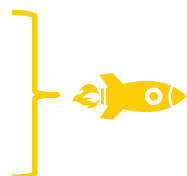




DETECTOR DE FAKE NEWS NO COMBATE A COVID-19 NO BRASIL

Felipe Gabriel, Lauro Namiki, Mel Iza, Moxú, Rodolpho Amadatsu.

"Uma mentira pode dar a volta ao mundo no mesmo tempo que a verdade leva para calçar seus sapatos."
Mark Twain



O que são as famosas *Fake News*? Onde vivem? E mais importante, como se reproduzem? O compartilhamento de Informações se tornou um mecanismo útil para diversos domínios do convívio social, permitindo avanços acelerados em pesquisas, acesso à conhecimentos antes restritos e criação de novas formas de interações humanas, como mensagens instantâneas e vídeo-chamadas.

Contudo, possibilitou também o compartilhamento de mensagens e informações de fontes desconhecidas com conteúdos duvidosos e/ou propositalmente falsos/não verídicos. Como define o autor Weedon *Artigos de notícias que parecem ser factuais, mas que contêm distorções intencionais de fatos com o propósito de provocar paixões, atrair audiência ou enganar.*

Como as notícias falsas podem se reproduzir tão rapidamente? Aliado à facilidade de compartilhamento junto com mensagens de procedência enganosa. Por se caracterizar por uma peculiaridade bem específica: são emocionalmente apelativas. Isso pode levar a um outro fenômeno, a desinformação. A desinformação é uma consequência de super espalhamento de notícias falsas.

A desinformação é um problema por vários motivos, principalmente porque:

- 1) permite a distribuição de notícias falsas, o que é um problema por si mesmo;

- 2) possibilita uma norma crescente em que as pessoas não se preocupam em conferir e nem verificar aquilo que acreditam;
- 3) por ser apelativa à reações emocionais imediatas e rápidas, pode ser compartilhada em uma frequência maior que informações verdadeiras.

Existe alguma maneira de combater desse ciclo? Do ponto mais geral, é um problema muito complexo. Mas se dividirmos um problema em variadas pequenas partes, é melhor para analisar e poder atuar. E dentro do escopo do nosso projeto, essa pequena parte a tomada de decisão.



PROBLEMA DE NEGÓCIO

O problema central que norteia o desenvolvimento do nosso trabalho é a conjunção de duas ideias principais. A primeira ideia é que pessoas que não conseguem distinguir afirmações (sejam elas veiculadas a notícias ou mensagens) falsas de verdadeiras podem ter mais facilidade a acreditar em afirmações falsas por apelo emocional, e por consequência compartilhar e contribuir para a rede de desinformação.

O perfil da nossa persona é definido pelas *pessoas que não conseguem distinguir notícias falsas de verdadeiras*, principalmente levando em consideração que as fontes de informações sobre a pandemia que essas pessoas costumam confiar são as redes sociais.

De acordo com um levantamento da Avaaz, as fontes mais consultadas pelos entrevistados são: mensagens instantâneas do Whatsapp (90%), O aplicativo de troca de mensagens é seguido pelo Facebook (83%), YouTube (71%), Instagram (64%) e Twitter (26%). Apenas 1% disse que não utiliza redes sociais. Abaixo para ilustrar melhor se encontram algumas notícias falsas apresentadas na pesquisa e as fontes mais citadas.

O novo coronavírus foi criado em um laboratório secreto na China	Total	Tomar grandes doses de vitamina C pode retardar ou até impedir a infecção do novo coronavírus	Total
Base	1513	Base	976
WhatsApp	47	WhatsApp	49
Facebook	46	Facebook	42
Conversas orais com família e amigos	42	Conversas orais com família e amigos	40

Especialistas em saúde recomendam beber água regularmente pois isso levará o novo coronavírus para seu estômago, onde a acidez irá matá-lo	Total	O novo coronavírus é como qualquer gripe, tem os mesmos sintomas e uma taxa de mortalidade igual ou inferior à gripe comuns	Total
Base	850	Base	1003
WhatsApp	51	Televisão	45
Facebook	38	WhatsApp	43
Televisão	37	Facebook	42

Figura 1. Estudo produzido pela Avaaz, 2020. Fonte: https://secure.avaaz.org/campaign/po/brasil_infodemia_coronavirus/

