

# Máquinas de Busca no Contexto da Pandemia COVID-19: um estudo de caso para aplicações Q&A

Equipe: Evolution

Membros:

Leila Fabiola Ferreira,

Mateus Cichelero da Silva

# Descrição e Objetivos

# Descrição e Objetivos

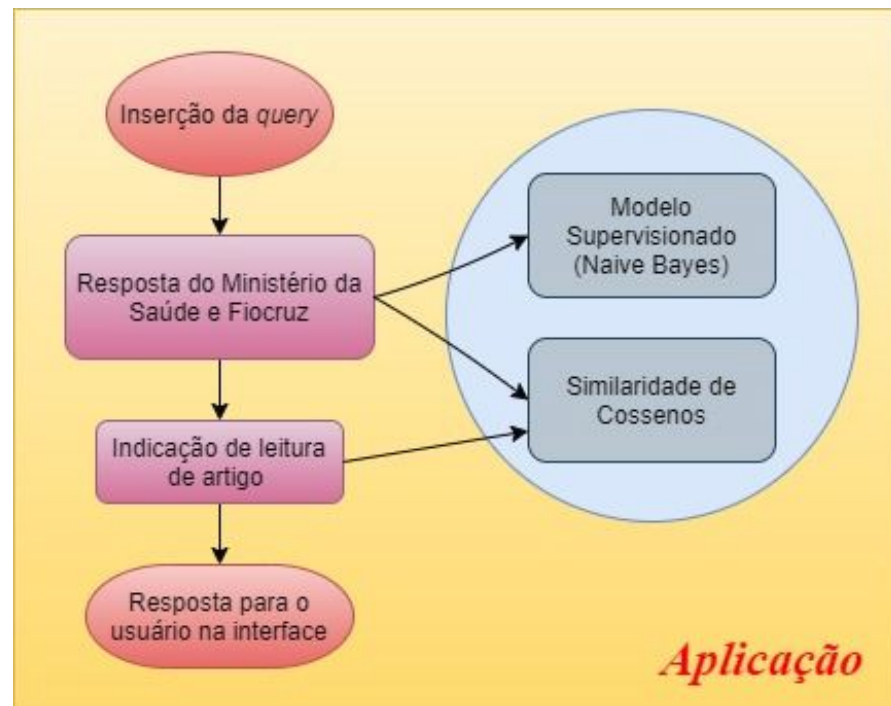
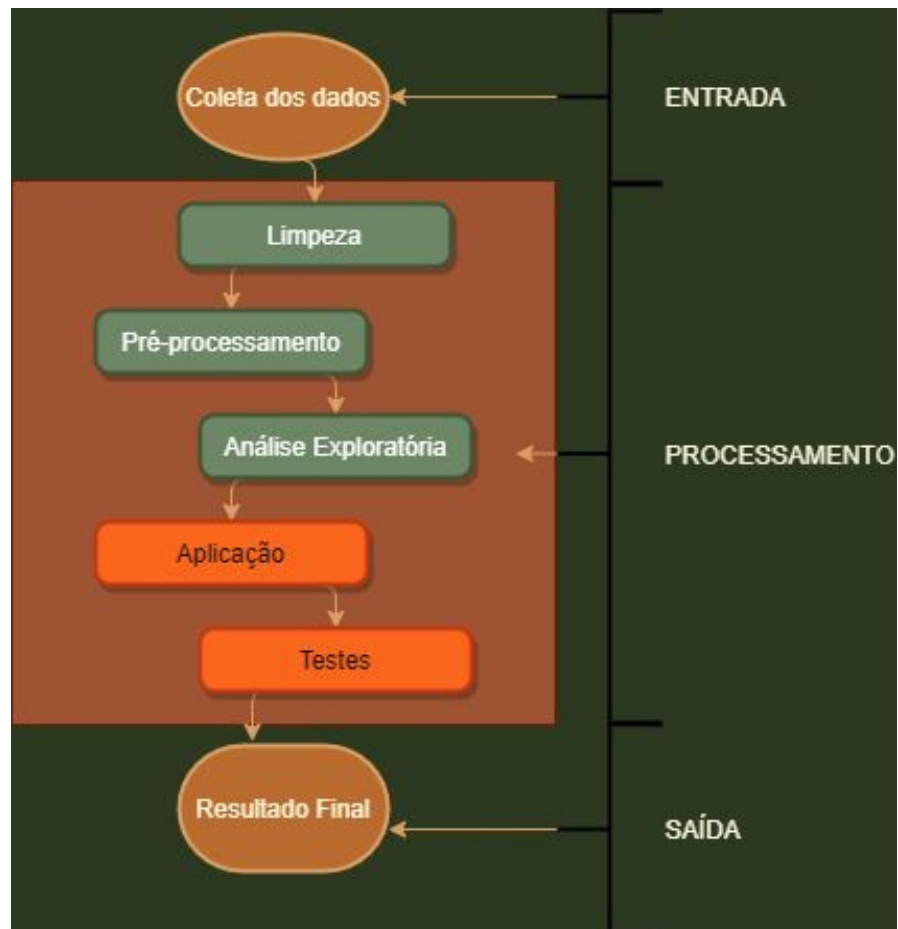
- Aplicação dos conhecimentos apresentados na disciplina (máquinas de busca, NLP, sistemas de recomendação) através de um estudo de caso no contexto da pandemia de COVID-19.
- Desenvolvimento de uma aplicação prova de conceito (PoC), que tomou como entrada questionamentos diversos sobre a pandemia e retorna tanto a resposta com maior relevância calculada pelo sistema como também uma sugestão de artigo científico sobre o tema.
- Por fim, por se tratar de um estudo de caso de aplicação de tecnologias, o artigo final resultante dará ênfase também nos conceitos teóricos que envolvem os processamentos realizados.

