



UNIVERSIDADE FEDERAL DE SANTA CATARINA - UFSC
CENTRO DE FILOSOFIA E CIÊNCIAS HUMANAS - CFH
PROGRAMA DE PÓS-GRADUAÇÃO EM DESASTRES NATURAIS - PPGDN

MICHEL PIRES DE ARAUJO

**CLASSIFICAÇÃO DA GRAVIDADE DE SÍNDROMES RESPIRATÓRIAS
BASEADA EM *MACHINE LEARNING* E O CASO DO APLICATIVO TRIAGEM
ON-LINE DA COVID-19**

Florianópolis, SC

2023

MICHEL PIRES DE ARAUJO

Classificação da Gravidade de Síndromes Respiratórias Baseada em *Machine Learning* e o Caso do Aplicativo Triagem On-Line da COVID-19

Dissertação submetida ao Programa de Pós Graduação em Desastres Naturais da Universidade Federal de Santa Catarina, como requisito parcial para a obtenção do título de Mestre em Desastres Naturais.

Orientador: Prof. Dr. Renato Fileto

Florianópolis, SC

2023

Araujo, Michel Pires de
Classificação da Gravidade de Síndromes Respiratórias
Baseada em Machine Learning e o Caso do Aplicativo Triagem
On-Line da COVID-19 / Michel Pires de Araujo ; orientador,
Renato Fileto, 2023.
103 p.

Dissertação (mestrado profissional) - Universidade Federal
de Santa Catarina, Centro de Filosofia e Ciências Humanas,
Programa de Pós-Graduação em Desastres Naturais,
Florianópolis, 2023.

Inclui referências.

1. Desastres Naturais. 2. Machine Learning. 3.
Síndromes Respiratórias. 4. Triagem On-Line da COVID-19.
5. Classificação de Gravidade. I. Fileto, Renato. II.
Universidade Federal de Santa Catarina. Programa de Pós-
Graduação em Desastres Naturais. III. Título.

MICHEL PIRES DE ARAUJO**Título: Classificação da Gravidade de Síndromes Respiratórias Baseada em *Machine Learning* e o Caso do Aplicativo Triagem On-Line da COVID-19**

O presente trabalho em nível de Mestrado foi avaliado e aprovado, em 13 de dezembro de 2023, pela banca examinadora composta pelos seguintes membros:

Prof. Renato Fileto, Dr.

Universidade Federal de Santa Catarina

Prof^a Adriana Marques Rossetto, Dra.

Universidade Federal de Santa Catarina

Prof. Jônata Tyska Carvalho, Dr.

Universidade Federal de Santa Catarina

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado adequado para obtenção do título de Mestre em Desastres Naturais.

Insira neste espaço a
assinatura digital

Coordenação do Programa de Pós-Graduação

Insira neste espaço a
assinatura digital

Prof. Renato Fileto, Dr.

Orientador

Florianópolis, 2023.

À Renata, pelo amor, pelo apoio e pela força
que me ofereceu ao longo desta jornada.

AGRADECIMENTOS

Agradeço a Deus pelo dom da vida e pela oportunidade de compartilhar e aprimorar conhecimento. Aos meus pais, por, de alguma forma, terem contribuído com quem sou hoje. Ao meu orientador, que não raras vezes foi escuta ativa, não medindo esforços para que o resultado do trabalho fosse condizente com o tamanho do meu propósito. Aos professores: Dra. Adriana Marques Rossetto e Dr. Jônata Tyska Carvalho pelos apontamentos, os quais desempenharam um papel crucial no aprimoramento deste trabalho. Aos meus amigos: Cristiano Francis Ferreira Costa, Frederico Steinmetz Alvarez e Jason Schreiner dos Santos pelas considerações que foram inspiradoras. Ao amor de minha vida, minha esposa, Renata Biana, por ser um convite diário ao melhoramento de mim. Pelas revisões, pelas pesquisas, pelas discussões filosóficas, pelo incentivo, pelo alimento, pelas leituras empolgadas e empolgantes. Por ser um presente de Jesus em minha existência.

RESUMO

A necessidade de resposta rápida ao aumento da demanda de pacientes contaminados por agentes patológicos em situações de pandemias deixa os serviços de saúde em alerta, uma vez que existe o risco de ocorrer um colapso nos atendimentos de urgência. Pandemias exigem uma ação ágil no que tange à identificação dos casos suspeitos, de modo que haja otimização dos recursos disponíveis. Nesse cenário imprevisível, o uso de aplicações baseadas em Inteligência Artificial pode ser eficaz na previsão de condições médicas em eventos adversos e pode ajudar profissionais da saúde na triagem de pacientes. Ressalta-se ainda que sistemas que possibilitem a realização desse processo de forma *on-line* podem auxiliar sobremaneira, pois evitam colocar pessoas não contaminadas em risco. Buscando automatizar a identificação dos casos graves de COVID-19, este estudo analisou as características clínicas de pacientes suspeitos da doença, ou outra síndrome respiratória, e explorou a aplicação de modelos de *Machine Learning* para prever diferentes gravidades dos casos suspeitos, utilizando uma base de dados anonimizada contendo 375.180 casos clínicos de pacientes atendidos pela Secretaria de Estado da Saúde de Santa Catarina no ano de 2020. Os dados foram divididos entre casos leves e graves e suas características subsidiaram o treinamento de classificadores baseados em *Machine Learning*, tais quais, *logistic regression*, *random trees* e SVM. Com base no resultado das análises dos modelos sobre suas efetividades (sensibilidade, acurácia, etc.), realizou-se a comparação entre o aplicativo desenvolvido e utilizado em Santa Catarina para realizar, remotamente, a triagem de casos suspeitos de COVID-19, durante o ano de 2020. Como resultado desse estudo restou comprovada a efetividade tanto dos parâmetros propostos por especialistas da área médica para o aplicativo em estudo. Os experimentos mostraram a efetividade/desempenho ligeiramente superior dos modelos de regressão logística, árvore de decisão e SVM, que foram treinados com grandes volumes de dados históricos, mas o desempenho do aplicativo se mostrou competitivo, e até superior, no que se refere à sensibilidade.

ABSTRACT

The need to respond quickly to the increased demand from patients contaminated by pathological agents in pandemic situations put health services on alert, as there is a risk of collapse in emergency care. Pandemics require agile action regarding the identification of suspected cases, so that available resources can be optimized. In this unpredictable scenario, the use of applications based on Artificial Intelligence can be effective in predicting medical conditions during adverse events and can help healthcare professionals screening patients. Systems that allow this process to be carried out online can greatly help, as they avoid putting uninfected people at risk. Seeking to automate the identification of serious cases of the disease, this study analyzed the clinical characteristics of patients suspected of COVID-19 or another respiratory syndrome and explored the application of Machine Learning models to predict different severities of suspected cases, using an anonymized database containing 375,180 clinical cases of patients treated by the State Department of Health of Santa Catarina in 2020. The data were divided between mild and severe cases and their characteristics supported the training of classifiers based on Machine Learning, such as logistic regression, random trees and SVM. Based on the results of the analysis of the models on their effectiveness (sensitivity, accuracy, etc.), a comparison was made between the application developed and used in Santa Catarina to carry out, remotely, the screening of suspected cases of COVID-19, during the year 2020. As a result of this study, the effectiveness of both the parameters proposed by medical experts for the application under study was proven. The experiments showed the slightly superior effectiveness/performance of the logistic regression, decision tree and SVM models, which were trained with large volumes of historical data, but the application's performance proved to be competitive, and even superior, in terms of sensitivity.

LISTA DE FIGURAS

- Figura 1 - Etapas do treinamento dos modelos.
- Figura 2 - Classificação dos casos.
- Figura 3 - Proporção das triagens realizadas.
- Figura 4 - Quantitativo de triagens ao longo dos dias.
- Figura 5 - Frequência dos sinais/sintomas nos casos suspeitos.
- Figura 6 - Frequência das comorbidades/fatores de risco nos casos suspeitos.
- Figura 7 - Distribuição dos casos suspeitos por faixa etária.
- Figura 8 - Casos moderados ou graves por município.
- Figura 9 - Gráfico da função sigmoide.
- Figura 10 - Dados desbalanceados.
- Figura 11 - Proposta de etapas para treinamento.
- Figura 12 - K-fold cross validation.
- Figura 13 - Frequência de sinais e sintomas em pacientes que tiveram de ser internados.
- Figura 14 - Frequência dos fatores de risco entre pacientes que tiveram de ser internados.
- Figura 15 - Distribuição por faixa etária dos casos graves.
- Figura 16 - Distribuição por faixa etária dos casos graves.
- Figura 17 - Métricas de desempenho.

LISTA DE QUADROS

- Quadro 1 - Pesos atribuídos para cada sinal ou sintoma.
- Quadro 2 - Matriz de confusão para classificação binária.
- Quadro 3 - Resultado da revisão bibliográfica sobre COVID-19.
- Quadro 4 - Quadro comparativo preliminar.
- Quadro 5 - Resultado consulta definitiva.
- Quadro 6 - Quadro comparativo definitivo.
- Quadro 7 - Análise qualitativa dos artigos selecionados.
- Quadro 8 - Base de dados encontradas.
- Quadro 9 - Parâmetros extraídos do dataset para o estudo.
- Quadro 10 - Métricas do Logistic Regression.
- Quadro 11 - Pesos atribuídos pelo modelo *Logistic Regression*.
- Quadro 12 - Métricas Decision Tree.
- Quadro 13 - Métricas Support Vector Machine.
- Quadro 14 - Métricas Aplicativo Triagem On-Line.
- Quadro 15 - Tabela Comparativa Final.
- Quadro 16 - Métricas do Logistic Regression.
- Quadro 17 - Métricas Decision Tree.
- Quadro 18 - Métricas Support Vector Machine.
- Quadro 19 - Comparativo de desempenho usando somente características (features) de sinais e sintomas (I) versus todas as características disponíveis, incluindo perfil de pacientes (II).
- Quadro 20 - Métricas do Logistic Regression.
- Quadro 21 - Métricas Decision Tree.
- Quadro 22 - Métricas Support Vector Machine.
- Quadro 23 - Comparativo de desempenho: (I) somente sinais e sintomas; (II) todas as características disponíveis, incluindo perfil de pacientes e (III) todas as características, mas somente registros de casos coletados até 2 de abril de 2020.
- Quadro 24 - Comparativo entre os resultados do Experimento III (desempenho dos modelos de aprendizado de máquina treinados com dados disponíveis no início da pandemia) e o desempenho do aplicativo Triagem On-Line.

LISTA DE ABREVIATURAS E SIGLAS

COBRADE - Classificação e Codificação Brasileira de Desastres.

CBMSC - Corpo de Bombeiros Militar de Santa Catarina.

DPOC - Doença Pulmonar Obstrutiva Crônica.

EM-DAT - *Emergency Disaster Data Base* (Banco de Dados Internacional de Desastres, do inglês).

ODS - Objetivos do Desenvolvimento Sustentável.

ONU - Organização das Nações Unidas.

PNPDEC - Política Nacional de Proteção e Defesa Civil.

SUS - Sistema Único de Saúde.

SVM - *Support Vector Machine* (Máquina de vetores de suporte).

SUMÁRIO

1. INTRODUÇÃO	13
1.1. Motivação	14
1.2. Pergunta de Pesquisa	15
1.3. Objetivos	16
1.3.2. Objetivos Específicos	16
1.4. Justificativa	17
1.5. Metodologia	18
2. FUNDAMENTOS	24
2.1. Saúde Pública Municipal Catarinense	24
2.2. Pandemia Causada pelo Coronavírus	25
2.3. Prevenção de Infecções na Assistência à Saúde	27
2.4. Triagem dos Usuários com Suspeita de Contágio pelo Coronavírus	28
2.5. Classificação da Gravidade de Síndromes Respiratórias a partir de Sinais, Sintomas e Perfis dos Pacientes	29
2.6. O Aplicativo Triagem On-Line	30
2.7. Medidas de Qualidade de Resultados de Classificadores	38
2.7.1. Matriz de confusão	38
2.7.2. Acurácia	39
2.7.3. Precisão	39
2.7.4. Sensibilidade	40
2.7.5. F1-Score	40
2.8. Classificadores Baseados em Aprendizagem de Máquina	40
2.8.1. Logistic Regression	42
2.8.2. Decision Tree	43
2.8.3. Support Vector Machine - SVM	44
3. REVISÃO DA LITERATURA	45
3.1 Classificação de Casos Suspeitos de COVID-19	45
3.2 Classificação de Gravidade de Síndromes Respiratórias	49
4. ETAPAS REALIZADAS	55
4.1 Compreensão do Problema	55
4.2 Preparação dos Dados	57
4.3 Treinamento dos Modelos	63
4.4 Avaliação dos Modelos	66
5. RESULTADOS	68
5.1 Análise Exploratória dos Dados	68
5.2 Experimento I - Somente sinais e sintomas	72
5.1.1 Análise Exploratória do Experimento I	77
5.3 Experimento II - Sinais, sintomas e dados de perfil dos pacientes	78
5.4 Experimento III - Dados disponíveis no início da pandemia	82
5.5 Discussões	86
6. CONCLUSÕES E TRABALHOS FUTUROS	90
8. REFERÊNCIAS BIBLIOGRÁFICAS	93
9. ANEXOS	101

1. INTRODUÇÃO

Síndromes Respiratórias são complicações associadas ao agravamento de alguma infecção viral. O paciente pode apresentar desconforto respiratório e queda no nível de saturação de oxigênio, entre outros sintomas. Essas síndromes são caracterizadas como infecções do trato respiratório superior, podendo também atingir o trato respiratório inferior e causar pneumonia, resultando em insuficiência respiratória, falência de múltiplos órgãos e morte (Girard *et al.*, 2010; Fontana e Strasfeld, 2019; Patel *et al.*, 2010).

Muitos patógenos podem estar relacionados ao surgimento de síndromes respiratórias, sendo os mais frequentemente encontrados: Influenza, Rinovírus, Vírus Sincicial Respiratório, Adenovírus, Metapneumovírus, Parainfluenza e o Coronavírus. O Coronavírus foi responsável por provocar epidemias recentes, como a COVID-19, que desencadeou pesquisas no mundo inteiro para combatê-la (Silva, 2021). Por se tratar de uma síndrome respiratória, o caso da pandemia da COVID-19 pode ser utilizado como estudo de caso para se aprimorar o enfrentamento de possíveis novas pandemias causadas por infecções respiratórias.

Com o advento da pandemia da COVID-19, houve necessidade, por parte do governo do Estado de Santa Catarina, de buscar medidas que minimizassem os efeitos da proliferação do vírus no âmbito do estado. Desta feita, com a previsão, por parte dos órgãos técnicos especializados, de que haveria grande número de pessoas infectadas pelo novo Coronavírus, já no início de 2020, o poder público necessitou buscar alternativas no fornecimento de alguns tipos de serviços, para que pessoas que apresentassem sintomas leves, por exemplo, não precisassem buscar atendimento presencial nas unidades de saúde, de modo a evitar uma possível sobrecarga no sistema (Santa Catarina, 2020).

Triagens objetivam, por meio de um processo sistemático, definir quais pacientes terão prioridade em um atendimento emergencial (Murray, 2003). Além disso, triagens podem diferenciar pacientes que necessitam ser atendidos com

prioridade se comparados com outros que podem aguardar por atendimento em segurança. (Albino; Grosseman; Riggenbach, 2007).

Desta forma, foi disponibilizado aos cidadãos catarinenses um *software* capaz de realizar uma triagem remota, por meio de um auto teste, de tal forma que o cidadão pudesse identificar quais medidas deveria tomar de acordo com sua situação de saúde. A ferramenta, denominada Triagem On-line, possibilitava que o cidadão, sem sair de casa, descrevesse os sinais, sintomas e fatores de risco de modo que o sistema, subsidiado por um algoritmo específico, identificasse a melhor maneira de direcionar aquele paciente (Santa Catarina, 2020).

Um estudo foi realizado sobre a utilização da ferramenta Triagem On-line, da concepção da ideia à efetiva utilização, que resultou em um trabalho de conclusão do curso de graduação em Administração Pública pela Universidade do Estado de Santa Catarina, apresentado em julho de 2021, cujo tema foi “Ferramenta Triagem On-line: contribuição para o enfrentamento da COVID-19 no âmbito do Estado de Santa Catarina.”. Tal estudo, com ênfase na descrição da rede de cooperação existente entre diversos órgãos do Estado, restou incompleto, uma vez que não foi possível, em razão do tempo, bem como da complexidade, realizar o cruzamento de dados entre as triagens realizadas e os casos efetivamente confirmados pelo sistema do SUS, objetivando uma possível mensuração da efetividade do processo de trabalho (estratégia adotada).

1.1. Motivação

O uso de ferramentas computacionais pôde desempenhar um papel importante no combate à pandemia causada por síndrome respiratória. Um *software* desenvolvido em Santa Catarina para realizar remotamente a triagem de casos suspeitos de COVID-19, no que tange à gravidade dos sinais/sintomas, permitia um “auto-teste”, a partir de sinais, sintomas e informações de perfil. O aplicativo retornava ao usuário a classificação da gravidade do caso, além de instruções sobre medidas a tomar de acordo com tal classificação. Todavia ainda não foi quantificada a eficácia do software e nem comparado o seu desempenho com o de outras

alternativas, tais como diversos modelos de aprendizado de máquina, que poderiam ter sido usados em diferentes estágios da recente pandemia de COVID-19 ou que possam ser apropriados em outras futuras situações de emergência em saúde, causadas por síndromes respiratórias.

1.2. Pergunta de Pesquisa

O uso da ferramenta computacional, como instrumento de enfrentamento à COVID-19, para realização de auto-teste pelos cidadãos catarinenses, gerou a oportunidade de estudar a relação existente entre os dados extraídos do sistema Triagem On-Line com os resultados dos exames laboratoriais realizados pelo Sistema Único de Saúde, no âmbito do estado de Santa Catarina; ambos disponibilizados em sítio eletrônicos de amplo acesso, de maneira anonimizada. Dessa forma, avaliou-se a efetividade do aplicativo, por meio do cruzamento de dados entre as triagens realizadas e os casos confirmados. Além disso, comparou-se a efetividade do aplicativo com modelos de aprendizado de máquina selecionados para situações análogas, treinados com os mesmos dados.

Portanto, entende-se útil a avaliação da efetividade do uso de sistemas informatizados para mitigar impactos causados por desastres naturais, levando assim ao seguinte questionamento:

Qual a efetividade (sensibilidade, acurácia, etc.) do aplicativo Triagem On-Line e de abordagens alternativas de aprendizado de máquina na classificação da gravidade de casos de síndrome respiratória aguda a partir de sinais, sintomas e dados de perfil relatados pelos pacientes?

A resposta a esta pergunta pode contribuir para determinar as abordagens mais efetivas e viáveis para triagem da gravidade de potenciais casos considerando:

- diferentes contextos de surtos e epidemias;

- alternativas que incluam:
 - critérios (e.g. pesos dos sinais, sintomas e dados de perfil definidos por especialistas de domínio) como usados pelo Triagem On-Line;
 - modelos como regressão logística, random trees e SVM.

1.3. Objetivos

O objetivo geral deste trabalho é avaliar a efetividade da ferramenta *Triagem On-Line* na classificação da gravidade de casos suspeitos de COVID-19 e compará-la com a de alternativas do estado da arte visando determinar as melhores abordagens para triagem de casos em eventuais situações futuras de surtos, epidemias e pandemias.

1.3.2. Objetivos Específicos

1. Estudar, entender e comparar diferentes modelos de classificação do estado da arte que possam ser usados na triagem on-line de possíveis casos de doenças a partir de sinais, sintomas e informações de perfil relatados pelos usuários/pacientes;
2. Confrontar as triagens realizadas pelo aplicativo Triagem On-Line com diagnósticos e dados sobre a severidade dos casos, de modo a mensurar sua efetividade (acurácia, precisão, etc.);
3. Avaliar a efetividade de modelos selecionados da literatura na classificação da gravidade de possíveis casos em diferentes períodos da pandemia de COVID-19, em que o conhecimento sobre a doença e a disponibilidade de tratamento e de dados para treinamento dos modelos preditivos são distintos, de modo a tentar extrapolar diretrizes para situações análogas no futuro.

1.4. Justificativa

A realização de triagens de pacientes, através de *softwares*, pode trazer benefícios para o sistema de saúde e os cidadãos durante a pandemia, pela redução de pessoas se dirigindo fisicamente a locais contaminados e comodidade, para que as pessoas se recuperassem do ciclo viral em suas residências, caso não apresentassem sintomas graves (Murray, 2003).

Desta feita, entende-se que a mensuração da efetividade da estratégia utilizada pelo sistema informatizado denominado Triagem On-line é uma forma de aproveitar os conhecimentos e capacidades advindos da atuação conjunta de diversos órgãos estaduais – CBMSC e Secretaria da Fazenda – no enfrentamento à pandemia da COVID-19, uma vez que de tal união foi possível vislumbrar uma solução computacional capaz de proporcionar, de maneira personalizada, a classificação da gravidade, com base em sinais e sintomas apresentados, e repassar instruções sobre medidas a tomar de acordo com tal classificação, aos cidadãos catarinenses. Primeiramente, a Secretaria de Saúde do Estado definiu o fluxograma de dados e os critérios para a triagem dos casos suspeitos. Com base neles, o autor deste trabalho desenvolveu o aplicativo Triagem On-Line, que ficou disponível no portal oficial de combate a pandemia da COVID-19 do governo do Estado de Santa Catarina.

Ressalta-se ainda que a mensuração da estratégia de uso de classificadores de gravidade de síndromes respiratórias, durante o enfrentamento de pandemias, pode colaborar diretamente com diversos Objetivos de Desenvolvimento Sustentável (ODS) da ONU (2015, grifo nosso), a saber:

Objetivo 1. Acabar com a pobreza em todas as suas formas, em todos os lugares [...]

1.4 Até 2030, garantir que todos os homens e mulheres, particularmente os pobres e vulneráveis, tenham direitos iguais aos recursos econômicos, bem como o acesso a serviços básicos, propriedade e controle sobre a terra e outras formas de propriedade, herança, recursos naturais, novas tecnologias apropriadas e serviços financeiros, incluindo microfinanças. [...]

Objetivo 3. Assegurar uma vida saudável e promover o bem-estar para todas e todos, em todas as idades [...]

3.3 Até 2030, acabar com as epidemias de AIDS, tuberculose, malária e doenças tropicais negligenciadas, e combater a hepatite, doenças transmitidas pela água, e outras doenças transmissíveis. [...]

3.8 Atingir a cobertura universal de saúde, incluindo a proteção do risco financeiro, o acesso a serviços de saúde essenciais de qualidade e o acesso a medicamentos e vacinas essenciais seguros, eficazes, de qualidade e a preços acessíveis para todos [...]

3.9.b Apoiar a pesquisa e o desenvolvimento de vacinas e medicamentos para as doenças transmissíveis e não transmissíveis, que afetam principalmente os países em desenvolvimento, proporcionar o acesso a medicamentos e vacinas essenciais a preços acessíveis, de acordo com a Declaração de Doha, que afirma o direito dos países em desenvolvimento de utilizarem plenamente as disposições do acordo TRIPS sobre flexibilidades para proteger a saúde pública e, em particular, proporcionar o acesso a medicamentos para todos. [...]

3.9.d Reforçar a capacidade de todos os países, particularmente os países em desenvolvimento, para o alerta precoce, redução de riscos e gerenciamento de riscos nacionais e globais de saúde.

1.5. Metodologia

O conjunto de passos, procedimentos, técnicas e ferramentas utilizadas para realizar a presente pesquisa é apresentado a seguir. Uma metodologia adequada é fundamental para atingir os objetivos dentro do tempo e com os recursos disponíveis, além de garantir a validade e confiabilidade dos resultados obtidos, proporcionando uma estrutura organizada e consistente para o desenvolvimento do trabalho científico.

1.5.1. Elaboração das perguntas e hipóteses de pesquisa.

Nesse primeiro passo, é necessário definir as perguntas de pesquisa no estudo. Em seguida, deve-se formular uma ou mais hipóteses para orientar o desenvolvimento da pesquisa.

1.5.2. Revisão bibliográfica e seleção de artigos sobre classificadores de gravidade de síndromes respiratórias baseados em *Machine Learning*.

Nesse passo, foi realizada uma revisão bibliográfica detalhada sobre o tema proposto para identificar artigos científicos, publicações e outras fontes relevantes que abordavam a utilização de classificadores baseados em machine learning para classificar a gravidade de síndromes respiratórias. O objetivo era identificar as abordagens que propiciaram os melhores resultados na classificação de gravidade de casos de síndromes respiratórias a partir do perfil dos pacientes, sinais e sintomas e comparar com os resultados obtidos pelo aplicativo Triagem On-Line. A seleção dos artigos deve levar em consideração a relevância, qualidade e contribuição para a pesquisa em questão.

1.5.3. Avaliação da qualidade metodológica das propostas e resultados selecionados da literatura e montagem de quadro comparativo.