A segunda ideia que constitui nosso problema de negócio é que o compartilhamento de notícias falsas relacionadas à SARS-CoV-2 (Covid-19) durante a pandemia é especialmente perigoso, pois pode apresentar consequências graves à saúde. Tomar decisões baseadas em falsidades no contexto da Covid-19 é prejudicial à saúde não somente individual, mas pública. Como por exemplo, tomar decisões com base na informação do tipo “covid é uma gripezinha”, ou “usar máscaras causam embolia pulmonar”, ou mesmo “lockdown não funciona e sim imunidade de rebanho” é extremamente prejudicial quando baseada em informações falsas como essas mencionadas acima. Na maioria das vezes elas são contra medidas básicas protetivas contra a Covid-19 (comprovadas cientificamente), como funcionamento do modo de transmissão da doença e uso de distanciamento social. Nesse cenário, as tomadas de decisão podem levar as pessoas a se contaminarem com mais facilidade, espalhar a doença, serem hospitalizadas ou virem a óbito.



IMPACTO

Nosso projeto é importante porque pretende a partir da análise dos dados de notícias falsas e verdadeiras coletadas, produzir um modelo que atuará como auxiliar no processo de tomada de decisão de compartilhar notícias. Por que isso é importante?

Porque é importante maximizar às pessoas que não conseguem distinguir o que é falso do que é verdadeiro, o acesso à informação de qualidade e ferramentas que possibilitem auxiliar o processo de tomada de decisão.

Nesse sentido, uma das funções de um detector de notícias falsas é ser uma ferramenta de verificação ao mesmo tempo que incentiva o desenvolvimento 1) do pensamento crítico e 2) da habilidade de confrontação e discussão de informações. Habilidades essas que, se desenvolvidas, ao longo prazo podem auxiliar o aperfeiçoamento de tomadas de decisão de modo crítico e evitar a desinformação.



DESENHO DA SOLUÇÃO

A solução do nosso problema foi dividida em 6 etapas: a problematização e desenvolvimento de hipóteses, a mineração dos dados de notícias falsas e verdadeiras, a limpeza e processamento dos dados, aplicação do aprendizado de máquina (divisão do dataset em treino e teste e treinamento de modelos), avaliação dos modelos e comparação de resultados, e por fim o deploy do modelo.



O primeiro passo foi refletir melhor sobre a problematização do assunto para redigir de modo claro e conciso as hipóteses norteadoras do projeto. A nossas hipóteses são:

1. Hipótese nula (H_0): Não é possível desenvolver um modelo de Machine Learning robusto e confiável para automatizar a detecção de fake News sobre Covid-19 e auxiliar no processo de tomada de decisão de pessoas que não sabem distinguir verdades de falsidades.
2. Hipótese alternativa (H_1): É possível desenvolver um modelo de Machine Learning robusto e confiável para automatizar a detecção de fake News sobre Covid-19 e auxiliar no processo de tomada de decisão de pessoas que não sabem distinguir verdades de falsidades.



DADOS

Apesar de muitas mensagens falsas circularem com mais frequência em redes sociais como Facebook e Twitter (BHATT, 2017), coletamos os dados - notícias falsas e notícias verdadeiras - retiradas de veículos jornalísticos de notícias e agências verificadoras. Os fundamentos utilizados para essa decisão se baseiam no fato de que dado o tempo hábil disponível para a entrega do projeto, não haveria possibilidade de verificar e analisar com profundidade a verdade de cada notícia. E a seleção das fontes foi realizada tendo em vista que existe uma equipe de jornalistas e profissionais que já realizam a checagem de antemão. Levando em consideração o viés da coleta de notícias, a estratégia utilizada para tentar

amenizar isso durante o período de mineração foi usar como fonte para as notícias verdadeiras grandes portais de notícias e para as falsas, agências checadoras de fatos. Dado o vasto número de notícias que existem na internet, os critérios de inclusão e exclusão para seleção das notícias seguem abaixo de acordo com a tabela.

Crítérios de inclusão	Crítérios de exclusão
Notícias de agências verificadoras em língua portuguesa	Notícias de agências verificadoras em outras línguas.
Notícias relacionadas ao assunto “Covid-19” e “Pandemia”	Notícias não relacionadas ao assunto “Covid-19” e “Pandemia”
Notícias com os seguintes dados: Título, texto, data de publicação e url.	Notícias que não possuem os seguintes dados: Título, texto, data de publicação e url.

Usamos também o *Data Tracking Sheet* como método de base para estabelecer quais dados seriam coletados, como e onde. As principais questões abordadas para a coleta dos dados foram: “Como identificar uma notícia falsa?” e “Quais são suas características?”.

