ou conhecer o trabalho.

O primeiro perfil de cliente de software envolvendo tanto o cientista como o analista de dados têm em seu escopo de atuação, acesso interno à aplicação, embora não restrito à parte interna. Suas principais funções envolvem acessar, visualizar, conferir e executar qualquer etapa do trabalho.

O segundo perfil de cliente de software será referenciado como pessoa interessada, englobando qualquer faixa etária (a partir de alfabetização) que se interesse pela aplicação. O segundo perfil de cliente deve possuir um conhecimento mínimo sobre as ferramentas do *Google Colaboratory*[5], *Jupyter Notebook*[6] ou Github[7] para

[5] https://colab.research.google.com/

acessar o trabalho, especificamente dentro do escopo aqui citado e do assunto a que se refere, e o primeiro perfil de cliente deve possuir um conhecimento maior.

[6] https://jupyter.org
[7] https://github.com

3 REQUISITOS DO SISTEMA

Os principais requisitos do sistema envolvem o processo de captação de dados e estruturação de um pipeline para que a o algoritmo classifique uma notícia com a probabilidade de ser verdadeira ou falsa dado um texto como entrada.

De modo geral, os requisitos são distribuídos de maneira que os clientes de software possam interagir com a aplicação. Esse contato será por meio de um texto inserido em parte do código em que o modelo recebe o texto de teste e retorne à probabilidade de ele ser verdadeiro ou falso.

3.1 REQUISITOS FUNCIONAIS

De modo geral, os requisitos funcionais podem ser definidos como requisitos que "[...] descrevem o que um sistema deve fazer. Eles dependem do tipo de software a ser desenvolvido, de quem são seus possíveis usuários e da abordagem geral adotada pela organização ao escrever os requisitos" (SOMMERVILLE, 2011, p. 59).

Como o trabalho visa desenvolver um projeto de software na área que envolve ciência de dados, aprendizagem de máquina e inteligência artificial, seus requisitos funcionais se baseiam na estrutura de desenvolvimento de um projeto do tipo – coleta de dados, processamento de dados, modelagem de dados e deploy de modelo.

Os requisitos funcionais de acordo com os critérios estabelecidos são:

Tabela 1. Requisitos Funcionais

Requisito Funcional (RF)	**Nome do Requisito Funcional**	**Descrição do Requisito Funcional**
RF01	Verificar confiança	O software deve determinar com confiança acima de 50% em qual

		categoria o texto se encaixa.
RF02	Processar dados	O software de deve possuir um algoritmo que processe dados estruturados e padronizados para teste e treino.
RF03	Efetuar classificação	O algoritmo deve ser capaz de realizar uma classificação binária (entre duas categorias – notícia falsa ou notícia verdadeira).
RF04	Estruturar pipeline	O software deve possuir um pipeline para automatização do processo de tratamento de dados.
RF06	Verificar inputs e outputs	O software deve possuir um texto como entrada e retornar à probabilidade de pertencimento a uma das classes.
RF07	Garantir estrutura de funcionamento	O software deve ser capaz de receber uma string de entrada (input), processar e padronizar os dados, classificar e indicar a classe pertencida nessa ordem.

3.2 REQUISITOS NÃO FUNCIONAIS

Os requisitos não funcionais, por outro lado podem ser definidos como requisitos “[...] que não estão diretamente relacionados com os serviços específicos oferecidos pelo sistema a seus usuários.” (SOMMERVILLE, 2011, p. 60).

Os requisitos não funcionais estão relacionados com as tarefas que devem ser executadas para garantir que a disponibilização e desempenho do trabalho. Os requisitos não funcionais de acordo com os critérios estabelecidos são:

Tabela 2. Requisitos Não funcionais

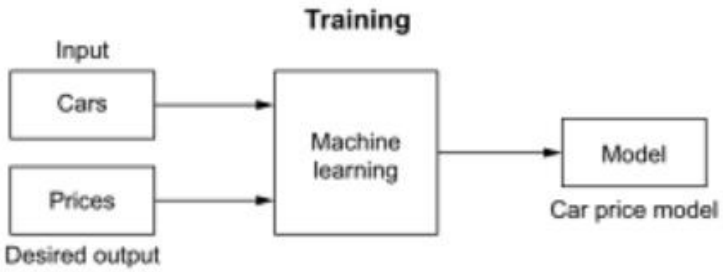
Requisito Não Funcional (RNF)	**Nome do Requisito Não Funcional**	**Descrição do Requisito Não Funcional**
RNF01	Separar dados	Os dados da base de teste não podem ser inseridos no conjunto de dados para treino
RNF02	Disponibilidade e facilidade de acesso	O modelo deve ser disponibilizado para usuários acessarem bem como para revisão de pares com conhecimento técnico básico
RNF03	Responsabilidade e transparência	O processo de aquisição de dados deve ser bem documentado e respeitar princípios éticos e legais, além de explicitar os vieses presentes.
RNF04	Disponibilizar trabalho	O produto final deve possuir um algoritmo

		disponibilizado em um ambiente para que seus clientes de software possam acessá-lo, como a plataforma GitHub e notebooks do Google Colab ou Jupyter Notebook na linguagem Python

4 DEFINIÇÃO DO PROJETO

O projeto consiste no desenvolvimento de algoritmos de aprendizado de máquina ou machine learning e sua aplicação ao problema de identificação de notícias falsas sobre vacinação. De modo geral, a estrutura do projeto segue o funcionamento de um projeto dentro da área de ciência de dados – a partir da etapa de coleta de dados com informações textuais sobre o tema de interesse, esse conjunto será passado aos algoritmos, eles aprenderão os padrões e características presentes no conjunto de treinamento, e farão uma previsão em dados não vistos anteriormente, os dados de teste.

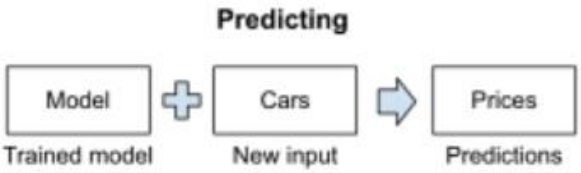
Figura 1. Funcionamento de um algoritmo de machine learning



Fonte: GRIGOREV, 2021

De acordo com um exemplo fornecido por (GRIGOREV, 2021) para ilustrar este tópico, a ideia principal funcionando por trás do aprendizado de máquina são exemplos. Se fornecermos um conjunto de dados com informações sobre carros (cor, tipo, quilometragem etc.) e passarmos o preço como output desejado, o sistema vai tentar por si só aprender quais características são importantes para o carro, e descobrir o output sem perguntar a um humano.

Figura 2. Exemplo sobre predição a partir de um modelo treinado



Fonte: GRIGOREV, 2021

Quando passamos à máquina o conjunto de dados com informações sobre carros, estamos fornecendo todos os exemplos para que ele aprenda – treine. Com isso em mente, formamos um novo conjunto de dados – os quais não sabemos os preços – e passamos ao modelo para que ele estenda as previsões baseadas nos aprendizados adquiridos durante o treino.

É importante destacar o quanto esse processo se difere de um processe tradicional de software:

> No aprendizado de máquina, damos ao sistema a entrada nos dados de saída e o resultado é um modelo (código) que pode transformar a entrada na saída. O trabalho difícil é feito pela máquina; precisamos apenas supervisionar o processo de treinamento para garantir que o modelo seja bom. Em contraste, nos sistemas tradicionais, primeiro encontramos os padrões nos dados e, em seguida, escrevemos o código que converte os dados no resultado desejado, usando os padrões descobertos manualmente.[8] (GRIGOREV, 2021, p. 4)

O processo de desenvolvimento de machine learning envolve a automação de tarefas, seguindo essa linha de raciocínio. Apesar de isso facilitar muito do processo na criação de um produto, por exemplo, essa parte é apenas um dos componentes do processo de aprendizado de máquina.

Existe uma metodologia criada a partir de um modelo de processo independente da indústria para mineração de dados (SCHRÖER, KRUSE, 2021). Essa metodologia é denominada CRISP-DM (*Cross-Industry Standard Process for Data Mining*) e consiste em seis fases iterativas desde o entendimento de negócio, até a implantação.

---

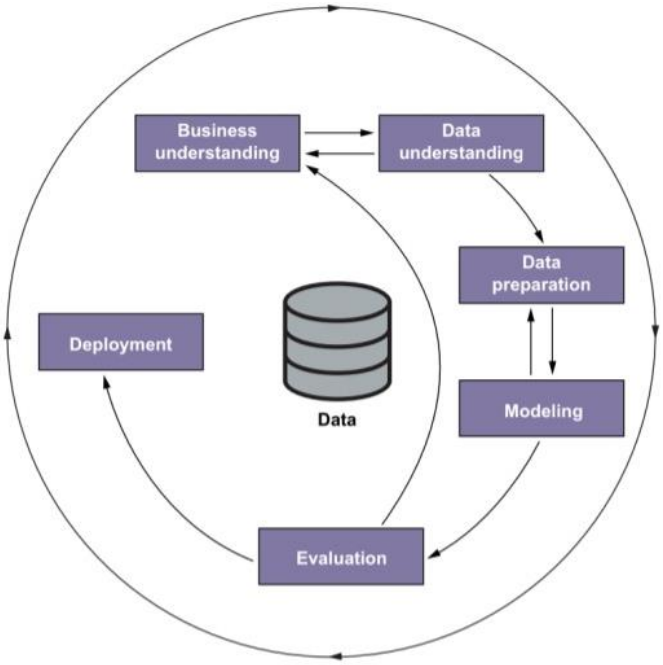
[8] Tradução livre nossa: “In machine learning, we give the system the input into the output data, and the result is a model (code) that can transform the input into the output. The difficult work is done by the machine; we need only supervise the trainning process to make sure that the model is good. In contrast, in traditional systems, we first find the patterns in the data ourselves and then write the code that converts the data to the desired outcome, using the mnually discovered patterns” (GRIGOREV, 2021, p. 4)

De acordo com a revisão sistemática da literatura sobre aplicação do modelo de processo CRISP-DM apresentada por Schröer e Kruse, as fases do CRISP-DM são divididas da seguinte maneira:

1. Entendimento de negócio: A situação do negócio deve ser avaliada para obter uma visão geral dos recursos disponíveis e necessários. A determinação de o objetivo da mineração de dados é um dos aspectos mais importantes nesta fase.
2. Entendimento dos dados: Coletar dados de fontes de dados, explorá-los e descrevê-los e verificar a qualidade dos dados são tarefas essenciais nesta fase. Para torná-lo mais concreto, o guia do usuário descreve a tarefa de descrição de dados usando análise estatística e determinando atributos e seus agrupamentos:
3. Preparação dos dados: A seleção dos dados deve ser realizada definindo critérios de inclusão e exclusão. A má qualidade dos dados pode ser tratada por dados de limpeza.
4. Modelagem: A fase de modelagem de dados consiste em selecionar a técnica de modelagem, construir o caso de teste e o modelo.
5. Avaliação: Na fase de avaliação, os resultados são verificados em relação aos objetivos de negócios definidos
6. Implantação: A fase de implantação é descrita geralmente no guia do usuário. Pode ser um relatório final ou um componente de software. O usuário guia descreve que a fase de implantação consiste em planejar a implantação, monitoramento e manutenção (Deployment) (SCHRÖER, KRUSE, 2021, p. 527)

A definição e estrutura do projeto são baseadas na CRISP-DM, no qual serão explorados 1) a problemática sobre como identificar notícias falsas na área da saúde – vacinação em tempos de desinformação sobre o tópico; 2) coleta de dados textuais em redes sociais, agências verificadoras de fatos e revisão bibliográfica, levando em conta os critérios de inclusão e exclusão adotados; 3) Aplicação de técnicas de processamento e limpeza de dados, como padronização, normalização, extração de features, e filtragem de assunto; 4) aplicação dos conjuntos de treino e teste à modelos de classificação; 5) avaliação e interpretação de métricas e por fim 6) deploy da aplicação.

Figura 3. Ilustração do funcionamento da Metodologia CRISP-DM



Fonte: GRIGOREV, 2021 1

4.1 POR QUE FALAR SOBRE NOTÍCIAS FALSAS É IMPORTANTE?

Nos comunicamos utilizando os recursos linguísticos disponíveis que possuímos em nossa língua materna. Entre os diversos tipos de comunicação que utilizamos, podemos destacar a comunicação oral e comunicação escrita - dentro da comunicação escrita, nos expressamos de acordo com as regras estabelecidas pela linguagem, por meio da sintaxe, semântica e regras estabelecidas.

Além da utilização da linguagem para interação com outras pessoas, expressamos nossas crenças e visões de mundo. E não só expressamos, como também argumentamos sobre elas e oferecemos argumentos que possam sustentar essas crenças. De forma geral:

> Justificar uma afirmação que se faz, ou dar razões para uma certa conclusão obtida, é algo que bastante importância em muitas situações [...]. A importância da boa justificativa vem do fato que muitas vezes cometemos erros de raciocínio, chegando a uma informação que simplesmente não decorre da informação disponível (MORTARI, 2001, p. 6)

Essa verificação é importante não apenas do ponto de vista da epistemologia, ou seja, do modo como nossas crenças estão estruturadas, mas em última análise, mas da realidade em si. Muitas vezes podemos fornecer informações errôneas, mas o que determina se ela é verdadeira, não é a vontade do próprio indivíduo, mas sim a realidade.