# Descrição e Objetivos

O usuário pode entrar com uma dúvida/ questionamento sobre a pandemia através do campo textual e a partir disso é apresentada:

- a resposta sugerida através da aplicação de um modelo supervisionado de classificação multiclases (Naive Bayes), treinado utilizando dados de respostas oficiais do Ministério da Saúde e Fundação Fio Cruz;
- a resposta sugerida de um modelo de recuperação de informação que usa como documentos as sentenças tokenizadas do mesmo Corpus de perguntas e respostas anterior e se baseia na distância de cossenos / TF-IDF;
- a sugestão de um artigo relacionado à questão, baseando-se na mesma estratégia do ponto anterior, mas utilizando como query a pergunta traduzida para o inglês e como documentos uma base de abstracts de mais de 160 mil artigos do CORD-19.

# Dados e Modelos



# Perguntas e Respostas

Publicado em 07/04/2021 23h35 | Atualizado em 04/05/2021 17h13

## ▼ Covid-19

### ▼ O que é o coronavírus?

Os coronavírus são uma grande família de vírus comuns em muitas espécies diferentes de animais, incluindo humanos. Raramente, os coronavírus que infectam animais podem infectar pessoas, como exemplo do MERS-CoV e SARS-CoV. Em dezembro de 2019, houve a transmissão de um novo coronavírus (SARS-CoV-2), o qual foi identificado em Wuhan, China, sendo em seguida disseminada e transmitida pessoa a pessoa.