BUSINESS QUESTIONS	DADOS	BASE OU CALCULADA	CÁLCULO	FONTE DO DADO
Como identificar uma notícia falsa?	Notícias em texto	-	Probabilidade / Classificação	Notícias verdadeiras e falsas mineradas da internet em conjunto com as bases de dados existentes
Quais são as características de uma notícia falsa?	Notícias falsas / verdadeiras	Base de dados	-	Estadão / G1 / Uol / Boatos.org / Agência Lupa / AFP Checamos / Terça Livre / Aosfatos.org
É possível desmentir uma notícia falsa?	Notícias falsas / verdadeiras	Base de dados	-	Estadão / G1 / Uol / Boatos.org / Agência Lupa / AFP Checamos / Terça Livre / Aosfatos.org

Fontes de dados

As fontes de onde as notícias foram retiradas são: Projeto Comprova, Uol, Estadão, G1: Fato ou Fake, Agência Lupa, E-farsas e Terça-livre. Para contribuir com a otimização do tempo e a fim de evitar dupla checagem dos dados coletados, foi adotado como método de busca o Web scrapping ou raspagem de dados juntamente com a conferência manual de cada integrante do grupo para conferir se os requisitos das notícias como título, texto, data e assunto estavam corretos; bem como uma verificação rápida da veracidade.

Foi necessário a criação de um raspador para cada site diferente – os códigos foram adaptados a cada um para captar as informações sobre o título da notícia, texto, data e salvar um arquivo em formato .csv. Essa foi a etapa de grande importância e decisiva para o projeto, pois todo o resto dependeria da qualidade do nosso dataset. Tivemos um pouco de dificuldade em precisar adaptar scrapers para sites diferentes, e as ferramentas foram

variadas para cada tipo de necessidade. Como utilizar a linguagem Python e R, bem como bibliotecas diferentes, a exemplo do Selenium e BeautifulSoup.

1a. Parte: Buscando os dados do site

```

In [2]: #Place where the files will be saved
dir_output = 'pages'

In [3]: # Procedure to create a directory to save the pages
try:
    os.mkdir(dir_output)
except OSError:
    print ("Directory %s already exists." % dir_output)
else:
    print ("Successfully created the directory %s " % dir_output)

Directory pages already exists.

In [4]: input_file = '2021_Uo1_04.txt'

In [5]: input_file
Out[5]: '2021_Uo1_04.txt'

Fazendo a coleta de dados

In [8]: from time import sleep
from random import randint

pages_collected = next(os.walk(dir_output))[2]

with open(input_file) as infile:
    for line in infile:
        url = line.strip()
        page_name = url.split('/')[-1]
        if page_name in pages_collected:
            print ("Page already collected.")
            continue
        elif page_name.split('.')[1] != "htm":
            print ("Link is not a pagelink não é uma página HTML")
            continue
        print (url)
        if url != [' ']:
            print('Conte até dez...')
            sleep(randint(7,15))
            try:
                page = requests.get(url)
            except HTTPError as e:
                print('Erro no HTTP do tipo ', e)
            except URLError as e:
                print('Servidor não pode ser encontrado', e)

            with open(dir_output + '/' + page_name, 'w') as f:
                f.write(page.text)
            print('Fim da lista')

```

Figura 2. Exemplo do Web Scraper em Python utilizado para coletas notícias do site Uol

A coleta foi dividida entre os membros do grupo durante o período de 01 de Abril a 06 de Maio e ao final do período estabelecido foram coletados o total de 5487 notícias, sendo 3263 verdadeiras e 2224 falsas. Esse é um exemplo de como foram coletados e organizados os dados para a próxima etapa – processamento e limpeza.

links	data	titulos	textos
<chr>	<chr>	<chr>	<chr>
https://www.boatos.org/saude/china-nao-teve-fase-3-coronavac-porque-brasileiros-sejam-cobaia.html	11/02/2021	China não teve fase 3 da Coronavac porque quer que brasileiros sejam cobaia #boato	"Saiu uma notícia de que a Coronavac não passou pela fase 3 de testes na China. Ora, os brasileiros são as maiores cobaia voluntárias da China nos testes da vacina. A fase 3, meus queridos, são vocês!".
https://www.boatos.org/saude/china-nao-teve-fase-3-coronavac-porque-brasileiros-sejam-cobaia.html	11/02/2021	China não teve fase 3 da Coronavac porque quer que brasileiros sejam cobaia #boato	"Coronavírus: CoronaVac não teve fase 3 de testagem na China".
https://www.boatos.org/saude/china-nao-teve-fase-3-coronavac-porque-brasileiros-sejam-cobaia.html	11/02/2021	China não teve fase 3 da Coronavac porque quer que brasileiros sejam cobaia #boato	"Claro que não, as cobaia são os Brasileiros!".