Após selecionar os artigos relevantes, foi necessário avaliar a qualidade metodológica das propostas e resultados apresentados. Isso envolveu analisar a validade dos métodos utilizados, a consistência dos resultados obtidos e a confiabilidade das conclusões. Com base nessa avaliação, foi possível montar um quadro comparativo, destacando as principais características e resultados dos estudos selecionados.

1.5.4. Busca e seleção de dados para experimentos.

Nesse passo, é necessário buscar e selecionar os dados necessários para realizar os experimentos. Isso envolverá a procura por conjuntos de dados públicos. Os dados devem ser representativos e adequados para treinar e testar os modelos de classificação.

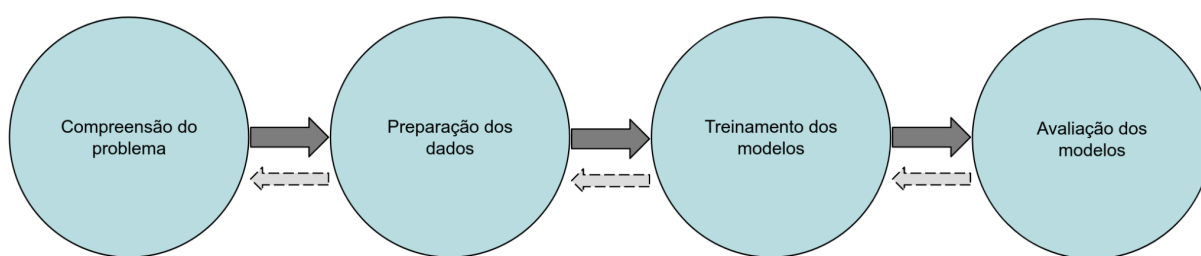
1.5.5. Construção dos modelos baseados em *Machine Learning*

Os projetos de modelos baseados em *Machine Learning* são, normalmente, de natureza exploratória. Por esta razão, faz-se mister utilizar de um processo para gerenciá-lo e garantir que etapas importantes não deixem de ser seguidas. O processo deve ser minucioso e rigoroso em sua abordagem, mas não tão rígido que o processo impeça a exploração. De acordo com Dietrich, Heller e Yang (2015), muitos problemas que parecem ser grandes e complexos à primeira vista podem ser divididos em partes menores ou etapas que podem ser resolvidas com mais facilidade. Ter uma sequência de passos pode facilitar o uso de um método abrangente e repetível para se conduzir análises. Além disso, pode economizar tempo e energia no início do processo para obter uma compreensão clara do problema que se pretende resolver. Um processo bem definido oferece uma estrutura comum para que outros pesquisadores adotem, de modo que os métodos e a análise possam ser repetidos no futuro. Dietrich, Heller e Yang (2015) propuseram um modelo de ciclo de vida de um projeto de análise de dados, composto por seis etapas. Esse modelo é iterativo, ou seja, cada etapa pode envolver retornos a etapas anteriores para ajustes e refinamentos, visando melhorar os resultados e garantir a qualidade do processo. Já Lantz (2019) sugere cinco passos para se aplicar *Machine Learning* em um conjunto de dados.

Considerando que essas metodologias são diretrizes gerais e podem ser adaptadas para atender às necessidades específicas de um projeto, o treinamento dos modelos utilizados neste trabalho seguiu uma abordagem sistemática própria, que envolverá a compreensão do problema, a preparação dos dados, a seleção e treinamento de modelos e uma avaliação do desempenho desses modelos. No

decorrer das etapas, a qualquer momento, poderia se retornar aos passos anteriores para realização de revisões, ajustes e melhorias. Trata-se de uma adaptação das metodologias proposta por Lantz (2019), Dietrich, Heller e Yang (2015) para a realidade desse estudo. A figura 1 apresenta as quatro etapas realizadas:

Figura 1 - Etapas do treinamento dos modelos



Fonte: adaptado de Lantz (2019), Dietrich, Heller e Yang (2015).

As quatro etapas serão minudenciadas a seguir:

a) Compreensão do problema - Esta fase visou compreender completamente o contexto do problema em questão, identificar os requisitos e objetivos da pesquisa, bem como definir métricas de avaliação apropriadas.

b) Preparação dos dados - Nessa fase os dados foram explorados e selecionados, para compreender suas características, identificar possíveis problemas ou inconsistências e realizar análises descritivas para obter um melhor entendimento dos mesmos. Os dados foram então preparados para a construção dos modelos de classificação. Isso envolveu a limpeza dos dados, tratamento de valores ausentes, codificação de variáveis categóricas, normalização ou padronização dos dados, entre outras etapas de pré-processamento. Faceli et al. (2011) explicam a importância do tratamento ou eliminação de dados incompletos, inconsistentes, duplicados ou ruidosos, para que esses não influenciem no processo de treinamento, pois podem provocar distorções no aprendizado.

c) Treinamento dos modelos - Nessa fase, foram construídos os modelos de classificação baseados em *machine learning*. Diferentes algoritmos foram testados e ajustados com base nas características dos dados e nos objetivos da pesquisa. Em seguida, os modelos foram treinados com parte dos dados disponíveis. Diferentes configurações e parâmetros dos modelos selecionados foram ajustados com o objetivo de identificar qual configuração forneceria resultados mais efetivos. Também foi importante utilizar técnicas de validação cruzada para melhorar a generalização dos modelos.

d) Avaliação dos modelos - Após o treinamento dos modelos, foi necessário avaliar seus desempenhos e compará-los. Esta etapa foi realizada com a construção de uma matriz de confusão, fornecendo condições para o cálculo de métricas de avaliação, como acurácia, precisão, recall (também chamada de sensibilidade ou cobertura), F1-score, entre outras. As avaliações finais foram realizadas com dados não utilizados durante o desenvolvimento da solução, para se certificar que os modelos estavam generalizados o suficiente.

1.5.6. Redação e publicação dos resultados.

Os principais produtos do trabalho foram a análise de desempenho do Triagem On-Line, comparação com abordagens de aprendizado de máquina e diretrizes para eventos futuros análogos à pandemia de COVID-19, bem como disponibilização do código-fonte dos modelos baseados em *Machine Learning*.

A seguir, será apresentada a estrutura do presente estudo:

No segundo capítulo deste trabalho, foram expostos os princípios essenciais necessários para compreender sua abordagem. Esse capítulo abarcou desde informações fundamentais sobre saúde pública e a pandemia de COVID-19 até os conceitos básicos do aprendizado de máquina, os modelos empregados no estudo e os métodos utilizados para avaliar seu desempenho. No terceiro capítulo, foram descritos os procedimentos de revisão bibliográfica, apresentando os resultados

obtidos e fazendo comparações entre a abordagem atual e trabalhos relacionados já existentes. O quarto capítulo ilustra a aplicação da metodologia CRISP-DM para conduzir o processo de mineração de dados e sua implementação nos modelos de aprendizado de máquina. O quinto capítulo relata os resultados alcançados nos experimentos realizados neste estudo. Esses resultados foram divididos em duas seções principais, separando a análise exploratória dos dados das fases de treinamento e avaliação dos modelos. Por fim, no sexto capítulo, destacam-se as conclusões alcançadas e são apresentadas sugestões para trabalhos futuros.

2. FUNDAMENTOS

Este capítulo apresenta os fundamentos necessários para o entendimento deste trabalho. Ele cobre desde informações básicas sobre saúde pública e pandemia de COVID-19 até fundamentos de aprendizado de máquina, modelos utilizados no trabalho e medidas para avaliação do seu desempenho.

2.1. Saúde Pública Municipal Catarinense

Devido à escassez de recursos, municípios de pequeno porte tendem a se agrupar, de modo que haja otimização na utilização dos recursos públicos destinados à saúde. Sá e Holanda (2017) detalham que a organização dos serviços de saúde, formando as Regiões de Saúde, além de propiciar uma melhor gestão dos custos e aproveitar ao máximo os recursos, seleciona municípios com maior capacidade técnica e informacional para servir como centros principais, pois eles têm uma economia de escala que se destaca no contexto urbano. Nem todos municípios têm as condições adequadas para oferecer todos os serviços de saúde devido a fatores sociais, demográficos, entre outros. Por isso, as atividades de alta complexidade são centralizadas nas principais cidades da macrorregião, especialmente aquelas que demandam mais investimentos ou que são menos comuns. Esta organização territorial visa aproveitar as vantagens de áreas com recursos externos que são atraentes para entidades privadas envolvidas na oferta de serviços mais especializados e infrequentes.

Esclarece-se, ainda, que dos 295 municípios existentes no Estado de Santa Catarina (IBGE, 2021) 139 não possuem leitos hospitalares do Sistema Único de Saúde – SUS (BRASIL, 2021).

2.2. Pandemia Causada pelo Coronavírus

A pandemia do Coronavírus assolou o mundo em grande velocidade, neste sentido, Guimarães *et al.* (2020) esclarecem que esta enfermidade evoluiu para uma questão global, funcionando como um vetor que afeta todo o sistema. Sua transmissão é altamente eficaz, com um ritmo de expansão acelerado e desafiador para os sistemas de saúde pública dos países. Estes sistemas, que ao longo do tempo já mostravam vulnerabilidades e sofriam com orçamentos limitados devido a abordagens neoliberais, encontram-se sobrecarregados diante do avanço da doença. Assim, os governos foram levados a adotar medidas de distanciamento social como resposta.

No início da pandemia pouco se sabia sobre a doença e, por tal razão, a comunicação entre os pesquisadores de todo o mundo se fez ainda mais necessária. De acordo com Bufren (2020), parte-se do princípio de que a informação desempenha um papel crucial na luta contra a pandemia, especialmente em contexto de doenças sem precedentes. Diante de questões como a evolução da doença, taxas de letalidade, diferentes formas de transmissão, tratamento, presença de outros efeitos ou sequelas em indivíduos infectados e outras necessidades emergentes relacionadas à COVID-19, pesquisadores e cientistas em todo o mundo, especialmente aqueles com recursos privilegiados e com uma coordenação governamental efetiva, mobilizaram-se para estimar não apenas os impactos da doença na saúde da população, bem como os impactos sociais e econômicos da crise gerada pela pandemia.

Uma forma reconhecidamente eficaz de se controlar a velocidade de propagação da COVID-19 é a aplicação do distanciamento social. Sobre a importância do distanciamento social, a Universidade Federal do Amazonas (2020) destaca que, nesse contexto, a principal estratégia não medicamentosa para frear a rápida propagação do vírus é o distanciamento social. Este é um método em saúde pública que encoraja indivíduos a se manterem afastados fisicamente durante epidemias de doenças transmissíveis, com o objetivo de diminuir o avanço da contaminação.

Há situações em que as pessoas necessitam buscar auxílio médico e, para isto, o distanciamento social ficaria prejudicado. Entretanto, para alguns casos, a telemedicina pode servir como alternativa para o acompanhamento médico, sem necessidade do deslocamento físico para uma unidade de saúde. A telemedicina, já utilizada anteriormente para tratar outras doenças, revelou-se indispensável para tratar os males causados pelo Coronavírus. Garcia e Garcia (2020) reforçam que a doença COVID-19 se disseminou rapidamente por todos os continentes, ultrapassando a capacidade dos sistemas de saúde em rastrear, testar e controlar a propagação do vírus. As práticas de telemedicina surgem como uma solução para evitar o contato próximo, reduzindo assim o risco de infecção pelo vírus da COVID-19. Além disso, a telemedicina acelera a disseminação de informações precisas, oferecendo plataformas de ensino e proporcionando acesso às opiniões de especialistas mesmo em locais remotos.

Ressalta-se, ainda, que a telemedicina teve sua aplicação ampliada, passando a ser chamada de telessaúde. Caetano *et al.* (2020) acrescentam que a telemedicina tradicionalmente se focava na adaptação das interações convencionais entre médico-paciente e médico-médico, enriquecida pela capacidade de vídeo e áudio bidirecionais. Com o tempo, as tecnologias de informação e comunicação (TIC) passaram a ser utilizadas também para respaldar serviços, capacitações e disseminação de informações em saúde, tanto para profissionais multidisciplinares quanto para pacientes. Esse cenário mais abrangente passou a ser conhecido como telessaúde.

Importante ressaltar que além de proteger pacientes não acometidos pela COVID-19, a telemedicina protege, também, os profissionais da área da saúde. Neste diapasão, Wen (2020) defende que, de acordo com especialistas, somente dessa maneira poderíamos iniciar o processo de redução dos impactos devastadores do COVID-19 no sistema de saúde do Brasil. Nesse contexto, a telemedicina se torna uma questão indispensável, tanto como uma solução logística para a prestação de assistência médica à distância, quanto como uma forma de proteção aos profissionais de saúde em grupos de risco e seus familiares durante a pandemia.

2.3. Prevenção de Infecções na Assistência à Saúde

O Coronavírus é extremamente transmissível. Sua grande capacidade de transmissão não pode ser negada, inclusive pela velocidade com que se espalhou pelo mundo. Ambientes de saúde que recebem pessoas acometidas da COVID-19 são, sem dúvidas, locais onde cuidados extras devem ser tomados, uma vez que unidades de saúde são reconhecidamente locais de transmissão de doenças. De acordo com Neto *et al.* (2010) as Infecções Relacionadas à Assistência à Saúde são um problema global que afeta tanto países desenvolvidos quanto países em desenvolvimento, representando uma carga significativa para os sistemas de saúde. Segundo dados da Organização Mundial da Saúde (OMS), em média, cerca de 8,7% de todos os pacientes hospitalizados apresentam algum tipo de infecção nosocomial. Além disso, as infecções adquiridas em hospitais estão entre as principais causas de morte e de prolongamento do tempo de internação.

Sobre a forma como as infecções hospitalares acontecem, Pereira *et al.* (2005) esclarecem que infecções hospitalares são aquelas que ocorrem depois que o paciente é internado e pode se mostrar enquanto ele ainda está no hospital ou após sua saída, desde que possa ser associada à sua estadia ou a procedimentos feitos no hospital. A maioria dessas infecções surge quando há um desequilíbrio entre a microbiota comum e as defesas naturais do organismo dos pacientes. Isso pode ser causado pela condição de saúde do paciente, por intervenções invasivas ou mudanças na microbiota, frequentemente causadas por antibióticos. Os microrganismos que causam essas infecções geralmente não são problemáticos em outras circunstâncias. Eles têm baixa capacidade de causar doenças, mas devido a fatores como a saúde do paciente, eles podem causar infecções. Cerca de dois terços dessas infecções são originadas da própria microbiota do paciente, podendo vir de fora ou dentro do hospital. Em ambos os casos, a infecção é geralmente precedida pela colonização, tornando difícil discernir se o paciente trouxe o microrganismo de fora ou se ele foi contaminado no hospital. O uso da telessaúde deve ser considerado na redução desses tipos de infecções.

2.4. Triagem dos Usuários com Suspeita de Contágio pelo Coronavírus

Triagens de classificação de risco são utilizadas em unidades de saúde para classificar os pacientes, de acordo com a gravidade do estado de saúde, para priorização de atendimento. De acordo com o Ministério da Saúde (Brasil, 2002), o processo de triagem classificatória de risco “deve ser realizado por profissional de saúde, de nível superior, mediante treinamento específico e utilização de protocolos pré-estabelecidos e tem por objetivo avaliar o grau de urgência das queixas dos pacientes, colocando-os em ordem de prioridade para o atendimento”.

Considerando que a telemedicina é uma realidade, a triagem à distância também poderia ser realizada em algumas situações. Schmitz *et al.* (2020) esclarecem essa situação, quando defendem a possibilidade de realização de triagem de pacientes de maneira remota, por meios digitais. Defendem que esse tipo de consulta pode variar desde uma simples triagem até um diagnóstico mais detalhado. O propósito da teleconsulta é fornecer um atendimento eficaz e conclusivo para ambas as partes envolvidas e decidir se é necessário um encontro presencial.

Além da tecnologia possibilitar o atendimento remoto por equipes de saúde, outras funcionalidades estão ganhando espaço na atualidade com o uso de sistemas automatizados. Campana e Oplustil (2011) descrevem os ganhos que se podem obter quando da utilização de sistemas automatizados na área da saúde. Para eles, o avanço da tecnologia médica e seu impacto na assistência à saúde e na economia são temas amplamente discutidos por várias partes envolvidas na área. A palavra "automação" tem origem no latim "*automatus*" e significa agir por si próprio. Existem várias definições disponíveis, mas, de forma resumida, a automação é a aplicação de técnicas computadorizadas ou mecânicas para tornar um processo mais eficiente, aumentando a produção com menor consumo de energia e proporcionando maior segurança. No contexto mencionado, o gasto de energia refere-se ao uso de mão de obra especializada em tarefas de baixo valor agregado, tempo desperdiçado e ineficiências em geral.

Ademais, sobre os benefícios do afastamento social, o Ministério da Saúde publicou a Recomendação nº 22 de 09 de abril de 2020 (BRASIL, 2020) que indica

que o estabelecimento de diretrizes de distanciamento social é uma tática crucial para reduzir a propagação do coronavírus e controlar o aumento acentuado de casos, circunstâncias que podem afetar significativamente a habilidade do sistema de saúde em atender pacientes com COVID-19 e outras complicações de saúde.

2.5. Classificação da Gravidade de Síndromes Respiratórias a partir de Sinais, Sintomas e Perfis dos Pacientes

A classificação da gravidade de síndromes respiratórias, a exemplo da COVID-19, pode ser feita por meio de informações clínicas (sinais, sintomas e dados de perfil de pacientes), por testes laboratoriais específicos, ou por exames de imagem. Ressalta-se que os sintomas da COVID-19 são semelhantes ao da Síndrome Gripal, sendo os mais comuns febre, tosse, falta de ar, fadiga, dores musculares, perda de paladar ou olfato, dor de cabeça, dor de garganta, coriza, náuseas, vômitos e diarreia. Na presença de dois ou mais sintomas, será considerada suspeita de COVID-19 (Lazarin; Mariano, 2022).

Definição do Problema

Para classificar a gravidade destas síndromes, a definição do problema é encontrar uma função de mapeamento dos potenciais casos em uma categoria.

Sejam:

- um conjunto P de potenciais casos, cada qual com seus, sinais, sintomas e dados de perfil de cada paciente em certo instante e
- um conjunto C de categorias de gravidade predefinidas ($|C| \geq 2$) que cada caso potencial pode ter.

Classificar a gravidade dos casos em P consiste em encontrar uma função de mapeamento $F : P \rightarrow C$. Tal função que mapeia cada potencial caso $p_i \in P$ em uma única categoria $c_j \in C$ é classificador de potenciais casos.

2.6. O Aplicativo Triagem On-Line

O sistema Triagem On-line foi desenvolvido utilizando conceitos de Engenharia de Usabilidade, por meio do desenvolvimento de interfaces simplificadas que pudessem ser utilizadas pelos mais variados tipos de usuários. Para Pádua (2012) “A Engenharia de Usabilidade visa o desenvolvimento da interação entre o usuário e sistemas informatizados. A engenharia de usabilidade tem por objetivo oferecer técnicas e métodos que possam ser utilizadas sistematicamente para assegurar um alto grau de usabilidade da interface de programas de computador.”.

O aplicativo Triagem On-Line, em uma primeira tela, solicitava informações pessoais do usuário e, em seguida, solicitava informações acerca da condição de saúde: sinais, sintomas e fatores de risco.

Um fluxograma do aplicativo Triagem On-Line foi empregado para estabelecer uma comunicação eficiente entre o desenvolvedor do *software* e os médicos infectologistas da Secretaria de Saúde do Estado. Por meio desse fluxograma, também foi possível que os médicos fizessem observações e sugestões para ajustes no programa, detalhando de maneira precisa em quais casos os pacientes deveriam permanecer em suas residências e em quais casos eles deveriam procurar uma unidade de saúde. Isso permitiu aprimorar o algoritmo utilizado na ferramenta computacional.

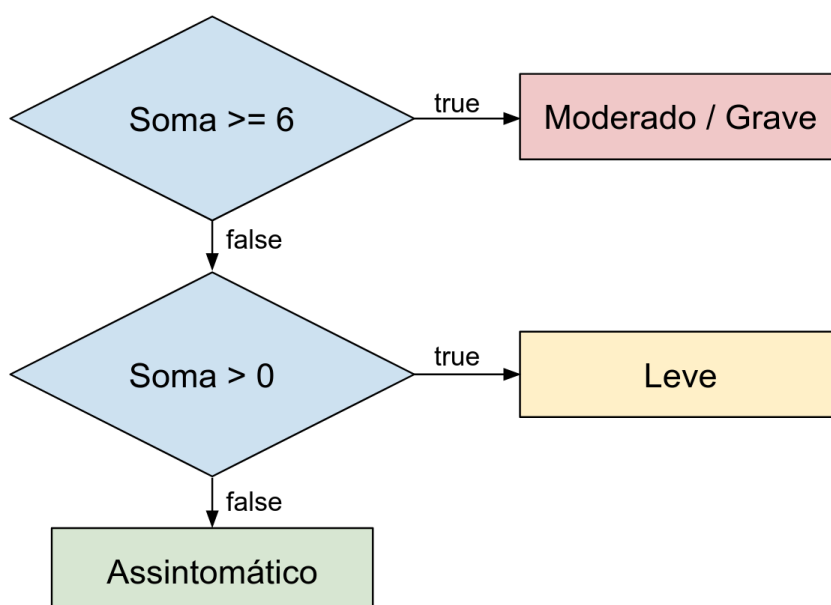
Para cada sinal ou sintoma, o algoritmo atribui um peso diferente, conforme apresentado no Quadro 1. Esses pesos foram propostos por médicos infectologistas da Secretaria de Saúde do Estado de Santa Catarina, no dia 2 de abril de 2020 (início da pandemia), de acordo com o conhecimento sobre COVID-19 que detinham à época. Destaca-se que dispneia (falta de ar) tem peso consideravelmente mais alto que os outros sintomas, pois já era sabido pelos médicos que tal sintoma é determinante na identificação de casos graves. Por outro lado, anosmia (perda de olfato) e disgeusia (perda de paladar), embora sejam sintomas frequentes nos casos de COVID-19, possuem um peso menor para o cálculo da gravidade. De acordo com Carrillo-Larco e Altez-Fernandez (2020), não há evidências suficientes para definir anosmia e disgeusia como sintomas preditores do agravamento da doença.

Quadro 1 - Pesos atribuídos para cada sinal ou sintoma.

Sinal, Sintoma	Peso
Febre	2
Tosse seca	2
Dor de garganta	1
Coriza	1
Dispneia	6
Cefaleia	1
Astenia	1
Mialgia	1
Diarreia	1
Tosse com sangue	2
Anosmia	0,5
Disgeusia	0,5

Fonte: do autor (2023).

Após o preenchimento do formulário, o programa soma os pesos referentes aos sintomas reportados. Daí o sistema usa tal soma para classificar o caso como assintomático, leve ou moderado/grave, conforme a regra expressa no fluxograma apresentado na Figura 2:

Figura 2 - Classificação dos casos.

Fonte: do autor (2023).

Em seguida, eram fornecidas instruções ao paciente, como procurar ajuda profissional por meio do contato telefônico com o setor de telemedicina da região em que o usuário se encontrava ou, presencialmente, em uma das unidades de saúde do município em que estava.

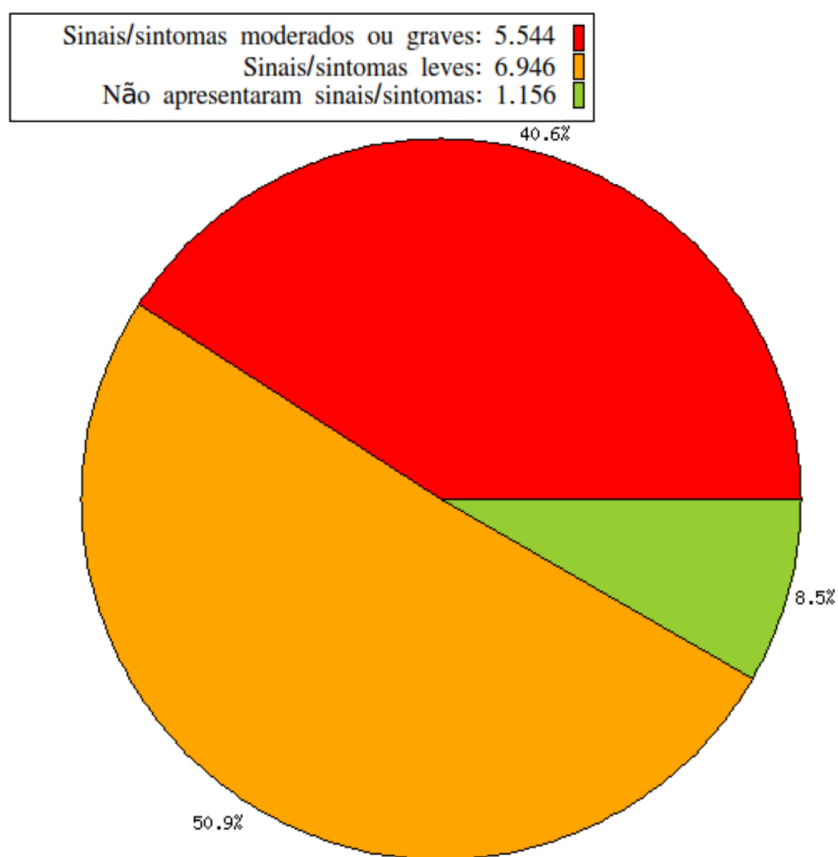
No processo de tomada de decisão sobre a gravidade dos casos, o sistema não levava em consideração os fatores de risco (comorbidades), nem informações de perfis apresentados. As informações sobre os fatores de risco e os dados pessoais eram utilizadas apenas para gerar dados estatísticos.