- ▶ O que fazer se eu estiver doente?
- ▶ Como ocorreram as primeiras infecções humanas pelo novo coronavírus?
- ▶ Qual a diferença entre COVID-19 e o novo coronavírus?

pergunta	resposta	topico
O que é o coronavírus?	Os coronavírus são uma grande família de vírus...	definicao_coronavirus
O que fazer se eu estiver doente?	Se estiver doente, com sintomas compatíveis co...	acoes_doente
Qual a diferença entre COVID-19 e o novo coron...	COVID-19 é a doença infecciosa causada pelo no...	covid_coronavirus
Qual a probabilidade de eu pegar a COVID-19?	O Brasil vivencia em muitos municípios a trans...	infeccao
Quem está mais vulnerável à COVID-19?	Pessoas idosas e pessoas com condições médicas...	grupo_risco

3. Siga as orientações médicas.  
4. Utilize máscara o tempo todo.  
5. Se for preciso cozinhar, use máscara de proteção, cobrindo boca e nariz todo o tempo.  
6. Depois de usar o banheiro, nunca deixe de lavar as mãos com água e sabão e sempre limpe vaso mantendo a tampa fechada, pia e demais superfícies com álcool, água sanitária ou outro produto recomendado pela Anvisa para desinfecção do ambiente.  
7. Separe toalhas de banho, garfos, facas, colheres, copos e outros objetos apenas para seu uso.  
8. O lixo produzido precisa ser separado e descartado.  
9. Evite compartilhar sofás e cadeiras e realize limpeza e desinfecção frequente com água sanitária ou álcool 70% ou outro produto recomendado pela Anvisa.  
10. Mantenha a janela aberta para circulação de ar do ambiente usado para isolamento e a porta fechada, limpe a maçaneta frequentemente com álcool 70%, água sanitária, ou outro produto recomendado pela Anvisa.  
11. Caso o paciente não more sozinho, recomenda-se que os demais moradores da residência durmam em outro cômodo, seguindo também as seguintes recomendações:  
12. Mantenha a distância mínima de 1 metro entre a pessoa infectada e os demais moradores.  
13. Limpe os móveis da casa frequentemente com água sanitária, álcool 70% ou outro produto recomendado pela Anvisa.  
14. Se uma pessoa da casa tiver diagnóstico positivo, todos os moradores devem ficar em distanciamento conforme orientação médica.  
15.  
16. Em 31 de dezembro de 2019, a constituição da Organização Mundial de Saúde (OMS) foi informada sobre

title	abstract	url
A predictive model for disease progression in ...	A predictive model for Corona Virus Disease 2019	<a href="https://www.ncbi.nlm.nih.gov/pubmed/32430433/">https://www.ncbi.nlm.nih.gov/pubmed/32430433/</a> ;
Suggestions for changes in professional proced...	Abstracts The COVID-19 (COrona Virus Disease 2...	<a href="https://doi.org/10.7416/ai.2021.2434">https://doi.org/10.7416/ai.2021.2434</a> ; <a href="https://...">https://...</a>
The Implications of COVID-19 in Radiation Onco...	The corona virus disease of 2019 (covid-19) ha...	<a href="https://doi.org/10.3747/co.27.7095">https://doi.org/10.3747/co.27.7095</a> ; <a href="https://www...">https://www...</a>
Understanding the fate of corona virus transmi...	We propose a simple model for understanding th...	<a href="https://arxiv.org/pdf/2003.10530v1.pdf">https://arxiv.org/pdf/2003.10530v1.pdf</a>
A Computer Simulation Study on novel Corona Vi...	The World Health Organization (WHO) on March 1...	<a href="http://medrxiv.org/cgi/content/short/2020.05.1...">http://medrxiv.org/cgi/content/short/2020.05.1...</a>

# Trabalhos Relacionados



COMMENT

OPEN



# Chatbots in the **fight** against the COVID-19 pandemic

Adam S. Miner <sup>1,2</sup>✉, Liliana Laranjo <sup>3</sup> and A. Baki Kocaballi <sup>3,4</sup>

## A Qualitative Evaluation of Language Models on Automatic Question-Answering for COVID-19

David Oniani

Mayo Clinic

Kern Center for the Science of Health Care Delivery

Rochester, MN, USA

oniani.david@mayo.edu

Yanshan Wang

Mayo Clinic

Division of Digital Health Sciences

Rochester, MN, USA

wang.yanshan@mayo.edu

ARTICLE

OPEN



## COVID-19 information retrieval with deep-learning based semantic search, question answering, and abstractive summarization