Figura 3. Resultado da coleta de dados do site Boatos.org utilizando Scraper com a linguagem R.

Dataset

Com a nossa base de dados consolidada tivemos um total de 5487 notícias distribuídas em 4 colunas. A coluna de URL do site, título da notícia, texto da notícia e classe que pode ser definida em verdadeira (representada pelo valor 1) ou falsa (representada pelo valor 0).

```
pd.set_option("display.max_columns", None)

# displaying the DataFrame
display(df)
df.classe.value_counts()
```

	url	titulo	noticia	classe
0	https://chequeado.com/verificaconfb/no-nostra...	nao ha registro da suposta profecia no livro d...	nostradamus previu o coronavirus	FALSO
1	https://plaul.foiha.uol.com.br/lupa/2021/02/05...	verificamos variante coronavirus manaus	a nova variante do sarscov e mais letal e nao ...	FALSO
2	https://maidita.es/maiditobulo/2020/04/03/ayud...	o site acessado nao oferece a possibilidade de...	mensagem que mostra um link que leva a um site...	FALSO
3	https://plaul.foiha.uol.com.br/lupa/2021/03/25...	verificamos inalacao agua sanitaria	inalacao de bicarbonato de sodio com agua sani...	FALSO
4	https://politica.estadao.com.br/blogs/estadao-...	nao ha evidencias de que nostradamus tenha esc...	postagens de uma profecia de nostradamus sobre...	FALSO
...
5497	https://economia.uol.com.br/noticias/redacao/2...	guedes diz que pode derrubar encargos trabalhi...	segundo ele a medida seria tomada la na frente...	VERDADEIRO
5498	https://www.camara.leg.br/noticias/680401-CAMA...	camara aprova mp que destinou r bilhoes para a...	najara araujocamara dos deputados deputados...	VERDADEIRO
5499	https://noticias.uol.com.br/ultimas-noticias/e...	espanha soma mil infectados pela covid em um d...	o pais acordou hoje com ruas vazias lojas fech...	VERDADEIRO
5500	https://www.camara.leg.br/noticias/647043-PROJ...	projeto preve alcool em gel gratuito para grup...	compartilhe versao para impressao ...	VERDADEIRO
5501	https://checamos.afp.com/esta-cena-de-um-hospi...	esta cena de um hospital apos um incendio foi ...	uma foto de pessoas deitadas em macas e aglome...	VERDADEIRO

5487 rows x 4 columns
VERDADEIRO 3263
FALSO 2224
Name: classe, dtype: int64

As colunas 'url', 'título' e 'notícia' são do tipo string e a variável alvo 'classe' que é o label se a notícia é verdadeira ou falsa é categórica.

```
[ ] df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5487 entries, 0 to 5486
Data columns (total 4 columns):
#   Column   Non-Null Count  Dtype
---  ---
0    url      5487 non-null   object
1    titulo   5487 non-null   object
2    noticia  5487 non-null   object
3    classe   5487 non-null   category
dtypes: category(1), object(3)
memory usage: 134.2+ KB
```



LIMPEZA DOS DADOS E MODELOS

Nesta etapa de limpeza e processamento dos dados, utilizamos técnicas de vetorização de textos com o objetivo de padronização das palavras, redução de ruídos dos textos, eliminação dos erros de interpretação e otimização do algoritmo.

As técnicas de vetorização fazem parte do amplo repertório de técnicas do processamento de linguagem natural (Natural Language Processing - NLP). Dentre elas, as que aplicamos foram a normalização, Tokenização, remoção de stop words e a lematização.