Levando em consideração que apliquemos esse raciocínio ao ambiente digital, em que muitas pessoas compartilham informações o tempo todo, “tornou-se mais fácil propagar qualquer informação para as massas em poucos minutos.” (BHATT, 2017, p.1) De acordo com as premissas acima, é importante falar sobre notícias falsas porque elas em última instância podem: 1) distorcer a visão que as pessoas possuem da realidade, 2) incentivar a promoção e propagação de mais notícias

falsas, 3) incentivar o descaso epistêmico[9], 4) incentivar comportamentos que possam ser prejudiciais à saúde e ao bem-estar social – principalmente quando se trata de saúde pública.

[9] “Trata-se de uma falta de preocupação descontraída quanto à questão de as nossas crenças terem ou não alguma base na realidade ou de as melhores provas disponíveis as apoiarem ou não.” (MURCHO, 2018)

4.1.1 VACINAÇÃO NO BRASIL

O Brasil se tornou referência em vacinação mundial pela criação do PNI (Plano nacional de imunização) em 1973 pelo Departamento Nacional de Profilaxia e Controle de Doenças (Ministério da Saúde) e da Central de Medicamentos (CEME - Presidência da República).[10] A partir do planejamento de campanhas do PNI ao longo do tempo, desde sua fundação, conseguiu a:

> Eliminação do sarampo e do tétano neonatal. A essas, se soma o controle de outras doenças imunopreveníveis como Difteria, Coqueluche e Tétano acidental, Hepatite B, Meningites, Febre Amarela, formas graves da Tuberculose, Rubéola e Caxumba em alguns Estados, bem como, a manutenção da erradicação da Poliomielite. (SAÚDE, 2023)

O Brasil então se tornou referência mundial em vacinação, tendo o PNI integrado ao Programa da Organização Mundial da Saúde com apoio da UNICEF além da criação dos CRIE (Centro de Referência para imunobiológicos Especiais).

A importância da discussão do tema ressurge a partir do momento em que o Brasil, que antes era referência, se encaixa na segunda posição de países com pior taxa de cobertura vacinal cobertura vacinal em bebês[11] (VARELLA, 2023). Não somente nesse faixa etária, mas alguns dados apontados pela FIOCRUZ no ano de 2022:

> Dados divulgados pelo Fundo das Nações Unidas para a Infância (Unicef) mostram que a taxa de vacinação infantil no Brasil vem sofrendo uma queda brusca: a taxa caiu de 93,1% para 71,49%. De acordo com a pesquisa, realizada em parceria com a Organização Mundial da Saúde (OMS), esse número coloca o Brasil entre os dez países com menor cobertura vacinal do mundo. (FIOCRUZ, 2023)

---

[10] Programa Nacional de Imunizações - Vacinação", [s.d.]). Disponível em: https://www.gov.br/saude/pt-br/acesso-a-informacao/acoes-e-programas/programa-nacional-de-imunizacoes-vacinacao#:~:text=Em%201973%20foi%20formulado%20o,pela%20reduzida%20área%20de%20cobertura.

[11] Relatório da UNICEF: The state of the world 's children 2023: For Every Child, Vaccination". Disponível em: https://www.unicef.org/media/108161/file/SOWC-2023-full-report-English.pdf

De acordo com algumas revisões bibliográficas, a baixa adesão à vacinação no Brasil em alguns casos, como a poliomielite e sarampo depois de muitos anos em controle, tem o risco de voltar a afetar crianças, adolescentes e adultos. Também se encontra na lista o risco da volta do HPV - IST causada pelo *Human papilomavírus* (PEREIRA; FERNANDES; CARNEIRO, 2021). A SBIM (Sociedade Brasileira de Imunização) aponta que a queda da vacinação no Brasil se iniciou em 2015, com expressivo aumento no período pós pandemia.[12]

De acordo com o levantamento realizado pela UNICEF, “a confiança da população nas vacinas caiu depois da pandemia: antes, 99,1% dos brasileiros confiavam nas vacinas infantis, taxa que hoje está em 88,8%” (UNICEF apud VARELLA, 2023). Muito do que se observa com relação ao comportamento de hesitação vacinal, foi intensificado no período pós pandêmico, em que a eficácia, utilização e efeito das vacinas – principalmente sobre o COVID-19 – foi posta em dúvida.

Não é possível estabelecer uma relação de causalidade apenas por essas fontes entre a pandemia e movimentos antivacinas. Mas podemos afirmar que de alguma forma, esses fenômenos se relacionam, pois a cobertura vacinal da população brasileira diminui substancialmente nos últimos anos, algumas doenças que foram erradicadas não são mais consideradas erradicadas, e muitas mensagens com cunho negativo sobre as vacinas começaram a circular nas redes sociais.

> Entre os motivos apontados pela Sociedade Brasileira de Imunizações (SBIm) para a queda estão falta de informação dos profissionais de saúde acerca do calendário vacinal; falta de informação da população; pouca confiança em governantes, instituições e profissionais de saúde; horário limitado de funcionamento dos postos de saúde; desinformação; comunicação falha; e crescimento do movimento antivacinista. (VARELLA, 2022)

[12] Disponível em: https://sbim.org.br/noticias/1790-ministerio-sbim-e-outras-entidades-cientificas-se-unem-em-prol-das-altas-coberturas

4.1.2 COMPARTILHAMENTO DE INFORMAÇÕES

As redes de interação dentro da internet, que denominamos redes sociais, enquanto ferramentas possibilitadoras de compartilhamento de informações e socialização mudou nossa forma de se relacionar com a tecnologia e abriu espaço para ambientes novos e diferentes. A partir do exame etimológico da própria construção da palavra rede, Zenha mostra como foi possível a criação desse espaço: “Essas e outras redes tecidas no espaço virtual só foram possíveis devido à união de três processos independentes: a expressão da diversidade, a comunicação e os avanços da tecnologia” (ZENHA, 2018, p.23).

Ainda sob o contexto que Zenha traz: “partir do ambiente virtual, as redes sociais têm como base a interação síncrona e assíncrona, nas quais os indivíduos que a realizam exercem papel de protagonista das e nas relações sociais que estabelecem na rede” (ZENHA, 2018, p.30). Portanto, as redes sociais se tornam um ambiente de trocas de ideias, experiências, aprendizados e proporciona uma oportunidade para pessoas que gostam das mesmas ideias e assuntos, se juntarem e compartilharem. Desde a criação de redes como Orkut[13], Facebook[14], Twitter[15], Instagram[16] entre outras, muitas pessoas se tornaram usuárias ativas. Muitos fatores são atrativos, mas se detacam:

> As redes sociais ganharam seu lugar de uma maneira vertiginosa nas trocas colaborativas na vida de adultos e jovens. As redes sociais proporcionam um aumento significativo das interações e da conectividade entre grupos sociais por serem um meio promissor de divulgação de conteúdo e de propagação de ideias. O diferencial das redes sociais está na facilidade que possuem para construir as mensagens; a facilidade na veiculação, o acesso rápido e em pontos distanciados que proporcionam as trocas de saberes disponibilizados pelos pontos na rede sociais (2018 ZENHA, apud RECUERO 2009).

[13] Orkut: https://pt.wikipedia.org/wiki/Orkut
[14] Fabeook: https://pt-br.facebook.com
[15] Twitter: https://twitter.com/
[16] Instagram: www.instagram.com/

O mesmo ambiente proporcionado para troca de informações, aprendizado, criação de novos laços e conhecimento, também foi o mesmo ambiente proporcionado para que informações como: “coronavirus como a gripe e uma doenca sazonal e as vacinas nao salvam vidas”[17] se espalhasse por milhares de grupos.

O compartilhamento em massa de informações falsas sobre determinado assunto pode ser considerado desinformação. De acordo com a OPAS (Organização Panamericana de Saúde), “Desinformação é uma informação falsa ou imprecisa cuja intenção deliberada é enganar.”[18] O compartilhamento online de desinformação tornou-se uma preocupação mundial com sérias consequências econômicas, políticas e sociais. Mais recentemente, a desinformação impediu aceitação de vacinas COVID-19 e medidas de mitigação[19]. (CEYLANA; ANDERSONB; WOOD, 2023).

Muitas informações falsas são compartilhadas e conseguem alcanço por parecer ter aspectos de certa forma verdadeiros à elas, quando na verdade não possuem:

> A desinformação pode prejudicar a saúde humana. Muitas histórias falsas ou enganosas são inventadas e compartilhadas sem que se verifique a fonte nem a qualidade. Grande parte dessas desinformações se baseia em teorias conspiratórias; algumas inserem elementos dessas teorias em um discurso que parece convencional (OPAS, 2020)

De acordo com um estudo publicado pela Universidade de Princeton em 2022 acerca do motivo de as pessoas compartilharem informações falsas em redes sociais, eles concluíram alguns pontos interessantes sobre os mecanismos de

---

[17] Retirado de uma amostra do conjunto de dados coletado

[18] Entenda a infodemia e a desinformação contra a COVID-19 (OPAS). Disponível em: https://iris.paho.org/bitstream/handle/10665.2/52054/Factsheet-Infodemic_por.pdf

[19] Tradução livre nossa: “The online sharing of misinformation has become a worldwide concern with serious economic, political, and social consequences. Most recently, misinformation has hindered acceptance of COVID-19 vaccines and mitigation measures”

compartilhamento de desinformação e principalmente sobre o comportamento dos usuários. Entre eles:

> 1. O compartilhamento habitual de desinformação não é inevitável.
> 2. Os usuários podem ser incentivados a criar hábitos de compartilhamento que os tornem mais sensíveis ao compartilhamento de conteúdo verdadeiro.
> 3. Reduzir efetivamente a desinformação exigiria a reestruturação dos ambientes online que promovem e apoiam seu compartilhamento. (CEYLANA; ANDERSONB; WOOD, 2023)[20]

4.2 INTELIGÊNCIA ARTIFICIAL E APRENDIZAGEM

A inteligência artificial pode ser definida como um “campo de conhecimento multidisciplinar que tenta não apenas compreender, mas construir entidades inteligentes” (RUSSEL; NORVIG, 2013, p.1). Quando mencionamos na introdução sobre um tipo de abordagem aos agentes inteligentes da ação humana, esse era apenas um dos tipos de abordagem. Levantar uma definição imutável sobre inteligência artificial ou agentes inteligentes não é uma tarefa trivial, ainda mais por se tratar de um tópico que não possui consenso com relação a isso.

Ainda que existam diversas definições, Russel e Norvig apresentam algumas possibilidades - por exemplo, estratégias diferentes de definição dão origem a concepções diferentes de agentes inteligentes. Elas são divididas em quatro principais campos: pensando como um humano, agindo como um humano, pensando racionalmente e agindo racionalmente. Cada concepção dessa avalia a inteligência de uma maneira diferente.

---

[20] Tradução livre nossa: “1. Habitual sharing of misinformation is not inevitable. 2.Users could be incentivized to build sharing habits that make them more sensitive to sharing truthful content. 3. Effectively reducing misinformation would require restructuring the online environments that promote and support its sharing.”

Após a apresentação inicial, os autores trabalham com a noção de quem um agente inteligente é um agente racional que toma a melhor decisão dado um cenário (RUSSEL; NORVIG, 2013). A racionalidade nesse sentido depende de 4 fatores: 1) a medida de desempenho define o critério, 2) o conhecimento prévio que o agente tem do ambiente, 3) as ações que o agente pode executar e 4) a sequência de percepções do agente até o momento. (RUSSEL; NORVIG, 2013).

A aplicação da racionalidade dentro de um ambiente iterativo proporciona o agente a adquirir experiência sobre as informações que coleta, armazena e transforma em desempenho. Desse modo:

> Para cada sequência de percepções possível, um agente racional dever selecionar uma ação que se espera venha a maximizar sua medida de desempenho, dada a evidência fornecida pela sequência de percepções e por qualquer conhecimento interno do agente (RUSSEL; NORVIG, 2013, p.64).

O ponto chave para a definição de agentes inteligentes que compões o core da inteligência artificial é que eles não apenas coletam informação do ambiente, mas aprendem com e sobre ela.

4.3 PROCESSAMENTO DE LINGUAGEM NATURAL

O processamento de linguagem natural “pode ser definido como um campo da ciência da computação que se preocupa em permitir que algoritmos de computador entendam, analisem e gerem linguagens naturais” (ROHAN C., ANIRUDDHA M. GODBOLE et al, 2020, p)

Para conseguir receber textos como entrada, interpretá-los e adquirir certo significado a partir deles, o processamento da língua escrita ou falada, deve fornecer algumas tarefas que sejam instrumentos que a máquina possa fazer.