Andre Esteva <sup>1,3</sup>✉, Anuprit Kale<sup>1,3</sup>, Romain Paulus<sup>1</sup>, Kazuma Hashimoto<sup>1</sup>, Wenpeng Yin<sup>1</sup>, Dragomir Radev <sup>1,2</sup> and Richard Socher<sup>1</sup>

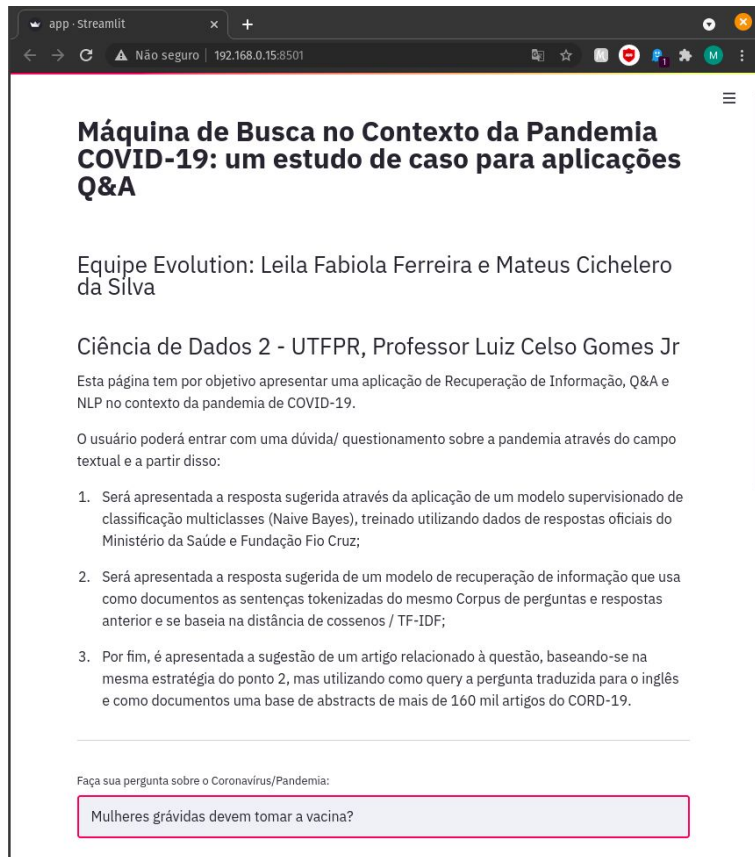
# Resultados

# Classificador Naive Bayes

accuracy 0.8482142857142857

	precision	recall	f1-score	support
definicao_coronavirus	0.60	1.00	0.75	3
acoes_doente	1.00	1.00	1.00	3
covid_coronavirus	1.00	1.00	1.00	2
infeccao	0.60	1.00	0.75	3
grupo_risco	0.60	1.00	0.75	3
virus_superficie	1.00	1.00	1.00	3
uso_mascara	1.00	1.00	1.00	3
seguranca_vacina	1.00	1.00	1.00	3
imunidade_vacina	1.00	1.00	1.00	2
vacina_mutacoes_virus	0.00	0.00	0.00	1
vacina_efeito_adverso	1.00	1.00	1.00	3
vacina_periodicidade	1.00	1.00	1.00	3
vacina_tempo_imunidade	0.00	0.00	0.00	2
vacinado_transmissao	1.00	1.00	1.00	3
quem_vacina_primeiro	1.00	1.00	1.00	3
vacina_quem_deve_tomar	0.00	0.00	0.00	2
vacina_adultos_crianças	0.00	0.00	0.00	2
vacina_infectados	1.00	1.00	1.00	3
vacina_obrigatoria	1.00	1.00	1.00	3
vacina_duas_doses	0.75	1.00	0.86	3
vacina_eficiencia	1.00	1.00	1.00	2
escolha_vacina	1.00	0.50	0.67	2
outras_vacinass_juntas	1.00	1.00	1.00	3
doses_diferentes_vacinass	1.00	1.00	1.00	2
vacinado_exterior	0.67	1.00	0.80	2