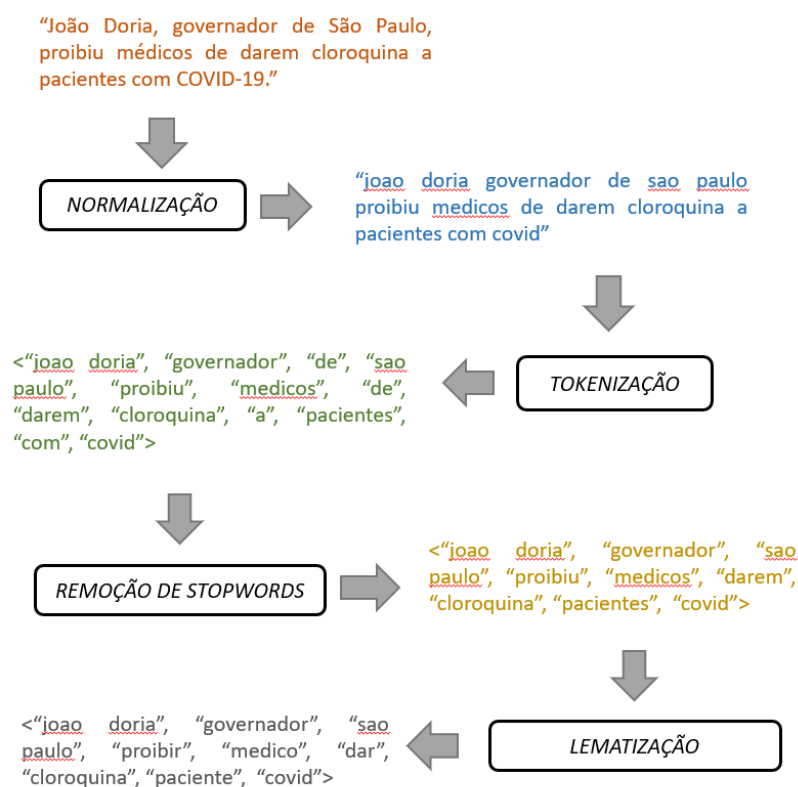
Antes de avançarmos para o estágio mais *hyped* de um projeto de ciência de dados, o treinamento do algoritmo, precisamos tratar nossos dados e depois transformá-los em uma

linguagem que o computador entenda, isso mesmo, números. Esse processo é composto por duas etapas, Limpeza e processamento dos dados, que possuem diversas tarefas associadas que serão detalhadas a seguir.

Limpeza de dados

Este processo visa reduzir o ruído nos textos, otimizando o algoritmo, pois deixa apenas palavras padronizadas e relevantes ao contexto para serem analisadas. Ele também evita erros de interpretação, como por exemplo, que duas palavras iguais sejam consideradas diferentes apenas por apresentar a primeira letra maiúscula.

- **Normalização:** Primeiro tratamento aplicado ao texto, nessa etapa o texto original da notícia já sofre uma grande transformação, é retirado as pontuações, os caracteres especiais (acentos, símbolos e números) são eliminados ou substituídos, todos os caracteres são padronizados para minúsculas.
- **Tokenização:** O texto da notícia é quebrado em palavras (tokens), é possível identificar palavras compostas como por exemplo “São Paulo”.
- **Remoção de StopWords:** Agora que já temos nossos tokens, vamos seleccionar apenas aqueles que nos interessam, para isso vamos eliminar palavras irrelevantes para nossa análise - As, e, os, de, para, com, sem e etc.
- **Lematização:** É o processo de flexionar uma palavra para determinar o seu lema, por exemplo as palavras: tiver, tenho, tinha, tem são do mesmo lema ‘ter’.



Processamento

Após a limpeza dos dados precisamos criar a representação numérica desses textos para que eles se tornem interpretáveis pelo computador e os algoritmos possam ser treinados na classificação das notícias.

Nós utilizamos basicamente duas técnicas, Bag of Words e TF-IDF:

- **Bag of Words:** Qualquer informação sobre gramática ou a ordem das palavras é desconsiderada, ele transforma o conjunto de palavras em um grupo de termos e suas respectivas frequências.
- **TF-IDF:** É utilizado para aplicar pesos às palavras, compara o número de vezes que aquela palavra aparece em uma notícia, com o número de notícias que contém a palavra.

BoW - COUNT VECTORIZER

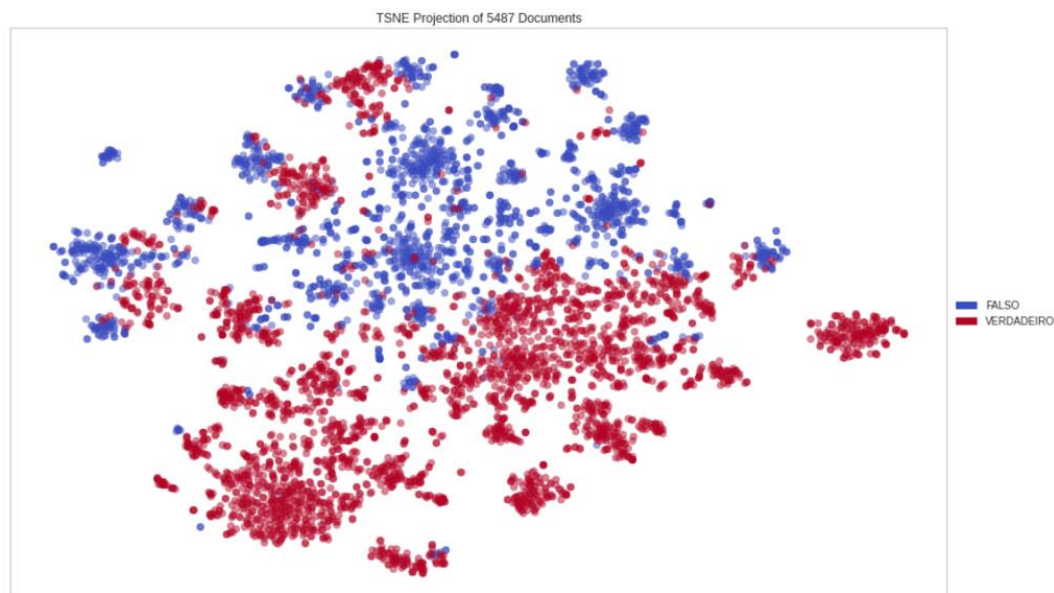
	VACINA	PRECOCE	CLOROQUINA	COVID
NOTÍCIA 1	1	0	0	1
NOTÍCIA 2	0	1	1	0