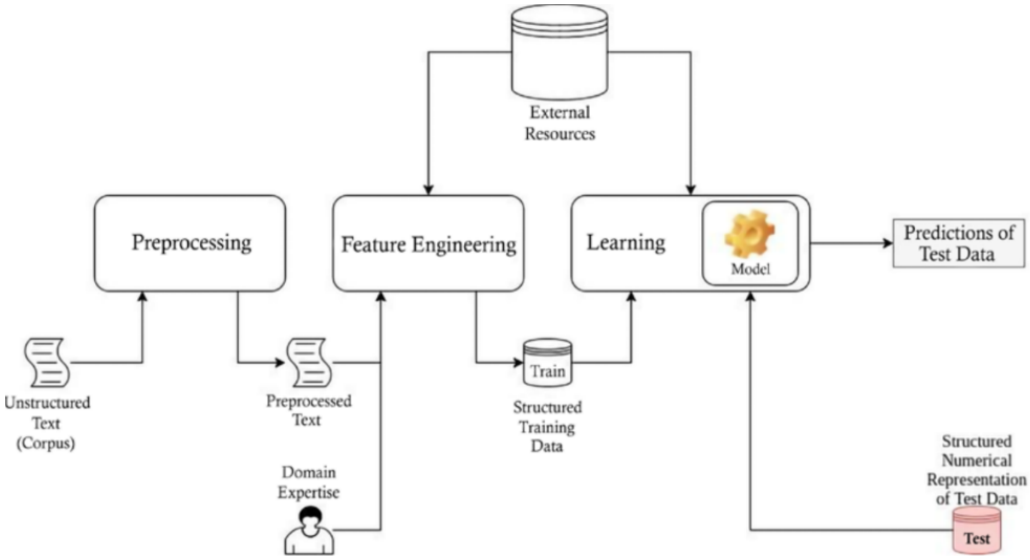
Dentre algumas das tarefas dentro do processamento de linguagem natural, ou NLP *(Natural Language Processing*), se destacam a análise de texto, *tokenização*, remoção de palavras sem significado, normalização e padronização. Cada uma dessas tarefas será abordada com seu devido cuidado nas próximas sessões.

Por ora, é importante, possuir uma compreensão holística sobre o funcionamento de um pipeline dentro do processamento de texto. De acordo com o CRISP-DM, um processo de análise ou ciência de dados tem os seis passos que foram mencionados acima. Na abordagem de análise de texto, são inclusas algumas particularidades. A primeira é a fase de pré-processamento, em que se pega dados não estruturados e os prepara para serem entendidos pela máquina. As tarefas mencionadas acima como padronização e normalização entram nessa fase.

A segunda fase é a de engenharia de características, ou comumente denominado, ‘feature engineering’ que envolve tarefas como tokenização, estematização, lematização e nuvem de palavras, bem como visualizações em gráficos a partir de frequências das palavras dentro de um córpus.

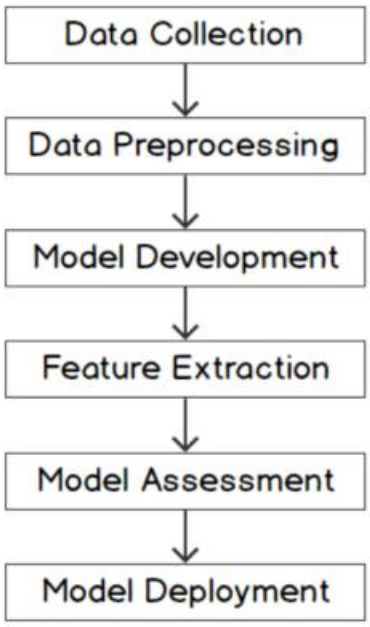
A partir da extração das principais características, os dados devem possuir uma representação numérica – geralmente em formato de matrizes esparsas ou densas – e assim estão prontos para serem treinados e classificados.

Figura 4. Abordagem geral da PNL clássica



Aplicando a estrutura clássica à logica de tarefas estabelecida pelo CRISP-DM, temos as seguintes etapas:

Figura 5. Fases de um projeto de PLN



4.4 DESCRIÇÃO DE DESENVOLVIMENTO DOS PROCESSOS

4.4.1 AQUISIÇÃO DE DADOS

A aquisição de dados pode ser realizada de variadas formas, envolvendo diversas ferramentas e técnicas, valendo tanto para dados estruturados como para dados não estruturados. Alguns métodos são: pesquisa de conteúdo, raspagem de dados em web, ou mesmo fornecimento de dados de empresas.

No trabalho utilizamos a aquisição de dados não estruturados do tipo textual. Foram coletados dados de 3 principais fontes:

1. Scrapping no Telegram;
2. Scrapping no Twitter;
3. Projetos já publicados na área.

Tabela 3. Quantidade de dados coletados

Fonte	Quantidade de dados coletados
Telegram	85.204
Twitter	14.781
Projetos	15.612
Total:	115.597

Os projetos com dados publicados e que foram utilizados como base foram os seguintes: Projeto Factck.Br, projeto Fake.Br NILC, projeto Veritas e projeto Fake Online. Os dados dos projetos foram incorporados ao trabalho respeitando as devidas fontes originais e processamentos já realizados.

Figura 6. Descrição dos dados coletados de projetos

Dataset	valor mensagem	Quantidade colunas	Quantidade linhas
telegram	falso	7	17987
veritas	verdadeiro	5	3263
veritas	falso	5	2224
fake online corpus	verdadeiro	7	1630
Factck br corpus	verdadeiro	8	129
Factck br corpus	falso	8	1166
Fake br corpus nilc	verdadeiro	3	3600

Fonte: Autoral

4.4.2 DICIONÁRIO DE DADOS

O período de coleta de dados foi realizado entre fevereiro e junho de 2023. Inicialmente os dados foram extraídos por meio de web scrapping ou raspagem de dados em grupos relacionados com o tema ‘antivacinas’ no Telegram. Os grupos utilizados para coleta de mensagens foram os seguintes com as respectivas caraterísticas:

Tabela 4. Dados coletados do Telegram

Nome do Grupo	Link de acesso	Quantidade de dados brutos coletados (linhas x colunas)	Quantidade de dados nulos
Vacinas o maior crime da história	https://t.me/vacinasomaiorcrimedahistoria/	31328 × 16	8.053
Antivaxx	https://t.me/antivaxxx	4962 x 8	1.361
Antivacinas	https://t.me/antivacinas	31328 × 16	4.017
Anti vacinas	https://t.me/anti_vacinas	17586 × 8	4827

Dentre os conjuntos de dados coletados, alguns possuíam 8 colunas e outros 16. Os dados que tinham 16 colunas, foram tratados e ficaram com as mesmas 8 colunas para fins de padronização. Vale mencionar que muitos dados contidos nessas colunas eram sensíveis, como nome de usuário e telefone. Esses dados foram

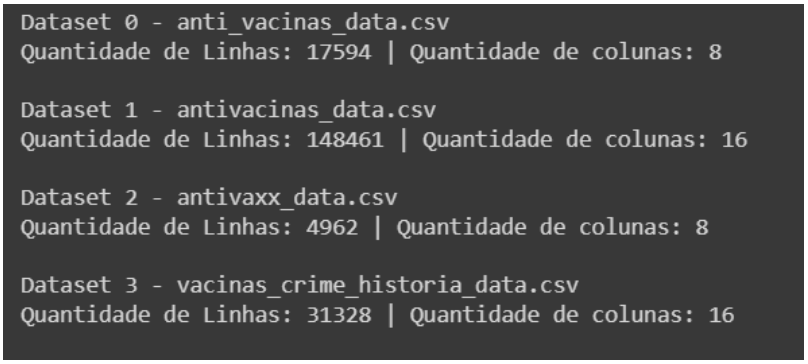
retirados do dataset para preservar os dados sensíveis e não foram utilizados. As colunas coletadas foram:

Tabela 5. Colunas dos dados coletados do Telegram

Coluna	Descrição
message.sender_id	Id da pessoa que enviou mensagem no grupo
Message.text	Texto da mensagem enviada
Message.date	Data de envio da mensagem
Message.id	ID da mensagem
Message.post_author	Se a pessoa que postou a mensagem é autora da mensagem enviada
Message.views	Quantidade de visualizações da mensagem
Message.peer_id_channel_id	ID do canal

Apenas 3 colunas foram para a versão do conjunto de dados final: texto da mensagem, data de envio e visualizações de mensagens. Ao total foram coletados 85.204 dados dos grupos antivacinas mencionados. O intervalo de tempo das mensagens consta como desde 2020 a 2023. Como todos os grupos eram públicos e declaradamente contra vacinação, foram automaticamente utilizados para o conjunto de mensagens caracterizadas com o valor falso, ou não verdadeiro (0).

Figura 7. Visualização de dados coletados do Telegram



Fonte: Autoral

Os dados coletados do Twitter, foram coletados com a filtragem por termos em uma query: “vacina OR vacinação OR imunizante OR imunização OR saúde OR vacinar”. Foram coletados dados de 16 perfis oficiais, declaradamente de divulgação científica ou perfis públicos de ciência e saúde, como o Ministério da Saúde e Fundação Fiocruz, por exemplo. O intervalo de tempo configurado foi o mesmo para o grupo do Telegram, do ano de 2020 ao ano de 2023. Em consonância com esses critérios, as mensagens foram automaticamente classificadas com o valor verdadeiro (1). Ao total, foram coletados 14.781 dados.

Tabela 6. Colunas dos dados coletados do Twitter

Coluna	Descrição
Data	Data de envio da mensagem
Username	@ do usuário (necessário para a query no caso)
Url	Link do tuíte
Texto	Texto do tuíte

Para complementar com os dados sobre revisão bibliográfica, foram utilizados: 7200 dados do Fake.br-Corpus com dados pré-processados de notícias falsas, sendo 3600 com notícias classificadas como falsas e 3600 com notícias classificadas como verdadeiras. Do projeto Veritas foram utilizados 5.487 (sendo 3263 dados verdadeiros e 2224 dados falsos). Do projeto Fake Online foram utilizados 1630 dados verdadeiros. E do projeto Factck foram utilizados 129 dados verdadeiros e 1166 falsos[21].

Vale mencionar que a maior parte das fontes desses projetos incluem dados de agências verificadoras de fato como Aos Fatos[22], Agência lupa Uol[23], Fato ou fake G1[24] entre outros. Ao total, somaram-se 44.780 dados brutos.

[21] Todos os projetos e repositórios podem ser acessados nas referências do trabalho.
[22] Aos Fatos disponível em: https://www.aosfatos.org
[23] Agência Lupa Uol disponível em: https://lupa.uol.com.br/
[24] Fato ou fake G1 disponível em: https://lupa.uol.com.br/

Após as filtragens por assunto – com exceção do Twitter, que já havia sido filtrado no ato da coleta – o total de dados pré-processados e limpos para o conjunto de dados foi de 42.665 – divididos em 20.990 para a classe falsa e 21675 para a classe verdadeira. Ainda sobre o formato final, o conjunto de dados possui 3 colunas: texto, fonte e label (0 e 1).

Figura 8. Conjunto de dados final

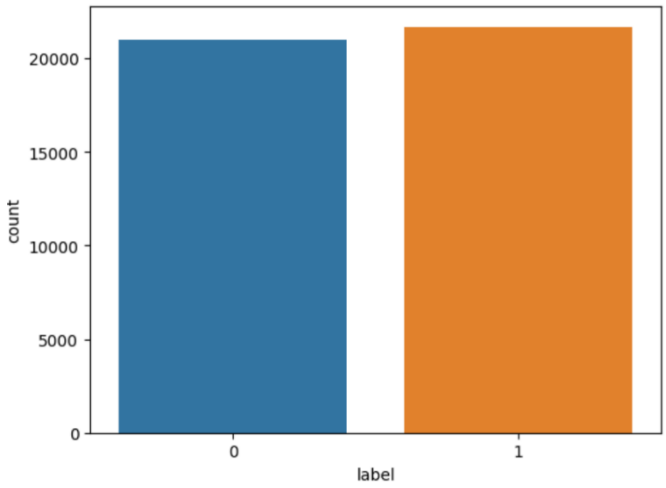
```
Datasets com valores falsos
--------------------------------------------------
Telegram | linhas: 15901 colunas: 3
corpus Veritas + corpus NILC | linhas: 5089 colunas: 3
quantidade de dados falsos: 20990
--------------------------------------------------


Datasets com valores verdadeiros
--------------------------------------------------
corpus veritas, corpus NILC, corpus Fake online e corpus Fake br
| linhas: 6894 colunas: 3
Twitter | linhas: 14781 colunas: 3
quantidade de dados verdadeiros: 21675
--------------------------------------------------
```

Fonte: Autoral

A proporção entre as classes ficou bem balanceada, de 49% e 50%.

Figura 9. Proporção de classes - verdadeira e falsa - do conjunto de dados



Fonte: Autoral

4.4.3 PRÉ-PROCESSAMENTO DE DADOS

Como muitas mensagens possuem conteúdo de diversos assuntos dentro dos dados e o objetivo deste trabalho concerne apenas ao tema de vacinação, foi necessária realizar uma filtragem sobre o tópico. Para isso foi utilizado o modelo Zero-shot para inferência sobre a classe. A classificação de texto zero-shot é uma tarefa no processamento de linguagem natural em que um modelo é treinado em um conjunto de exemplos rotulados, mas é capaz de classificar novos exemplos de classes inéditas. (HUGGING FACE, 2023)

As etapas de normalização e padronização dos dados se referem respectivamente a “coloca os dados no intervalo entre 0 e 1 ou -1 e 1 caso haja valores negativos, sem distorcer as diferenças nas faixas de valores”. (PARTNER, 2021). Já a padronização “colocamos a média dos dados em 0 e o desvio padrão em 1”. (PARTNER, 2021). A partir disso, também foram aplicadas algumas outras transformações nos dados: eles foram colocados em letras minúsculas, sem acentos e sinais de pontuação, foram retirados emojis e links. Números também foram retirados e algumas substituições de texto por valores como ‘falso’ por 0 foram realizados. As técnicas foram aplicadas por meio do standard scaler e outras manipulações dentro das bibliotecas pandas[25], numpy[26] e scikit-learn[27].