Dados Colhidos pelo Aplicativo Triagem On-Line

As avaliações realizadas pelo aplicativo Triagem On-Line foram todas armazenadas em um banco de dados relacional, de modo a facilitar o acesso por outros sistemas computacionais e possibilitar a geração de estatísticas e gráficos. Importante esclarecer que, por questões burocráticas/legais, não foi possível ter acesso à base de dados gerada pelo aplicativo Triagem On-Line, o que impediu, conseqüentemente, a realização de experimentos com tais dados.

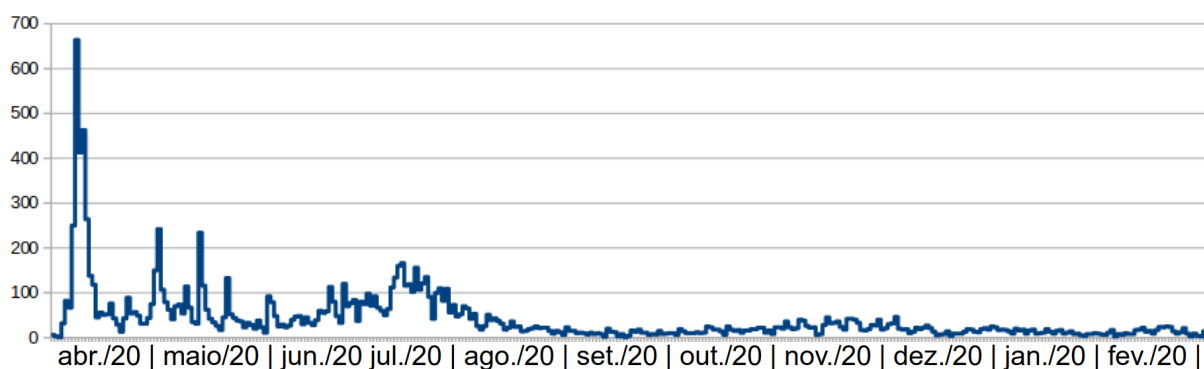
Assim sendo, apresentar-se-á, neste tópico, um compilado de informações extraídas de relatórios do uso do aplicativo.

Como resultado da utilização do Triagem On-Line, até 29/03/2021, foram registradas 13.646 avaliações. Dessas, 6.946 foram feitas por pessoas com sintomas leves, 5.544 por pessoas com sintomas moderados ou graves, e 1.156 por pessoas que não apresentaram nenhum sinal ou sintoma relacionado à COVID-19. A Figura 3 mostra a proporção das avaliações realizadas, ficando claro que a maioria delas foi classificada como casos leves.

Figura 3 - Proporção das triagens realizadas.

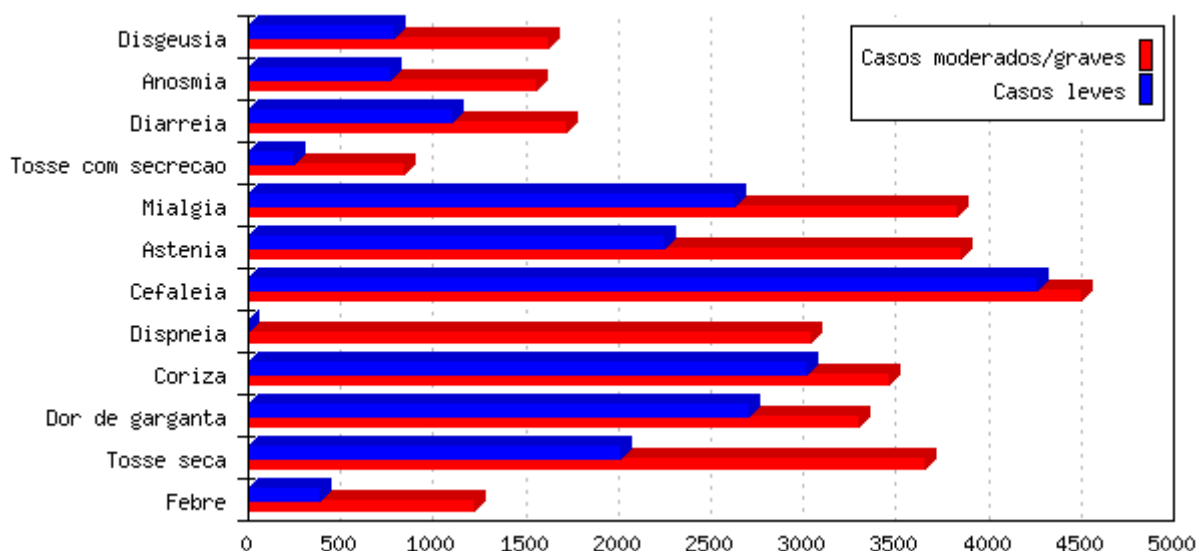
Fonte: do autor (2023).

Durante o período em que o aplicativo Triagem On-Line estava disponível, houve dias em que ocorreu uma maior demanda por parte dos usuários. Essas datas coincidiram com os dias em que o aplicativo foi divulgado em programas de rádio e televisão. Além disso, observou-se que nos primeiros meses em que o aplicativo estava disponível, um número maior de avaliações foi realizado. É provável que isso tenha ocorrido devido à pouca familiaridade com a doença naquela época, resultando em uma maior preocupação com os primeiros sinais e sintomas que pudessem surgir. O dia em que a ferramenta foi mais procurada para uso foi em 12 de abril de 2020, quando o governador do Estado a divulgou durante o programa “Boletim Diário da COVID-19” sobre o combate à pandemia, transmitido em alguns canais de televisão. Nesse dia, foram realizadas 664 avaliações por meio do referido *software*. A Figura 4 mostra a quantidade de avaliações realizadas entre 4 de abril de 2020 e 8 de março de 2021.

Figura 4 - Quantitativo de triagens ao longo dos dias.

Fonte: do autor (2023).

A Figura 5 apresenta as frequências dos sinais e sintomas da COVID-19 reportados pelos usuários do aplicativo Triagem On-line nos casos classificados como leves ou graves pelo aplicativo. É importante ressaltar que o sintoma de dispneia não está presente em nenhum caso classificado como leve, uma vez que a presença desse sintoma por si só já eleva a gravidade da avaliação para moderada/grave.

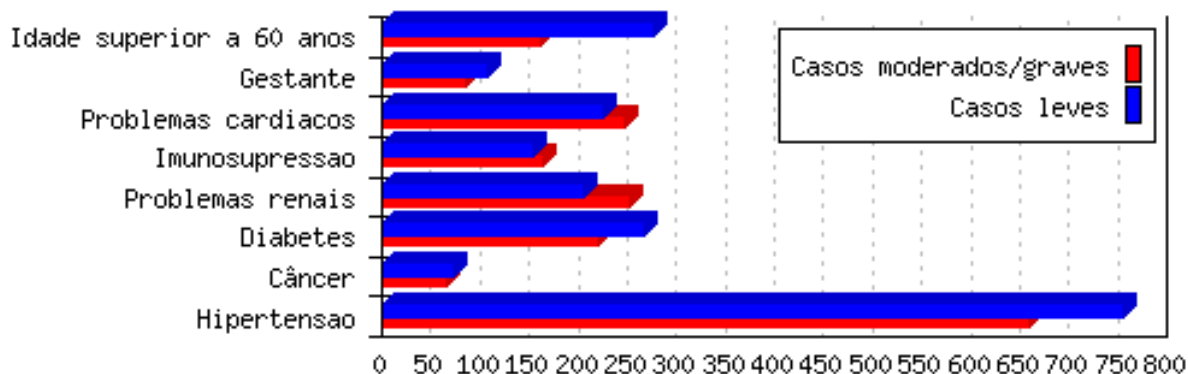
Figura 5 - Frequência dos sinais/sintomas nos casos suspeitos.

Fonte: do autor (2023).

A distribuição dos fatores de risco e comorbidades pode ser observada na Figura 6. O registro das comorbidades permitiu a comparação dessas condições

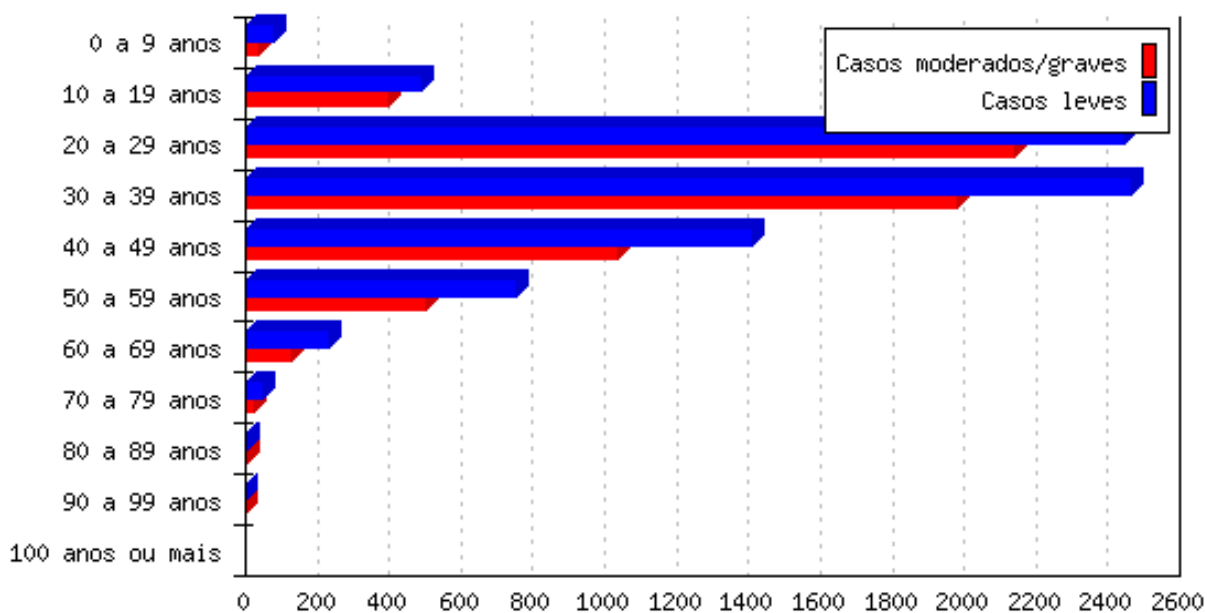
com a gravidade dos sinais e sintomas. Além disso, foi identificado que a hipertensão foi o fator de risco mais comum tanto nos casos leves quanto nos casos moderados e graves.

Figura 6 - Frequência das comorbidades/fatores de risco nos casos suspeitos.



Fonte: do autor (2023).

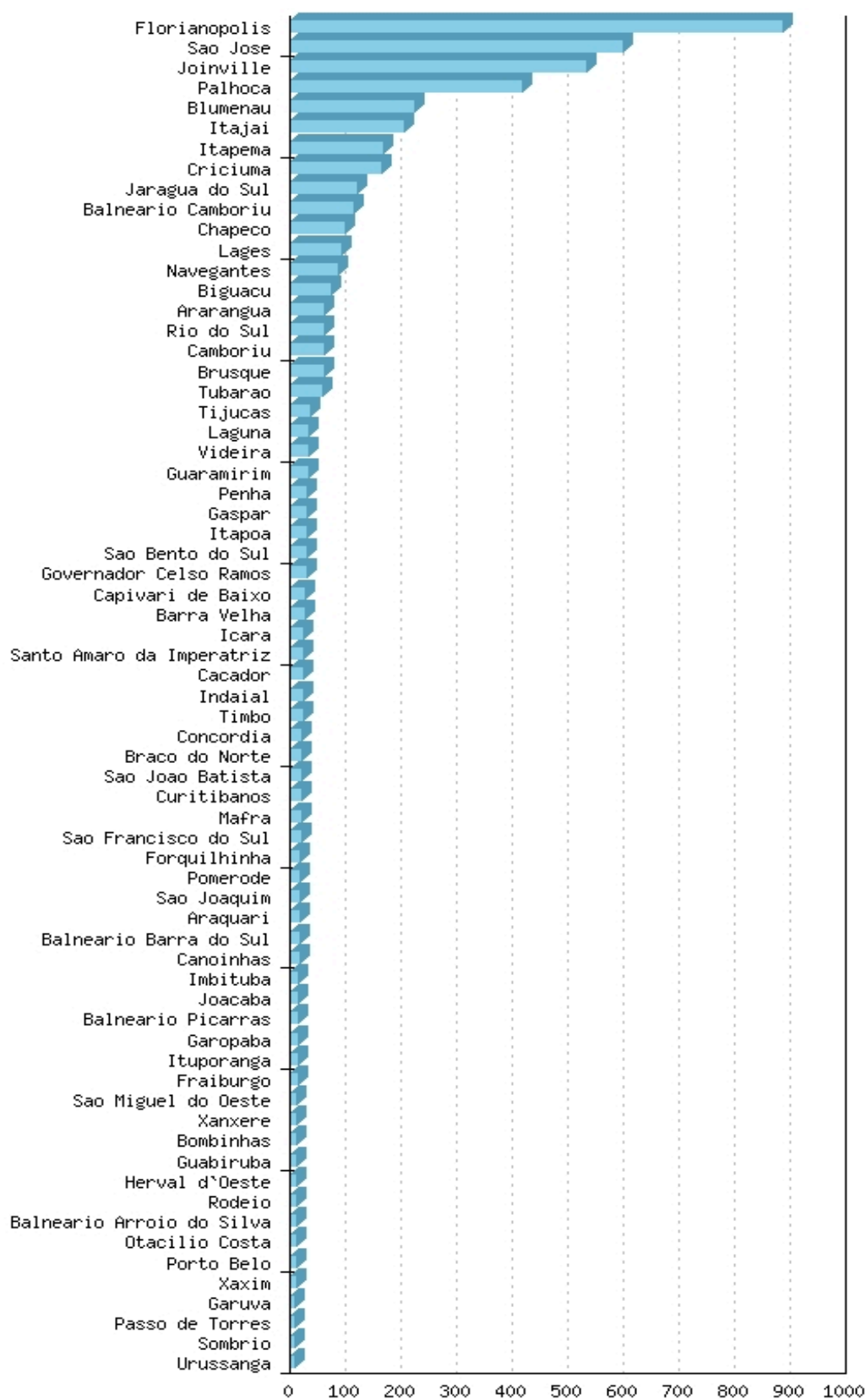
A Figura 7 apresenta a distribuição dos casos suspeitos colhidos pelo aplicativo por faixa etária. Obviamente, esta distribuição não retrata a distribuição da população nem a quantidade de casos suspeitos real por faixa etária, mas as quantidades de casos reportados pelo aplicativo. Além disso, é importante destacar que não houve nenhum caso suspeito cadastrado em pessoas com 100 anos ou mais. A maioria dos casos envolveu pacientes com idades entre 20 e 39 anos.

Figura 7 - Distribuição dos casos suspeitos por faixa etária.

Fonte: do autor (2023).

O aplicativo possibilitou amostrar e estimar a disseminação das infecções pelo Coronavírus em diversas regiões do Estado de Santa Catarina. Na Figura 8 são mostradas as ocorrências de casos classificados pelo aplicativo como moderados ou graves por município. Para facilitar a visualização e por limitação de espaço, os municípios com menos de 10 casos de contaminação pelo vírus foram suprimidos.

Figura 8 - Casos moderados ou graves por município.



Fonte: do autor (2023).

2.7. Medidas de Qualidade de Resultados de Classificadores

2.7.1. Matriz de confusão

A matriz de confusão é uma tabela que descreve o desempenho de um modelo de classificação em termos das suas previsões em relação às classes reais dos dados. É uma ferramenta útil para avaliar o desempenho de um algoritmo de aprendizado de máquina e entender os erros e acertos cometidos pelo modelo.

Uma matriz de confusão típica tem dimensões de $N \times N$, onde N é o número de classes ou categorias presentes no problema de classificação. Cada célula da matriz representa o número de amostras que foram classificadas em uma determinada combinação de classe real e classe prevista.

Vamos considerar um exemplo comum em classificação binária, onde temos duas classes: "positivo" e "negativo". Nesse caso, a matriz de confusão terá uma estrutura 2×2 , com quatro células, conforme pode ser visualizado no Quadro 2:

Quadro 2 - Matriz de confusão para classificação binária.

	Classe Predita: Positivo	Classe Predita: Negativo
Classe Real: Positivo	Verdadeiro Positivo (TP)	Falso Negativo (FN)
Classe Real: Negativo	Falso Positivo (FP)	Verdadeiro Negativo (TN)

Fonte: Adaptado de Garcia (2016).

Os elementos da matriz de confusão são definidos da seguinte forma:

- Verdadeiro Positivo (TP): O número de amostras que foram corretamente classificadas como positivas.
- Falso Positivo (FP): O número de amostras que foram erroneamente classificadas como positivas, quando na verdade são negativas.
- Falso Negativo (FN): O número de amostras que foram erroneamente classificadas como negativas, quando na verdade são positivas.

- Verdadeiro Negativo (TN): O número de amostras que foram corretamente classificadas como negativas.

A partir da matriz de confusão, é possível calcular diversas métricas de avaliação do modelo, como precisão, *recall*, F1-score e acurácia. Essas métricas são calculadas com base nos valores presentes na matriz de confusão (Fawcett, 2006).

2.7.2. Acurácia

A acurácia é uma medida de quão corretas são todas as previsões feitas pelo modelo, independentemente da classe. Ela é calculada dividindo o número total de previsões corretas (TP + TN) pelo número total de amostras na base de dados (TP + TN + FP + FN), onde TP é o número de verdadeiros positivos, TN é o número de verdadeiros negativos, FP é o número de falsos positivos e FN é o número de falsos negativos. A acurácia mede a proporção de previsões corretas em relação ao total de amostras, sendo definida pela fórmula::

$$\text{acurácia} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

2.7.3. Precisão

A precisão é uma medida de quão precisas são as previsões positivas feitas pelo modelo. Ela é calculada dividindo o número de verdadeiros positivos (TP) pelo número total de positivos previstos (TP + FP), onde TP é o número de amostras corretamente classificadas como positivas e FP é o número de amostras incorretamente classificadas como positivas. Portanto, a precisão mede a proporção de previsões positivas corretas em relação ao total de previsões positivas:

$$\text{precisão} = \text{TP} / (\text{TP} + \text{FP})$$

2.7.4. Sensibilidade

A sensibilidade, também conhecida como *recall*, é uma medida de quão bem o modelo consegue identificar corretamente as amostras positivas. Ele é calculado dividindo o número de verdadeiros positivos (TP) pelo número total de amostras positivas presentes na base de dados (TP + FN), onde TP é o número de amostras corretamente classificadas como positivas e FN é o número de amostras erroneamente classificadas como negativas. O *recall* mede a proporção de amostras positivas corretamente identificadas em relação ao total de amostras positivas. A fórmula para o cálculo do *recall* é:

$$\text{recall} = \text{TP} / (\text{TP} + \text{FN})$$

2.7.5. F1-Score

O F1-score é uma métrica que combina precisão e *recall* em uma única medida. Ele é calculado como a média harmônica entre precisão e *recall*. O F1-score fornece uma medida balanceada entre a precisão e o *recall*, sendo útil quando há um desequilíbrio entre as classes do problema. A fórmula para o cálculo do F1-score é a seguinte:

$$\text{F1} = 2 * (\text{precisão} * \text{recall}) / (\text{precisão} + \text{recall}).$$

2.8. Classificadores Baseados em Aprendizagem de Máquina

Aprendizado de máquina, também conhecido como *machine learning*, é uma subárea da inteligência artificial (IA) que se concentra no desenvolvimento de técnicas e algoritmos que possibilitam que os sistemas computacionais aprendam a partir de dados e melhorem seu desempenho ao longo do tempo, sem serem explicitamente programados (Russel; Orvig, 2002).

Em vez de seguir um conjunto determinado de instruções, por exemplo, os modelos de aprendizado de máquina podem ser alimentados com dados de

treinamento, tais como exemplos com suas respostas corretas. O modelo então aprende a generalizar a partir desses exemplos e pode fazer previsões ou tomar decisões sobre novos dados não vistos anteriormente (Bishop, 2006).

Aprendizado de Máquina pode ser supervisionado ou não supervisionado. O Aprendizado Supervisionado treina um modelo usando dados rotulados, ou seja, dados de entrada que já possuem uma saída desejada conhecida. O objetivo é ensinar ao modelo como mapear corretamente os dados de entrada para as saídas desejadas. O processo de treinamento envolve apresentar ao modelo exemplos de entrada com as saídas corretas correspondentes; permitindo que ele aprenda a fazer previsões ou tomar decisões com base nesses exemplos.

O Aprendizado Não Supervisionado, por outro lado, constrói os modelos usando dados não rotulados, ou seja, dados de entrada que não possuem as saídas desejadas (respostas) conhecidas. O objetivo é encontrar padrões, estruturas ocultas ou grupos nos dados de entrada sem a necessidade de rótulos pré-existentes. O modelo deve descobrir informações significativas por conta própria, identificando semelhanças, diferenças ou relações entre os dados (Lorena; Carvalho, 2007).

Além desses dois principais tipos, há outras possibilidades, como o Aprendizado Semi-Supervisionado (que combina elementos de ambos os tipos previamente mencionados) e o Aprendizado por Reforço, onde um agente aprende a tomar ações em um ambiente dinâmico com o objetivo de maximizar uma recompensa cumulativa (Silva *et al.*, 2016).

Neste trabalho explorou-se principalmente tarefas de aprendizado supervisionado, dentre as quais regressão e classificação são as principais. A regressão é usada quando o objetivo é prever um valor numérico contínuo com base nos dados de entrada. O modelo de Aprendizado de Máquina é treinado usando dados rotulados em que as entradas são mapeadas para um valor numérico correspondente.

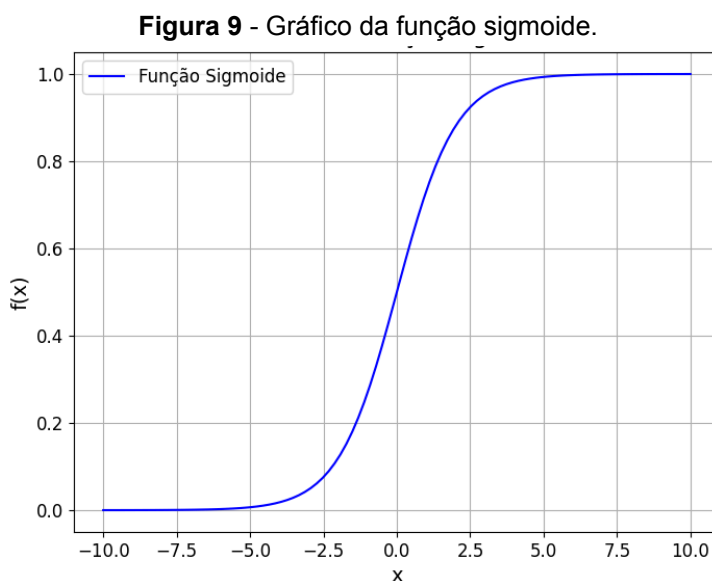
Já a classificação é usada quando o objetivo é atribuir uma determinada categoria ou classe a uma entrada, mapeando as variáveis de entrada em categorias discretas. O modelo de Aprendizado de Máquina é treinado usando dados de entrada rotulados, nos quais cada exemplo possui uma classe ou

categoria associada. O objetivo é aprender a mapear corretamente as características dos dados de entrada para suas respectivas classes. A classificação pode ser binária ou multiclasse.

A seguir serão apresentados os modelos baseados em *Machine Learning* utilizados neste trabalho.

2.8.1. *Logistic Regression*

A *Logistic Regression* (ou Regressão Logística) é um algoritmo de classificação que busca prever a probabilidade de um evento ocorrer. Bao, Lianju e Yue (2019) afirmam que, embora tenha a palavra "regressão" no nome, ele se demonstra mais apropriado para tarefas de classificação binária. A ideia por trás da Regressão Logística é usar uma função sigmoide para transformar a saída de uma regressão linear, restringindo-a ao intervalo entre 0 e 1, conforme pode ser observado na Figura 9.



Fonte: do autor (2023).

Destaca-se ainda o fato da função sigmoide ser também diferenciável, fornecendo um gradiente suave, evitando saltos nos valores de saída. A representação matemática da função é apresentada através da Equação 1.

Equação 1 - função sigmoide.

$$f(x) = \frac{1}{1 + e^{-x}}$$

Durante o treinamento, o modelo estima parâmetros que maximizam a verossimilhança dos dados observados. O entendimento básico por trás do uso de probabilidade máxima para ajustar um modelo de regressão logística é buscar estimativas para os parâmetros de entrada de modo que a probabilidade prevista seja o mais próximo possível das características da classe analisada (James *et al.*, 2013). Na fase de previsão, o modelo combina linearmente os recursos de entrada e passa esse valor pela função sigmoide, interpretando a saída como a probabilidade de enquadrar na classe positiva.

