



## Orientação para novos alunos – CoDatMo Brasil

19 de abril de 2021

## Sumário

<b>1</b>	<b>Sobre o Projeto:</b>	<b>2</b>
1.1	O que é o CoDatMo . . . . .	2
1.2	Objetivos do Projeto . . . . .	3
<b>2</b>	<b>Dados</b>	<b>3</b>
2.1	O que temos até agora? . . . . .	3
2.2	Exemplos de tweets rotulados . . . . .	4
2.3	Acesso aos dados . . . . .	5

## 1 Sobre o Projeto:

### 1.1 O que é o CoDatMo

O CoDatMo ou the (Co)vid (Dat)a (Mo)deling project é um projeto global de monitoramento de sintomas de COVID-19 (COVID-19 Symptom Watcher) da Universidade de Liverpool que consiste em monitorar menções em redes sociais sobre sintomas de COVID-19, rotulá-los e inseri-los num modelo estatístico. O modelo estatístico proposto usa inferência Bayesiana (pelo software Stan) usando dados dos tweets juntamente com outras fontes de dados gerando intervalos de credibilidade para parâmetros de interesse importantes no monitoramento da pandemia de COVID-19, notoriamente o número de mortes e  $R_0$  (o número reprodutivo—um indicador de quão contagiosa uma doença é, ou quão facilmente ela se espalha de pessoa para pessoa em um ambiente). Este projeto tem como objetivo catalogar e rotular os tweets brasileiros juntamente com as fontes de dados brasileiras de COVID-19 (Ministério de Saúde e Consórcio dos Veículos de Imprensa) e desenvolver, em conjunto com a Universidade de Liverpool, o modelo estatístico para o Brasil. Os resultados do projeto poderão ser utilizados no apoio à tomada de decisão e formulação de políticas públicas para o combate à pandemia COVID-19 potencialmente salvando vidas e reduzindo o número de mortes e complicações sérias.

## 1.2 Objetivos do Projeto

O objetivo do projeto é desenvolver um modelo probabilístico com estatística Bayesiana, com base em dados rotulados e classificados de menções realizadas no Twitter sobre sintomas associados a COVID-19, de forma ser possível identificar potenciais pessoas infectadas com o novo coronavírus. As estimativas do modelo podem ser utilizadas no apoio à tomada de decisão e formulação de políticas públicas para o combate à pandemia do novo coronavírus e a outras crises sanitárias. Potencialmente evitando o colapso do sistema de saúde, identificando novos sintomas associados à doença, casos leves até complicações que levem risco de morte.

## 2 Dados

### 2.1 O que temos até agora?

No momento os alunos de IC (Iniciação Científica) da Uninove estão rotulando tweets de acordo com as seguintes critérios:

0. Tweet não pertinente (não fala de COVID)
1. Usuário tem sintomas no momento atual
2. Usuário teve sintomas no passado
3. Alguma outra pessoa (não o usuário) tem sintomas no momento atual
4. Alguma outra pessoa (não o usuário) tem sintomas no passado
5. Teoria da Conspiração
6. Alguma outra pessoa (não o usuário) morreu (relacionado à morte)

Os alunos que estão atualmente no projeto fizeram a rotulação de 1000 tweets usados para o primeiro teste do modelo. mais abaixo mostraremos um exemplo de cada tweet rotulado.

## 2.2 Exemplos de tweets rotulados

## 2.2 Exemplos de tweets rotulados

Abaixo temos uma imagem com um exemplo de cada tweet rotulado.

Dia/Hora da Criação	Texto	Usuário	Local	Local do Usuário	País	Coordenadas	Indefinido	Indefinido	Indefinido	Label
2020-06-21T16:16:26Z	eu to febril e com dor de cabeça <a href="https://t.co/m0nggEs1vY">https://t.co/m0nggEs1vY</a>	1,156E+18	NA	ela/dela	NA	NA	NA	1	0	1
2020-08-11T23:45:46Z	agora to achando que a gripe que eu tive depois do carnaval era covid	3,39E+09	NA	Volta Redonda, Brasil	NA	NA	NA	0	0	2
2020-10-24T06:21:02Z	Nem preciso explicar o retrocesso que nos tomamos.	75380145	NA	Isac Mota	NA	NA	NA	0	0	3
2021-01-11T15:54:55Z	o uber tossindo e eu e minha mae olhando uma pra outra em pânico mt bom	3,043E+09	NA	udia • A	NA	NA	NA	0	0	4
2020-10-25T14:59:47Z	@VitorioEvandro @jaibolsonaro É até óbvio isso. Mas todos os indícios apontam que o vírus foi mesmo fabricado. Uma bomba biológica. E nós provavelmente não precisaremos de vacina. A evolução e a cura dos infectados superam 95 por cento. E a taxa de letalidade já está igual a de uma gripe comum.	1,525E+09	NA	NA	NA	NA	NA	0	0	5

No caso de um tweet possuir dois rótulos, a rotulação será separada por “,” ex.: 1,2 no campo label. e no caso de um tweet vir em um idioma que não o português do Brasil, o tweet receberá a label “99”.

## **2.3 Acesso aos dados**

Nossos dados estão em uma pasta do google drive e no GitHub da CoDatMo. obs.: Para obter acesso, solicitar ao professor José Eduardo Storopoli.