sem_carteira_sus	1.00	1.00	1.00	3
controle_vacinacao	1.00	1.00	1.00	2
capacidade_vacinacao_sus	1.00	1.00	1.00	3
comorbidade_atestado	1.00	0.67	0.80	3
comprar_vacina	1.00	1.00	1.00	2
uso_emergencial	1.00	1.00	1.00	3
vacina_gravida	1.00	1.00	1.00	2
tempo_imunidade	1.00	0.67	0.80	3
variantes_virus	0.50	0.67	0.57	3
identificacao_variantes	0.38	1.00	0.55	3
reinfeccao	1.00	0.67	0.80	3
origem_coronavirus	1.00	1.00	1.00	2
animais_coronavirus	0.60	1.00	0.75	3
sintomas	1.00	0.75	0.86	4
antibioticos	1.00	1.00	1.00	2
hidroxicloroquina	1.00	1.00	1.00	3
aumentar_imunidade	0.00	0.00	0.00	2
testes_coronavirus	0.00	0.00	0.00	2
micro avg	0.85	0.85	0.85	112
macro avg	0.78	0.81	0.78	112
weighted avg	0.82	0.85	0.82	112

# Aplicação Web App Streamlit <https://ptbr-duvidas-covid19.streamlit.app/>



app - Streamlit

Não seguro | 192.168.0.15:8501

## Máquina de Busca no Contexto da Pandemia COVID-19: um estudo de caso para aplicações Q&A

Equipe Evolution: Leila Fabiola Ferreira e Mateus Cichelero da Silva

Ciência de Dados 2 - UTFPR, Professor Luiz Celso Gomes Jr

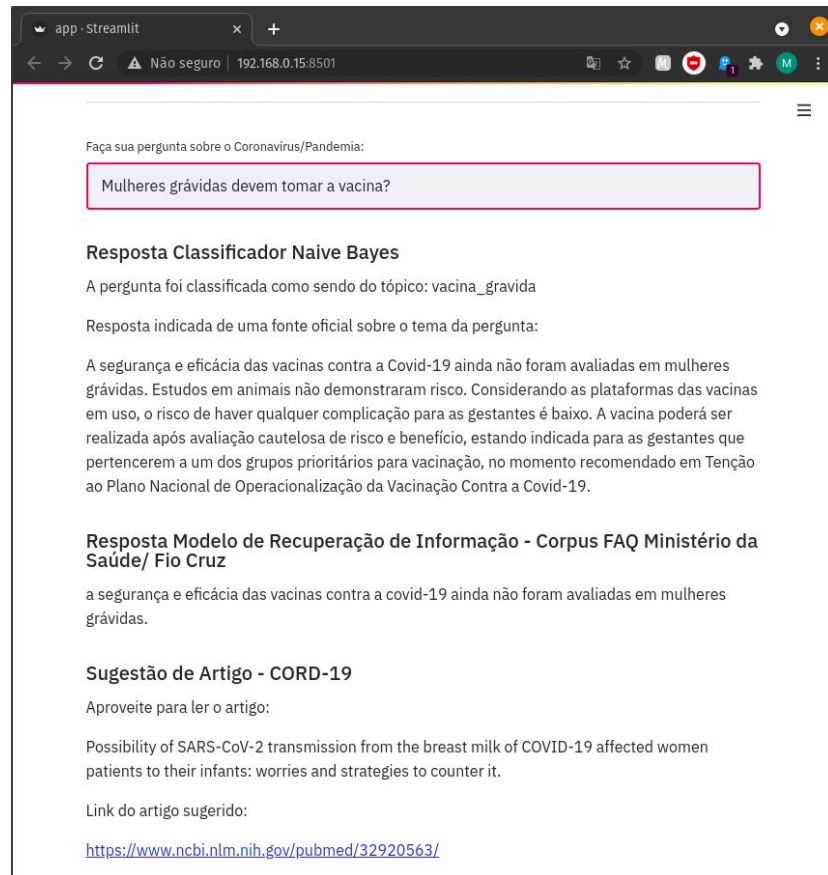
Esta página tem por objetivo apresentar uma aplicação de Recuperação de Informação, Q&A e NLP no contexto da pandemia de COVID-19.