2.8.2. *Decision Tree*

Decision Trees (ou Árvores de Decisão) são modelos que fazem previsões com base em decisões sequenciais relacionadas aos recursos de entrada e são frequentemente adotados em tarefas de mineração de dados. Durante o processo de treinamento, o algoritmo seleciona de forma recursiva o recurso que proporciona a maior redução em heterogeneidade, que pode ser medida por critérios como o ganho de informação ou a impureza de Gini (Santhanam, 2017). Este processo de divisão continua até que se atinja um critério de parada, como a profundidade máxima da árvore ou um número mínimo de amostras em um nó folha. Os nós finais, ou folhas, da árvore representam a decisão tomada pelo modelo.

Para Körting *et al.* (2013), a popularidade dos algoritmos baseados em árvore de decisão ocorre por conta de sua eficiência em termos de tempo, à clareza de como se apresentam, e pelas regras de classificação diretas que implementam. Estas regras não se baseiam em pressupostos sobre distribuições estatísticas ou na independência das classes. Além disso, os atributos usados pelo modelo podem ser facilmente reconhecidos e podem abranger tanto valores numéricos quanto categorias.

2.8.3. *Support Vector Machine* - SVM

Support Vector Machine (ou Máquina de Vetores de Suporte), ou SVM, é um algoritmo de aprendizado supervisionado que é amplamente reconhecido por sua capacidade em tarefas de classificação (Lantz, 2019). De acordo com Haykin (2001), a principal ideia por trás do SVM é encontrar um hiperplano que melhor separa as classes no espaço de características. O objetivo é maximizar a margem entre as classes, sendo esta margem definida como a distância entre o hiperplano e os pontos de dados mais próximos de cada classe, conhecidos como vetores de suporte. Para situações em que os dados não são linearmente separáveis, o SVM utiliza uma técnica chamada "kernel trick", que mapeia os dados para um espaço de características de dimensão superior onde se tornam separáveis (Mahesh, 2020). Uma vez treinado, o SVM classifica novas amostras com base em qual lado do hiperplano elas se encontram.

3. REVISÃO DA LITERATURA

Este capítulo contém o processo de revisão bibliográfica, apresenta os seus resultados e compara a presente proposta com trabalhos relacionados. Ele está dividido em duas seções, referentes a pesquisas bibliográficas sobre classificação de casos suspeitos de COVID-19 (para tentar prever se o paciente é portador do vírus) e classificação da gravidade de casos de síndrome respiratória, respectivamente. Em ambos os casos o objetivo é encontrar trabalhos relacionados em classificação de casos a partir de dados de perfil, sinais e sintomas reportados por pacientes.

3.1 Classificação de Casos Suspeitos de COVID-19

Inicialmente, tentou-se vislumbrar a possibilidade de construir um classificador que pudesse diferenciar casos positivos de contaminação pelo Coronavírus. Para tanto fora desenvolvida uma revisão sistemática da literatura — instrumento que permite uma seleção planejada de trabalhos sobre o tema, por meio de métodos explícitos e replicáveis de busca, seleção e análise, conforme Mendes-Da-Silva (2019).

Foram adotados aspectos recomendados pelo modelo PRISMA (Moher *et al.*, 2009), notadamente os critérios de elegibilidade, análise e comunicação da pesquisa aplicáveis à revisões sistemáticas.

Os critérios de elegibilidade aplicados são:

- a) **Tipo de estudo:** as referências a serem selecionadas devem acessar um dos termos em inglês COVID-19, SARS-CoV-2, Corona Virus Disease, Coronavirus Disease e tratar sobre screening app symptoms, screening questions, app symptoms predictor, prediction symptoms, machine learning, artificial intelligence, KNN, *Naive Bayes*, *Logistic Regression*, *Support Vector Machines*, *Decision Trees* e classification problem. Como critério de exclusão, trabalhos que relacionam a

COVID-19, SARS-CoV-2, Corona Virus Disease, Coronavirus Disease a outras áreas que não a de interesse da pesquisa não são mantidos, tais como reconhecimento de padrão em imagem radiográfica, baseados em resultados de exames laboratoriais, entre outras;

- b) **Tópico:** a identificação e seleção dos trabalhos é feita por uma leitura nos títulos e resumos;
- c) **Design de pesquisa:** são elegíveis estudos teóricos e empíricos que abordem a identificação das abordagens que têm propiciado os melhores resultados na classificação de gravidade de casos de síndromes respiratórias a partir não apenas do perfil dos pacientes, como também dos sinais e sintomas;
- d) **Recorte temporal e idioma:** sem recortes temporais; inglês;
- e) **Status da publicação:** artigos científicos publicados em *journals* ou capítulos de livros.

No que tange às fontes de informação, iniciou-se a busca nas bases indicadas no Quadro 3, no mês de janeiro de 2023, com o intuito de mapear trabalhos relacionados:

Consulta 1: ('COVID-19' OR 'SARS-CoV-2' OR 'Corona Virus Disease' OR 'Coronavirus Disease') **AND** ('screening app symptoms' OR 'screening questions' OR 'app symptoms predictor' OR 'prediction symptoms') **AND** ('machine learning' OR 'artificial intelligence' OR 'KNN' OR 'Naive Bayes' OR 'Logistic Regression' OR 'Support Vector Machines' OR 'Decision Trees' OR 'classification problem')

Consulta 2: ('COVID-19' OR 'SARS-CoV-2' OR 'Corona Virus' OR 'Coronavirus') **AND** ('screening app' OR 'predictor' OR 'classifier') **AND** ('symptoms' OR 'screening questions')

As quantidades de artigos resultantes das consultas e selecionados depois de aferir título e resumo de cada um são listadas no Quadro 3. As consultas às bases de artigos resultaram em total de 10.808 artigos dos quais, após a filtragem por títulos restaram 32 artigos. Após realizar a leitura dos resumos remanesceram 8

artigos.

Quadro 3 - Resultado da revisão bibliográfica sobre COVID-19.

Bases	Resultados consulta 1	Resultados consulta 2	Selecionados por título	Selecionados por resumo
Springer Nature	8 artigos	825 artigos	2 artigos	2 artigos
IEEE Xplore	1 artigo	29 artigos	10 artigos	4 artigos
Scopus (Elsevier)	3 artigos	1.880 artigos	1 artigo	zero
ACM Digital Library	12 artigos	184 artigos	9 artigos	1 artigo
BDBComp	zero	zero	zero	zero
Google Scholar	140 artigos	5.240 artigos	3 artigos	zero
SciELO.org	zero	12 artigos	zero	zero
Web of Science	353 artigos	2.074 artigos	5 artigos	1 artigo
EBSCO (Computers & Applied Sciences Complete)	4 artigos	43 artigos	3 artigos	1 artigo
Spell	zero	zero	zero	zero

Fonte: do autor (2023).

O Quadro 4 compara os diversos trabalhos, as diferentes classes utilizadas nas classificações, modelos de aprendizagem de máquina utilizados, tamanho e período de coleta das amostras e medidas de desempenho obtidas, tais como acurácia, precisão, *recall* e F1-score.

Quadro 4 - Quadro comparativo preliminar.

Trabalho	Classes	Modelo(s)	Amostras período	Acurácia	Precisão	Recall	F1-score
Mohapatra, Mohanty e Mohanty, 2021	negativo positivo	SVM	2.889	92,00	-	96	-
		DNN	June 14th 20	97,00	-	98	-
Soui, <i>et al.</i> 2021.	negativo positivo	MultiLayer Perceptron	99.232 Feb 20 - Aug 20	89,36	89,36	89,36	89,36
		SVM		92,62	92,62	92,62	92,62
		Logistic Regression		92,88	92,88	92,88	92,88
		Decision tree		85,96	85,96	85,96	85,96
		Gradient boosting		92,41	92,41	92,41	92,41
		Random forest		89,36	89,36	89,36	89,36
		XGboost		92,36	92,36	92,36	92,36
		AdaBoost		89,35	89,35	89,35	89,35

Oliveira, <i>et al.</i> 2022	negativo positivo	Decision Tree	13.434 March 20 - January 21	96,25	99,91	93,42	-
		Random Forest		91,36	91,97	93,42	-
		Random Tree		68,26	98,99	50,80	-
		Naïve Bayes		86,32	98,80	79,09	-
		Naïve Bayes-Kernel		87,45	98,36	82,57	-
		Deep Learning		89,07	97,57	89,37	-
Mendoza <i>et al.</i> , 2022.	negativo positivo	LightGBM	1.434.868 Jun 20 - Jan 21	75,21	-	9,21	-
		AdaBoost		75,21	-	9,21	-
		Gaussian Naïve-Bayes		75,21	-	9,21	-
		MultiLayer Perceptron		75,72	-	9,02	-
		Quadratic Discriminant		75,21	-	9,21	-
		Decision Tree		75,36	-	9,21	-
		K-mode		5.966 período não informado	71,00	-	77,00
Adam <i>et al.</i> , 2021.	negativo positivo	SVC		76,00	-	79,00	-
Sun, Ding e Zhou, 2021.	negativo positivo outros	SVM	4.704.597 June 18th 21	89,72	86,16	89,72	87,37
		Logistic Regression		89,55	85,36	89,55	86,31
		Naïve Bayesian		89,39	85,58	89,39	86,91
		Decision Tree		89,71	86,05	89,71	87,20
		Random Forest		89,73	85,94	89,73	87,00
		Neural Network		89,72	86,21	89,72	87,41
	negativo positivo	SVM		91,32	89,38	91,32	89,80
		Logistic Regression		91,21	88,58	91,21	88,59
		Naïve Bayesian		91,33	88,95	91,09	89,44
		Decision Tree		91,33	89,31	91,33	89,68
		Random Forest		91,35	89,20	91,35	89,49
		Neural Network		91,39	89,36	91,39	89,64
		Yeramalli, Intodia e Shambharkar, 2022.		negativo positivo	Naïve Bayes	5.434 período não informado	75,27
gradient boosting	98,46		99,00		-		-
KNN	98,36		99,00		-		-
SVM	98,47		99,00		-		-
Decision tree	98,30		99,00		-		-
Logistic Regression	97,03		97,00		-		-
Random Forest	98,39		99,00		-		-
Villavicencio, <i>et al.</i> , 2022.	negativo positivo	J48 DT	8.766 ? - 30 June 2021	98,40	-	100,00	-
		Random Forest		98,75	-	100,00	-
		SVM		98,75	-	100,00	-
		KNN		98,75	-	100,00	-
		Naïve Bayes		95,00	-	96,12	-
		ANN		99,00	-	100,00	-

Fonte: do autor (2023).

Em um desses artigos estudados, com o título “*A Machine Learning Approach in Evaluating Symptom Screening in Predicting COVID-19*”, Mendoza *et al.* (2022) concluíram que a baixa sensibilidade mostra claramente que os sintomas não se correlacionam com os resultados do teste de RT-PCR. Assim, concluiu-se que a triagem de sintomas não é um processo clinicamente adequado para determinar se um indivíduo tem COVID-19. Isso significa que a utilização de classificadores de casos positivos ou negativos de COVID-19, por apresentarem baixa sensibilidade, poderiam trazer falsos negativos, contribuindo para a disseminação da doença. Pacientes avaliados incorretamente por esses classificadores poderiam deixar de manter os afastamentos de outras pessoas por pensarem não estar contaminados, colocando outros em risco. Dessa maneira, optou-se por mudar o escopo do estudo para classificadores de gravidade de casos de COVID-19 com base em sinais, sintomas e fatores de risco.

3.2 Classificação de Gravidade de Síndromes Respiratórias

A novas expressões de consulta realizadas na revisão bibliográfica, visando identificar artigos sobre classificação de gravidade de síndromes respiratórias, nos moldes apresentados no item 3.1, foram as seguintes:

Consulta 1: ('COVID-19' OR 'SARS-CoV-2' OR 'Corona Virus Disease' OR 'Coronavirus Disease') **AND** ('screening app symptoms' OR 'screening questions' OR 'app symptoms predictor' OR 'prediction symptoms') **AND** ('machine learning' OR 'artificial intelligence' OR 'KNN' OR 'Naive Bayes' OR 'Logistic Regression' OR 'Support Vector Machines' OR 'Decision Trees' OR 'classification problem') **AND** ('severity' OR 'gravity')

Consulta 2: ('COVID-19' OR 'SARS-CoV-2' OR 'Corona Virus' OR 'Coronavirus') **AND** ('screening app' OR 'predictor' OR 'classifier') **AND** ('symptoms' OR 'screening questions') **AND** ('severity' OR 'gravity')

Essas novas consultas retornaram os resultados listados no Quadro 5. As novas consultas às bases de artigos resultaram em 1.406 artigos dos quais, após a filtragem por títulos restaram 29 artigos. Após realizar a leitura dos resumos sobraram 5 artigos que seriam estudados na íntegra, conforme demonstrado no Quadro 5.

Quadro 5 - Resultado consulta definitiva.

Bases	Resultados consulta 1	Resultados consulta 2	Selecionados por título	Selecionados por resumo
Springer Nature	8	123	zero	zero
IEEE Xplore	9	3	4	1
Scopus (Elsevier)	1	1	zero	zero
ACM Digital Library (WebMedia)	12	149	1	zero
BDBComp	zero	zero	zero	zero
Google Scholar	164	334	7	zero
SciELO.org	zero	3	zero	zero
Web of Science	80	514	18	4
EBSCO (Computers & Applied Sciences Complete)	1	4	zero	zero
Spell	zero	zero	zero	zero
SBC.org (SBD/BRACIS)	zero	zero	zero	zero

Fonte: do autor (2023).

Dos trabalhos selecionados foram analisados os modelos baseados em *Machine Learning* que foram utilizados, bem como as métricas dos desempenhos de cada um. O Quadro 6 apresenta um comparativo entre as métricas dos diversos modelos utilizados nos artigos selecionados na pesquisa bibliográfica.

Quadro 6 - Quadro comparativo definitivo.

Trabalho	Classes	Modelo(s)	Amostras período	Acurácia	Precisão	Recall	F1-Score
Rochmawati <i>et al.</i> 2020.	nenhum leve moderado grave	J48	31.740	83,60	84,60	83,60	-
		Hoeffding Tree	13 Aug 20 (66% train 34% test)	83,07	84,90	83,10	-
Gupta e Sinha, 2022.	nenhum leve moderado grave	Logistic Regression	316.800 2020	24,91	24,91	24,91	24,91
		Decision Tree		74,83	66,38	99,99	79,79
		SVM		80,07	75,33	90,73	82,31
Lian, <i>et al.</i> , 2021.	não grave grave	Logistic Regression	1.046 July to September in 2020	"AUC values were greater than 0.9, indicating high accuracy"	-	82,70	-

Yusuf, Rosli e Yusop, 2022.	assintomático leve grave	Random Forest	478 Jan 20 - Feb 20	78,40	91,00	99,00	95,00
		Gradient Boosting		77,50	91,00	99,00	95,00
		Naïve Bayes		76,10	89,00	100	94,00
Suma, <i>et al.</i> , 2021.	leve moderado grave	<i>Logistic Regression</i>	65.000 2020 (70% train 30% test)	97,82	96,00	97,00	-

Fonte: do autor (2023).

Assim, decidiu-se comparar os resultados dos 5 trabalhos relacionados que usam tais modelos para predizer casos graves de síndrome respiratória aguda com os modelos treinados com o conjunto de dados da Secretaria da Saúde do Estado de Santa Catarina e, posteriormente, comparar com a estratégia utilizada pelo aplicativo Triagem On-Line. Uma análise qualitativa dos trabalhos relacionados está apresentada no Quadro 7:

Quadro 7 - Análise qualitativa dos artigos selecionados

Trabalho	Metodologia	Resultados	Sinais e Sintomas	Critério de Classificação	Principais Descobertas
Rochmawati et al., 2020.	O estudo utilizou um conjunto de dados de sintomas clínicos para classificar os sintomas usando um algoritmo de Árvore de Decisão. As árvores de decisão utilizadas na pesquisa são J48 e Hoeffding Tree. A pesquisa foi realizada por meio do aplicativo Weka, e o conjunto de dados usado é proveniente do site Kaggle. O pré-processamento dos dados envolveu limpeza e seleção de atributos.	A Árvore de Decisão é um dos métodos de classificação mais populares devido à sua facilidade de interpretação por humanos. O resultado do J48 foi ligeiramente melhor do que a árvore de Hoeffding em termos de exatidão, precisão e recuperação. A Árvore Hoeffding é mais simples e tem menos nós do que J48.	Febre, cansaço, tosse, tosse seca vespertina, dificuldade para respirar, dor de garganta, assintomático, dores, congestão nasal, coriza, diarreia e idade.	O classificador avalia a presença ou não dos sinais ou sintomas. Sobre a idade, ele verifica se pertence ou não a determinadas faixa-etárias. Posteriormente e classifica como: nenhum, leve, moderado ou grave.	A Árvore de Decisão J48 teve um desempenho ligeiramente melhor do que a Árvore Hoeffding em termos de classificação.

Gupta e Sinha, 2022.	O estudo visa desenvolver um modelo de previsão multiclasse baseado em modelos de aprendizado de máquina, incluindo regressão logística, máquina de vetores de suporte e árvores de decisão, bem como um modelo de rede neural profunda recorrente para comparação de desempenho com os modelos baseados em Autoencoder (AE)	A regressão logística baseada em autoencoder, árvores de decisão e máquina de vetor de suporte mostrou melhoria no desempenho e precisão da previsão para antecipar a disseminação do coronavírus. Os modelos de aprendizado profundo baseados em AE superaram algumas das técnicas existentes de aprendizado de máquina e aprendizado profundo, alcançando uma precisão máxima de 99,66%.	O modelo proposto considera sintomas como febre, tosse seca, cansaço, dificuldade para respirar, dores, dor de garganta, congestão nasal, diarreia e coriza	O modelo de previsão pode classificar a gravidade do coronavírus em diferentes estados, incluindo: - Não sofre de coronavírus. - Sintomas leves de coronavírus. Gravidade moderada que pode exigir internação do paciente. - Sintomas altamente graves que requerem cuidados urgentes.	Os resultados experimentais concluem que modelos de aprendizado profundo baseados em AE propostos combinados com regressão logística, SVM e árvores de decisão são mais adequados para previsão multiclasse de dados centrados no paciente em grande escala com melhor precisão em comparação com as contrapartes existentes.
Lian, <i>et al.</i> , 2021.	O estudo analisou dados clínicos de 1046 pacientes com COVID-19 internados em um hospital designado de julho a setembro de 2020. Os pacientes foram classificados em grupos de gravidade: assintomáticos, leves, comuns, graves e críticos. Foram coletadas características clínicas dos pacientes e realizada uma análise de regressão logística incondicional multivariada para determinar os fatores de risco.	A análise de regressão logística identificou que a idade, complicações com doença pulmonar obstrutiva crônica (DPOC), tosse e trombose venosa foram fatores de risco significativos para pacientes com COVID-19 desenvolverem condições graves durante a hospitalização. Quando as complicações não foram consideradas, a idade dos pacientes, o número de doenças e as doenças subjacentes foram identificados como fatores de risco.	Os sintomas clínicos mais relevantes associados ao desenvolvimento de condições graves incluíram febre, tosse, fadiga, sintomas gastrointestinais, sintomas do trato respiratório superior, dores musculares, dor de cabeça e aperto no peito.	A classificação clínica dos pacientes foi baseada na "Orientação clínica chinesa para diagnóstico e tratamento de pneumonia por COVID-19 (7ª Edição)" e incluiu categorias como assintomáticos, leves, comuns, graves e críticos.	A análise da curva ROC mostrou que a área sob a curva (AUC) que previu a gravidade dos pacientes com COVID-19 na admissão foi de 0,943, indicando um alto poder preditivo do modelo. O estudo concluiu que as mudanças na condição de pacientes graves com COVID-19 estão relacionadas a vários fatores, como idade, sintomas clínicos e doenças subjacentes.
Yusuf, Rosli e Yusop, 2022.	O estudo utilizou uma metodologia híbrida que combina os métodos cascata e ágil, composta por cinco fases: Requisito, Projeto, Desenvolvimento, Teste e Avaliação. Na fase de desenvolvimento, foram desenvolvidos modelos preditivos usando técnicas de aprendizado de máquina. Os dados brutos foram submetidos a um processo de pré-processamento, seleção de recursos, treinamento do modelo	Foram discutidos o desempenho de três técnicas de aprendizado de máquina para prever a gravidade do COVID-19: Random Forest (RF), Naïve Bayes (NB) e Gradient Boosting (GB). O RF obteve a maior pontuação de desempenho, que foi de 78,4%, em comparação com NB e GB.	Os resultados sugeriram que sete sintomas são substanciais: Febre, Tosse, Fadiga, Dispneia (dificuldade de respirar), Dor de cabeça, H-Temperatur e (temperatura mais alta do paciente), Sexo (embora não seja um sintoma, foi listado como	O estudo propõe um sistema de triagem que utiliza informações clínicas em registros eletrônicos de saúde (EHR) para prever a gravidade do estágio do COVID-19. As classificações foram: assintomático, leve ou grave.	O aprendizado de máquina pode ser aplicado com sucesso para determinar tratamento específico e triagem eficaz. No estudo em comento o RF obteve a melhor pontuação de desempenho. O sistema proposto ajuda os médicos a priorizar pacientes críticos, tornando o processo de triagem mais eficiente e melhorando a qualidade de vida do paciente. Sexo foi considerado como uma característica

	com dados de treinamento e teste, e avaliação do modelo.		uma das características relevantes).		relevantes para se considerar a gravidade.
Suma, <i>et al.</i> , 2021.	O estudo utilizou aprendizado de máquina para analisar um conjunto de dados clínicos composto por 65.000 registros de pacientes com 26 recursos. Foi derivado um conjunto ótimo de características deste conjunto de dados usando uma variante proposta do algoritmo de otimização de colônias artificiais de abelhas. Diferentes modelos foram testados, sendo que a máquina de vetores de suporte apresentou a maior precisão de 96%.	O modelo conseguiu prever três estados de gravidade para pacientes com COVID-19: leve, moderado e grave. A matriz de confusão e os valores de recuperação de precisão (0,96 e 0,97) do classificador binário indicam a eficiência do classificador em prever casos positivos corretamente. A curva de operação do receptor gerada para o modelo de previsão de gravidade mostra uma precisão de 96,0% para classe 1 e 85,0% para pacientes de classe 2.	Os sintomas utilizados foram: Febre, cansaço, tosse seca, dificuldade respiratória, dor de garganta, dores, congestão nasal, coriza e diarreia.	O estudo utilizou um classificador binário modelado com uma máquina de vetores de suporte para a triagem de pacientes com COVID-19. Também foi realizada a previsão de gravidade em pacientes positivos para COVID, classificando-os em leve, moderado ou grave.	O aprendizado de máquina é uma ferramenta eficiente para processar grandes volumes de dados, incluindo sintomas clínicos. O modelo proposto no estudo é eficaz em prever a gravidade da doença em pacientes com COVID-19, ajudando os médicos a tomar decisões informadas sobre o tratamento.

Fonte: do autor (2023)

Os estudos analisados exploram diferentes metodologias para prever a gravidade da COVID-19 em pacientes, variando desde técnicas tradicionais de aprendizado de máquina até abordagens mais avançadas de aprendizado profundo. Os estudos “Covid Symptom Severity Using Decision Tree” (Rochmawati et al., 2020) e “A Screening System for COVID-19 Severity using Machine Learning” (Yusuf, Rosli e Yusop, 2022) compartilham uma abordagem semelhante ao utilizar árvores de decisão. No primeiro, a Árvore de Decisão J48 mostrou-se ligeiramente superior à Árvore Hoeffding. No segundo, o Random Forest (uma extensão das árvores de decisão) obteve a maior pontuação de desempenho em comparação com outras técnicas.

O “Multi-class autoencoder-ensembled prediction model for detection of COVID-19 severity” (Gupta e Sinha, 2022) destaca-se por sua abordagem de aprendizado profundo, alcançando uma precisão impressionante de 99,66%. Este resultado sugere que modelos baseados em aprendizado profundo podem ser particularmente eficazes na previsão da gravidade da COVID-19.

O estudo “The Prediction Model of Risk Factors for COVID-19 Developing into Severe Illness Based on 1046 Patients with COVID-19” (Lian, et al., 2021) adota uma abordagem mais clínica, identificando fatores de risco específicos, como idade

e complicações com DPOC, que podem predispor os pacientes a desenvolver condições graves.

Por fim, o “Nature inspired optimization model for classification and severity prediction in COVID-19 clinical dataset” (Suma, et al., 2021) utiliza uma abordagem de otimização inspirada na natureza, demonstrando que a máquina de vetores de suporte pode alcançar uma alta precisão de 96% em um conjunto de dados clínicos.

