Plan Overview

A Data Management Plan created using DMP Tool

Title: COVID Sympton Watcher Brazil

Creator: Andre Santos

Affiliation: Universidade Nove de Julho (UNINOVE)

Funder: São Paulo Research Foundation (fapesp.br)

Template: Digital Curation Centre (português)

Project abstract:

Introdução: A pandemia de COVID-19 devastou a saúde pública no mundo, por isso é importante identificar maneiras práticas de alfabetização em saúde, arenas políticas, mídias tradicionais e novas mídias, como as mídias sociais. O ponto forte das mídias sociais para o controle e monitoramento de surtos é a troca de informações geradas pelos usuários. Contudo, apesar das tentativas de monitoramento e predição da pandemia de COVID-19 no Brasil não há, na literatura, estudos que incorporem dados de mídias sociais em modelos epidemiológicos de COVID-19. Para preencher essa lacuna, propomos um modelo *Susceptible-Infectious-Recovered-Terminally-III-Deceased*(SIRTD) com dados do Sistema Único de Saúde (SUS) e menções de sintomas de COVID-19 no Twitter. Com isso, pretende-se responder à seguinte pergunta de pesquisa: como as mídias sociais podem ser usadas no apoio às estratégias de comunicação em saúde eficazes no combate à pandemias e outras crises sanitárias? **Objetivo**: propor um modelo de comunicação em saúde e desenvolver aplicações de incorporação de dados de mídias sociais em modelos epidemiológicos Bayesianos para a COVID-19. **Métodos**: este projeto possui um paradigma de ciência de dados no qual serão utilizados dados extraídos do Twitter e dos sistemas de informação em saúde do SUS desde Janeiro/2019 até Dezembro/2021. Serão utilizadas técnicas de: (1) processamento de linguagem natural; (2) aprendizagem de máquina; (3) análise de redes sociais; e (4) modelos epidemiológicos Bayesianos.

Start date: 08-01-2021

End date: 06-20-2023

Last modified: 07-08-2024

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of

the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

COVID Sympton Watcher Brazil

Coleta de Dados

Que dados serão coletados ou criados?

Serão coletados dados do Twitter e dados da vigilância de síndrome respiratória aguda grave (SRAG) no Brasil.

As bases de dados SRAG disponibilizadas pelo Sistema de Informação de Saúde do SUS, passam por tratamento que envolve a anonimização, em cumprimento a legislação vigente. A vigilância de SRAG é realizada em todos os hospitais do país que possuem capacidade de assistência aos casos de SRAG, de rede pública ou privada. Esses hospitais estão aptos para notificar os casos de SRAG e/ou óbitos por SRAG, coletar amostras clínicas, de maneira universal, seguindo fluxos estabelecidos à vigilância de síndromes respiratórias agudas e, agora, **incluindo a vigilância dos casos e óbitos de SRAG suspeitos para a COVID-19** (fonte: openDataSUS).

DEFINIÇÕES OPERACIONAIS DO PLANO DE VIGILÂNCIA DE SRAG

[CASOS SUSPEITOS]

DEFINIÇÃO 1: SÍNDROME GRIPAL (SG)
Indivíduo com quadro respiratório agudo, caracterizado por pelo menos dois (2) dos seguintes sinais e sintomas: febre (mesmo que referida), calafrios, dor de garganta, dor de cabeça, tosse, coriza, distúrbios olfativos ou distúrbios gustativos.
Observações:

- **Em crianças**: além dos itens anteriores considera-se também obstrução nasal, na ausência de outro diagnóstico específico.
- **Em idosos**: deve-se considerar também critérios específicos de agravamento como sincope, confusão mental, sonolência excessiva, irritabilidade e inapetência.
- Na suspeita de COVID-19, a febre pode estar ausente e sintomas gastrointestinais (diarreia) podem estar presentes.

DEFINIÇÃO 2: SÍNDROME RESPIRATÓRIA AGUDA GRAVE (SRAG)
Indivíduo com SG que apresente: dispneia/desconforto respiratório OU pressão ou dor persistente no tórax OU saturação de O2 menor que 95% em ar ambiente OU coloração azulada (cianose) dos lábios ou rosto.
Observações:

- **Em crianças**: além dos itens anteriores, observar os batimentos de asa de nariz, cianose, tiragem intercostal, desidratação e inapetência;
- Para efeito de notificação no Sivep-Gripe, devem ser considerados os casos de SRAG hospitalizados ou os óbitos por SRAG independente de hospitalização.