4.4.4 EXTRAÇÃO DE FEATURES

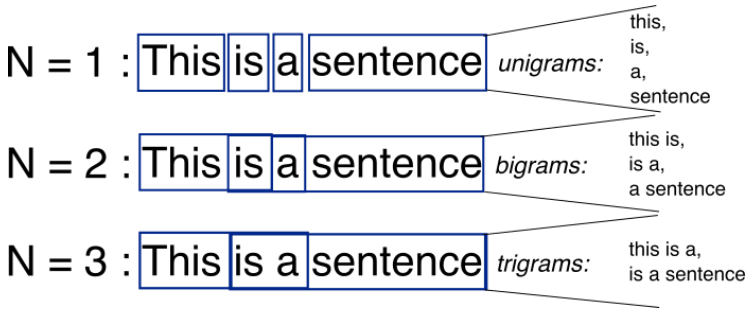
Algumas técnicas de extração de features incluem a tokenização com divisão de n-gramas, *stemming*, *lemmatizing,* remoção de stop words, BOW (Bag of words), TFIDF e Word cloud. Dentre essas técnicas foram utilizadas a tokenização e BOW por meio da implementação do Count vectorizer, remoção de stop words e nuvem de palavras.

---

[25] Pandas: disponível em: https://pandas.pydata.org
[26] Numpy disponível em: https://numpy.org
[27] Scikit-learn disponível em: https://scikit-learn.org/stable/

a tokenização (transformar as palavras em tokens) “é o processo de dividir uma frase em palavras ou tokens individuais. Durante esse processo, pontuações e caracteres especiais são completamente removidos. “(ALURA, 2021) e as n-gramas são “um tipo de modelo probabilístico usado para prever o próximo item de uma sequência na forma de um modelo de Markov. Em um contexto linguístico, o n-grams se refere a uma sequência n de palavras”. (ALURA, 2021). Essas sequências variam e podem ser definidas unitariamente, binariamente ou em mais de 3 termos. Para exemplificar melhor:

Figura 10. N-grams



Fonte: N-Grams – Deep AI

A técnica de remoção stop words é o ato de remover palavras que não acrescentam nenhum significado às frases, como artigos e pronomes (exemplos: a,o,é,que,etc). Desse modo, é facilitado o processamento – pois texto é eliminado.

O BOW (Bag of Words) “nos permite representar o texto com a ocorrência de cada palavra, sem levar em conta a ordem das palavras ou sua estrutura no texto. É realmente como se todas as palavras fossem colocadas dentro de um saco” (ALURA, 2021). Se tratando de frequência de palavras, também temos a nuvem de palavras que é uma simplificação visual das palavras mais frequentes no texto, dispostas em ordem de importância.

A partir do pré-processamento dos dados e montagem do conjunto com as três colunas limpas – texto, label (categoria) e fonte – com os dados já filtrados por

assunto - os dados foram divididos em treino e teste para que pudessem ser treinados e testados.

Vale mencionar que os modelos foram rodados na ferramenta Google Colaboratory Pro[28] para utilização de alta memória – com a RAM do sistema com capacidade de 51.0 GB e 225.8 GB de Disco disponíveis. Em alguns casos se utilizou a GPU do tipo T4 para acelerar o tempo de treinamento dos modelos. Ainda assim o conjunto de dados inicial – 42.665 teve que ser diminuído, para que todos os modelos pudessem ser treinados e testados sem interrupções por falta de memória. Alguns modelos como o SVM chegaram a ter mais de 10 horas de tempo de execução apenas para um ciclo treino/teste.

Sendo assim. o primeiro conjunto de dados testado foi com 21.332 (sample aleatória do conjunto de dados principal, mantendo a proporção das classes) dados filtrados por assunto com o modelo do tipo Zero-Shot para inferência apenas - previamente o modelo já havia sido treinado em língua portuguesa e disponibilizado[29], para o presente trabalho, apenas foi realizado o carregamento e inferência. A proporção de teste e treino foi de 0,33 com random state a 42.

É importante mencionar que o segundo conjunto de dados foi criado a partir de uma filtragem de assunto por aplicação do algoritmo LDA (Latent Dirichlet Allocation), mas não foi utilizado no trabalho para análises devido ao curto tempo de desenvolvimento e alto tempo de treinamento para modelos. Ao longo do desenvolvimento surgiu a intenção de realizar uma análise comparativa de desempenho entre as técnicas para filtragem dos dados, mas como não havia tempo disponível, isso fica como opção para próximos passos. Ainda assim, seguem os dados sobre os conjuntos de dados filtrados:

[28] Disponível em: https://colab.research.google.com/signup
[29] Disponível em https://huggingface.co/Mel-Iza0/zero-shot

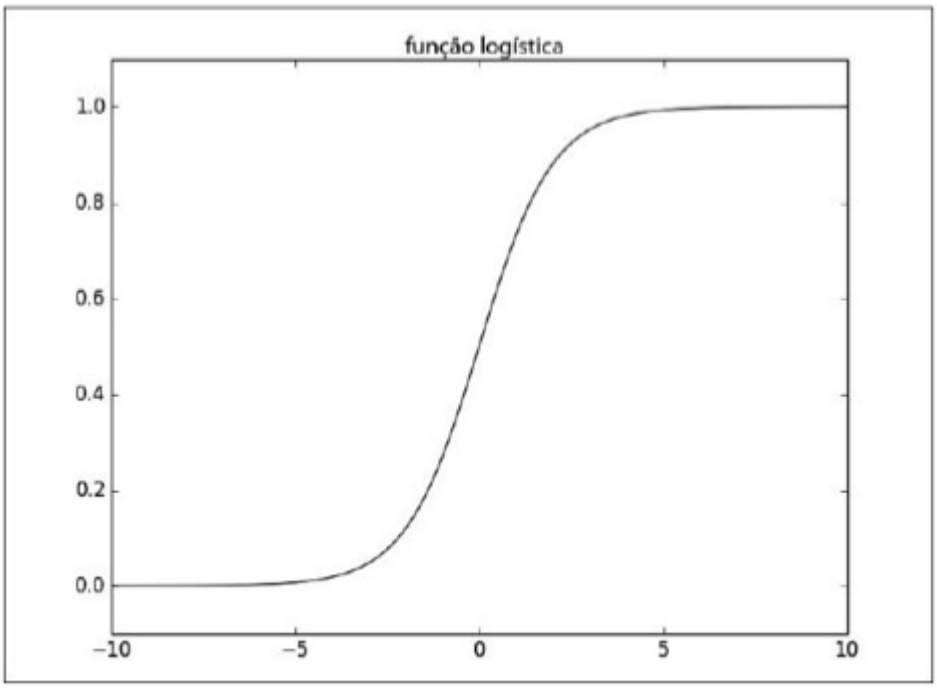
Tabela 7. Descrição da construção do conjunto de dados e técnicas aplicadas

Conjunto de dados	Técnica Utilizada	Total
1. Zero-Shot filter	Modelo zero-shot para inferência de classes pré estabelecidas – [vacina, vacinação, imunização, saúde hobbie, compras]	21.332
2. LDA filter	Modelo LDA (Latent Dirichlet Allocation) para encontro de tópicos no texto. Tópicos foram selecionados após a modelagem + técnica de oversampling para aumentar balanceamento de classes selecionadas	26.122

4.4.5 DESENVOLVIMENTO MODELO

A parte da modelagem dentro um projeto é a aplicação de uma função do algoritmo selecionado ao conjunto de dados. De acordo com Grus, um modelo é “uma especificação de uma relação matemática (ou probabilística) existente entre variáveis diferentes” (GRUS, 2016, p.204). No caso especificado, o que está sendo investigado é a relação entre características textuais de uma mensagem sobre vacina e quais dessas características são importantes para determinar o valor de verdade de seu conteúdo. A principal função para se aplicar a esse caso é a função logística:

Figura 11: Função Logística ou Regressão logística



Fonte: GRUS, 2016

De acordo com uma definição apresentada pela IMB, uma função logística é:

> Um tipo de modelo estatístico (também conhecido como modelo logit) é frequentemente usado para classificação e análise preditiva. A regressão logística estima a probabilidade de ocorrência de um evento, como um voto, com base em um determinado conjunto de dados de variáveis independentes. Como o resultado é uma probabilidade, a variável dependente é limitada entre 0 e 1. Na regressão logística, uma transformação logit é aplicada com base nas probabilidades, ou seja, a probabilidade de sucesso dividida pela probabilidade de falha. Isso também é comumente conhecido como "log odds", ou logaritmo natural de probabilidades (IBM, 2023)

Foram escolhidos 6 modelos de classificação: regressão logística, XGBoost, Naive Bayes, Decision Tree, AdaBoost e SVM. Todos os modelos treinados tiveram o resultado obtido a partir da média de aplicação de um cross-validation score[30]

[30] Cross validation: é uma técnica muito utilizada para avaliação de desempenho de modelos de aprendizado de máquina. O CV consiste em particionar os dados em conjuntos(partes), onde um

De acordo com a definição das métricas de avaliação com um artigo da Microsoft sobre serviços de linguagem:

> **Precisão**: mede a precisão/exatidão do modelo. É a taxa entre os positivos identificados corretamente (verdadeiros positivos) e todos os positivos identificados. A métrica de precisão revela quantas das classes previstas estão rotuladas corretamente.
>
> **Recall**: mede a capacidade do modelo de prever classes positivas reais. É a taxa entre os verdadeiros positivos previstos e o que foi realmente marcado. A métrica de recall revela quantas das classes previstas estão corretas.
>
> **Medida F1**: a medida f é uma função de Precisão e Recall. Ela é necessária quando você busca um equilíbrio entre Precisão e Recall. (AAHILL, 2023)

Por última, a métrica acurácia pode ser definida como “o número de acertos (positivos) divido pelo número total de exemplos. Ela deve ser usada em dados com a mesma proporção de exemplos para cada classe, e quando as penalidades de acerto e erro para cada classe forem as mesmas.” (FILHO, 2018)

Com uma média de 3 folds[31]. As métricas utilizadas também foram escolhidas a partir de seu balanceamento disponível: acurácia balanceada, F1-score ponderado, Recall ponderado e precisão ponderada.

---

conjunto é utilizado para treino e outro conjunto é utilizado para teste e avaliação do desempenho do modelo. (RABELLO, 2019)

[31] “K-fold consiste em dividir a base de dados de forma aleatória em K subconjuntos (em que K é definido previamente) com aproximadamente a mesma quantidade de amostras em cada um deles” (RABELLO, 2019)

4.4.6 DEPLOY DO MODELO

De modo geral, o ato de fazer o deploy de um modelo é proporcionar sua implantação em algum ambiente para que as pessoas consigam acessar a aplicação. Em conformidade com esse raciocínio, a Tera explica: “O deploy é a etapa de preparação de um modelo para ser usado no dia a dia, ou seja, adaptação a uma aplicação maior para que usuários façam uso do algoritmo.” (2021)

> Deploy é o processo de finalização de um projeto, em que se gera um código para exportar a aplicação para ser usada por outras pessoas no dia a dia. No exemplo de um modelo de ML em um projeto de Data Science, é importante decidir como as pessoas usarão aquele modelo para realizar predições ou para identificar padrões e circunstâncias. (TERA, 2021)

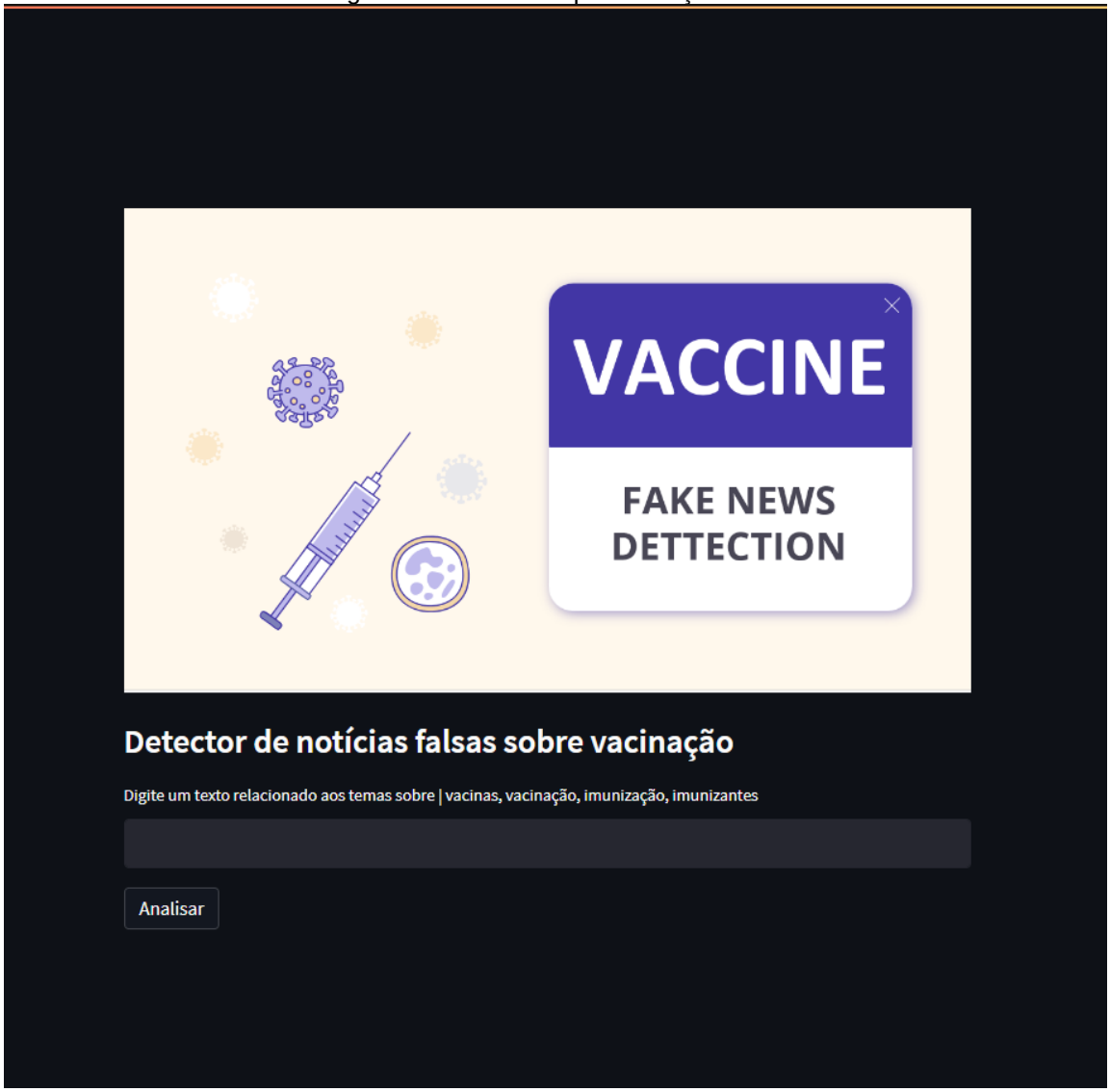
O deploy do modelo foi realizado utilizando a plataforma Streamlit. O Streamlit é uma biblioteca Python de código aberto que facilita a criação e o compartilhamento de aplicativos da Web personalizados para aprendizado de máquina e ciência de dados.[32] (STREAMLIT, 2023)

[32] Streamlit Docs. Disponível em: <https://docs.streamlit.io>.