TF-IDF VECTORIZER

	VACINA	PRECOCE	CLOROQUINA	COVID
NOTÍCIA 1	0.707107	0	0	0.707107
NOTÍCIA 2	0	0.707107	0.707107	0

Figura 4. Bag of Words e TF-IDF vectorizer

Depois de toda manipulação dos dados, você não está curioso em saber como os vetores ficaram?! Nós também ficamos e para possibilitar uma confrontação visual dos vetores das notícias verdadeiras x notícias falsas foi utilizado o t-SNE, que é um método estatístico para visualização de dados altamente dimensionais em um mapa, neste caso, 2D.



Treinamento do algoritmo e avaliação

Antes de iniciar o processo iterativo de treinamento de algoritmos, foi realizado a divisão de dados entre treino e teste, na proporção de 67% dos dados para treino e 33% para teste, nos dados de treino ainda foi realizado um *cross validation* (k folds = 3), a técnica de *cross validation* consiste em dividir os dados em subgrupos – k folds – onde um subgrupo é separado para teste e os demais são utilizados para treino do algoritmo, esse processo é repetido até que cada subgrupo tenha sido utilizado como teste uma vez.

Assim, testamos diferentes abordagens na busca pelo algoritmo com melhor desempenho, combinamos as técnicas de processamento *Bag of Words* e *TF-IDF* com algoritmos de classificação como *Regressão Logística*, *Random Forest* e *XGBoost Classifier*.

Existem diversas métricas para avaliação dos modelos de classificação como a Precisão, F1-score e a Curva ROC, porém as métricas selecionadas para comparação de performance entre os modelos foram a matriz de confusão, acurácia e recall. No qual:

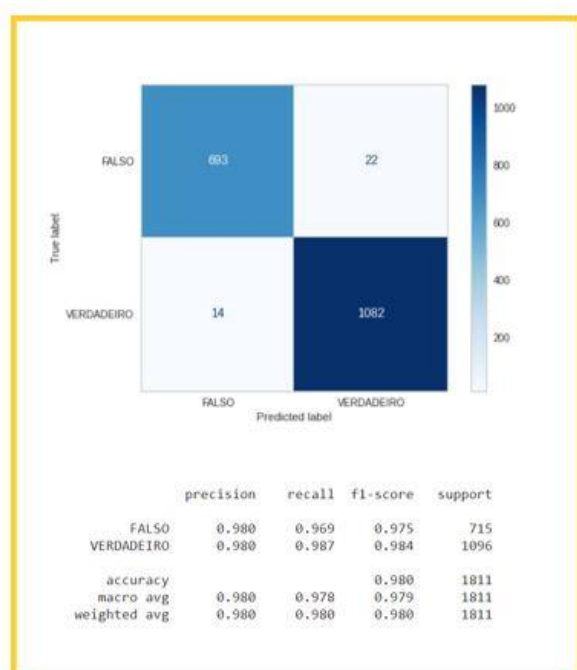
- **Matriz de confusão:** A partir dela são calculadas todas as outras métricas, é uma tabela que mostra as frequências de classificação para cada classe do modelo;
- **Acurácia:** Indica uma performance geral do modelo. Dentre todas as classificações, quantas o modelo classificou corretamente;
- **Sensibilidade (Recall pra classe positiva):** Dentre todas as situações de situação positiva (fake news) como valor esperado, quantas estão corretas.
- **Especificidade (Recall pra classe negativa):** Dentre todas as situações de situação negativa (notícia verdadeira) como valor esperado, quantas estão corretas.