Em resumo, enquanto as árvores de decisão e suas variantes, como o Random Forest, são frequentemente utilizadas e mostram-se eficazes, as abordagens de aprendizado profundo e otimização inspirada na natureza também apresentam resultados promissores. A identificação de fatores de risco clínicos específicos, como feito no estudo de Gupta e Sinha (2022), oferece *insights* valiosos para a triagem e tratamento de pacientes. A combinação dessas diferentes metodologias pode oferecer uma abordagem holística para prever e gerenciar a gravidade da COVID-19 em pacientes.

Em síntese, depreende-se do estudo dos trabalhos selecionados (dos 5 da última análise), que três modelos de aprendizagem de máquina se destacaram em relação à sensibilidade (*recall* - crucial em saúde):

- *Decision Tree* (frequência=2)
- *Logistic Regression* (frequência=3)
- *Support Vector Machine* - SVM (frequência=1).

Faz-se mister destacar que em razão do tamanho da amostra do artigo “Screening System for COVID-19 Severity using Machine Learning” (Yusuf, Rosli e Yusop, 2022), mesmo estando ele entre os trabalhos selecionados, optou-se por não se utilizar dos modelos nele defendidos.

4. ETAPAS REALIZADAS

Os modelos utilizados neste trabalho foram desenvolvidos em linguagem Python, utilizando-se a IDE PyCharm no Ubuntu 20.04.6 LTS. Foi utilizada a biblioteca Scikit-Learn 1.2.0, por ser uma das bibliotecas mais populares e amplamente utilizadas para machine learning em Python (IGUAL e SEGUÍ, 2017). Foram utilizadas, também, as bibliotecas Pandas 1.5.2 e numpy 1.24.1.

O *hardware* utilizado para o treinamento e testes foi um computador da marca Acer, modelo Nitro 5 AN515-54-75FJ, com o processador Intel^(R) Core^(TM) i7-9750H, 9th Gen, 2.60GHz, com 16GB de RAM e placa gráfica GeForce GTX 1650 Mobile.

O código-fonte dos experimentos realizados neste trabalho, bem como os conjuntos de dados utilizados estão disponíveis no **GitHub**¹. Os experimentos também podem ser acessados e executados diretamente no **Google Colab**².

Os experimentos realizados neste trabalho foram previamente definidos e seguiram uma adaptação das metodologias proposta por Lantz (2013), Dietrich, Heller e Yang (2015) para a realidade desse estudo. A definição prévia de um processo para gerenciar um projeto se faz necessária para garantir que etapas importantes não deixem de ser seguidas e permita que os experimentos possam ser repetidos por outros pesquisadores. A seguir são descritas as atividades executadas em cada uma das quatro etapas deste experimento.

4.1 Compreensão do Problema

Nesta etapa inicial, foram identificados os problemas a serem resolvidos, como também as metas a serem alcançadas.

¹ <https://github.com/michelaraujoo/classificadores/>

² <https://colab.research.google.com/drive/1tE6WC0khYRkgY-XYeuUXux2wH84-Cdhp>

Considerando a pretensão de análise de um classificador de gravidade de possíveis casos de COVID-19, baseados em sinais e sintomas, fez-se necessário o acesso a um conjunto de dados para treinamento.

Assim sendo, encontrou-se uma base de dados que contivesse sinais, sintomas e gravidade dos casos para servir de regra ouro no treinamento e na avaliação de modelos de aprendizado de máquina.

Os resultados colhidos no procedimento acima descrito foram comparados com o desempenho da estratégia utilizada no aplicativo Triagem On-Line.

Uma busca na *Web* permitiu encontrar algumas fontes de dados anonimizadas que foram minuciosamente avaliadas, objetivando atender a demanda pretendida. Como resultado do esforço, uma tabela comparativa foi construída para comparar as características dessas diferentes bases de dados, a qual é apresentada no Quadro 8.

Quadro 8 - Base de dados encontradas.

Base de dados	Conclusão da análise
Diagnosis of COVID-19 and its clinical spectrum	Não apresentava sinais ou sintomas dos casos.
Dados Abertos de Covid-19 em Alagoas	Não apresentava a gravidade dos casos: apenas se recuperado ou óbito
Dados COVID Beneficência Portuguesa de SP	Não apresentava sinais ou sintomas dos casos.
Dados COVID Hospital das Clínicas da Faculdade de Medicina da USP	Não apresentava sinais ou sintomas dos casos.
Symptoms and COVID Presence	Não apresentava gravidade: apenas se caso era positivo ou negativo para a COVID-19.
Dados COVID Grupo Fleury	Não apresentava sinais ou sintomas dos casos.
Dados COVID Hospital Israelita Albert Einstein	Não apresentava sinais ou sintomas dos casos.
Dados COVID do Estado do Espírito Santo	Não apresentava gravidade: apenas se caso era suspeito, confirmado ou descartado.
Dados COVID Hospital Sírio-Libanês	Não apresentava sinais ou sintomas dos casos.
Dados COVID do Estado de Santa Catarina	1.878.916 registros com sinais, sintomas, fatores de risco, informações de perfis e se houve internação em cada caso.

Fonte: do autor (2023).

Verificou-se, após a análise dos *datasets* disponíveis, que nenhum deles atendia totalmente aos quesitos pretendidos. Entretanto, embora os [Dados COVID do Estado de Santa Catarina](#), não contivessem a gravidade dos casos, informavam, para cada um deles, se houve ou não internação do paciente em comento. A partir da análise desse dado é possível a inferência da gravidade apresentada por cada um dos pacientes.

Desta feita, no presente estudo, os pacientes que não foram internados após avaliação médica são caracterizados como casos leves. Por outro lado, aqueles que tiveram internação após avaliação, são caracterizados como moderados/graves. Considerando a limitação de não ser possível, nesses casos, a diferenciação dos casos moderados dos casos graves, ambos permaneceram agrupados, como sendo casos graves.

4.2 Preparação dos Dados

Há de se destacar que o *dataset* selecionado contém, em formato estruturado e de forma anonimizada, a relação dos casos confirmados de COVID-19 no Estado de Santa Catarina, os quais foram inseridos pela Diretoria de Vigilância Epidemiológica (DIVE), vinculada à Superintendência de Vigilância em Saúde da Secretaria de Estado da Saúde. Tais dados são mantidos no sistema de *big data* do Estado de Santa Catarina (Santa Catarina, 2020). No Anexo 1 é possível acessar informações sobre cada parâmetro por meio do dicionário de dados. Destaca-se ainda que este conjunto de dados tem a limitação de não incluir os sintomas: tosse com sangue, anosmia (perda de olfato) e disgeusia (perda de paladar).

Nesta fase do processo, os dados foram preparados para análise. Isso envolveu atividades como limpeza dos dados, seleção de variáveis relevantes, transformação em um formato adequado, balanceamento e criação de conjuntos de dados de treinamento e teste. As mudanças mais significativas são apresentadas a seguir:

- Foram excluídas as colunas (município, nom_laboratorio, regional_saude,

latitude, longitude, etc), uma vez que não contempladas pelo objetivo da pesquisa;

- As colunas “sintomas” e “comorbidades” traziam os atributos separados por vírgula. Para cada um dos sintomas existentes foi criada uma coluna cujo valor poderia ser “0” ou “1”;
- Para as idades dos indivíduos, foram criadas faixas etárias com base no padrão de grupos etários da OMS;
- Foi criado um campo “gravidade” que pode ser
 - “0”, para casos leves, ou
 - “1” para casos moderados/graves (quando houve internação);
- Foram selecionados apenas os registros cujos sintomas iniciaram no ano de 2020: ano em que o sistema Triagem On-Line operou;

Da execução das operações supracitadas resultou uma tabela com 375.180 registros. O Quadro 9 apresenta os 29 atributos booleanos selecionados para o estudo, que indicam a presença ou ausência de sinais, sintomas ou fatores de riscos dos pacientes no momento da coleta das informações.

Quadro 9 - Parâmetros extraídos do *dataset* para o estudo.

Atributo	Detalhamento	Valores possíveis
tosse	Paciente apresentava tosse.	1/0 (sim ou não)
febre	Paciente apresentava febre.	1/0 (sim ou não)
mialgia	Paciente relatava dor muscular.	1/0 (sim ou não)
dispneia	Paciente sofria de falta de ar.	1/0 (sim ou não)
diarreia	Paciente com diarreia.	1/0 (sim ou não)
cefaleia	Paciente relatava dores de cabeça.	1/0 (sim ou não)

cansaco	Paciente relatava astenia.	1/0 (sim ou não)
coriza	Paciente com coriza.	1/0 (sim ou não)
dor_corpo	Paciente relatava dores corporais.	1/0 (sim ou não)
dor_garganta	Paciente relatava dor de garganta.	1/0 (sim ou não)
idade_0_9	Paciente com idade entre 0 e 9 anos.	1/0 (sim ou não)
idade_10_19	Paciente com idade entre 10 e 19 anos.	1/0 (sim ou não)
idade_20_24	Paciente com idade entre 20 e 24 anos.	1/0 (sim ou não)
idade_25_59	Paciente com idade entre 25 e 59 anos.	1/0 (sim ou não)
idade_60_mais	Paciente com 60 anos ou mais.	1/0 (sim ou não)
gravidez	Paciente grávida.	1/0 (sim ou não)
diabetes	Paciente com quadro de diabetes.	1/0 (sim ou não)
obesidade	Paciente obeso.	1/0 (sim ou não)
imunossupressao	Paciente imunossuprimido.	1/0 (sim ou não)
renal	Paciente com doença renal.	1/0 (sim ou não)
hipertensao	Paciente hipertenso.	1/0 (sim ou não)
cardiaco	Paciente com problemas cardíacos.	1/0 (sim ou não)
cancer	Paciente apresentava quadro de câncer.	1/0 (sim ou não)
hepatica	Paciente com problemas hepáticos.	1/0 (sim ou não)
pneumatica	Paciente com problemas respiratórios.	1/0 (sim ou não)
asma	Paciente asmático.	1/0 (sim ou não)
neurologica	Paciente com problemas neurológicos.	1/0 (sim ou não)
masculino	Paciente do sexo masculino.	1/0 (sim ou não)

gravidade	Paciente foi internado. Considerado como caso grave neste estudo.	1/0 (sim ou não)
-----------	---	------------------

Fonte: do autor (2023).

O último atributo (gravidade) é a variável que tentar-se-á prever.

Da divisão dos dados - *data split*

Seguindo a lógica do que é proposto por Ng (2018), para se desenvolver e testar modelos baseados em *Machine Learning*, faz-se imprescindível a utilização de um conjunto de dados para treinamento e outro conjunto de dados avaliação da efetividade do modelo treinado. Os dados utilizados na avaliação não devem ser utilizados em nenhuma parte do desenvolvimento da aplicação, justamente para que seja possível a verificação da capacidade do modelo em se “generalizar” com qualquer novo conjunto de dados. Dessa forma, o *dataset* escolhido para essa pesquisa será inicialmente dividido em duas partes: uma parte maior, para o desenvolvimento e outra menor, para avaliação do trabalho.

Para execução da divisão dos dados, foram levados em consideração alguns fatores. Para Ng (2018), o conjunto de dados de teste deve ser grande o suficiente para proporcionar alta confiança no desempenho geral do sistema desenvolvido. Recomenda, também, que se tente extrair os conjuntos de desenvolvimento e teste de uma mesma distribuição. Além disso, ressaltou que, embora houvesse uma heurística popular de se utilizar uma divisão de 30% dos dados para o conjunto de testes, isso funciona bem quando se tem um número modesto de exemplos, entre 100 e 10.000 exemplos e reforçou: “A antiga heurística de uma divisão de treinamento/teste de 70%/30% não se aplica a problemas onde há muitos dados; os conjuntos de desenvolvimento e teste podem representar muito menos que 30% dos dados.”.

Para subsidiar a validação dos treinamentos que foram realizados neste trabalho, foi indispensável, portanto, realizar previamente uma separação dos dados que seriam posteriormente utilizados para testar o comportamento dos modelos. Tal divisão permite gerar métricas com dados que não participaram no treinamento ou ajustes dos modelos, demonstrando o quão generalizados ficaram os modelos para

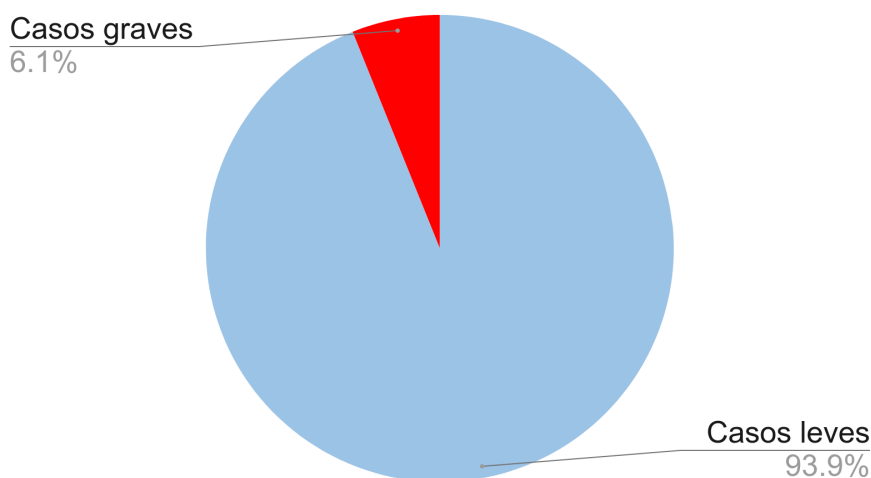
lidar em uma situação real. Considerando que o presente dataset possui 375.180, optou-se por limitar a proporção para avaliação em 10% do conjunto de dados e 90% para o desenvolvimento dos modelos. A divisão foi feita de forma aleatória, sem balanceamento dos dados, e sem o uso de recurso de separação estratificada. Por fim, Ng (2018) conclui que não há necessidade de se ter conjuntos de desenvolvimento/teste excessivamente grandes além do necessário para avaliar o desempenho de algoritmos.

A mesma estratégia será utilizada para a divisão do restante dos dados que serão utilizados para o desenvolvimento: 90% serão utilizados para treinamento (*train*) e 10% para validação (*test*), nesse caso, com os dados balanceados.

Do balanceamento dos dados

Dados balanceados referem-se ao conjunto de dados onde as classes ou categorias de interesse estão representadas em quantidades aproximadamente iguais. Isso significa que há um equilíbrio adequado entre as amostras pertencentes a diferentes classes, evitando viés e melhorando o desempenho do modelo. A utilização de dados balanceados é fundamental ao treinar modelos baseados em *machine learning*. (Haixiang *et al.*, 2017). Dos 337.661 registros separados para o desenvolvimento dos modelos, apenas 20.554 eram de casos graves. O *dataset* estava, portanto, desbalanceado, conforme demonstrado na Figura 10.

Figura 10 - Dados desbalanceados.



Fonte: do autor (2023).

Ao pretender utilizar dados desbalanceados como no caso em tese, ou seja, quando uma classe tem uma representação muito maior do que a outra, o modelo tende a aprender a classificar a classe majoritária (de casos leves, no caso) com maior precisão quando em comparação com as das classes minoritárias (a classe dos casos graves). Isso ocorre porque o modelo é influenciado pela distribuição dos dados e acaba favorecendo a classe mais frequente. Com dados balanceados, o modelo é incentivado a considerar todas as classes igualmente importantes, evitando esse viés indesejado (Haixiang *et al.*, 2017).

Para gerar um *dataset* balanceado, costumam ser usadas as seguintes técnicas:

- **Oversampling**: cria novas observações da classe minoritária a partir das informações contidas nos dados originais.
- **Undersampling**: reduz o desbalanceamento do *dataset* mediante redução das observações da classe majoritária.

De acordo com Chawla *et al.* (2002), utilizar a técnica de *undersampling* tem sido proposto como um bom meio de se aumentar a sensibilidade de um classificador em se tratando da classe minoritária. Por essa razão, optou-se por aplicar a técnica de *undersampling* objetivando o balanceamento das amostras.

A redução da classe majoritária pode ser feita de duas maneiras:

- de forma aleatória: selecionando e removendo aleatoriamente os casos majoritários até que se tenha a mesma quantidade das duas classes;
- pelo uso de algoritmos que removem os casos da classe majoritária utilizando determinadas estratégias.

Foi identificado, por meio de experimentos realizados por Lööv (2020), que o algoritmo de *undersampling* com o melhor desempenho médio é o *NearMiss*. Como o objetivo principal do *NearMiss* é tornar o modelo mais sensível à classe minoritária, era esperado uma melhora na capacidade do modelo de detectar corretamente instâncias dessa classe durante o treinamento e a avaliação.

O resultado da aplicação do algoritmo de *undersampling* foi um *dataset* de desenvolvimento com 45.668 tuplas, dos quais metade eram casos leves e metade

casos graves. Ressalta-se que os 10% dos dados previamente separados para a validação não foram balanceados.

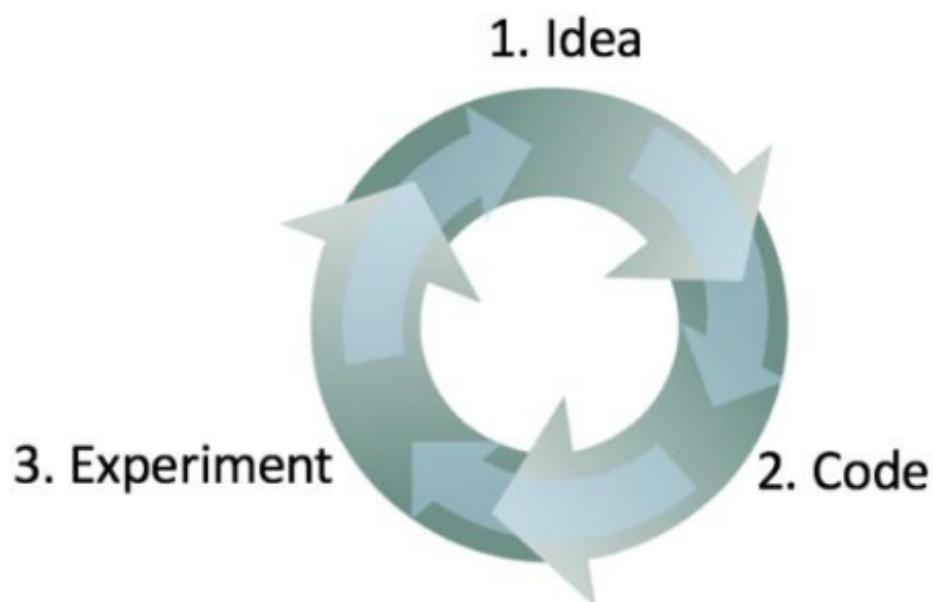
4.3 Treinamento dos Modelos

Ng (2018) sugere um conjunto de passos, os quais foram utilizados neste estudo:

1. Começar com alguma ideia sobre como construir o sistema.
2. Implementar a ideia em código.
3. Realizar experimentos e tentar compreender se a ideia funcionou bem. Com base nesses aprendizados, gerar mais ideias e continuar iterando.

A representação dessa proposta de etapas para treinamento de modelos baseados em *Machine Learning* podem ser observados na Figura 11:

Figura 11 - Proposta de etapas para treinamento.



Fonte: Ng, 2018.

Como afirmado anteriormente, os três modelos escolhidos para serem desenvolvidos, por serem os mais encontrados nos trabalhos selecionados são:

- *Logistic Regression*;
- *Decision Tree*;
- *Support Vector Machine - SVM*;

Na fase de treinamento dos modelos foi fundamental compreender as capacidades de aprendizado dos algoritmos. Foi necessário encontrar um equilíbrio para evitar o *overfitting* e o *underfitting*. O *overfitting* ocorre quando o modelo se ajusta muito bem ao conjunto de treinamento disponível, mas produz resultados insatisfatórios ao ser aplicado a um conjunto de teste. Por contraste, o *underfitting* é observado quando o algoritmo generaliza demais o aprendizado, resultando em uma capacidade fraca de classificar novos casos de forma precisa, tratando todos eles como pertencentes a uma única classe (Reis, 2019). Portanto, se buscou treinar os modelos evitando tanto a especialização excessiva quanto a generalização excessiva, encontrando um equilíbrio adequado entre os dois.

Conforme descrito anteriormente, esse trabalho utilizou validação cruzada (K-Fold Cross Validation). Para encontrar uma divisão apropriada para as proporções (valor de k), uma pesquisa foi realizada e identificou-se que a escolha de k=10 é popular e muitas vezes usada como padrão em muitos contextos de pesquisa e aplicação prática, facilitando a comparação de resultados entre diferentes estudos e experimentos. Por ser uma prática padrão na área de machine learning, devido ao seu equilíbrio entre viés e variância, sua eficiência computacional e sua capacidade de fornecer uma avaliação robusta do desempenho do modelo optou-se por utilizar a divisão em 10 segmentos para a realização da validação cruzada. Witten *et al.* (2005) demonstraram que, após a execução de experimentos amplos, conduzidos em diversas coleções de dados, usando uma variedade de abordagens de aprendizado, demonstraram que k=10 produzem bons resultados, tendo bases teóricas que sustentam essa conclusão e, na prática, k=10 se consolidou como a abordagem convencional mais comumente utilizada.

Dessa forma, para cada modelo, foram realizados 10 treinamentos (K=10), alterando o *fold* de teste, conforme demonstrado na Figura 12.



Fonte: adaptado de Sun e Braatz (2021).

Ressalta-se que o dataset foi balanceado logo após a separação do conjunto de dados para validação. Uma alternativa seria realizar esse balanceamento dentro de cada fold. Essa abordagem pode introduzir alguma variabilidade nos resultados, uma vez que cada fold pode ter uma distribuição ligeiramente diferente das classes.

Para todos os modelos, na tentativa de avaliar se os treinamentos tiveram *overfitting* ou *underfitting*, após a otimização de hiperparâmetros, os melhores desempenhos com o conjunto de desenvolvimento foram registrados para posteriormente serem comparados com o desempenho com o conjunto dos dados de teste. Conforme Ng (2018), se o desempenho do conjunto de desenvolvimento for muito superior que o desempenho dos dados de teste significa que o modelo se ajustou demais ao conjunto de desenvolvimento, nesse caso um novo conjunto de desenvolvimento seria necessário.

Ajustes de Hiperparâmetros

O modelo base, criado sem o auxílio da otimização de hiperparâmetros, apresentou resultados significativos quando comparado aos resultados apresentados em outros trabalhos - estado da arte.

Ainda assim, foram envidados esforços com o intuito de aprimorar os desempenhos dos modelos desenvolvidos utilizando técnicas de busca automática de hiperparâmetros.

A biblioteca *scikit-learn* possui duas funções para executar tal tarefa: `GridSearchCV` e `RandomizedSearchCV`. A `GridSearchCV` é uma abordagem de busca exaustiva em que são especificados os hiperparâmetros que se almeja serem ajustados, com posterior execução pelo algoritmo, utilizando-se, na etapa de treinamento e avaliação do modelo, para cada combinação possível dos hiperparâmetros encontrados, a validação cruzada. Já o `RandomizedSearchCV` aborda a limitação computacional do `GridSearchCV` ao fazer uso de um número fixo de combinações aleatórias dos hiperparâmetros a partir de distribuições especificadas. Dessa forma, devem ser especificados intervalos ou distribuições para cada hiperparâmetro, com posterior seleção, pelo algoritmo, das combinações aleatórias encontradas, combinações essas que serão utilizadas no treinamento e avaliação do modelo.

Considerando a disponibilidade de capacidade de processamento, optou-se por utilizar a função `GridSearchCV`, uma vez que, dessa forma, seria executada uma exploração completa do espaço de hiperparâmetros, conferindo à investigação um caráter de robustez, já que foi possível encontrar e melhor combinação de hiperparâmetros contidos no universo de análise.

Depois de treinados, os desempenhos desses classificadores baseados em *Machine Learning* foram comparados com o desempenho do aplicativo Triagem On-Line.

4.4 Avaliação dos Modelos

Nesta etapa, os modelos desenvolvidos foram avaliados quanto à sua qualidade e desempenho. Isso envolveu a realização de testes adicionais, com dados que não foram utilizados no desenvolvimento, e a análise dos resultados para determinar a adequação dos modelos e identificar possíveis melhorias.

Após a realização dos experimentos, serão calculadas as médias das métricas de avaliação. O desempenho do aplicativo Triagem On-Line foi mensurado

em relação ao mesmo conjunto de dados utilizado no treinamento dos classificadores baseados em *Machine Learning*.

5. RESULTADOS

Este capítulo apresenta primeiramente ([seção 5.1](#)) uma análise exploratória dos dados utilizados no estudo. Em seguida, apresenta os resultados obtidos em três experimentos. No Experimento I ([seção 5.2](#)), os modelos baseados em *Machine Learning* foram treinados apenas com os sinais e sintomas de cada caso clínico, ou seja, com as mesmas características (*features*) utilizadas pelo aplicativo Triagem On-Line para classificar os casos. Já no Experimento II ([seção 5.3](#)) tentou-se utilizar, além dos sinais e sintomas, todas as demais informações disponíveis de cada caso, ou seja, de perfil dos pacientes. O objetivo foi verificar a máxima efetividade que os modelos conseguiriam alcançar com a adição de tais características de perfil. Por fim, no Experimento III ([seção 5.4](#)), tentou-se estimar o quão efetivo poderiam ser os modelos baseados em *Machine Learning* em fases iniciais de pandemias de síndromes respiratórias. Para isso, os modelos foram treinados e testados com os dados disponíveis na base de dados da Secretaria de Saúde do Estado de Santa Catarina até o dia 2 de abril - dia em que os médicos infectologistas criaram a estratégia utilizada no aplicativo Triagem On-Line.