4.4.6.1 DISPONIBILIZAÇÃO DO PROJETO

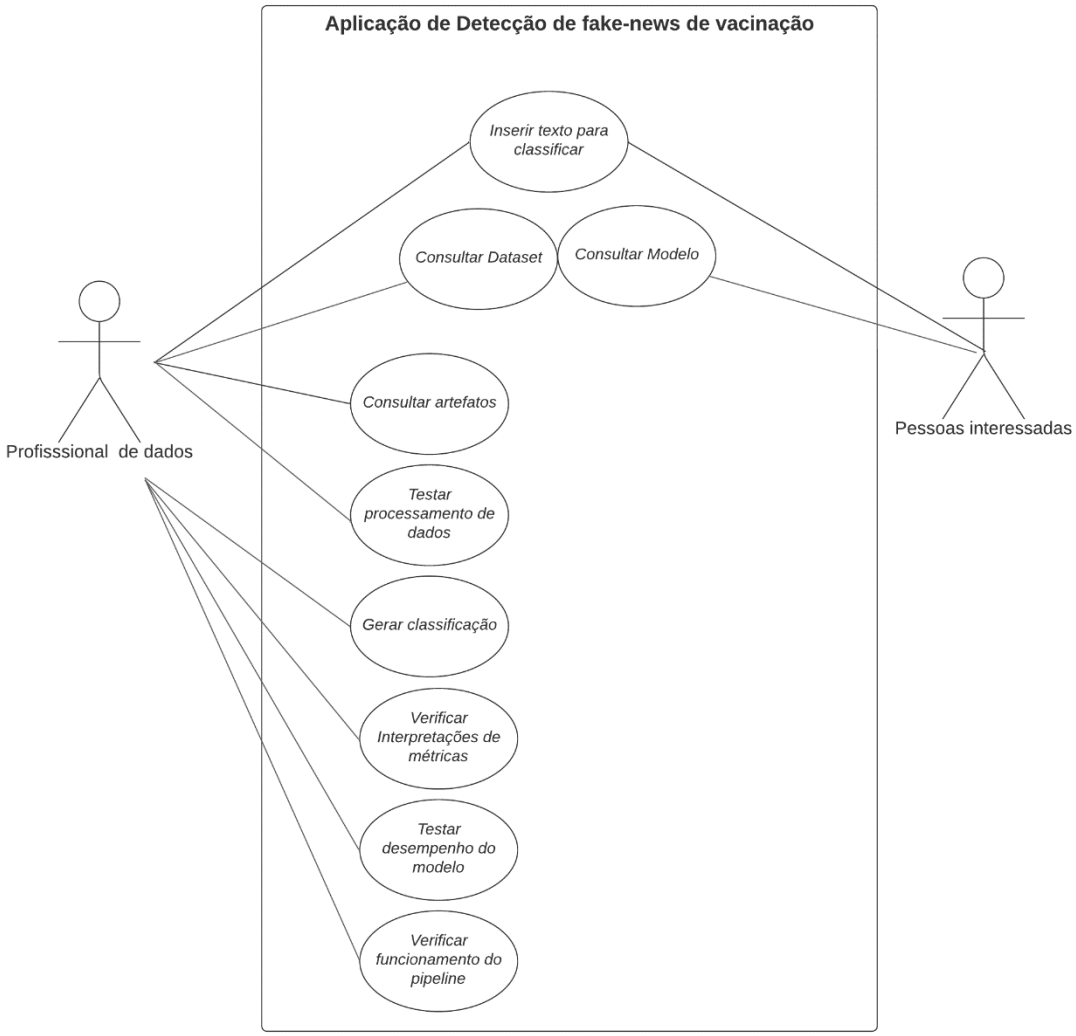
O modelo foi disponibilizado na plataforma Streamlit. A aplicação possui uma única tela recebendo um texto de entrada e retornando à probabilidade da classificação. A aplicação está disponível no link https://vaccine.streamlit.app

Figura 12. Tela da disponibilização do modelo



4.5 DIAGRAMA DE CASO DE USO

Figura 13. Diagrama de caso de uso



4.6 DOCUMENTO DE CASOS DE USO

Cada caso de uso apresentado possui uma descrição detalhada, qual ator realiza a ação, pré-condições para a ação e como ela é realizada no cenário principal e alternativo. Também se existe alguma pós condição após a realização da ação principal.

Tabela 8. Caso de Uso (Inserir texto para classificar)

01. INSERIR TEXTO PARA CLASSIFICAR	
Descrição:	Usuário acessará aplicação e vai inserir texto para classificar
Ator:	Profissional de dados e Pessoas interessadas
Pré-Condição: Usuário precisa saber utilizar um navegador e ter acesso à internet.	
Cenário Principal: 1. Abrir o navegador 2. Entrar no link da aplicação: https://vaccine.streamlit.app 3. Inserir texto na caixa de input 4. Apertar botão “Analisar”	
Cenário Alternativo: Não se aplica.	
Pós-Condição: Não se aplica.	

Tabela 9. Caso de Uso (Consultar Dataset)

02.CONSULTAR DATASET	
Descrição:	Usuário acessará repositório do GitHub e consultará o conjunto de dados
Ator:	Profissional de dados e pessoas interessadas
Pré-Condição: Usuário precisa saber utilizar um navegador, ter acesso à internet, possuir uma conta no GitHub e conhecimentos mínimos sobre machine learning	
Cenário Principal: 1.Abrir repositório principal do projeto 2. Entrar na pasta ‘data’ para acessar arquivo de dados 3. Abrir arquivo dataset.csv	
Cenário Alternativo: Não se aplica.	
Pós-Condição: Não se aplica.	

Tabela 10. Caso de Uso (Consultar Modelo)

03.CONSULTAR MODELO	
Descrição:	Usuário acessará repositório do GitHub e consultará o notebook de modelagem
Ator:	Profissional de dados e pessoas interessadas
Pré-Condição: Usuário precisa saber utilizar um navegador, ter acesso à internet, possuir uma conta no GitHub e conhecimentos mínimos sobre machine learning	
Cenário Principal: 1.Abrir repositório principal do projeto 2. Entrar na pasta ‘Notebooks’ 2. Entrar na pasta ‘Modelos’ 3. Abrir arquivo Modelagem. ipynb	
Cenário Alternativo: Não se aplica.	
Pós-Condição: Não se aplica.	

Tabela 11. Caso de Uso (Gerar Classificação)

04.GERAR CLASSIFICAÇÃO	
Descrição:	Usuário acessará repositório do GitHub, consultará o notebook de modelagem e gerará classificação a partir do pipeline
Ator:	Profissional de dados
Pré-Condição: Usuário precisa saber utilizar um navegador, ter acesso à internet, possuir conta no GitHub, possuir conhecimentos básicos/intermediários de machine learning e conhecimentos básicos do Google Colab	
Cenário Principal: 1.Abrir repositório principal do projeto 2. Entrar na pasta ‘Notebooks’ 2. Entrar na pasta ‘Modelos’ 3. Abrir arquivo Modelagem.pynb 4. Conectar ao ambiente de execução 5. Executar imports de bibliotecas 6. Executar pipeline 7. Verificar resultado da classificação	
Cenário Alternativo: Não se aplica.	

Pós-Condição: Não se aplica.

Tabela 12. Caso de Uso (Testar Processamento de Dados)

05.TESTAR PROCESSAMENTO DE DADOS	
Descrição:	Usuário acessará repositório do GitHub, consultará o notebook de modelagem e testará o processamento dos dados por meio do pipeline
Ator:	Profissional de dados
Pré-Condição: Usuário precisa saber utilizar um navegador, ter acesso à internet, possuir conta no GitHub, possuir conhecimentos básicos/intermediários de machine learning e conhecimentos básicos do Google Colab.	
Cenário Principal: 1.Abrir repositório principal do projeto 2. Entrar na pasta ‘Notebooks’ 2. Entrar na pasta ‘Modelos’ 3. Abrir arquivo Modelagem.pynb 4. Conectar ao ambiente de execução 5. Executar imports de bibliotecas 6. Executar pipeline	
Cenário Alternativo: Não se aplica.	
Pós-Condição: Não se aplica.	

Tabela 13. Caso de Uso (Consultar/Acessar Artefatos)

06.CONSULTAR ARTEFATOS	
Descrição:	Usuário acessará repositório do GitHub e consultará a pasta de modelos disponibilizados
Ator:	Profissional de dados
Pré-Condição: Usuário precisa saber utilizar um navegador, ter acesso à internet e possuir conta no GitHub	

Cenário Principal: 1.Abrir repositório de deploy do projeto 2. Entrar na pasta ‘Artifacts’ 2. Conferir artefatos do projeto
Cenário Alternativo: Não se aplica.
Pós-Condição: Não se aplica.

Tabela 14. Caso de Uso (Testar Desempenho do Modelo)

07.TESTAR DESEMPENHO DO MODELO	
Descrição:	Usuário acessará repositório do GitHub, consultará o notebook de modelagem e testará o desempenho do modelo por meio do pipeline
Ator:	Profissional de dados
Pré-Condição: Usuário precisa saber utilizar um navegador, ter acesso à internet, possuir conta no GitHub, possuir conhecimentos básicos/intermediários de machine learning e conhecimentos intermediários do Google Colab.	
Cenário Principal: 1.Abrir repositório principal do projeto 2. Entrar na pasta ‘Notebooks’ 2. Entrar na pasta ‘Modelos’ 3. Abrir arquivo Modelagem.pynb 4. Conectar ao ambiente de execução 5. Executar imports de bibliotecas 6. Executar pipeline 7. Testar desempenho em dados de teste	
Cenário Alternativo: Não se aplica.	
Pós-Condição: Não se aplica.	

Tabela 15. Caso de Uso (Verificar Interpretação de Métricas)

08.VERIFICAR INTERPRETAÇÃO DE MÉTRICAS	
Descrição:	Usuário acessará repositório do GitHub, consultará o notebook de modelagem, testará o desempenho do modelo por meio do pipeline e verificará o desempenho das métricas do modelo
Ator:	Profissional de dados
Pré-Condição: Usuário precisa saber utilizar um navegador, ter acesso à internet, possuir conta no GitHub, possuir conhecimentos básicos/intermediários de machine learning e conhecimentos intermediários do Google Colab.	
Cenário Principal: 1.Abrir repositório principal do projeto 2. Entrar na pasta ‘Notebooks’ 2. Entrar na pasta ‘Modelos’ 3. Abrir arquivo Modelagem.pynb 4. Conectar ao ambiente de execução 5. Executar imports de bibliotecas 6. Executar pipeline 7. Testar desempenho em dados de teste 8. Interpretar métricas 9. Verificar interpretação de métricas	
Cenário Alternativo: Não se aplica.	
Pós-Condição: Não se aplica.	

Tabela 16. Caso de Uso (Verificar Funcionamento do Pipeline)

09.VERIFICAR FUNCIONAMENTO DO PIPELINE	
Descrição:	Usuário acessará repositório do GitHub, consultará o notebook de modelagem, e testará o funcionamento do pipeline
Ator:	Profissional de dados
Pré-Condição: Usuário precisa saber utilizar um navegador, ter acesso à internet, possuir conta no GitHub, possuir conhecimentos básicos/intermediários de machine learning e conhecimentos intermediários do Google Colab.	