MODELO	ACURÁCIA	PRECISÃO		RECALL		F1-SCORE	
	GERAL	VERDADEIRO	FALSO	VERDADEIRO	FALSO	VERDADEIRO	FALSO
Regressão Logística + BOW	0,988	0,998	0,974	0,983	0,997	0,990	0,985
Regressão Logística + TF-IDF	0,989	0,993	0,983	0,989	0,989	0,991	0,986
Random Forest + TF-IDF	0,987	0,999	0,969	0,979	0,999	0,989	0,983
CGBost + TF-IDF	0,985	0,994	0,971	0,981	0,992	0,988	0,981
Regressão Logística + TF-IDF (dados novos)	0,840	0,889	0,812	0,727	0,929	0,800	0,867
Random Forest + TF-IDF (dados novos)	0,920	1,000	0,875	0,818	1,000	0,900	0,933
CGBost + TF-IDF (dados novos)	0,800	0,800	0,800	0,727	0,857	0,762	0,828
Regressão Logística + TF-IDF - Lematizado	0,910	0,872	0,995	0,997	0,776	0,931	0,872
Regressão Logística + TF-IDF - Truncado	0,969	0,954	0,994	0,996	0,926	0,975	0,959
Regressão Logística + TF-IDF - Balanceado	0,981	0,973	0,993	0,995	0,958	0,984	0,975
Regressão Logística + TF-IDF - Balanceado Manual	0,980	0,980	0,980	0,987	0,969	0,984	0,975
Regressão Logística + TF-IDF - Sem datas	0,911	0,953	0,856	0,898	0,931	0,924	0,892

Após a avaliação das métricas dos primeiros modelos surgiu a hipótese de que o comprimento da notícia estava servindo de proxy para sua classificação, ou seja, em vez de aprender a diferenciar as notícias falsas das verdadeiras o modelo estava associando que as notícias com poucas palavras eram falsas e por consequência as que tinham mais palavras eram verdadeiras. Para contornar esse problema foi necessário balancear o comprimento entre as notícias, então das 5487 notícias originais, foram utilizadas apenas 1811 notícias.

Também se revelou a necessidade de aplicar pesos diferentes para as notícias, já que o dataset final estava com um desbalanceamento moderado de 57.2% de notícias verdadeiras e 42.8% de notícias falsas.

Comparando os resultados após o truncamento (redução do comprimento das notícias por força bruta, equalizando o tamanho das notícias através da limitação de caracteres) foi observado que o modelo com melhor performance foi a combinação da Regressão Logística com TF-IDF e balanceado manualmente. Este modelo obteve resultados melhores, tendo um ótimo desempenho na identificação tanto das notícias verdadeiras como as falsas, esse modelo ainda é interpretável.



PERFORMANCE DO MODELO

- **Acurácia: 98.0%;**
- **Sensibilidade: 98.7%**
- **Especificidade: 96.9%**

Além de apenas detectar as notícias falsas, o que também motivou o desenvolvimento deste projeto foi a possibilidade de encontrar padrões, de buscar caracterizar como são as notícias falsas, e para isso utilizamos a biblioteca LIME para fazer o raio – x do nosso modelo.

Através do LIME retiramos as palavras que mais impactaram para uma notícia ser classificada como falsa ou verdadeira.

PESOS - FALSAS

-21.1040 covid
-13.9851 vacina
-13.2641 diminuir
-12.5774 colombia
-11.5524 distanciamento
-11.4035 legendar
-10.9290 descartar
-10.5511 whatsapp
-10.4138 madri
-10.3081 isolamento
-10.2076 conteudos
-10.1358 espanha
-10.1289 oms
-9.6966 buzios
-9.3750 afirmacoes
-9.2557 governador
-9.0654 supostamente
-8.9268 aliexpress
-8.7953 espanhol
-8.7629 mascara

PESOS - VERDADEIRAS

21.9491 atualizado
21.2955 fake
18.7916 deputar
18.5396 compartilhar
18.3724 segundar
17.4463 ultima
15.6157 falecer
15.1303 evento
14.5337 proposto
14.5236 quarta-feira
14.3126 entanto
14.2349 vezar
14.0304 marcar
13.9511 sexta-feira
13.9332 atar
13.8163 comentarios
13.3343 medir
13.3208 semana
13.2701 anunciar
13.1990 riscar

É possível também ter uma visualização de como o modelo faz a análise da notícia para classificá-la.