5.1 Análise Exploratória dos Dados

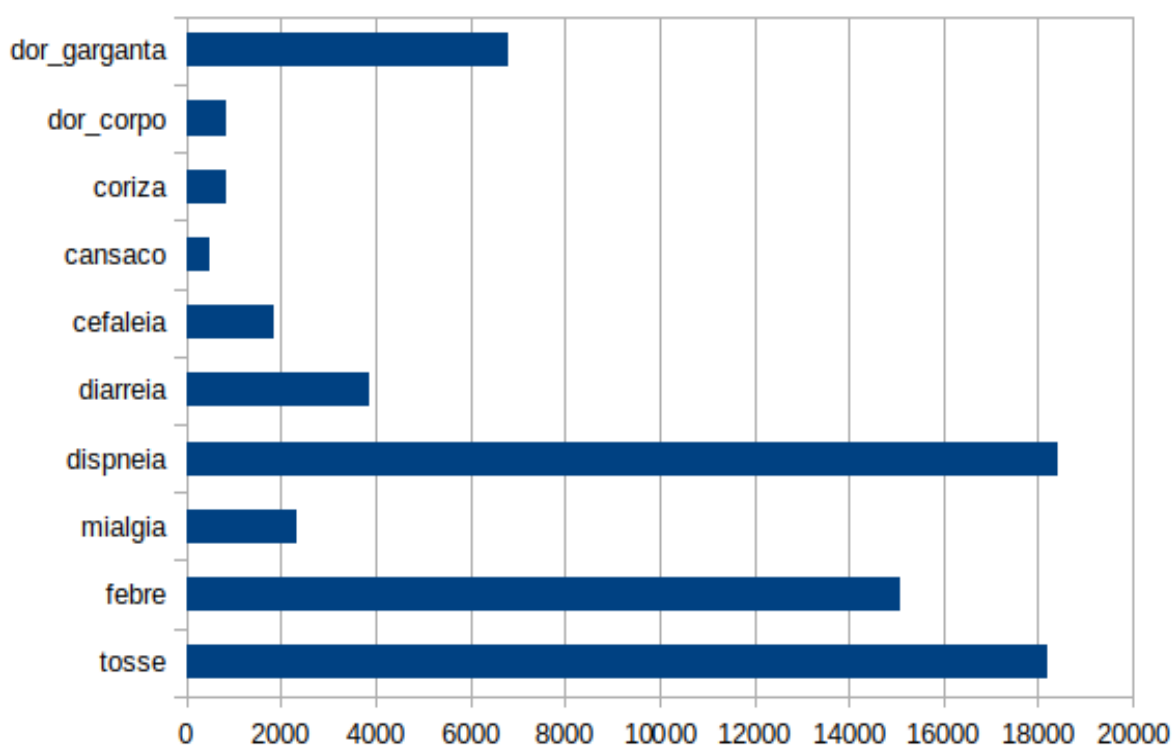
Esta seção apresenta uma análise exploratória dos dados dos casos confirmados de COVID-19 em Santa Catarina que foram disponibilizados pela Diretoria de Vigilância Epidemiológica. Algumas atividades de preparação foram realizadas, como a seleção dos casos cujos sintomas iniciaram em 2020, a exclusão de colunas irrelevantes, a criação de colunas binárias para sinais, sintomas, fatores de risco e categorização de idades em faixas etárias da OMS. Além disso, foi criada uma coluna, também binária, denominada 'gravidade', que identifica os casos de pacientes que necessitam internação por conta da gravidade da doença.

O resultado dessas operações foi uma tabela com 375.180 registros contendo 29 atributos booleanos que indicam a presença ou ausência de sinais, sintomas ou fatores de risco dos pacientes no momento da coleta das informações. É importante

relembrar que o conjunto de dados não informa a presença de tosse com sangue, anosmia (perda de olfato) e disgeusia (perda do paladar), sintomas esses que eram utilizados pelo aplicativo Triagem On-Line para determinar a gravidade dos casos.

Verificou-se, na preparação dos dados, antes da separação das porções que seriam utilizadas para desenvolvimento e testes, que a base estava desbalanceada em relação aos casos graves e não graves da doença. Dos 375.180 registros resultantes da preparação dos dados, apenas 22.834 eram de casos graves. A Figura 13 apresenta as frequências dos sinais e sintomas encontrados nos casos em que houve a internação dos pacientes.

Figura 13 - Frequência de sinais e sintomas em pacientes que tiveram de ser internados.



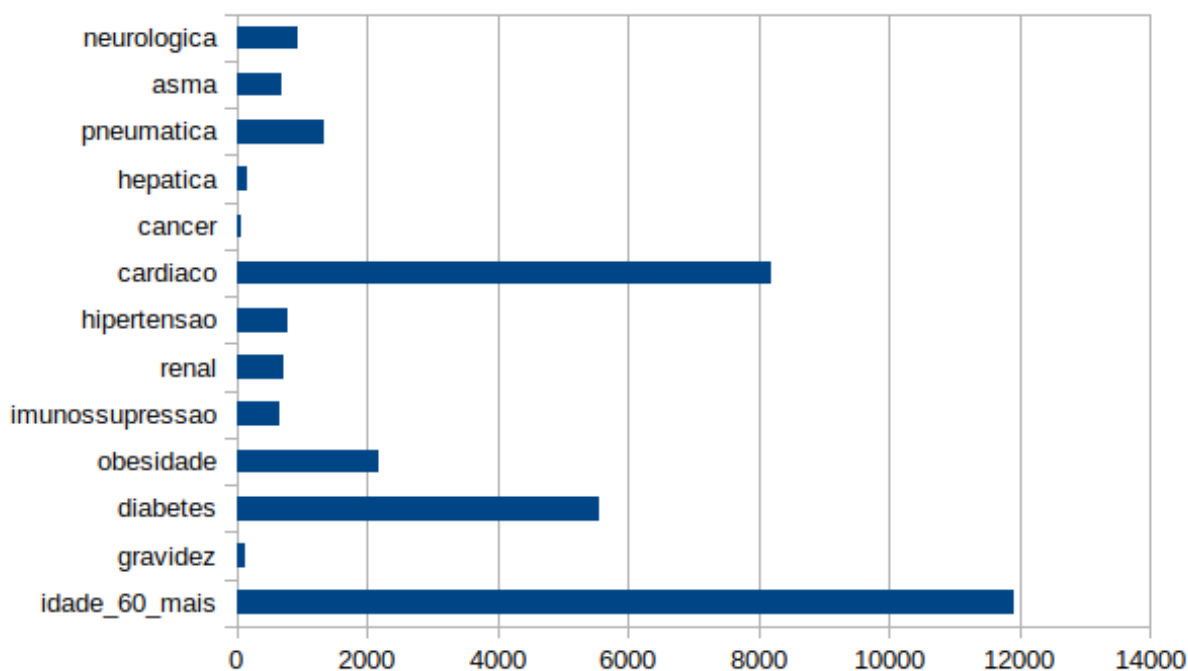
Fonte: do autor (2023).

Depreende-se, da Figura 13, que os sintomas mais frequentes são dispnéia (falta de ar), febre e tosse, o que ratifica a coerência na atribuição dos pesos, pelos médicos infectologistas, utilizados na classificação de gravidade, apresentados no Quadro 1. Ademais, o fluxograma de funcionamento do aplicativo, apresentado na Figura 2, revela que ao atingir a pontuação 6 um caso é classificado como

moderado/grave, o que revela que ao apresentar dispneia, sintoma ao qual fora atribuído o peso 6, o paciente era imediatamente recolhido à internação.

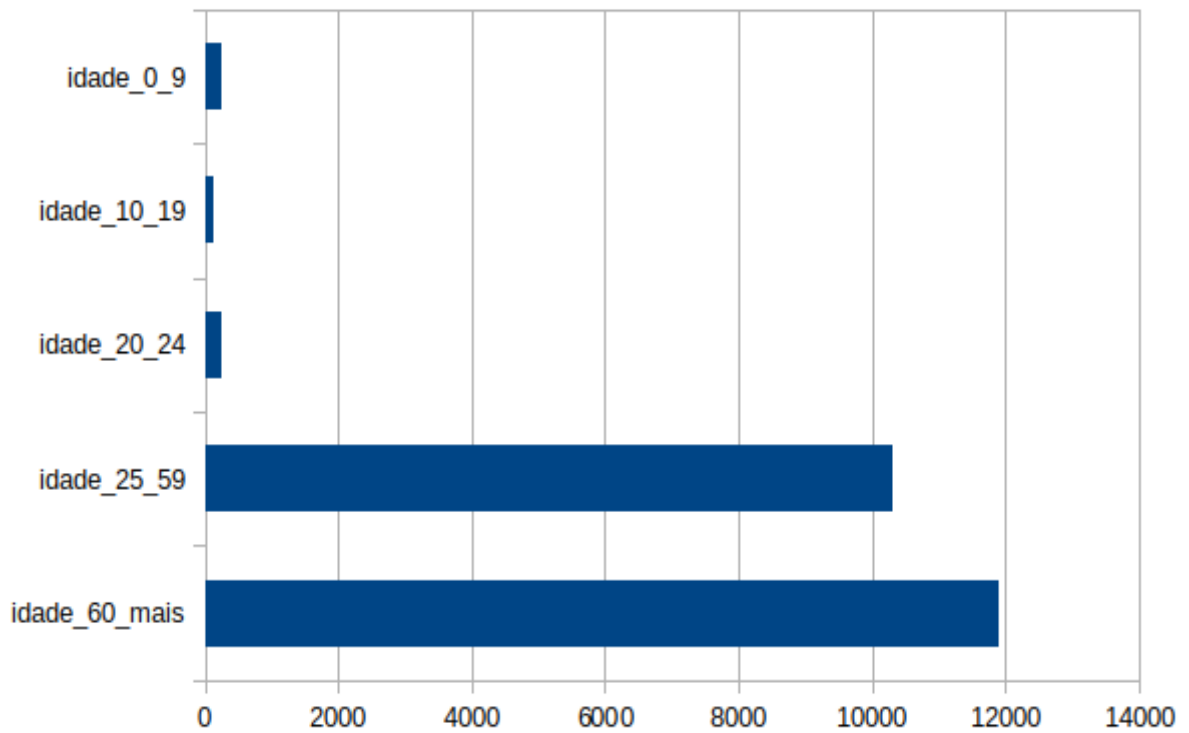
A Figura 14 apresenta as frequências dos fatores de risco encontrados no conjunto de dados dos pacientes que tiveram de ser internados. Note que os fatores de risco mais frequentes entre os pacientes internados foram senilidade, cardíacos e diabetes.

Figura 14 - Frequência dos fatores de risco entre pacientes que tiveram de ser internados.



Fonte: do autor (2023).

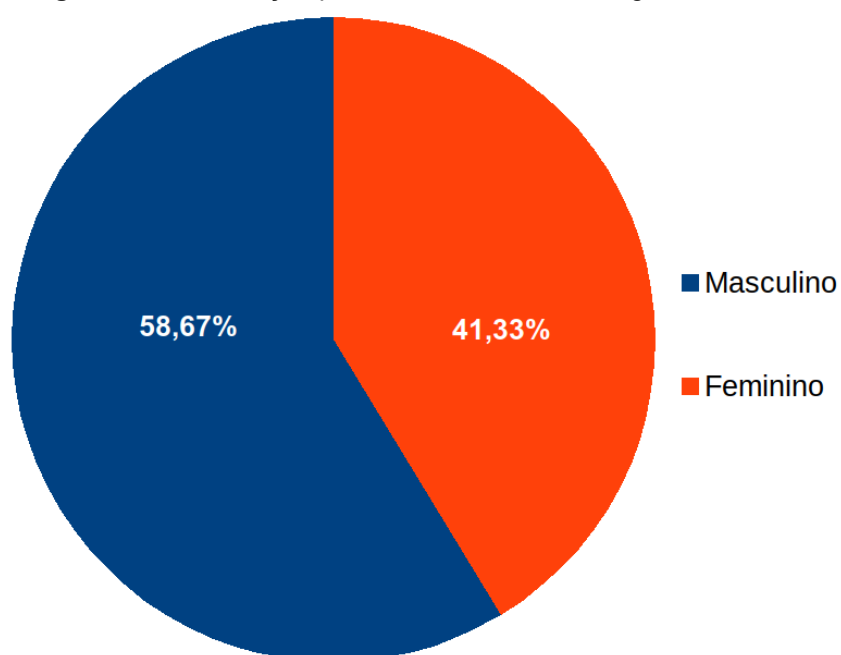
A Figura 15 apresenta as distribuições dos casos graves por faixa etária. Note que a faixa etária dos pacientes graves internados encontra-se entre 25 anos ou mais. Este achado pode ser utilizado como evidência científica para novos estudos relacionados ao aprimoramento de planos de vacinação.

Figura 15 - Distribuição por faixa etária dos casos graves.

Fonte: do autor (2023).

A Figura 16 apresenta a proporção por sexo dos indivíduos que foram internados. Valer ressaltar o maior número de internações de pessoas de sexo masculino, entretanto não há como apresentar conclusões que extrapolem a análise descritiva dos achados.

Figura 16 - Distribuição por faixa etária dos casos graves.



Fonte: do autor (2023).

5.2 Experimento I - Somente sinais e sintomas

Este primeiro experimento teve o objetivo de comparar o desempenho dos modelos baseados em *Machine Learning* com a estratégia utilizada pelo aplicativo Triagem On-Line. Desta feita, foram utilizados os mesmos parâmetros utilizados pelo aplicativo, cujas descrições estão presentes no Quadro 9, a saber:

- 'tosse';
- 'febre';
- 'mialgia';
- 'dispneia';
- 'diarreia'
- 'cefaleia';
- 'cansaco';
- 'coriza';
- 'dor_corpo';
- 'dor_garganta' e
- 'gravidade'.

A seguir são apresentadas as métricas encontradas tanto com os dados de treinamento (durante a validação cruzada), como do conjunto de testes (que não foram utilizados durante o treinamento dos modelos). Ressalta-se que os parâmetros utilizados para a realização dos treinamentos são somente aqueles utilizados pelo aplicativo Triagem On-Line, para que se possa comparar o desempenho das classificações nas mesmas condições. O primeiro treinamento foi em um modelo baseado em *Logistic Regression* (Modelo 1). Depois em árvore de decisão (Modelo 2) e, por fim, com o SVM (Modelo 3).

Modelo 1 - *Logistic Regression*

O Quadro 10 apresenta as métricas geradas pelo treinamento e testes de um modelo baseado em *Logistic Regression*. Para o treinamento do modelo foi utilizada a validação cruzada. O hiperparâmetro C, calculado pela função GridSearchCV, para este conjunto específico de dados, teve o valor igual a 0,01. Para a geração das métricas, tanto o treinamento quanto os testes utilizaram um conjunto de dados distintos entre si.

Quadro 10: Métricas do *Logistic Regression*

	Métricas da validação cruzada	Métricas no conjunto de teste
Acurácia	0,8143	0,8282
Precisão	0,8143	0,6116
Sensibilidade (Recall)	0,8144	0,8179
F1-Score:	0,8143	0,6343

Fonte: do autor (2023).

Dos três modelos utilizados neste estudo, o *Logistic Regression* é o que mais se assemelha à estratégia utilizada no aplicativo Triagem On-Line: ambos definem pesos para cada *feature* (cada sinal/sintoma) e possuem uma abordagem linear para modelar a relação entre as variáveis independentes e a probabilidade de cada classe. Por essa razão os pesos de cada sintoma calculado pelo modelo *Logistic*

Regression foram disponibilizados no Quadro 11 para uma posterior análise. O valor de interceptação encontrado pelo treinamento do modelo foi de -1.8858. O termo de interceptação é um componente que é adicionado à combinação dos pesos e ajusta a função logística de tal forma que ela não passe pela origem ainda que todas as *features* tenham valor igual a zero.

Quadro 11 - Pesos atribuídos pelo modelo Logistic Regression.

Sinal, Sintoma	Peso
Tosse	0,5636
Febre	0,6273
Mialgia	-0,4145
Dispneia	2,7272
Diarreia	0,9899
Cefaleia	-0,3553
Cansaço	0,2551
Coriza	-0,3845
Dor no corpo	-0,1625
Dor de Garganta	-0,6814

Fonte: do autor (2023).

Modelo 2 - *Decision Tree*

O Quadro 12 apresenta as métricas geradas pelo treinamento e testes de um modelo baseado em *Decision Tree*. Para o treinamento do modelo foi utilizada a validação cruzada. O hiperparâmetro `max_depth`, calculado pela função `GridSearchCV`, para este conjunto específico de dados, teve o valor igual a 9.

Para a geração das métricas, tanto o treinamento quanto os testes utilizaram um conjunto de dados distintos entre si, conforme realizado com o modelo anterior.

Quadro 12: Métricas *Decision Tree*

	Métricas da validação cruzada	Métricas no conjunto de teste
Acurácia	0,8184	0,7642
Precisão	0,8186	0,5834

Sensibilidade (Recall)	0,8181	0,7856
F1-Score:	0,8183	0,5785

Fonte: do autor (2023).

Modelo 3 - Support Vector Machine

O Quadro 13 apresenta as métricas geradas pelo treinamento e testes de um modelo baseado em *Support Vector Machine*. Para o treinamento do modelo foi utilizada a validação cruzada. Os hiperparâmetros C e gamma, calculados pela função GridSearchCV, para este conjunto específico de dados, tiveram seus valores iguais a 0,1 e 10, respectivamente. Para a geração das métricas, tanto o treinamento quanto os testes utilizaram um conjunto de dados distintos entre si, conforme realizado com os modelos anteriores.

Quadro 13: Métricas Support Vector Machine

	Métricas da validação cruzada	Métricas no conjunto de teste
Acurácia	0,8201	0,7338
Precisão	0,8164	0,5749
Sensibilidade (Recall)	0,8259	0,7736
F1-Score:	0,8211	0,5566

Fonte: do autor (2023).

Análise dos Resultados Parciais

Observou-se uma discreta diferença entre as métricas extraídas dos dados de treinamento em relação às métricas realizadas no conjunto de testes - que foram utilizadas somente para este propósito. Pode-se considerar o resultado como dentro do esperado para modelos corretamente treinados. Isso significa que os modelos ficaram ajustados para trabalhar com dados desconhecidos: ficaram generalizados o suficiente para classificar novos dados (Ng, 2018). Ressalta-se aqui que os dados de testes foram separados antes do balanceamento.

Métricas do Aplicativo Triagem On-Line

Para que se pudesse comparar o desempenho do aplicativo Triagem On-Line com os modelos anteriormente implementados, necessitar-se-ia aplicar o algoritmo do aplicativo sobre o mesmo conjunto de dados. Para isso, os dados utilizados para testes, aqueles que foram utilizados somente para a validação dos modelos anteriores, foram aplicadas fórmulas com os pesos definidos pelos médicos infectologistas para cada sintoma, de tal forma que se pudesse criar uma matriz de confusão e calcular as mesmas métricas - acurácia, precisão *recall* e F1-Score. O Quadro 14 apresenta o resultado das métricas calculadas.

Quadro 14: Métricas Aplicativo Triagem On-Line

	Métricas do conjunto de testes
Acurácia	0,7619
Precisão	0,1824
Sensibilidade (Recall)	0,8152
F1-Score	0,2981

Fonte: do autor (2023).

Tabela Comparativa

O Quadro 15 apresenta o comparativo entre as métricas de testes dos modelos baseados em *Machine Learning* e as métricas calculadas pela estratégia dos médicos infectologistas da Secretaria de Estado da Saúde de Santa Catarina que foi implementada no aplicativo Triagem On-Line. Ressalta-se que todos esses valores foram levantados sobre o mesmo conjunto de dados.

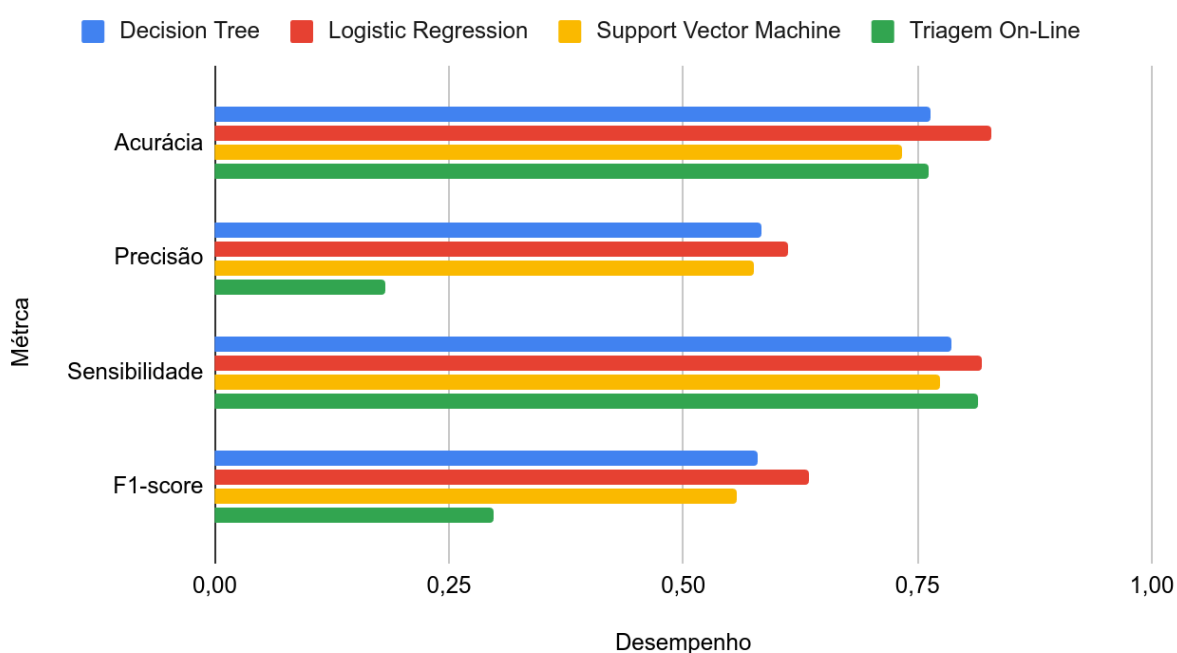
Quadro 15: Tabela Comparativa Final

Modelo/Aplicativo	Acurácia	Precisão	Sensibilidade	F1-score
<i>Decision Tree</i>	0,7642	0,5834	0,7856	0,5785
<i>Logistic Regression</i>	0,8282	0,6116	0,8179	0,6343
<i>Support Vector Machine</i>	0,7338	0,5749	0,7736	0,5566
Triagem On-Line	0,7619	0,1824	0,8152	0,2981

Fonte: do autor (2023).

A Figura 17 apresenta graficamente os resultados encontrados. Percebe-se que o modelo baseado em *Machine Learning*, *Logistic Regression*, obteve o melhor desempenho em todas as métricas.

Figura 17 - Métricas de desempenho.



Fonte: do autor (2023).

5.1.1 Análise Exploratória do Experimento I

Após análise dos resultados, verificou-se que os modelos baseados em *Machine Learning* tiveram um rendimento aproximado entre si no conjunto de dados

disponíveis. Na tentativa de atingir resultados superiores, uma análise mais detalhada do *dataset* foi realizada e alguns fatores mereceram atenção:

- 10 pacientes apresentavam apenas dor de cabeça e foram internados;
- 14 pacientes apresentavam apenas dores no corpo e foram internados;
- 395 pacientes apresentavam apenas febre e foram internados;
- 109 pacientes apresentavam apenas diarreia e foram internados;
- 599 pacientes apresentavam apenas tosse e foram internados;
- 5 pacientes apresentavam apenas coriza e foram internados;
- 8 pacientes apresentavam apenas cansaço e foram internados;
- 45 pacientes apresentavam apenas dor de garganta e foram internados;
- 59.767 pacientes apresentavam dispneia (falta de ar) e não foram internados.

Infere-se, dos fatores supracitados, que ao basear a análise exclusivamente em sinais e sintomas, sem que sejam consideradas as comorbidades, ou fatores de risco, há uma classificação equivocada no que tange à atribuição de gravidade do caso. Ou seja, um paciente com comorbidades pré-existentes pode induzir o médico a, com base em apenas um sintoma, pela decisão de internação, extrapolando a conclusão que a análise tão somente de sinais e sintomas (desconsiderando comorbidades e fatores de risco) ofereceria quando da tomada de decisão. Inclusive depreende-se da análise dos dados que um número considerável (59.767) de pacientes que relataram falta de ar não foram internados, o que pode revelar uma potencial fragilidade para o treinamento de modelos baseados em aprendizado de máquina.

Assim sendo, tornou-se necessária a realização de um novo experimento, denominado Experimento II, que extrapolasse a análise de sinais e sintomas.