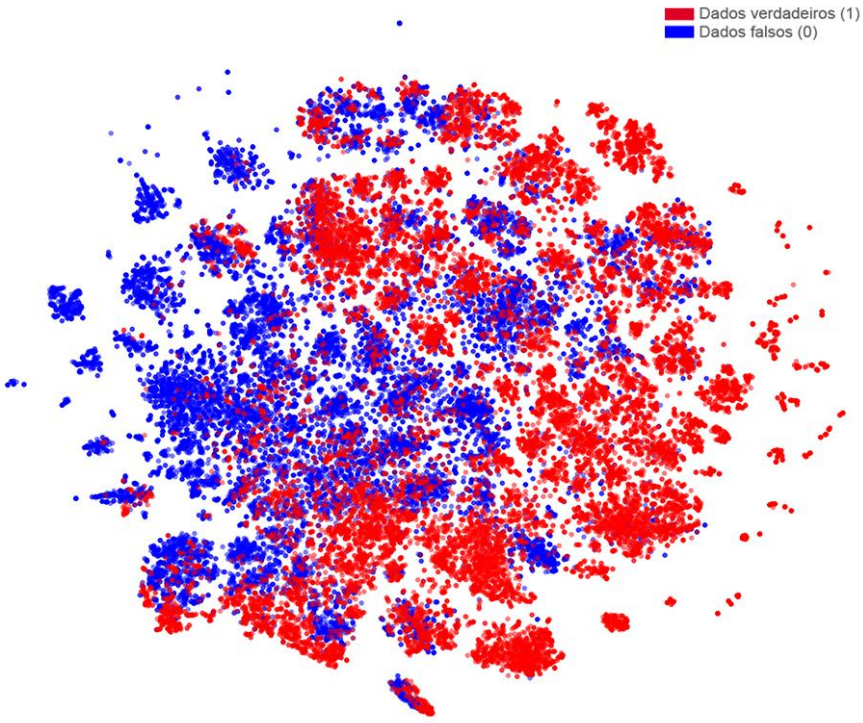
Cenário Principal: 1.Abrir repositório principal do projeto 2. Entrar na pasta ‘Notebooks’ 2. Entrar na pasta ‘Modelos’ 3. Abrir arquivo Modelagem.pynb 4. Conectar ao ambiente de execução 5. Executar imports de bibliotecas 6. Executar pipeline 7. Verificar inputs e seus respectivos tipos 8. Verificar outputs retornados 9. Verificar se os outputs estão coerentes de acordo com as funções do pipeline
Cenário Alternativo: Não se aplica.
Pós-Condição: Não se aplica.

5 ANÁLISES E RESULTADOS

Os resultados obtidos durante a extração de features dos dados foram compilados em formato de imagens, incluindo as nuvens de palavras, frequência de palavras por conjuntos falsos e conjuntos verdadeiros dentro do dataset. Seguem algumas análises e resultados sobre esta etapa:

A primeira técnica utilizada foi o TSNE[33], em que os dados foram plotados em um gráfico. Observa-se que há uma clara tendência entre as características de cada tipo de mensagem – entre as falsas, coloridas em azul e as verdadeiras, coloridas em vermelho. Mas é interessante pontuar o quanto há uma zona mista nas áreas mais centrais do gráfico. Isso acentua a dificuldade, até para a máquina, de encontrar certos padrões entre as mensagens falsas e verdadeiras, pois em algum momento elas são muito parecidas.

Figura 14. Visualização TSNE de dados



Fonte: Autoral

[33] Disponível em https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html

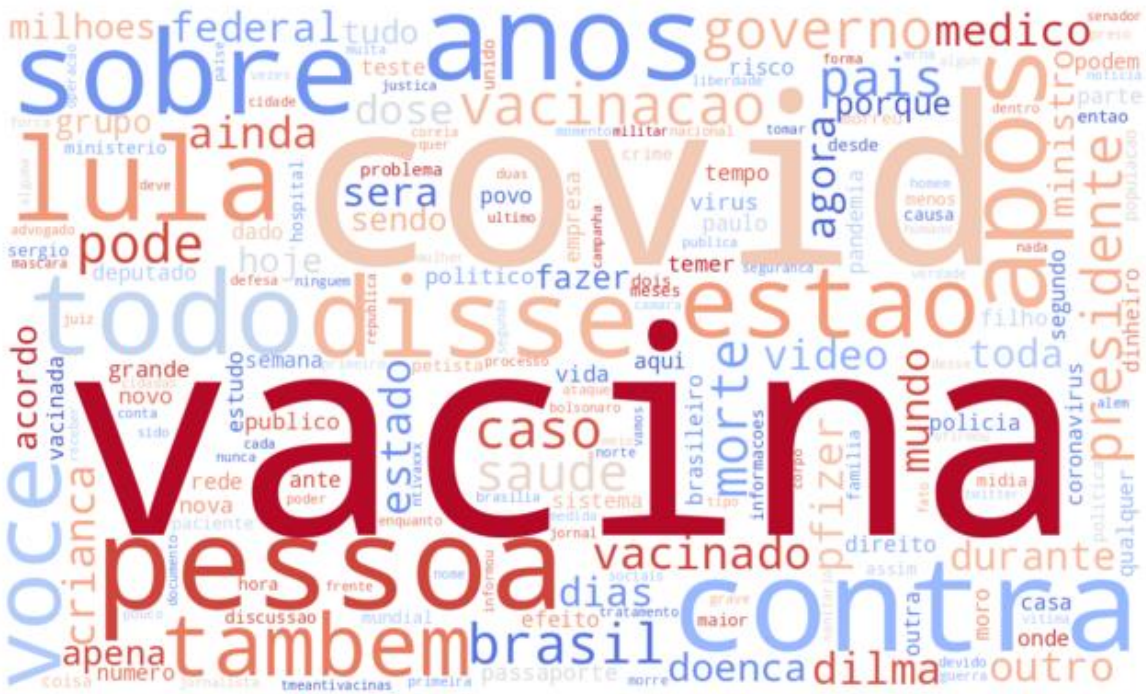
As próximas visualizações são as nuvens de palavras. A primeira nuvem é do conjunto completo de dados – vemos com bastante ênfase as palavras: vacina, saúde, covid, sobre, pessoa, coronavírus, ministério, estado entre outras.

Figura 15. Nuvem de palavras do conjunto de dados completo

Fonte: Autoral

Abaixo realizei duas visualizações – uma para os dados verdadeiros e outra para os dados falsos.

Figura 16. Nuvem de palavras dos dados falsos



Fonte: Autoral

Figura 17. Nuvem de palavras dos dados verdadeiros

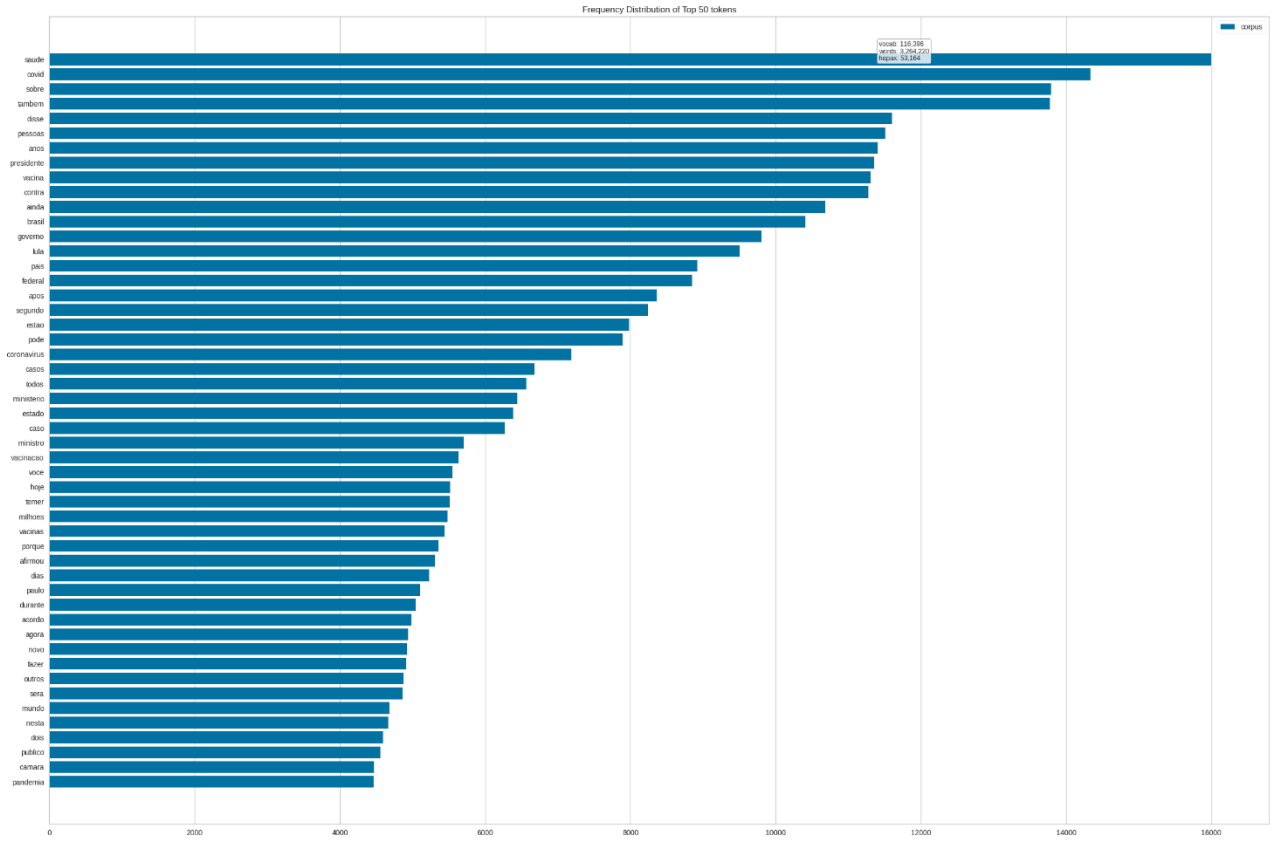
Fonte: Autoral

É interessante notar que as palavras em mais destaques são diferentes: da nuvem de palavras falsas, a maior palavra foi vacina e da nuvem verdadeira a maior palavra foi saúde. Enquanto no conjunto falso, a palavra vacina é a maior, ela no conjunto verdadeiro aparece em um ranking bem menor.

Entre as palavras que se destacam na nuvem falsa estão: covid, contra, pessoa, sobre, anos e até lula. Já na nuvem, de verdadeiras: saúde, também, caso, sobre, estadocovid, presidente e também lula.

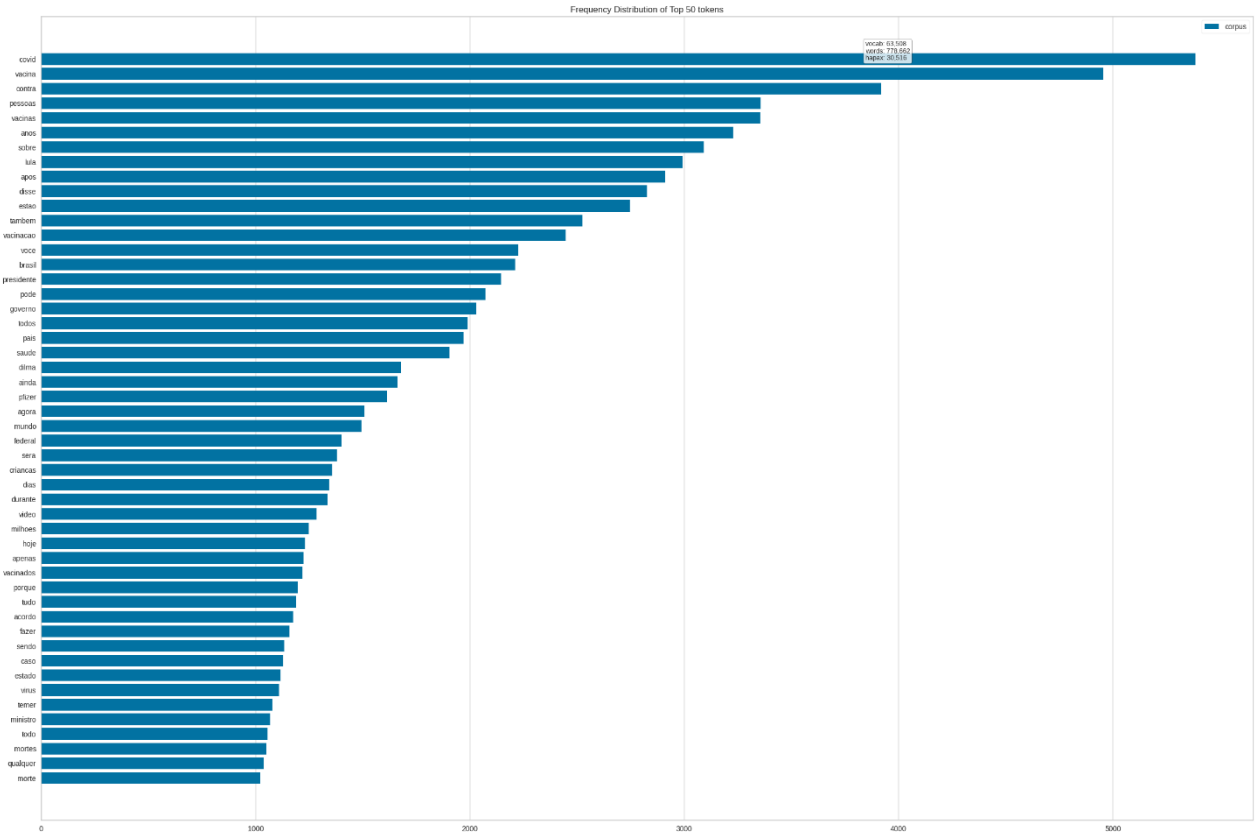
As próximas visualizações são as frequências de palavras, em que as palavras com maiores frequências no texto ocupam as primeiras posições e as com menor frequência ficam abaixo. Os mesmos padrões podem ser notados de uma forma diferente:

Figura 18. Frequência de palavras do conjunto todo



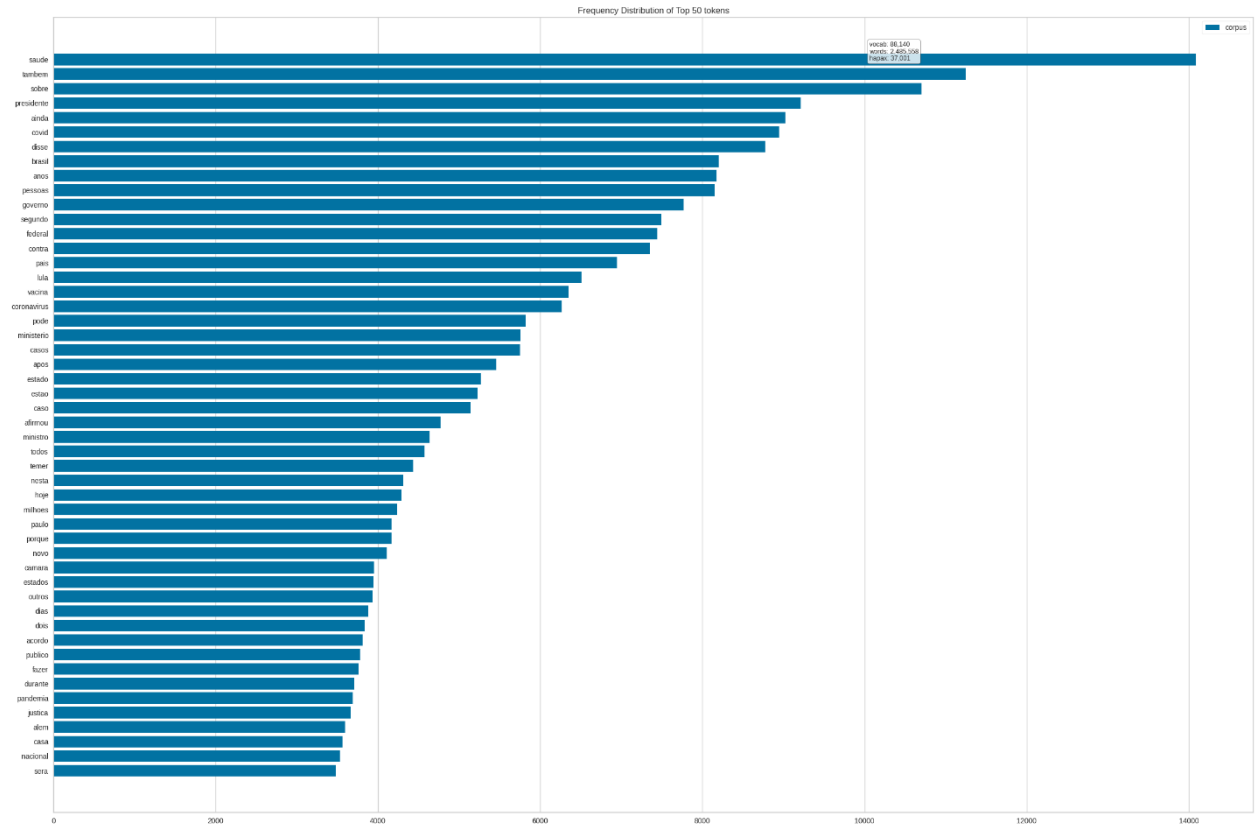
Fonte: Autoral

Figura 19. Frequência de palavras dos dados falsos



Fonte: Autoral

Figura 20. Frequência de palavras dos dados verdadeiros

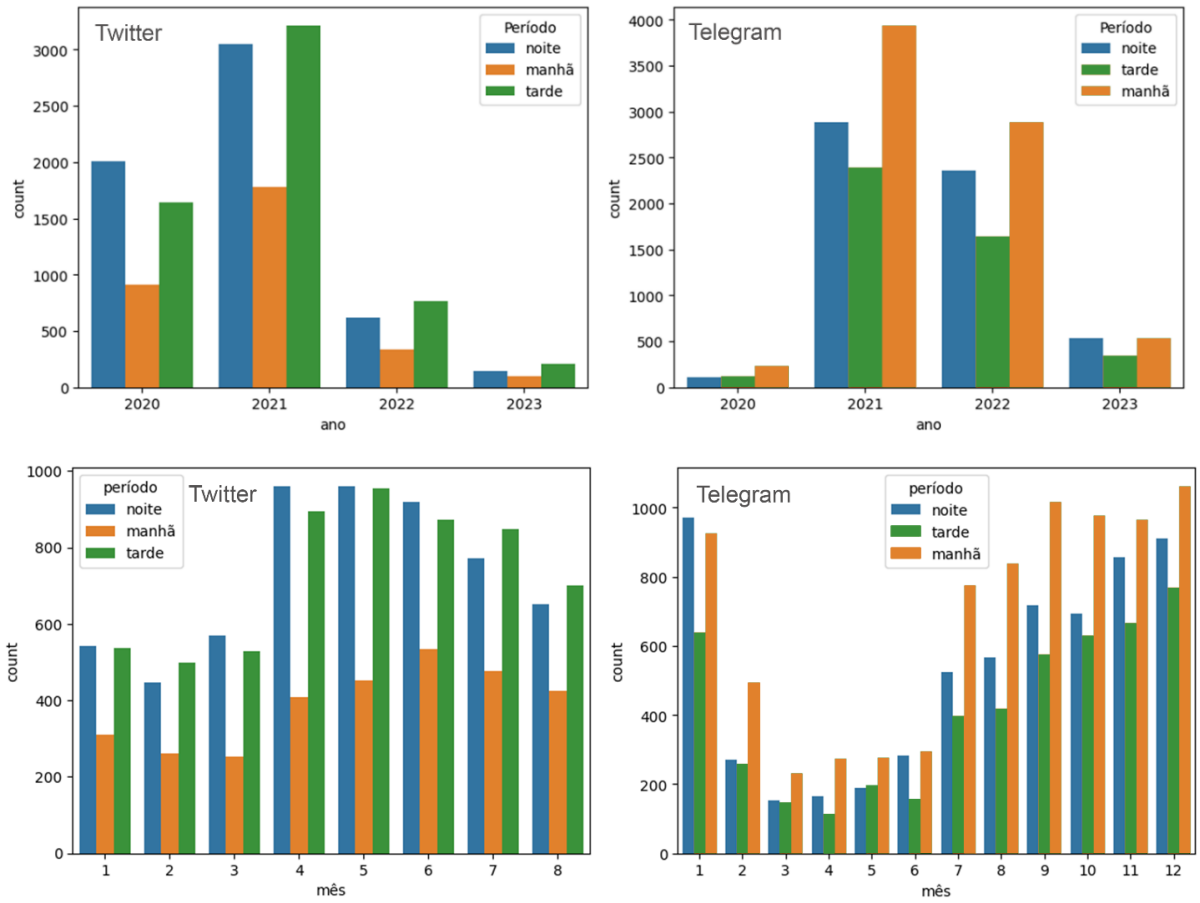


Fonte: Autoral

É muito interessante observar que a palavra vacina em si só aparece no conjunto verdadeiro em 17ª posição, enquanto no conjunto falso ela aparece em 2º lugar. Será que essa pode ser uma característica de menção direta às vacinas como uma técnica de mensagens que são falsas sobre o assunto? Vale investigar com mais calma e levantar mais hipóteses a respeito.

Uma última visualização que foi realizada foi a partir do tratamento das colunas de data dos dados – tanto no Twitter como no Telegram, os dados possuíam uma coluna que identificava a data que a mensagem havia sido enviada. A partir do tratamento desses dados, pudemos verificar um padrão de atividade por período. No grupo do Twitter foi observado que os períodos majoritários de atividade eram no período noturno enquanto no Telegram os períodos majoritários eram diurnos:

Figura 21. Comparação de período de envio de mensagens entre o Twitter e Telegram nos anos de 2020 a 2023



5.1 DESEMPENHO E AVALIAÇÃO DO MODELO

Os testes realizados com os modelos variaram com a aplicação de duas principais técnicas: a primeira foi com relação à implementação de um pipeline e a não implementação de um pipeline. A segunda diz respeito ao tipo de processamento realizado nos dados – foi feito um processamento base em que os textos estavam iguais e com as mesmas técnicas aplicadas e a única diferença aplicada foi um conjunto com as stop words e sem as stop words.

Texto filtrado com modelo Zero Shot (treinado em português) - sem pipeline						
Modelos	Regressão Logística	XGBoost	Naive Bayes	Decision Tree	SVM	AdaBoost
Texto padronizado						
Pré-processamento: Tokenização (Count vectorizer), normalização (Standard scaler), texto padronizado (em minúsculas, sem acentuação, emojis, caracteres especiais e links.)						
Acurácia	0,855	0,944	0,716	0,885	0,722	0,955
F1 Score	0,861	0,944	0,718	0,888	0,77	0,956
Recall	0,863	0,944	0,726	0,882	0,788	0,956
Precision	0,868	0,944	0,741	0,887	0,802	0,956

Texto sem stop words						
Pré-processamento: Tokenização (Count vectorizer), normalização (Standard scaler), texto padronizado (em minúsculas, sem acentuação, emojis, caracteres especiais e links) e texto sem stop words						
Acurácia	0,837	0,939	0,717	0,852	0,725	0,958
F1 Score	0,846	0,94	0,717	0,856	0,781	0,961
Recall	0,849	0,94	0,724	0,86	0,798	0,961
Precision	0,857	0,94	0,74	0,857	0,808	0,962

Texto filtrado com modelo Zero Shot (treinado em português) - com pipeline						
Modelos	Regressão Logística	XGBoost	Naive Bayes	Decision Tree	SVM	AdaBoost
Texto padronizado						
Pré-processamento: Tokenização (Count vectorizer), normalização (Standard scaler), texto padronizado (em minúsculas, sem acentuação, emojis, caracteres especiais e links.)						
Acurácia	0,855	0,944	0,709	0,882	0,734	0,955
F1 Score	0,861	0,944	0,71	0,884	0,783	0,956
Recall	0,863	0,944	0,719	0,885	0,801	0,956
Precision	0,873	0,944	0,734	0,884	0,821	0,957

As métricas de avaliação trouxeram resultados altos, tendo como modelo com melhor desempenho o Ada Boost em todos os casos, apenas atrás do XGBoost. Os maiores valores de acertos variaram entre 95% e 96% de acerto. De acordo com a acurácia por exemplo, isso significa que de 100 amostras o modelo acertou corretamente 95 ou 96 delas.

Para investigar melhor a questão dos desempenhos, valeria uma coleta com dados novos apenas para teste e verificação de desempenho dos modelos com dados novos. Também a expansão das métricas de avaliação, incluindo análise de um classification report[34] por exemplo.

[34] Disponível em https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html

5.2 VIESES

Em um trabalho com dados, não se pode se não lembrar da questão dos vieses, pois eles existem e estão sempre presentes ainda que medidas sejam tomadas para que eles sejam os menores possíveis. Dito isso, vale mencionar alguns pontos importantes para se discutir sobre o desenvolvimento do trabalho.

O primeiro é com relação à atuação da métrica de acurácia. A acurácia precisa de um cuidado com sua implementação por conta de conjunto de dados desbalanceados – não foi o caso no presente trabalho, mas ainda assim um sinal de atenção. Ela pode confundir na interpretação quando possui resultados muito altos. Como mais bem explicitado por essa definição:

> Em linhas gerais, o Paradoxo da Acurácia pode ser definido como a situação contraditória na qual uma acurácia elevada em seu modelo de classificação pode evidenciar uma falha do seu próprio modelo em realizar predições de fato significativas (AZANK, 2020).

Também é importante lembrar do viés de seleção na hora da coleta dos dados e os tipos de filtragem que foram realizados por grupos – (Telegram – Antivacina - falsidade) e (Twitter – Perfis divulgadores científicos - Verdade). Pode ser que nem todos os dados em cada um desses conjuntos sejam verdadeiros de acordo com definições diferentes de notícias falsas, por exemplo, podem estar fora de contexto, fora de época entre outros. Mas como não havia tempo hábil para explorar essas questões, os critérios de seleção atuais foram bem definidos e podem estar abertos à mudanças futuramente.

Por último, um ponto de atenção seria coletar mais dados para testes mais robustos com relação ao desempenho dos modelos. Também é um ponto de melhoria futura e a ciência dele, faz esse ponto se tornar importante para melhorias que serão implementadas.

6 ARQUITETURA DA SOLUÇÃO

Levando em consideração a natureza do problema abordado no presente trabalho, algumas decisões foram tomadas considerando em princípio, técnicas adequadas para que todas as etapas de resolução pudessem ser organizadas da forma mais eficiente. As tecnologias utilizadas bem como suas justificativas são as seguintes:

A linguagem de programação utilizada foi Python. Python é uma linguagem muito versátil, ainda mais quando se trata de manipulação, tratamento e análise de dados. Dado o tipo de problema de classificação do trabalho, foi escolhido um algoritmo capaz de realizar uma classificação binária utilizando a função logística como base. Entre as tecnologias utilizadas na modelagem e escolha dos modelos, processamento e tratamento de dados, visualização de dados e métricas de avaliação e deploy foram:

Biblioteca	Link
Scikit-learn: Machine Learning in Python	https://scikit-learn.org/stable/
NLTK: Natural Language Toolkit	https://www.nltk.org
Numpy: The fundamental package for scientific computing with Python	https://numpy.org
Pandas: Python Data Analysis Library	https://pandas.pydata.org
Matpolotlib: Visualization with Python	https://matplotlib.org
Seaborn: statistical data visualization	https://seaborn.pydata.org
Gensim: Topic modelling for humans	https://radimrehurek.com/gensim/
Streamlit: A faster way to build	https://streamlit.io

and share data apps	
PIL: Python Imaging Library	https://pypi.org/project/Pillow/
Snscrape: scraper for social networking services (SNS)	https://github.com/JustAnotherArchivist/snscrape
Telethon	https://docs.telethon.dev/en/stable/

As ferramentas para desenvolvimento de código, organização do projeto e desenvolvimento de diagramas de comunicação visual que foram utilizadas são: Google ColabPro, VSCode, Git e Github e Lucidchart.

