

MAE5904 - Projeto: Etapa 1

Nome do grupo: Grupo 7

Membros do grupo: Lucas Bortolucci, Bruno de Castro Paul Schultze, Thiago Henrique Martinoni, Luiz Afonso Glatzl Junior, Lucas Monteiro Bianchi

Descrição do problema: Durante a pandemia de Covid-19, ficou-se evidente a importância da testagem em massa, entretanto, muitos testes necessitam de equipamentos sofisticados e de execução especializada. Diante disso, os testes rápidos ganharam certa notoriedade e vem sendo largamente utilizados. Esses testes são geralmente sorológicos, capazes de identificar os anticorpos produzidos quando em contato com o antígeno. A metodologia mais utilizada por esse tipo de teste é o imunocromatografia, no qual, se consiste em produzir uma coloração através da reação entre o antígeno e o anticorpo, sendo capazes de detectar se o organismo esta produzindo anticorpos (IgG e IgM). Embora as vantagens desse tipo de teste são claras, os testes rápidos possuem baixa sensibilidade. Podemos definir sensibilidade como a probabilidade do teste resultar em positivo quando o indivíduo de fato tem a doença. Em outras palavras, um teste com baixa sensibilidade pode resultar em um elevado número de falsos negativos, gerando estatísticas que não expressam a realidade e consequentemente dificultam o processo de tomada de decisões.

Descrição do conjunto de dados: Recentemente (2020), a vigilância da COVID-19, a infecção humana causada pelo novo Coronavírus, que vem causando uma pandemia, foi incorporada na rede de vigilância da Influenza e outros vírus respiratórios. Os dados são provenientes do site openDataSUS. Esta página tem como finalidade disponibilizar o legado dos banco de dados (BD) epidemiológicos de SRAG, da rede de vigilância da Influenza e outros vírus respiratórios, desde o início da sua implantação (2009) até os dias atuais (2020), com a incorporação da vigilância da COVID-19.

A população definida para esse estudo são pessoas residentes do estado de São Paulo cujo foram diagnosticadas com SRAG-Covid 19. Segundo o Ministério da Saúde, temos as seguintes definições:

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 síncope, confusão mental, sonolência excessiva, irritabilidade e inapetência.
- Na suspeita de COVID-19, a febre pode estar ausente e sintomas gastrointestinais (diarreia) podem estar presentes.

Síndrome Respiratória Aguda Grave (SRAG): Indivíduo com SG que apresente: dispneia/desconforto respiratório OU pressão persistente no tórax OU saturação de O_2 menor que 95% em ar ambiente OU coloração azulada dos lábios ou rosto.

Além disso, um caso de covid-19 pode ser confirmado por:

- Por critério clínico: Caso de SG ou SRAG com confirmação clínica 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.

- Por critério clínico-imagem: Caso de SG ou SRAG ou óbito por SRAG que não foi possível confirmar por critério laboratorial E que apresente pelo menos uma (1) das seguintes alterações tomográficas: opacidade em vidro fosco (periférico ou multifocal) e sinal de halo reverso.

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.

As variáveis disponíveis nesse banco de dados cujo o qual predente-se utilizar são:

- Resultado: Variável binária que representa o resultado do teste, podendo ser classificada em positivo se o teste do indivíduo deu positivo para COVID-19 ou negativo caso contrário.
- Sintomas: Variável representando os sintomas que o indivíduo apresentava separado por vírgulas, pretendemos separar em variáveis binárias para cada sintoma futuramente.
- Profissional da saúde: Variável binária com o valor Sim, se o indivíduo é um profissional da área da saúde e não caso contrário
- Condições: Variável representando condições de saúde que o indivíduo apresenta que podem causar complicações ao caso, separado por vírgulas, pretendemos separar em variáveis binárias para cada sintoma futuramente.
- Sexo: Variável binária com os valores feminino ou masculino.
- Idade: variável numérica representando a idade do indivíduo.
- Evolução: Variável categórica representando a evolução do indivíduo em relação ao seu caso de Síndrome Gripal, com as seguintes classificações: cancelado, ignorado, óbito, cura, internado, internado em UTI, em tratamento domiciliar.

Avaliação do problema: Este projeto pode ser compreendido como de classificação e supervisionado. De classificação, pois o objetivo é classificar os indivíduos em positivo ou negativo para covid-19. Isso será feito utilizando os indivíduos que foram submetidos ao teste padrão ouro, RT-PCR para treinar o modelo. Uma vez treinado e validado o modelo, será feita a sua testagem com os indivíduos que se submeteram ao teste rápido, ou seja, dados desconhecido para o modelo. Como o conjunto de dados traz o resultado dos testes, ou seja, compararemos a saída atual do modelo com o valor observado, buscando uma função que minimize a perda esperada, caracterizando assim como aprendizado supervisionado. Entende-se que em cenários onde o resultado esperado é diferente do resultado obtido pelo modelo como erro. Entretanto, o objetivo é estimar o percentual de falsos negativos, ou seja, casos onde o teste rápido foi negativo, quando classificados pelo modelo como positivo, entendendo isso não como erro de classificação do modelo, mas dos testes rápidos devido à baixa sensibilidade.

Possíveis limitações: Classificar indivíduos como positivos ou negativos para uma determinada doença considerando um limitado conjunto de informações originários de uma fonte secundária, bem como, a não disponibilidade de informações como carga viral, intensidade dos sintomas e qual anticorpo foi observado (IgM ou IgG) podem limitar a capacidade de classificar adequadamente os indivíduos.

Endereço: <https://opendatasus.saude.gov.br/dataset/bd-srag-2020>