5.3 Experimento II - Sinais, sintomas e dados de perfil dos pacientes

Este experimento foi realizado com os mesmos modelos baseados em *Machine Learning*, com os mesmos casos clínicos, mas levando em consideração

não somente os sinais e sintomas utilizados pelo aplicativo Triagem On-Line. Considerará, desta vez, todas as demais informações dos pacientes como idade, sexo, comorbidades e outros fatores de risco. Nos primeiros experimentos, portanto, foram utilizados os parâmetros:

- 'tosse';
- 'febre';
- 'mialgia';
- 'dispneia';
- 'diarreia'
- 'cefaleia';
- 'cansaco';
- 'coriza';
- 'dor_corpo';
- 'dor_garganta' e
- 'gravidade'.

No Experimento II foram acrescentados os parâmetros:

- 'idade_0_9';
- 'idade_10_19';
- 'idade_20_24';
- 'idade_25_59';
- 'idade_60_mais';
- 'gravidez';
- 'diabetes';
- 'obesidade';
- 'imunossupressao';
- 'renal';
- 'hipertensao';
- 'cardiaco';
- 'cancer';
- 'hepatica';
- 'pneumatica';

- 'asma';
- 'neurologica' e
- 'masculino'.

A seguir serão apresentados os resultados dos treinamentos dos três modelos com o acréscimo do conjunto de novos parâmetros.

Modelo 1 - *Logistic Regression*

O Quadro 16 apresenta as métricas geradas pelo treinamento e testes de um modelo baseado em *Logistic Regression* com todos os parâmetros pessoais disponíveis de cada caso. Para o treinamento do modelo foi utilizada a validação cruzada. O hiperparâmetro C, calculado pela função GridSearchCV, para este conjunto específico de dados, teve o valor igual a 17.8864. Para a geração das métricas, tanto o treinamento quanto os testes utilizaram um conjunto de dados distintos entre si, conforme efetuado no Experimento I.

Quadro 16 - Métricas do Logistic Regression

	Métricas da validação cruzada	Métricas no conjunto de teste
Acurácia	0,8732	0,8896
Precisão	0,8862	0,6635
Sensibilidade (Recall)	0,8564	0,8762
F1-Score:	0,8710	0,7113

Fonte: do autor (2023).

Modelo 2 - *Decision Tree*

O Quadro 17 apresenta as métricas geradas pelo treinamento e testes de um modelo baseado em *Decision Tree* com todos os parâmetros pessoais disponíveis de cada caso. Para o treinamento do modelo foi utilizada a validação cruzada. O

hiperparâmetro `max_depth`, calculado pela função `GridSearchCV`, para este conjunto específico de dados, teve o valor igual a 10. Para a geração das métricas, tanto o treinamento quanto os testes utilizaram um conjunto de dados distintos entre si, conforme realizado com o modelo anterior.

Quadro 17 - Métricas *Decision Tree*

	Métricas da validação cruzada	Métricas no conjunto de teste
Acurácia	0,8600	0,8703
Precisão	0,8709	0,6424
Sensibilidade (Recall)	0,8452	0,8560
F1-Score:	0,8579	0,6824

Fonte: do autor (2023).

Modelo 3 - *Support Vector Machine*

O Quadro 18 apresenta as métricas geradas pelo treinamento e testes de um modelo baseado em *Support Vector Machine* com todos os parâmetros pessoais disponíveis de cada caso. Para o treinamento do modelo foi utilizada a validação cruzada. Os hiperparâmetros `C` e `gamma`, calculados pela função `GridSearchCV`, para este conjunto específico de dados, tiveram seus valores iguais a 1,0 e 10, respectivamente. Para a geração das métricas, tanto o treinamento quanto os testes utilizaram um conjunto de dados distintos entre si, conforme realizado com os modelos anteriores.

Quadro 18 - Métricas *Support Vector Machine*

	Métricas da validação cruzada	Métricas no conjunto de teste
Acurácia	0,8522	0,8301
Precisão	0,8228	0,6199
Sensibilidade (Recall)	0,8976	0,8607
F1-Score:	0,8586	0,6448

Fonte: do autor (2023).

No Quadro 19 são apresentados os resultados incrementais dos experimentos para facilitar a comparação dos resultados do Experimento I com os resultados do Experimento II.

Quadro 19 - Comparativo de desempenho usando somente características (*features*) de sinais e sintomas (I) versus todas as características disponíveis, incluindo perfil de pacientes (II).

Modelo	I	II	I	II	I	II	I	II
	Acurácia	Acurácia	Precisão	Precisão	Recall	Recall	F1-score	F1-score
Decision Tree	0,7642	0,8703	0,5834	0,6424	0,7856	0,8560	0,5785	0,6824
Logistic Regression	0,8282	0,8896	0,6116	0,6635	0,8179	0,8762	0,6343	0,7113
Support Vector Machine	0,7338	0,8301	0,5749	0,6199	0,7736	0,8607	0,5566	0,6448

Fonte: do autor (2023).

Os testes iniciais, com número menor de características, foram denominados “Experimento I”. O segundo teste, com todas as informações dos pacientes, foram caracterizados como “Experimento II”. Foi possível identificar, em todos os modelos, uma significativa melhora nas métricas, como era esperado.

5.4 Experimento III - Dados disponíveis no início da pandemia

O Experimento III foi realizado para se estimar o quão efetivo poderiam ser os modelos baseados em *Machine Learning* em fases iniciais de pandemias de síndromes respiratórias. Dessa maneira, diferentemente dos experimentos anteriores, os modelos foram treinados e testados somente com os dados disponíveis na base de dados da Secretaria de Saúde do Estado de Santa Catarina do início da pandemia até o dia 2 de abril de 2020 - dia em que os médicos infectologistas criaram a estratégia utilizada no aplicativo Triagem On-Line. Os dados desse período totalizaram 1.628 casos. Nesse treinamento, foram utilizados todos os sinais, sintomas e fatores de risco para se avaliar o quão efetivo seriam os classificadores baseados nos modelos de *Machine Learning* apresentados neste estudo nessa condição. Depois de treinados, os modelos foram testados com a

mesma porção do dataset utilizada nos testes dos experimentos anteriores (10% de todos os dados de 2020, sem balanceamento, que não participaram dos treinamentos dos modelos). A seguir serão apresentados os resultados dos treinamentos e testes efetuados em cada modelo baseado em *Machine Learning*.

Modelo 1 - *Logistic Regression*

O Quadro 20 apresenta as métricas geradas pelo treinamento e testes de um modelo baseado em *Logistic Regression* com todos os parâmetros pessoais disponíveis dos casos registrados até o dia 2 de abril de 2020. Para o treinamento do modelo foi utilizada a validação cruzada. O hiperparâmetro C, calculado pela função GridSearchCV, para este conjunto específico de dados, teve o valor igual a 1,9179. Para a geração das métricas, tanto o treinamento quanto os testes utilizaram um conjunto de dados distintos entre si, conforme efetuado no Experimento I.

Quadro 20 - Métricas do *Logistic Regression*

	Métricas da validação cruzada	Métricas no conjunto de teste
Acurácia	0,8293	0,4003
Precisão	0,8796	0,5268
Sensibilidade (Recall)	0,7630	0,6091
F1-Score:	0,8172	0,3417

Fonte: do autor (2023).

Modelo 2 - *Decision Tree*

O Quadro 21 apresenta as métricas geradas pelo treinamento e testes de um modelo baseado em *Decision Tree* com todos os parâmetros pessoais disponíveis dos casos registrados até o dia 2 de abril de 2020. Para o treinamento do modelo foi utilizada a validação cruzada. O hiperparâmetro max_depth, calculado pela função GridSearchCV, para este conjunto específico de dados, teve o valor igual a 10. Para

a geração das métricas, tanto o treinamento quanto os testes utilizaram um conjunto de dados distintos entre si, conforme realizado com o modelo anterior.

Quadro 21 - Métricas *Decision Tree*

	Métricas da validação cruzada	Métricas no conjunto de teste
Acurácia	0,7891	0,5404
Precisão	0,900	0,5300
Sensibilidade (Recall)	0,6506	0,6327
F1-Score	0,7552	0,4227

Fonte: do autor (2023).

Modelo 3 - Support Vector Machine

O Quadro 22 apresenta as métricas geradas pelo treinamento e testes de um modelo baseado em *Support Vector Machine* com todos os parâmetros pessoais disponíveis dos casos registrados até o dia 2 de abril de 2020. Para o treinamento do modelo foi utilizada a validação cruzada. Os hiperparâmetros C e gamma, calculados pela função GridSearchCV, para este conjunto específico de dados, tiveram seus valores iguais a 0,1 e 1,0, respectivamente. Para a geração das métricas, tanto o treinamento quanto os testes utilizaram um conjunto de dados distintos entre si, conforme realizado com os modelos anteriores.

Quadro 22 - Métricas *Support Vector Machine*

	Métricas da validação cruzada	Métricas no conjunto de teste
Acurácia	0,8453	0,1856
Precisão	0,8675	0,5061
Sensibilidade (Recall)	0,8152	0,5131
F1-Score	0,8405	0,1805

Fonte: do autor (2023).

O Quadro 23 apresenta os resultados do Experimento I, do Experimento II e do Experimento III para facilitar o comparativo do desempenho dos modelos.

Quadro 23: Comparativo de desempenho: (I) somente sinais e sintomas; (II) todas as características disponíveis, incluindo perfil de pacientes e (III) todas as características, mas somente registros de casos coletados até 2 de abril de 2020.

Modelo	I	II	III	I	II	III	I	II	III	I	II	III
	Acurácia	Acurácia	Acurácia	Precisão	Precisão	Precisão	Recall	Recall	Recall	F1-score	F1-score	F1-score
Decision Tree	0,7642	0,8703	0,5404	0,5834	0,6424	0,5300	0,7856	0,8560	0,6327	0,5785	0,6824	0,4227
Logistic Regression	0,8282	0,8896	0,4003	0,6116	0,6635	0,5268	0,8179	0,8762	0,6091	0,6343	0,7113	0,3417
Support Vector Machine	0,7338	0,8301	0,1856	0,5749	0,6199	0,5061	0,7736	0,8607	0,5131	0,5566	0,6448	0,1805

Fonte: do autor (2023).

O Quadro 24 apresenta o comparativo resultante do Experimento III com o resultado do desempenho do do aplicativo Triagem On-Line. A análise desses dados possibilitará identificar qual a melhor estratégia a ser utilizada em início de pandemias: se utilizar classificadores baseados em *Machine Learning* ou classificadores baseados em estratégias definidas por médicos.

Quadro 24 - Comparativo entre os resultados do Experimento III (desempenho dos modelos de aprendizado de máquina treinados com dados disponíveis no início da pandemia) e o desempenho do aplicativo Triagem On-Line

Modelo/Aplicativo	Acurácia	Precisão	Sensibilidade	F1-score
<i>Decision Tree</i>	0,5404	0,5300	0,6327	0,4227
<i>Logistic Regression</i>	0,4003	0,5268	0,6091	0,3417
<i>Support Vector Machine</i>	0,1856	0,5061	0,5131	0,1805
Triagem On-Line	0,7619	0,1824	0,8152	0,2981

Fonte: do autor (2023).

5.5 Discussões

Inicialmente, por meio do Experimento I que utilizou somente sinais e sintomas dos pacientes, tentou-se identificar a efetividade do aplicativo Triagem On-Line em relação a diferentes modelos de classificação do estado da arte. Para isso, separou-se uma porção de 10% do dataset que seria utilizado apenas para para o comparativo final da efetividade de cada modelo. Foram treinados algoritmos baseados em *Machine Learning* dos modelos *logistic regression*, *decision tree* e SVM. Ressalta-se que nesses treinamentos/testes foram utilizados apenas os sintomas usados na estratégia implementada no aplicativo Triagem On-Line, a saber:

- 'tosse';
- 'febre';
- 'mialgia';
- 'dispneia';
- 'diarreia'
- 'cefaleia';
- 'cansaco';
- 'coriza';
- 'dor_corpo';
- 'dor_garganta' e
- 'gravidade'.

Uma tabela comparativa foi construída e foi possível observar que o aplicativo Triagem On-Line teve um desempenho bastante competitivo em relação a métrica Sensibilidade (*recall*), métrica essa considerada a mais importante em aplicações médicas, conforme explanado anteriormente. Por outro lado o aplicativo demonstrou um rendimento baixo na métrica precisão que, embora seja tolerável um rendimento inferior em aplicações médicas, não é desejável, em situações de pandemias, pois o encaminhamento de casos leves de doenças para ambientes hospitalares já

sobrecarregados também poderia colocar vidas em risco. No Experimento I, o modelo *Logistic Regression* obteve melhor desempenho em todas as métricas calculadas. As métricas calculadas por esse modelo foram apresentadas no Quadro 11. Observou-se que o modelo atribuiu peso maior ao sintoma dispneia (falta de ar), assim como fora atribuído pelos médicos infectologistas no sistema Triagem On-Line. Por outro lado, mialgia, cefaleia, coriza, dor no corpo e dor de garganta obtiveram valores negativos. Isso significa que o modelo encontrou relação inversa dessas *features* com casos graves da doença.

Ressalta-se que os pesos propostos pelo *Logistic Regression* não podem ser diretamente substituídos no sistema Triagem On-Line, uma vez que este utiliza um valor de corte para a soma dos pesos; enquanto o *Logistic Regression* faz uso de uma função sigmóide, conforme explicado na [seção 2.8.1](#).

Para que seja possível a utilização dos pesos encontrados pelo *Logistic Regression*, no aplicativo Triagem On-Line, deve ser aplicada uma função sigmóide com o intuito de estimar a gravidade, conforme apresentado na Equação 2.

Equação 2 - função sigmoide utilizado pelo modelo *Logistic Regression*.

$$f(\text{gravidade}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_{10} \cdot X_{10})}}$$

Fonte: do autor (2023).

Onde:

β_0 é o termo de interceptação (-2,0596)

$\beta_1 \dots \beta_{10}$ pesos dos sintomas disponíveis no Quadro 11;

$X_1 \dots X_{10}$ terão valor igual a 0 ou 1 indicando a presença do sintoma.

Os pesos apresentados na Equação 2 devem ser substituídos pelo conjunto individual daqueles apresentados no Quadro 11. O valor calculado pela equação será um número real entre zero e um. Quando a função retornar um valor igual ou superior a 0,5 tratar-se-á de um caso grave. Quando este valor for inferior a 0,5 referenciar-se-á, então, um caso não grave.

Depois, por meio do Experimento II, que utilizou também os dados de perfil dos pacientes, tentou-se verificar qual seriam as métricas máximas que poderiam ser alcançadas se fossem utilizados todos os parâmetros disponíveis sobre cada caso. Ou seja, os modelos baseados em *Machine Learning* seriam treinados também com os parâmetros:

- 'idade_0_9';
- 'idade_10_19';
- 'idade_20_24';
- 'idade_25_59';
- 'idade_60_mais';
- 'gravidez';
- 'diabetes';
- 'obesidade';
- 'imunossupressao';
- 'renal';
- 'hipertensao';
- 'cardiaco';
- 'cancer';
- 'hepatica';
- 'pneumatica';
- 'asma';
- 'neurologica' e
- 'masculino'.

Percebeu-se que, com o uso de todos os parâmetros, obteve-se um ganho de desempenho em todas as métricas, de todos os modelos. No Experimento II, o modelo *Logistic Regression* novamente obteve o melhor desempenho em todas as métricas calculadas.

Por fim, no Experimento III, que utilizou apenas os dados dos primeiros meses de 2020, tentou-se identificar a possibilidade de se desenvolver classificadores de gravidade de síndromes respiratórias baseados em *Machine Learning* no início de pandemias, situação em que há poucos dados disponíveis para o treinamento de

modelos. Este experimento visou comparar as métricas dos modelos com a estratégia utilizada no aplicativo Triagem On-Line em situação de igualdade. Utilizou-se para o experimento os dados disponíveis na base de dados da Secretaria de Estado da Saúde de Santa Catarina no dia em que os médicos infectologistas determinaram os pesos que dariam para cada atributo utilizado no desenvolvimento do aplicativo. Conforme foi apresentado no Quadro 23, constatou-se que o modelo baseado em *decision tree* alcançou os melhores desempenhos na precisão e F1-score. Entretanto, o aplicativo Triagem On-Line, teve os melhores desempenhos de acurácia e de sensibilidade, métrica essa considerada a mais importante em aplicativos dessa natureza.

6. CONCLUSÕES E TRABALHOS FUTUROS

Os esforços intentados visaram à preparação do campo para a análise do desempenho da ferramenta Triagem On-Line, na categorização da gravidade de casos de COVID-19, em comparação com outras abordagens do estado da arte.

Outrossim, por meio da revisão sistemática da literatura foi possível mapear trabalhos relacionados à temática, com o intuito de aprimorar o campo de pesquisa desbravado até então; cujo resultado foi utilizado para comparar o desempenho do aplicativo Triagem On-Line com diagnósticos reais.

Depois de analisar o comportamento de modelos baseado em *Machine Learning* em outros trabalhos acadêmicos, verificou-se métricas que poderiam ser alcançadas em aplicações semelhantes. É evidente que cada pesquisa se utilizou de bases de dados diferentes, bem como sinais e sintomas diversos. Por essa razão, embora não se pudesse encontrar valores idênticos, foi possível utilizar tais resultados para compará-los com os resultados dos experimentos da presente pesquisa, e identificar, assim, a coerência do estudo. Depois de treinados os modelos mais encontrados na literatura e calculado suas métricas com a base de dados da Secretaria de Saúde do Estado de Santa Catarina foi possível chegar a resultados semelhantes àqueles encontrados em outros trabalhos. Com base nisso foi possível comparar o desempenho desses modelos baseados em *Machine Learning*, utilizando-se da mesma base de dados, com rendimento semelhante ao da estratégia utilizada no aplicativo Triagem On-Line.

Embora a precisão do aplicativo tenha ficado abaixo das métricas dos modelos baseados em *Machine Learning*, a métrica de sensibilidade teve um rendimento bastante competitivo - o segundo melhor desempenho, mesmo considerando não haver três sintomas utilizados para a triagem pelo aplicativo: tosse com sangue, anosmia e disgeusia.

O baixo rendimento da precisão significa que a estratégia definida pelos médicos encaminhou muitos pacientes com casos leves da doença a procurarem auxílio em unidades de saúde. Por outro lado, como foi visto anteriormente, esse tipo de erro é um erro tolerável para aplicações médicas, uma vez que é melhor

errar por excesso de cautela com casos leves de uma doença do que perder totalmente um diagnóstico (Kumar, 2023).

Restou, portanto, comprovado, que ferramentas computacionais de triagem remota de pacientes são possibilidades a serem consideradas durante o enfrentamento de pandemias de síndromes respiratórias: tanto ferramentas baseadas em *Machine Learning* quanto aquelas que utilizam estratégias definidas por médicos.

Importante ressaltar que o aplicativo Triagem On-Line, que fora desenvolvido no início da pandemia da COVID-19, teve uma sensibilidade semelhante ao de modelos baseados em *Machine Learning* treinados com um conjunto de dados expressivo, que englobou o período de um ano de pandemia.

Conclui-se, portanto, que caso uma nova pandemia de síndrome respiratória surja, aplicativos de triagem remota podem auxiliar na mitigação de doenças: tanto aqueles baseados em *machine learning* quanto aqueles que implementem estratégias definidas por médicos. A definição de qual o melhor modelo para determinada pandemia dependerá do comparativo de métricas com os dados que estiverem disponíveis no momento. Este trabalho demonstrou como calcular tais métricas e como analisar os resultados para auxiliar na tomada de decisão.

Não há como negar, neste sentido, que Santa Catarina foi um dos estados que se destacou nacionalmente na adoção de medidas preventivas no que concerne ao enfrentamento da pandemia da Covid-19, restando evidente que a criação da ferramenta de triagem, de maneira preventiva, corroborou para o ótimo desempenho do Estado. A importância da análise, e posterior registro de eventuais esforços envidados pelos órgãos públicos permitirá ações, por parte dos agentes públicos, cada vez mais céleres e eficientes. Neste sentido sugere-se para estudos futuros: (I) experimentos com dados de outras doenças para aplicação de triagens à distância (como malária, dengue, doença de chagas), com o intuito de avaliar e expandir cada vez mais o arcabouço tecnológico à disposição dos gestores e tomadores de decisão; (II) análise da gravidades de casos de doenças por faixa etária, entre outros dados de perfil que se mostraram relevantes na classificação dos casos, visando priorizar atendimentos, bem como para identificar a necessidade de vacinação de grupos com baixo risco de desenvolver casos graves; (III) experimentos utilizando

outras bases de dados visando compreender os pesos negativos propostos pelo modelo baseado em *Logistic Regression*. A possibilidade de se utilizar menos sintomas em triagens com o objetivo de mensurar gravidades de doenças possibilitaria a redução do número de campos para preenchimento por parte dos usuários/pacientes. Isso pode significar uma simplificação nas interfaces, o que melhoraria a usabilidade de ferramentas com esse propósito; (IV) experimentos com modelos de redes neurais profundas (*deep learning*) para a triagem de casos de doenças. Tais modelos têm propiciado desempenho superior em vários problemas de classificação e dispensam engenharia de características minuciosas em muitos empreendimentos. Todavia, podem requerer maior quantidade de dados de treinamento e são modelos “caixa preta”, usualmente menos explicáveis que o utilizado no aplicativo Triagem On-Line, regressão logística e árvores de decisão.

8. REFERÊNCIAS BIBLIOGRÁFICAS

Adam, A. A.; Elhag, N. A. A.; Saeed, F. A.; Yagob, M.; Mohamed, F.; Eassa, N.; Abdalaziz, H. S.; Ahmed, M. A.; Babiker, S. F. **An Ensemble Machine Learning Model to Investigate the Screening System for Identification of Potential Patients with COVID-19 in Sudan**. 2021 International Congress of Advanced Technology and Engineering (ICOTEN), 2021, pp. 1-5, doi: 10.1109/ICOTEN52080.2021.9493517.

ALBINO, Rubia Maria; GROSSEMAN, Suely; RIGGENBACH, Viviane. **Classificação de risco: uma necessidade inadiável em um serviço de emergência de qualidade**. Artigo. Arquivos Catarinenses de Medicina - Vol. 36, no. 4, de 2007.

ALMEIDA, Luciana. **Covid-19: aluno e professor do IFTM de Ituiutaba desenvolvem aplicativo para suporte no diagnóstico da doença através de inteligência artificial**. Revista Newslab. Disponível em: <https://newslab.com.br/covid-19-aluno-e-professor-do-iftm-de-ituiutaba-desenvolvem-aplicativo-para-suporte-no-diagnostico-da-doenca-atraves-de-inteligencia-artificial>. Acesso em 13 de abril de 2022.

BAO, W.; LIANJU, N.; YUE, K. **Integration of unsupervised and supervised machine learning algorithms for credit risk assessment**. Expert Systems with Applications, 2019. Exp. Syst. Appl. 2019;128:301–315. doi: 10.1016/j.eswa.2019.02.033.

BISHOP, C. M. **Pattern recognition and machine learning**. [S.l.]: springer, 2006. Disponível em https://cds.cern.ch/record/998831/files/9780387310732_TOC.pdf. Acesso em 12 de abril de 2023.

BRASIL. Ministério da Saúde. **Portaria nº 2048, de 5 de novembro de 2002**. Disponível em: https://bvsms.saude.gov.br/bvs/saudelegis/gm/2002/prt2048_05_11_2002.html. Acesso em 8 de outubro de 2021.

BRASIL, Ministério da Saúde. **Quantidade de municípios sem leitos SUS**, uf: SC, ano: 2021. Banco de dados do Sistema Único de Saúde – DATASUS. Cadastro Nacional de Estabelecimentos de Atenção à Saúde – CNES. Disponível em <https://auditasus.com.br/leitos-cnes/sem-leitos-sus/qtd-de-municipios-sem-leitos-sus-uf?ano=2021&uf=SC>. Acesso em 10 de outubro de 2021.

BRASIL. Ministério da Saúde. **Recomendação nº 022, de 9 de abril de 2020**. Disponível em: <http://conselho.saude.gov.br/recomendacoes-cns/1112-recomendac-a-o-n-022-de-09-de-abril-de-2020>. Acesso em 29 de setembro de 2021.

BRUCE, R. (2001). **A bayesian approach to semi-supervised learning**. In M. Ishizuka & A. Satter (Eds.), Proceedings of the Sixth Natural Language Processing Pacific Rim Symposium — NLPRS-2001, Tokyo, Japan, pp. 57–64. Springer Verlag.

BUFREM, Leilah Santiago. **A Pandemia da Covid-19 no Brasil: informações e contradições na atual conjuntura**. Artigo. Universidade Federal do Paraná. DOI: <https://doi.org/10.21721/p2p.2020v7n1.p101-120>

CAETANO, R., SILVA, A. B., GUEDES, A. C. C. M., PAIVA, C. C. N. RIBEIRO, G. R., SANTOS, D. L., SILVA, R. M. **Desafios e oportunidades para telessaúde em tempos da pandemia pela COVID-19: uma reflexão sobre os espaços e iniciativas no contexto brasileiro**. Cad. Saúde Pública, Rio de Janeiro, v. 36, n. 5, e00088920, 2020.