[CASOS CONFIRMADOS DE COVID-19]

POR CRITÉRIO CLÍNICO
Caso de SG ou SRAG associado a anosmia (disfunção olfativa) OU ageusia (disfunção gustatória) aguda sem outra causa pregressa.
POR CRITÉRIO CLÍNICO-EPIDEMIOLÓGICO
Caso de SG ou SRAG com histórico de contato próximo ou domiciliar, nos 14 dias anteriores ao aparecimento dos sinais e sintomas com caso confirmado para COVID-19

Os dados do **Twitter** serão menções feitas na rede sobre 56 palavras-chave referentes à sintomas de COVID-19 em Língua Portuguesa. São elas:

1. Adinamia;
2. Ageusia;
3. Anosmia;
4. Boca azulada;
5. Calafrio;
6. Cansaço;
7. Cefaleia;
8. Cianose;
9. Coloração azulada no rosto;
10. Congestão nasal;
11. Conjuntivite;
12. Coriza;
13. Desconforto respiratório;
14. Diarreia;
15. Dificuldade para respirar;
16. Diminuição do apetite;
17. Dispneia;
18. Distúrbio gustativo;
19. Distúrbio olfativo;
20. Dor abdominal;
21. Dor de cabeça;
22. Dor de garganta;
23. Dor no corpo;
24. Dor no peito;
25. Dor persistente no tórax;
26. Erupção cutânea na pele;
27. Fadiga;
28. Falta de ar;
29. Febre
30. Gripe;
31. Hiporexia;
32. Inapetência;
33. Infecção respiratória;
34. Lábio azulado;
35. Mialgia;
36. Nariz entupido;
37. Náusea;
38. Obstrução nasal;
39. Perda de apetite;
40. Perda do olfato;
41. Perda do paladar;
42. Pneumonia;
43. Pressão no peito;
44. Pressão no tórax;
45. Prostração;
46. Quadro gripal;
47. Quadro respiratório;
48. Queda da saturação;
49. Resfriado;
50. Rosto azulado;
51. Saturação baixa;
52. Saturação de O2 menor que 95%;

53. Síndrome respiratória aguda grave;
54. SRAG;
55. Tosse;
56. Vômito.

Essas palavras-chaves foram selecionados com base em fontes oficiais (Ministério da Saúde e WOS) e orientações de epidemiologistas sobre os termos relacionados à COVID-19.

Como os dados serão coletados ou criados?

Os dados serão obtidos direto nos sites Twitter e openDataSUS por meio de interfaces de programação de aplicações,
do inglês Application Programming Interface - API, com o auxílio da linguagem de programação Python.

Documentação e Metadados

Que documentação e metadados acompanharão os dados?

- Dicionário de Dados SRAG.

Ética e Conformidade Legal

Como você administrará qualquer questão ética?

Os dados coletados nesta pesquisa estão sobre a licença *Creative Commons*:

- Os dados são tecnicamente abertos, o que significa que serão publicados em formatos eletrônicos legíveis por máquina e não proprietários, para que qualquer pessoa possa acessar e usar os dados usando ferramentas de software comuns e disponíveis gratuitamente.
- O uso e tratamento desses dados estão sob a Lei Geral de Proteção de Dados Pessoais (Lei n. 13.709, de 14 de agosto de 2018) da República Federativa do Brasil.

Como você vai gerenciar os direitos autorais e os direitos de propriedade intelectual (IP / IPR)?

Todos direitos são reservados ao LabCidades UNINOVE:

- Laboratório de Cidades Inteligentes e Gestão do Conhecimento da Universidade Nove de Julho - UNINOVE.

Armazenamento e Backup

Como os dados serão armazenados e terão backup durante a pesquisa?

Os dados serão armazenados e estarão disponíveis para consultas públicas nos sites:

- Github
- Zotero

Como você vai gerenciar o acesso e a segurança?

Os acessos e segurança da informação são gerenciados pelos próprios sites onde estão armazenados os dados e relatórios são frequentemente disponibilizados para os gestores do projeto de forma automática pelos sistemas.

Seleção e Preservação

Quais dados são de valor a longo prazo e devem ser mantidos, compartilhados e / ou preservados?

Todos dados são de valor a longo prazo, assim podem ser mantidos e compartilhados em outras pesquisas.

Qual é o plano de preservação a longo prazo do conjunto de dados?

Ao longo deste projeto todos conjuntos de dados serão mensalmente atualizados.

Compartilhamento de Dados

Como você vai compartilhar os dados?

Os dados serão compartilhados em repositórios de dados para ciência aberta:

- Github
- Zotero

Existem restrições ao compartilhamento de dados requeridos?

Para compartilhamento e utilização dos dados siga o que estabelece a licença Creative Commons e a Lei Geral de Proteção de Dados.

Responsabilidades e Recursos

Quem será responsável pelo gerenciamento de dados?

O gerenciamento dos conjuntos de dados está sob a responsabilidade de André Luis Marques Ferreira dos Santos:

- Lattes
- andrelmfsantos@gmail.com

Quais recursos você precisará para entregar seu plano?

Serão necessários os seguintes recursos para entregar o plano de dados:

- Computador pessoal.
- Linguagem de programação científica Python.
- Acesso à Internet.
- Conta de usuário nos sites Github e Zotero.

Planned Research Outputs

Data paper - "Brazilian Portuguese COVID-19 Tweets"

Artigo sobre expressões relacionadas à sintomas sobre COVID-19 em Língua Portuguesa (Brasil) no Twitter.

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Brazilian Portuguese COVID-19 Tweets	Data paper	Unspecified	Open	None specified	4 GB	Creative Commons Attribution 4.0 International	None specified	No	No