Categoria: FALSO --> Predição: FALSO | Probabilidade: 99.683979 %

FALSO

VERDADEIRO

Text with highlighted words

o **oms** **garantir** que o **coronavirus** nao e transmitir pela **ar** e e por isso que nao e **necessario** usar **mascara** em **espacos publicos**

oms
0.01
mascara
0.01
espacos
0.01
necessario
0.00
garantir
0.00
publicos
0.00
ar
0.00
coronavirus
0.00
nao
0.00
usar
0.00



CONCLUSÃO

A regressão logística foi o modelo com as melhores métricas - com a acurácia de 98% - então primeiro nós podemos dizer que, os dados são linearmente separáveis, e segundo, que temos um detector de fake news sobre a Covid-19 confiável. É possível também que existam temas associados à procedência de uma notícia, através da análise dos maiores pesos aplicados as palavras pelo algoritmo nota-se que há referências a países (Colômbia, Espanha), cidades (Búzios, Madrid) e medidas preventivas (distanciamento, isolamento, máscara) que valem uma investigação mais profunda.

O processo de mineração de notícias foi de extrema importância pois se os dados estivessem enviesados não seria possível a criação do detector.

O tamanho da notícia influenciava o aprendizado do algoritmo, foi necessário equalizar o comprimento entre as notícias falsas e verdadeiras;

O registro de todo o desenvolvimento foi muito importante para compreensão de todos do grupo do que está sendo feito e para correção e/ou aperfeiçoamento do trabalho.

Desenvolver um projeto de dados de ponta a ponta sobre um assunto tão complexo como as *fake news*, que também envolvem outros fenômenos, foi de extremo aprendizado. Tanto se tratando da parte técnica de mineração e processamento os dados, quanto da produção do modelo e realização de testes.

Gostaríamos de agradecer à Tera por essa incrível oportunidade, e aos professores, experts e facilitadores que nos acompanham durante essa jornada. Em especial – André, Bia, Bidu, Edu e Marcos. Obrigada, sem o apoio e incentivo de vocês, esse projeto não seria possível. Também agradecemos aos colegas de grupo e de turma.



COLOCANDO SOLUÇÃO À PROVA

Utilizaremos as plataformas Streamlit e Heroku para realizar o deploy do modelo e nossa proposta de interface contém alguns elementos essenciais: o primeiro deles é uma caixa de input de texto. O usuário colocará o texto nessa caixa e a partir dele terá a resposta se a notícia é falsa ou não. Outro elemento essencial é a adição de uma cartilha para complementar a estratégia em ajudar as pessoas no processo de tomada de decisão. Juntamente a essa ideia da cartilha, teremos algumas referências e informações disponíveis sobre a pandemia e fontes confiáveis de consulta para auxiliar na conscientização sobre notícias falsas. A cartilha ficará disponível no aplicativo e também no Github.



Figura. Proposta de interface para o Streamlit



Figura. Proposta para Cartilha



IMPLICAÇÕES E PRÓXIMOS PASSOS

Pensando no impacto que pode ser gerado, nossos próximos passos seriam:

01

Implementar indicador no app

Aplicar questionário para medir a influência do detector no processo decisório de compartilhamento de notícias;

02

Técnicas avançadas de processamento

Utilização de *Word Embeddings* para vetorização das notícias;

03

Otimização de hiperparâmetros

Utilização de técnicas como o *GridSearch* ou *Randomized Search*.

Visto que a proposta da nossa solução está diretamente ligada com o processo de tomada de decisão dos usuários, vale lembrar que o objetivo é propor uma solução que possa auxiliar nesse processo. Vale lembrar que é de responsabilidade do usuário suas crenças, decisões e comportamentos. Uma outra ideia de melhoria seria justamente poder ter um feedback do usuário da plataforma, através de um questionário simples, com a utilização de escala *likert*, poder mensurar o impacto do detector no seu comportamento, se por exemplo, este usuário deixou de compartilhar uma possível notícia falsa após usar o detector e se na sua própria opinião notou uma melhora no desenvolvimento do pensamento crítico em relação ao recebimento de notícias de terceiros.

A nível técnico, existem outras técnicas de NLP que gostaríamos de ter experimentado no nosso projeto, como as *Word Embeddings*, essa técnica se utiliza de vetores (pré-treinados) densos de tamanho fixo que são capazes de armazenar informações sobre o contexto e significado, assim palavras que ocorrem em enredos similares vão estar próximas no espaço. E também implementar técnicas de otimização de hiperparâmetros (parâmetros arbitrários configurados antes do início do treinamento do modelo), *GridSearch* ou *RandomizedSearch*, sistematizando a busca pelos melhores parâmetros afim de melhorar ainda mais as métricas obtidas.

Por último, poderia ser interessante melhorar a qualidade dos dados coletados incluindo mais variáveis e buscar outras fontes de coleta, como diretamente de redes sociais por exemplo, para realizar um estudo sobre o alcance e propagação dessas notícias falsas.