CAMPANA, G. A.; OPLUSTIL, C. P. **Conceitos de automação na medicina laboratorial: revisão de literatura**. J Bras Patol Med Lab. v. 47. n. 2. p. 119-127. Abril de 2011.

CARRILLO-LARCO, R. M., ALTEZ-FERNANDEZ, C. (2020). **Anosmia and dysgeusia in COVID-19: A systematic review**. Wellcome Open Research, 5(94). Disponível em <https://doi.org/10.12688/wellcomeopenres.15917.1>

CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHABAZA, T.; REINARTZ, T.; HEARER, C.; WIRTH, R. **CRISP-DM 1.0: step-by-step data mining guide**. Cross Industry Standard Process for Data Mining, 2000. Disponível em: <https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf>. Acesso em 18 de setembro de 2023.

Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, P. (2000). **SMOTE: Synthetic Minority Over-sampling TEchnique**. In International Conference of Knowledge Based Computer Systems, pp. 46–57. National Center for Software Technology, Mumbai, India, Allied Press.

Dietrich, D. Heller, B. Yang, B. (2015). **Data Science and Big Data Analytics: Discovering, Analysing, Visualizing and Presenting Data**. EMC Education Service. Indianapolis, Indiana: John Wiley & Sons, Inc.

EKSTEN, Rickard. **Lund University launches COVID Symptom Tracker app**. European Regions Research and Innovation Network (ERRIN). 2020. Disponível em: <https://errin.eu/node/8479>. Acesso em 17 de abril de 2022.

FACELI, K. et al. **Inteligência artificial: uma abordagem de aprendizado de máquina**. 2ª Edição. LTC: Grupo Gen. Rio de Janeiro, 2011. ISBN: 9788521637349.

FAWCETT, T. **An Introduction To ROC Analysis**, v. 27, pp. 861–874, 2006. doi: 10.1016/j.patrec.2005.10.010.

FONTANA, L.; STRASFELD, L. **Respiratory Virus Infections of the Stem Cell Transplant Recipient and the Hematologic Malignancy Patient**. *Infectious Disease Clinics of North America*, v. 33, n. 2, p. 523–544, 2019.

GARCIA, Luís Paulo Faina. **Noise detection in classification problems**. 2016. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2016. doi:10.11606/T.55.2016.tde-29112016-155215.

GARCIA, Marcos Vinicius Fernandes, GARCIA, Marco Aurélio Fernandes. **Telemedicina, segurança jurídica e COVID-19: onde estamos?** *Jornal Brasileiro de Pneumologia*. 2020; 46(4):e20200363. Disponível em https://jbp.org.br/detalhe_artigo.asp?id=3402. Acesso em 12 de outubro de 2021.

GIRARD, M. P.; TAM, J. S.; ASSOSSOU, O. M.; KIENY, M. P. **The 2009 A (H1N1) influenza virus pandemic: A review**. *Vaccine*, v. 28, n. 31, p. 4895–4902, 2010. Elsevier Ltd. Disponível em: <http://dx.doi.org/10.1016/j.vaccine.2010.05.031>. Acesso em 22 de outubro de 2023.

GUIMARÃES, V. L., CATRAMBY, T., MORAES, C. C. de A. SOARES, C. A. L. (2020). **A pandemia COVID-19 e a educação superior em Turismo no Estado do Rio de Janeiro (Brasil): notas preliminares de pesquisa**. *Rosa dos Ventos Turismo e Hospitalidade*, 12 (3 - Especial Covid-19). Disponível em: <http://dx.doi.org/10.18226/21789061.v12i3a09>. Acesso em 13 de fevereiro de 2023.

Gupta, M., Sinha, A. **Multi-class autoencoder-ensembled prediction model for detection of COVID-19 severity**. *Evolutionary Intelligence*, 2022. <https://doi.org/10.1007/s12065-022-00744-9>

H AIXIANG , G.; Y IJING , L.; S HANG , J.; M INGYUN , G.; Y UANYUE , H. **Learning from class-imbalanced data: Review of methods and applications**. *Expert Systems with Applications*, 73:220–239, 2017.

HAYKIN, S. **Redes Neurais: Princípios e prática**, 2ª Edição, Porto Alegre: Editora Brookman, 2001.

IGUAL, L., & SEGUÍ, S. **Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications**. 1. Ed. Springer, 2017. E-book. Disponível em: <https://nibmehub.com/opac-service/pdf/read/Introduction%20to%20Data%20Science-%202017.pdf>. Acesso em 23 de outubro de 2023.

Infermedica. **Symptomate: Your guide through health**. 2022. Disponível em: <https://symptomate.com/about/>. Acesso em 17 de abril de 2022.

Instituto Brasileiro de Geografia e Estatística - IBGE. **Produto Interno Bruto dos Municípios**. Disponível em <https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/9088-produto-inter>

no-bruto-dos-municipios.html?t=resultados&c=4216602. Acesso em 5 de outubro de 2021.

Lian, Z.; Li, Y.; Wang, W.; Ding, W.; Niu, Z.; Yang, X.; Wu, C. **The Prediction Model of Risk Factors for COVID-19 Developing into Severe Illness Based on 1046 Patients with COVID-19**. Emerg. Med. Int. 2021.

Mendoza, J. A. A.; Solano, G. A.; Pontiveros, M. J.; Di Caro, J.; Gomez, P. M. D.; Manuel, C. G.; Rosell-Ubial, P. J. B.; Tee, M. **A Machine Learning Approach in Evaluating Symptom Screening in Predicting COVID-19**. 2022 International Conference on Artificial Intelligence in Information and Communication (ICAIC), 2022, pp. 188-193, doi: 10.1109/ICAIC54071.2022.9722659.

JAMES, Gareth; WITTEN, Daniele; HASTIE, Trevor; TIBSHIRANI, Robert. **An Introduction to statistical learning** – with applications in R. New York: Springer, 2013.

KUMAR, Ajitesh. **Accuracy, Precision, Recall & F1-Score – Python Examples**.

Disponível em:

<https://vitalflux.com/accuracy-precision-recall-f1-score-python-example/> Acesso em 15 de janeiro de 2023.

LANTZ, Brett. **Machine learning with R: expert techniques for predictive modeling**. Packt publishing ltd, 2019.

LAZARIN, A. C.; MARIANO, R. C. Z. **Orientações para a Rede Pública de Atenção à Saúde de Campinas no enfrentamento à pandemia de infecção humana pelo novo coronavírus (SARS-COV-2) - COVID-19**. Secretaria de Saúde de Campinas. Fevereiro, 2022.

Lööv, Simon. **Comparison of Undersampling Methods for Prediction of Casting Defects Based on Process Parameters**. Student thesis, Högskolan i Skövde, Institutionen för ingenjörsvetenskap, 2021.

<http://urn.kb.se/resolve?urn=urn:nbn:se:his:diva-20596>.

LORENA, A. C.; CARVALHO, A. C. P. L. F. **Uma Introdução às Support Vector Machines**. Revista de Informática Teórica e Aplicada, vol.14, nº 2, pp 43-67, 2007. Disponível em: https://seer.ufrgs.br/index.php/rita/article/view/rita_v14_n2_p43-67. Acesso em 30 de junho de 2023.

MAHESH, B. **Machine learning algorithms** - a review. International Journal of Science and Research, v. 9, n. 1, p. 381-386, 2020.

MALHOTRA, Naresh K. **Pesquisa de marketing: uma orientação aplicada**. 3. ed. Porto Alegre: Bookman, 2001.

Mendes-Da-Silva, W. (2019). **Editorial. Contribuições e limitações de revisões narrativas e revisões sistemáticas na área de negócios**. Revista de

Administração Contemporânea. Disponível em:
<https://rac.anpad.org.br/index.php/rac/article/view/1320/1348>. Acesso em 20 de outubro de 2023.

MOHAPATRA, S. K.; MOHANTY, M. D.; DAS, A.; MOHANTY, M. N. **A Deep Learning Approach for COVID-19 Infection Probability Prediction**. 2021 19th OITS International Conference on Information Technology (OCIT), 2021, pp. 245-249, doi: 10.1109/OCIT53463.2021.00057.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). **Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement**. PLoS Med, 6(7), 1-6. Disponível em:
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2707599/>. Acesso em 20 de outubro de 2023.

MURRAY, M. J. **The Canadian Triage and Acuity Scale: A Canadian perspective on emergency department triage**. Emerg Med 2003.

NETO, E. A.; RESQUE, H. R.; MÓDENA, J. L. P.; CASTILHO, J. G.; RÁCZ, M. L.; CRIADO, M. F.; CARUSO, T. A. R.; MUNFORD, V. **Microbiologia Clínica para o Controle de Infecção Relacionada à Assistência à Saúde**. Agência Nacional de Vigilância Sanitária. 2010.

Oliveira, D., Ferreira, D., Abreu, N. et al. Prediction of COVID-19 diagnosis based on openEHR artefacts. Sci Rep 12, 12549 (2022).
<https://doi.org/10.1038/s41598-022-15968-z>

PÁDUA, Clarindo Isaías Pereira da Silva. **Engenharia de Usabilidade: Material de Referência**. Departamento de Ciência da Computação. Universidade Federal de Minas Gerais. Belo Horizonte, MG. 2012.

PATEL, M. et al. **Pandemic (H1N1) 2009 influenza**. British Journal of Anaesthesia, v. 104, n. 2, p. 128–142, 2010.

PEREIRA M. S., SOUZA A. C. S., TIPPLE A. F. V., PRADO M. A. **A infecção hospitalar e suas implicações para o cuidar da enfermagem**. Texto e Contexto – enferm. v.14 n.2 Florianópolis abr./jun. 2005

PRODANOV, Cleber Cristiano. **Metodologia do trabalho científico** [recurso eletrônico]: métodos e técnicas da pesquisa e do trabalho acadêmico / Cleber Cristiano Prodanov, Ernani Cesar de Freitas. – 2. ed. – Novo Hamburgo: Feevale, 2013.

REIS, Rita Soares. **Plataforma inteligente de apoio à decisão médica no transplante de órgãos**. Dissertação de Mestrado. Universidade do Minho. Portugal, 2019.

Rochmawati N., Hidayati H.B., Yamasari Y., Yustanti W., Rakhmawati L., Tjahyaningtjas H.P.A., Anistyasari Y. **Covid Symptom Severity Using Decision Tree**. Proceedings of the 2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE); Surabaya, Indonesia, 2020.

RUSSELL, S.; NORVIG, P. **Artificial intelligence: a modern approach**. [S.l.]: Prentice Hall, 2002.

SÁ, Maria Cássia, HOLANDA, Virgínia Célia Cavalcante. **O uso seletivo do território na distribuição dos serviços de saúde: um estudo de Sobral – Ceará**. Artigo. Ateliê Geográfico. Goiânia-GO, v. 11, n. 2, ago./2017, p. 95-111.

SARAIVA, Aléxia. **5 apps e sites dão orientação em casos suspeitos de Covid-19**. Gazeta do Povo. 2020. Disponível em: <https://www.gazetadopovo.com.br/gazz-conecta/5-apps-e-sites-dao-orientacao-em-casos-suspeitos-de-covid-19/>. Acesso em 13 de abril de 2022.

Santa Catarina. **Dados anonimizados casos confirmados COVID-19**. Portal de Dados abertos do estado de Santa Catarina. Secretaria estadual de Saúde. Diretoria de Vigilância Sanitária. COVID-19 – casos confirmados. Disponível em: <http://dados.sc.gov.br/dataset/covid-19-dados-anonimizados-de-casos-confirmados> Acesso em 17 de agosto de 2022.

Santa Catarina. **Governo do Estado completa 120 dias de enfrentamento à pandemia e mantém melhor desempenho do país**. Secretaria estadual de Saúde. Notícias, 2020. Disponível em <https://www.saude.sc.gov.br/index.php/noticias-geral/todas-as-noticias/1652-noticias-2020/11455-governo-do-estado-completa-120-dias-de-enfrentamento-a-pandemia-e-mantem-melhor-desempenho-do-pais>. Acesso em: 23 de outubro de 2023.

SANTHANAM, T.; SUNDARAM, S. **Application of CART Algorithm in Blood Donors Classification**. Journal of Computer Science ,p. 548-552, 2010. Disponível em: <http://dx.doi.org/10.3844/jcssp.2010.548.552>. Acesso em: 21 de outubro de 2023.

SANTOS, Cinara de Jesus. **Avaliação do uso de classificadores para verificação de atendimento a critérios de seleção em programas sociais**. Programa de Pós-graduação em Modelagem Computacional, Universidade Federal de Juiz de Fora. Juiz de Fora, 2017.

SCHMITZ, Carlos André Aita, SCHERER, Charleni Inês, NASCIMENTO , Diane Moreira, VEADRIGO, Francine, SOARES, Konrad Gutterres, MENDONÇA, Marcos Vinícius Ambrosini, DOTTO, Mariana Kreutz, NUNES, Paula Martina da Silva Araújo,FRANK, Thiago. **Manual de teleconsulta na APS**. Programa de Pós-Graduação em Epidemiologia. TelessaúdeRS (TelessaúdeRS-UFRGS). Porto Alegre, jun. 2020.

SHEARER, C. **The CRISP-DM Model: The New Blueprint for Data Mining**. Journal of

Data Warehousing, v. 5, n. 4, p. 13-22, 2000.

SILVA, Nadia F., COLETTA, Luiz, HRUSCHKA, Eduardo R., **A Survey and Comparative Study of Tweet Sentiment Analysis via Semi-Supervised Learning**. ACM Comput. Surv. , 2016. 49(1): 15:1-15:26.

SILVA, Lena Veiga. **Estratégias para análise de subnotificações de mortalidade em epidemias ocorridas nos países em desenvolvimento: Um estudo de caso da COVID-19 no Brasil**. Tese de doutorado em Engenharia Elétrica. Universidade Federal do Pará. Belém, 2021.

Soui, M., Mansouri, N., Alhamad, R. et al. **NSGA-II as feature selection technique and AdaBoost classifier for COVID-19 prediction using patient's symptoms**. Nonlinear Dyn 106, 1453–1475 (2021). <https://doi.org/10.1007/s11071-021-06504-1>

Suma, L.S., Anand, H.S., Vinod chandra, S.S. **Nature inspired optimization model for classification and severity prediction in COVID-19 clinical dataset**. Journal of Ambient Intelligence and Humanized Computing, 2021. <https://doi.org/10.1007/s12652-021-03389-1>

Sun, W., Braatz, R.D. **Smart process analytics for predictive modeling**. Computers & Chemical Engineering. Volume 144, 4 January 2021, 107134

UFSC. Biblioteca Universitária. Grupo de trabalho para levantamento das ações de enfrentamento à COVID-19 na BU/UFSC. **Ações de enfrentamento à COVID-19**. Florianópolis, 2020.

UNIVERSIDADE FEDERAL DO AMAZONAS. Comitê Interno de Enfrentamento do Surto epidemiológico de Coronavírus da Ufam. **Plano de Contingência da Universidade Federal do Amazonas frente à pandemia da doença pelo SARS-COV-2 (COVID-19)**. Manaus: Universidade Federal do Amazonas, 2020.

University of Cincinnati. **COVID Check App**. 2020. Disponível em: <https://www.uc.edu/publichealth/testing/covid-check-app.html>. Acesso em 17 de abril de 2022.

VERGARA, Sylvia C. **Projetos e relatórios de pesquisa em administração**. 3.ed. Rio de Janeiro: Atlas, 2000.

Villavicencio, C.N.; Macrohon, J.J.; Inbaraj, X.A.; Jeng, J.-H.; Hsieh, J.-G. **Development of a Machine Learning Based Web Application for Early Diagnosis of COVID-19 Based on Symptoms**. Diagnostics 2022, 12, 821. <https://doi.org/10.3390/diagnostics12040821>

WEN, Chao Lung. **Telemedicina: cuidado aos pacientes e proteção para os profissionais da saúde**. Artigo. Associação Nacional de Hospitais Privados. São Paulo, 2020. Disponível em: <https://www.anahp.com.br/noticias/covid-19/artigo-telemedicina-cuidado-aos-pacient>

es-e-protecao-para-os-profissionais-da-saude/. Acesso em 11 de outubro de 2021.

WITTEN, Ian H. et al. **Data Mining**: Practical machine learning tools and techniques. Morgan Kaufmann, 2005. Disponível em http://www.academia.dk/BiologiskAntropologi/Epidemiologi/DataMining/Witten_and_Frank_DataMining_Weka_2nd_Ed_2005.pdf. Acesso em 23 de outubro de 2023.

WLSL. **University of Lynchburg using app to track students, staff with COVID-19 symptoms**. Disponível em: <https://www.wsls.com/news/local/2020/07/28/university-of-lynchburg-using-app-to-track-students-staff-with-covid-19-symptoms/>. Acesso em 17 de abril de 2022.

Y. Yeramalli, S. Intodia and P. G. Shambharkar. **COVID-19 Diagnosis Using Machine Learning Techniques**. 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT), 2022, pp. 1-6, doi: 10.1109/IC3IoT53935.2022.9767856.

Yusuf, A.M.I.A., Rosli, M.M., Yusop, N.S.M. **A Screening System for COVID-19 Severity using Machine Learning**. International Journal of Advanced Computer Science and Applications. Vol. 13, No. 7, 2022.

ZANELLA, Liane Carly Hermes. **Metodologia de pesquisa**. UFSC, 2013. Disponível em: http://arquivos.eadadm.ufsc.br/EaDADM/UAB_2014_2/Modulo_1/Metodologia/material_didatico/Livro%20texto%20Metodologia%20da%20Pesquisa.pdf. Acesso em 17 de outubro de 2021.

Zhanyang Sun, SZY, SUN*, Rui Ding, DR, DING*, and Xinyu Zhou, ZXY, ZHOU*. 2021. **Machine Learning Applications in Forecasting of COVID-19 Based on Patients' Individual Symptoms**. In 2021 the 3rd International Conference On Intelligent Science And Technology (ICIST) (ICIST 2021), September 25–27, 2021, Tokyo, Japan. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3507959.3507966>

9. ANEXOS

Anexo I - Dicionário de dados

Campo	Formato	Codificação	Descrição	Observação
data_publicacao	Texto	AAAA-MM-DD HH:MM:SS	Data de publicacao do conjunto de dados no portal de dados abertos	
recuperados	Texto	SIM / NAO	Indicação de que o paciente foi recuperado	Caso confirmado de COVID-19 que não evoluiu a óbito, que tenha registro de alta hospitalar e que tenham decorridos pelo menos 14 dias do início dos sintomas.
data_inicio_sintomas	Texto	AAAA-MM-DD	Data do início dos sintomas	
data_coleta	Texto	AAAA-MM-DD	Data da coleta da amostra	Registros de confirmação com valor "NÃO SE APLICA" são de confirmações por critério clínico ou vínculo epidemiológico
estado	Texto	Texto separado por vírgula	Sintomas do paciente	Todos os sintomas registrados para o caso estão agrupados no mesmo campo
comorbidades	Texto	Texto separado por vírgula	Comorbidades do paciente	Todas as comorbidades registradas para o caso estão agrupadas no mesmo campo
gestante	Texto	IDADE GESTACIONAL / PUERPERA	Indica os casos de gestantes ou púérpera	
internacao	Texto	INTERNADO / NAO INTERNADO	Indicação de que o paciente está internado	

internacao_uti	Texto	INTERNADO UTI / NAO INTERNADO UTI	Indicação de que o paciente está internado em UTI	Os pacientes internados em UTI também estão registrados no campo "internacao"
sexo	Texto	FEMININO / MASCULINO	Indicação de sexo biológico do paciente	
municipio	Texto	Nome do município	Município de residência do paciente	Quando a residência do paciente for de outro estado ou outro país serão agrupados em OUTROS ESTADOS e OUTROS PAISES respectivamente
obito	Texto	SIM / NAO	Indicação de que o paciente veio a óbito	
data_obito	Texto	AAAA-MM-DD	Data do óbito do paciente	Quando não se aplica, o campo recebe valor nulo
idade	Numérico	999	Idade do paciente	
regional	Texto	Nome da mesoregião do estado	Mesoregião de residência do paciente	Quando a residência do paciente for de outro estado ou outro país serão agrupados em OUTROS ESTADOS e OUTROS PAISES respectivamente
raca	Texto	AMARELA, BRANCA, PARDA, PRETA, NAO INFORMADO, IGNORADO	Raça do paciente	
data_resultado	Timestamp	AAAA-MM-DD HH:MM:SS	Data e hora da confirmação	
codigo_ibge_municipio	Numérico	9999999	Código do IBGE do município de residência do paciente	

latitude	Numérico	(-)99,9999999	Latitude do município de residência do paciente	Quando a residência do paciente for de outro estado ou outro país o campo recebe valor nulo. Coordenada geográfica em frações decimais
longitude	Numérico	(-)99,9999999	Longitude do município de residência do paciente	Quando a residência do paciente for de outro estado ou outro país o campo recebe valor nulo. Coordenada geográfica em frações decimais
estado	Texto	Nome do estado	Nome do estado de residência do paciente	Quando a residência do paciente for de outro estado ou outro país serão agrupados em OUTROS ESTADOS e OUTROS PAISES respectivamente
critério_confirmacao	Texto	LABORATORIAL / CLINICO-EPIDEMIOLÓGICO	Critério utilizado para confirmação do caso	
tipo_teste	Texto	BIOLOGIA MOLECULAR (RT-PCR), IMUNOLÓGICO (TESTE RÁPIDO), IMUNOLÓGICO (SOROLÓGICO), NAO SE APLICA	Tipo de teste utilizado para confirmação do caso	Quando o critério utilizado para confirmação do caso for o Clínico Epidemiológico o campo indicará NAO SE APLICA
municipio_notificacao	Texto	Nome do município	Município onde a notificação foi registrada	Quando o município de notificação for de outro estado ou outro país serão agrupados em OUTROS ESTADOS e OUTROS PAISES respectivamente
codigo_ibge_municipio_notificacao	Numérico	9999999	Código do IBGE do município onde a notificação foi registrada	

latitude_notificacao	Numérico	(-)99,9999999	Latitude do município onde foi registrada a notificação	Coordenada geográfica em frações decimais
longitude_notificacao	Numérico	(-)99,9999999	Longitude do município onde foi registrada a notificação	Coordenada geográfica em frações decimais
classificacao	Texto	CONFIRMADO	Classificação de confirmação de caso positivo	Este conjunto de dados é composto apenas por casos confirmados
origem_esus	Texto	SIM / NAO	Indica que a origem da informação encontra-se no e-SUS VE	
origem_sivep	Texto	SIM / NAO	Indica que a origem da informação encontra-se no SIVEP Gripe	SIVEP Gripe - SISTEMA DE INFORMAÇÃO DE VIGILÂNCIA EPIDEMIOLÓGICA DA GRIPE utilizado para notificação de hospitalização de pacientes com Síndrome Respiratória Aguda Grave (SRAG)
origem_lacen	Texto	SIM / NAO	Indica que a origem da informação encontra-se no LACEN/SC	
origem_laboratorio_privado	Texto	SIM / NAO	Indica que a origem da informação encontra-se em um Laboratório Privado	
nome_laboratorio	Texto	Nome do laboratório	Quando campo origem_laboratorio_privado for preenchido, neste campo constará o nome do laboratório	
fez_teste_rapido	Texto	SIM / NAO	Indicativo se o paciente (sem informação do paciente) fez teste rápido	
fez_pcr	Texto	SIM / NAO	Indicativo se o paciente (sem informação do paciente) fez PCR	

data_internacao	Texto	AAAA-MM-DD	Data da internação informada no SIVEP Gripe	SIVEP Gripe - SISTEMA DE INFORMAÇÃO DE VIGILÂNCIA EPIDEMIOLÓGICA DA GRIPE utilizado para notificação de hospitalização de pacientes com Síndrome Respiratória Aguda Grave (SRAG)
data_entrada_uti	Texto	AAAA-MM-DD	Data da internação UTI informada no SIVEP Gripe	SIVEP Gripe - SISTEMA DE INFORMAÇÃO DE VIGILÂNCIA EPIDEMIOLÓGICA DA GRIPE utilizado para notificação de hospitalização de pacientes com Síndrome Respiratória Aguda Grave (SRAG)
regional_saude	Texto	AAAA-MM-DD	Data da evolução do caso informada no SIVEP Gripe	SIVEP Gripe - SISTEMA DE INFORMAÇÃO DE VIGILÂNCIA EPIDEMIOLÓGICA DA GRIPE utilizado para notificação de hospitalização de pacientes com Síndrome Respiratória Aguda Grave (SRAG)
data_evolucao_caso	Texto	AAAA-MM-DD	Data da internação informada no SIVEP Gripe	SIVEP Gripe - SISTEMA DE INFORMAÇÃO DE VIGILÂNCIA EPIDEMIOLÓGICA DA GRIPE utilizado para notificação de hospitalização de pacientes com Síndrome Respiratória Aguda Grave (SRAG)
data_saida_uti	Texto	Nome regional de saúde	da Regional de saúde do município de notificação	
bairro	Texto	Nome do bairro	Bairro do Paciente	Para municípios com mais de 100 mil habitantes