O usuário poderá entrar com uma dúvida/ questionamento sobre a pandemia através do campo textual e a partir disso:

1. Será apresentada a resposta sugerida através da aplicação de um modelo supervisionado de classificação multiclases (Naive Bayes), treinado utilizando dados de respostas oficiais do Ministério da Saúde e Fundação Fio Cruz;
2. Será apresentada a resposta sugerida de um modelo de recuperação de informação que usa como documentos as sentenças tokenizadas do mesmo Corpus de perguntas e respostas anterior e se baseia na distância de cossenos / TF-IDF;
3. Por fim, é apresentada a sugestão de um artigo relacionado à questão, baseando-se na mesma estratégia do ponto 2, mas utilizando como query a pergunta traduzida para o inglês e como documentos uma base de abstracts de mais de 160 mil artigos do CORD-19.

Faça sua pergunta sobre o Coronavírus/Pandemia:

Mulheres grávidas devem tomar a vacina?



app - Streamlit

Não seguro | 192.168.0.15:8501

Faça sua pergunta sobre o Coronavírus/Pandemia:

Mulheres grávidas devem tomar a vacina?

### Resposta Classificador Naive Bayes

A pergunta foi classificada como sendo do tópico: vacina\_gravida

Resposta indicada de uma fonte oficial sobre o tema da pergunta:

A segurança e eficácia das vacinas contra a Covid-19 ainda não foram avaliadas em mulheres grávidas. Estudos em animais não demonstraram risco. Considerando as plataformas das vacinas em uso, o risco de haver qualquer complicação para as gestantes é baixo. A vacina poderá ser realizada após avaliação cautelosa de risco e benefício, estando indicada para as gestantes que pertencerem a um dos grupos prioritários para vacinação, no momento recomendado em Tensão ao Plano Nacional de Operacionalização da Vacinação Contra a Covid-19.

### Resposta Modelo de Recuperação de Informação - Corpus FAQ Ministério da Saúde/ Fio Cruz

a segurança e eficácia das vacinas contra a covid-19 ainda não foram avaliadas em mulheres grávidas.

### Sugestão de Artigo - CORD-19

Aproveite para ler o artigo:

Possibility of SARS-CoV-2 transmission from the breast milk of COVID-19 affected women patients to their infants: worries and strategies to counter it.

Link do artigo sugerido:

<https://www.ncbi.nlm.nih.gov/pubmed/32920563/>



Limitações/trabalhos futuros

- Apesar das boas métricas encontradas para o classificador, deve-se considerar que são ainda poucos exemplos de treinamento para cada tópico. O ideal seria continuar expandindo a base rotulada, talvez aplicando algum método mais automatizado, como bootstrapping a partir do que já criamos manualmente;
- Otimizações dos modelos;
- Trazer pontos como explicabilidade das classificações realizadas;
- Aplicar diferentes métodos para avaliação dos modelos de recuperação de informação;
- Trabalhar na otimização da aplicação dos modelos serializados. Os artefatos são “pesados” e tornam a aplicação externalizada lenta, principalmente na recuperação dos artigos sugeridos, de base maior;

# Obrigado a todos!