7 CONSIDERAÇÕES FINAIS

Tendo em vista todo o trabalho desenvolvido, pensando no objetivo final de gerar um classificador de notícias falsas sobre vacinação, pode-se dizer que o trabalho cumpriu sua função. Todas as etapas durante o processo foram cheias de muito aprendizado – desde a própria coleta de dados em diversas fontes diferentes, o processamento dos dados, modelagem e aplicação de diversas técnicas, até o produto – deploy do melhor modelo em funcionamento. O trabalho cumpriu sua missão dentro do que foi proposto, ainda levando em consideração todos os requisitos iniciais.

Ainda existem muitos pontos de melhoria mas todos eles incluem próximos passos para serem trabalhados em um futuro próximo. Uma dessas possibilidades é um teste de hipótese acerca do da comparação de desempenho entre as técnicas de filtragem de dados apresentadas. Outra possibilidade é expandir o escopo de modelos e/ou de testes. Outra possibilidade é a coleta de mais dados para verificar o real desempenho do modelo em comparação à dados novos e como ele lida com isso – São possibilidades muito atraentes mas que este trabalho já fornece uma boa base para que tudo isso seja possível.

REFERÊNCIAS

ALURA, **Guia de NLP - conceitos e técnicas**. 2021. Disponível em: <https://www.alura.com.br/artigos/guia-nlp-conceitos-tecnicas>. Acesso em: 11 jun. 2023.

AAHILL. **Métricas de avaliação de classificação de textos personalizada** - Azure Cognitive Services. Disponível em: <https://learn.microsoft.com/pt-br/azure/cognitive-services/language-service/custom-text-classification/concepts/evaluation-metrics>. Acesso em: 13 jun. 2023.

ANÁLISE, REDE. **Equipe**. Disponível em: <https://redeanalise.com.br/equipe/>. Acesso em: 12 jun. 2023.

AZANK, FELIPE. **Medium: O paradoxo da acurácia**. 2020. Disponível em: <https://medium.com/turing-talks/paradoxo-da-acur>. Acesso em: 13 jun. 2023.

BHATT, GAURAV; SHARMA, SHIVAM; NAGPAL, ANKUSH; RAMAN, BALASUBRAMANIAN; MITTAL, ANKUSH. **On the Benefit of Combining Neural, Statistical and External Features for Fake News Identification**. ArXiv, Indian Institute of Technology, Roorkee, India, 11 dez. 2017. Disponível em: https://arxiv.org/abs/1712.03935. Acesso em 10 maio 2023

CEYLANA, GIZEM; ANDERSONB, Ian A; WOOD, Wendy. **Sharing of misinformation is habitual, not just lazy or biased**. 2022. PNAS, Princetion University, 17 jan. 2023.

DEEP AI. **N-Grams: Whats is na n-gram?**. Disponível em: <https://deepai.org/machine-learning-glossary-and-terms/n-gram>.

FILHO, M. **As Métricas Mais Populares para Avaliar Modelos de Machine Learning**. Disponível em: <https://mariofilho.com/as-metricas-mais-populares-para-avaliar-modelos-de-machine-learning/#acur>. Acesso em: 13 jun. 2023.

FIOCRUZ. **Vacinação infantil sofre queda brusca no Brasil**. Disponível em: <https://portal.fiocruz.br/noticia/vacinacao-infantil-sofre-queda-brusca-no-brasil>. Acesso em: 12 jun. 2023.

GRIGOREV, ALEXEY. **Machine Learning Bookcamp: build a portifolio of real-life projects.** Shelter Island: Manning, 2021.

GRUS, JOEL. **Data Science do Zero: Primeiras regras com o Python**. Rio de Janeiro: Alta Books, 2016.

HARTMANN, I. A.; IUNES, J. **Fake news no contexto de pandemia e emergência social: os deveres e responsabilidades das plataformas de redes sociais na moderação de conteúdo online entre a teoria e as proposições legislativas**. Direito Público, [S. l.], v. 17, n. 94, 2020. Disponível em: https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/4607. Acesso em: 10 jun. 2023

HUGGING, FACE. **What is Zero-Shot Classification?** - Hugging Face. Disponível em: <https://huggingface.co/tasks/zero-shot-classification>. Acesso em: 13 jun. 2023.

IBM. **O que é regressão logística?** | IBM. Disponível em: <https://www.ibm.com/br-pt/topics/logistic-regression>.

KAHNEMAN, Daniel. **Rápido e devagar: duas formas de pensar**. Rio de Janeiro: Objetiva, 2012.

KULSHRESTHA, R. **Latent Dirichlet Allocation(LDA).** Disponível em: <https://towardsdatascience.com/latent-dirichlet-allocation-lda-9d1cd064ffa2>.

LISBÔA, J. V. C. de O.; CAETANO, M. B. L.; FERMOSELI, A. F. de O.; DE OLIVEIRA, J. S. **Reincidência epidêmica do sarampo no brasil como consequência da pouca adesão popular à vacinação**. Caderno de Graduação - Ciências Biológicas e da Saúde - UNIT - ALAGOAS, [S. l.], v. 7, n. 1, p. 149, 2021. Disponível em: https://periodicos.grupotiradentes.com/fitsbiosaude/article/view/9042. Acesso em: 12 jun. 2023.

MONTEIRO R.A., SANTOS R.L.S., PARDO T.A.S., DE ALMEIDA T.A., RUIZ E.E.S., VALE O.A. (2018) **Contributions to the Study of Fake News in Portuguese: New Corpus and Automatic Detection Results**. In: Villavicencio A. et al. (eds) Computational Processing of the Portuguese Language. PROPOR 2018. Lecture Notes in Computer Science, vol 11122. Springer, Cham. Disponível em: https://github.com/roneysco/Fake.br-Corpus. Acesso em: maio de 2023

MORENO, JOÃO, BRESSAN GRAÇA. 2019. **FACTCK.BR: a new dataset to study fake news.** In Proceedings of the 25th Brazillian Symposium on Multimedia and the Web (WebMedia '19). Association for Computing Machinery, New York, NY, USA, 525–527. https://doi.org/10.1145/3323503.3361698. Disponível em: https://github.com/jghm-f/FACTCK.BR Acesso em: maio de 2023

MORTARI, Cezar A. **Introdução à Lógica**. São Paulo: Editora Unesp, 2001.

MURCHO, DESIDÉRIO. **Crítica na rede: Descaso epistémico**. 2018. Disponível em: <https://criticanarede.com/descaso.html>. Acesso em: 12 jun. 2023

NILC. **Resources and tools**. Disponível em: <https://sites.google.com/view/nilc-usp/resources-and-tools>. Acesso em: 12 jun. 2023.

O'CONNOR, CAILIN; WEATHERALL, JAMES OWEN. **The misinformation age: how false beliefs spread**. United States: Yale University Press, 2019.

OPAS, **Entenda a infodemia e a desinformação contra a COVID-19** (OPAS). 2020. Disponível em: https://iris.paho.org/bitstream/handle/10665.2/52054/Factsheet-Infodemic_por.pdf. Acesso em junho de 2023.

PARTNER, I. G. A **importância da normalização e padronização dos dados em Machine Learning**. Disponível em: <https://medium.com/ipnet-growth-partner/padronizacao-normalizacao-dados-machine-learning-f8f29246c12>. Acesso em: 13 jun. 2023.

PEREIRA, J. F. O.; FERNANDES, Q. H. R. F.; CARNEIRO, R. T. DE O. **Baixa adesão ao esquema vacinal anti-HPV por crianças e adolescentes**. Revista Família, Ciclos de Vida e Saúde no Contexto Social, v. 9, n. 4, 9 ago. 2021.

SANDRINI BEZERRA, L.; GIBERTONI, D. **As mídias sociais durante a pandemia do COVID-19: análise comportamental dos usuários durante este período e as possibilidades para o futuro**. Revista Interface Tecnológica, [S. l.], v. 18, n. 2, p. 144–156, 2021. DOI: 10.31510/infa.v18i2.1239. Disponível em: https://revista.fatectq.edu.br/interfacetecnologica/article/view/1239. Acesso em: 10 jun. 2023.

SCHRÖER, CHRISTOPH; KRUSE, FELIX; GÓMEZ, JORGE MARX. **A Systematic Literature Review on Applying CRISP-DM Process Model**. Procedia Computer Science, [s. l.], p. 526-534, 2021. Disponível em: https://www.sciencedirect.com/science/article/pii/S1877050921002416. Acesso em: 10 jun. 2023.

SEBASTIAN RASCHKA, MIRJALILI VAHID. **Python Machine Learning: Machine Learning and Deep Learning With Python, scikit Learn and TensorFlow 2**. Third Edition: Packt Publishing Ltd, 2019.

SILVA, RENATO M., SANTOS R.L.S, ALMEIDA T.A, AND PARDO T.A.S. (2020). **Towards Automatically Filtering Fake News in Portuguese**.Expert Systems with Applications, vol 146, p. 113199.

SOMMERVILLE, IAN. **Engenharia de Software**. Tradução: Ivan Bosnic e Kalinka G. de O. Golçalves. Revisão Técnica: Kechi Hirama. 9ª ed. São Paulo: Pearson Education, 2011

SOUZA, D. M. R. DE. **m-oxu/dscbc_2021_01-fakenews**. Disponível em: <https://github.com/m-oxu/dscbc_2021_01-fakenews>. Acesso em: 12 jun. 2023.

STREAMLIT DOCUMENTATION: **Create an app**. [S. l.], 1 jan. 2023. Disponível em: https://docs.streamlit.io/library/get-started/create-an-app. Acesso em: 10 jun. 2023.

_____. **host your Streamlit app for free**. Disponível em: <https://blog.streamlit.io/host-your-streamlit-app-for-free/#:~:text=Connect%20your%20account%20to%20GitHub>.

RABELLO, E. B. **Cross Validation: Avaliando seu modelo de Machine Learning**. Disponível em: <https://medium.com/@edubrazrabello/cross-validation-avaliando-seu-modelo-de-machine-learning-1fb70df15b78>.

RASCHKA, SEBASTIAN; MIRJALILI, Vahid. **Python Machine Learning: Machine learning and deep learning with Python, scikit-learn, and tensorFlow 2**. 3ª. ed. Birmingham: Packt, 2019.

R7.COM. **Baixa procura pela vacina bivalente contra a Covid preocupa Ministério da Saúde**. Disponível em: <https://noticias.r7.com/jr-24h/boletim-jr-24h/videos/baixa-procura-pela-vacina-bivalente-contra-a-covid-preocupa-ministerio-da-saude-28032023>. Acesso em: 12 jun. 2023.

RODRIGO. **FakeOnlineTCC-Corpus**. Disponível em: <https://github.com/Rodrigoguigo/FakeOnlineTCC-Corpus>. Acesso em: 12 jun. 2023.

ROHAN CHOPRA, ANIRUDDHA M. GODBOLE, et al. **The Natural Language Processing Workshop: Confidently design and build your own NLP Projects with this easy-to-understand practical guide**.Packt Publishing Ltd, 2020.

RUSSEL, STUART; NORVIG, PETER**. Inteligência artificial**. 3 ed. ed. Rio de janeiro: Elsevier, 2013

SAÚDE, MINISTÉRIO. **Programa Nacional de Imunizações - Vacinação.** Disponível em:<https://www.gov.br/saude/pt-br/acesso-a-informacao/acoes-e-programas/programa-nacional-de-imunizacoes-vacinacao#:~:text=Em%201973%20foi%20formulado%20o>.

SBIM. **Ministério, SBIm e outras entidades científicas se unem em prol das altas coberturas - SBIm.** Disponível em: <https://sbim.org.br/noticias/1790-ministerio-sbim-e-outras-entidades-cientificas-se-unem-em-prol-das-altas-coberturas>. Acesso em: 12 jun. 2023.

TERA**. Entenda o que é deploy de modelos em Machine Learning**. Disponível em: <https://blog.somostera.com/data-science/deploy-o-que-e>.2021 Acesso em: 13 jun. 2023.

VARELLA, D. D. **Brasil ocupa 2a posição entre países com a pior taxa de vacinação em bebês da América Latina**. Disponível em: <https://drauziovarella.uol.com.br/coluna-2/brasil-ocupa-segunda-posicao-entre-paises-com-a-pior-taxa-de-vacinacao-em-bebes-da-america-latina/>.

WALLACH, WENDELL; ALLEN, COLIN. **Moral Machines: teaching robots right from wrong**. New York: Oxford University Press, 2009.

ZENHA, L. **Redes sociais online: o que são as redes sociais e como se organizam?** Caderno de Educação, n. 49, p. 19–42, 27 mar. 2018.

ZUBAREV, VASILY. **Machine Learning for Everyone In simple words. With real-world examples. Yes, again: The map of the machine learning world**. [S. l.], 2019. Disponível em: https://vas3k.com/blog/machine_learning/. Acesso em: jun. 2023.