

MODELO DE *MACHINE LEARNING* PARA PREDIÇÃO DE PROBLEMAS RESPIRATÓRIOS

Gustavo Torres Belini, Danilo da Silva Ataíde e Johannes Von Lochter
Centro Universitário Facens
18087-125, Sorocaba, São Paulo, Brasil

RESUMO

A Síndrome Respiratória Aguda Grave (SRAG) representa um desafio persistente para a saúde pública, especialmente em contextos pandêmicos, como observado na COVID-19. Este estudo explora o uso de técnicas de *Machine Learning* para a predição e diagnóstico de doenças respiratórias, visando otimizar o gerenciamento de recursos e o tratamento de pacientes. Foram aplicados modelos de aprendizado supervisionado, incluindo Regressão Logística, KNN, *Naive Bayes*, SVM, *Gradient Boosting*, Árvores de Decisão e Floresta Aleatória, com avaliação de métricas como AUC-ROC, F-medida, precisão, revocação e acurácia. O *Gradient Boosting* destacou-se com os melhores resultados em termos de precisão e eficiência computacional. A metodologia CRISP-DM orientou a coleta, tratamento e modelagem dos dados, garantindo rigor no processo. Esta pesquisa reforça o papel da Inteligência Artificial na saúde, demonstrando a eficácia de modelos preditivos no monitoramento e controle de SRAG.

PALAVRAS-CHAVE

Machine Learning, Síndrome Respiratória Aguda Grave, Modelos Preditivos, *Gradient Boosting*, Saúde Pública

1. INTRODUÇÃO

A pandemia de COVID-19 em 2020 apresentou desafios sem precedentes ao Brasil, exigindo respostas rápidas e baseadas em informações precisas sobre a disseminação do vírus. Um dos sistemas centrais para coletar e analisar esses dados foi o Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVEP-Gripe).

Criado em 2000, o SIVEP-Gripe monitorava a Síndrome Respiratória Aguda Grave (SRAG), principalmente causada pelo vírus da Influenza, utilizando uma rede de vigilância sentinela para a Síndrome Gripal (SG). A SRAG, definida pela deterioração da função respiratória com hospitalização sem causa específica, pode ser causada por diversos agentes patogênicos, como os vírus Influenza A e B, Vírus Sincicial Respiratório (VSR), SARS-CoV-2, e por bactérias, fungos e outros agentes. O rápido diagnóstico e tratamento da SRAG são essenciais para reduzir complicações graves, como insuficiência respiratória que, em alguns casos, pode evoluir para óbito.

No SIVEP-Gripe estão informações como número de casos confirmados e suspeitos, distribuição geográfica, características demográficas dos pacientes, como faixa etária e sexo, gravidade dos casos e óbitos.

Esses dados foram fundamentais para a implementação de medidas estratégicas pelo governo brasileiro. Essas informações orientaram decisões relacionadas ao distanciamento social, quarentena, uso de máscaras, priorização de grupos na vacinação e alocação de recursos para hospitais (Brasil, 2020). A coleta de dados precisa e constante sobre a incidência da doença e a mortalidade foi crucial para a proteção da população e a gestão do sistema de saúde.

Paralelo a isso, a Inteligência Artificial (IA) começava a desempenhar um papel cada vez mais relevante na saúde, tanto no diagnóstico quanto na gestão de recursos hospitalares. Estudos como os de Braga et al. (2019) mostram que a IA é eficaz na previsão de tempo de internação e desfechos clínicos, permitindo uma alocação mais eficiente de recursos, como leitos e equipes médicas. Já no campo do diagnóstico, como apontam Magalhães et al. (2024), a IA tem potencial para melhorar a precisão e a rapidez na análise de imagens e exames médicos, reduzindo custos e otimizando o atendimento aos pacientes.

Neste trabalho é proposto um modelo de *Machine Learning* (ML) com potencial de melhorar a compreensão dos fatores que influenciam a propagação de doenças respiratórias e a eficácia das intervenções de saúde pública. A análise sistemática e a abordagem preditiva propostas podem ajudar a elaborar estratégias mais eficazes de vigilância epidemiológica e prevenção, ajudando a proteger a saúde da população em futuros surtos de doenças respiratórias.

O restante deste trabalho está organizado da seguinte forma: na Seção 2 são apresentados os trabalhos relacionados; na Seção 3 é apresentada a metodologia empregada na realização dos experimentos; já os resultados são apresentados e discutidos na Seção 4; finalmente, na Seção 5 são oferecidas a conclusão e sugestões para trabalhos futuros.

2. TRABALHOS RELACIONADOS

A detecção precoce e o controle de surtos de doenças respiratórias, especialmente em pandemias, são desafios constantes para os sistemas de saúde. A capacidade de prever a evolução dessas doenças é essencial para a implementação de medidas de prevenção e o uso adequado de recursos de saúde. Nesse sentido, o desenvolvimento de modelos preditivos, baseados em dados epidemiológicos e clínicos, pode fornecer às autoridades de saúde informações valiosas para antecipar surtos e mitigar seus impactos (Braga et al., 2019).

Os dados utilizados neste trabalho foram obtidos no portal OpenDATASUS, uma plataforma de dados de saúde do Ministério da Saúde que abrange informações sobre estatísticas hospitalares, procedimentos médicos e indicadores de saúde. No Brasil, a SRAG é notificada no SIVEP-Gripe quando o paciente apresenta Síndrome Gripal (SG) com sinais de gravidade, como dispneia, desconforto respiratório, saturação de oxigênio inferior a 95%, entre outros sintomas. A notificação dos casos deve ser feita em até 24 horas após a hospitalização, incluindo óbitos que ocorrem fora do ambiente hospitalar. O SIVEP-Gripe classifica os casos de SRAG como relacionados a Influenza, COVID-19, outros vírus respiratórios, outros agentes etiológicos ou SRAG não especificada (Altino et al., 2020).

Durante a pandemia, as hospitalizações por SRAG aumentaram significativamente, como evidenciado por Bastos et al. (2020), que observaram um crescimento acentuado dos casos nas primeiras semanas de 2020. A pressão sobre o sistema de saúde foi intensificada pela necessidade de cuidados intensivos e leitos hospitalares para tratar pacientes com complicações graves da COVID-19.

Diversos agentes etiológicos podem causar SRAG, sendo o vírus Influenza um dos mais comuns, responsável por casos graves de gripe, especialmente em crianças, idosos e pessoas com doenças crônicas. Outros agentes incluem o Vírus Sincicial Respiratório (VSR), que afeta principalmente crianças, o adenovírus, que causa infecções respiratórias e gastrointestinais, e o hantavírus, transmitido por roedores e que pode causar Síndrome Pulmonar por Hantavírus (SPH), uma condição grave e muitas vezes fatal.

A tese intitulada "Análise de Sobrevivência e Fatores Prognósticos Associados à Mortalidade em Pacientes com SRAG por COVID-19 Hospitalizados em UTI na Paraíba" (Faria, 2021) analisou as taxas de sobrevivência e os fatores prognósticos relacionados à mortalidade em pacientes com SRAG devido à COVID-19 que necessitaram de internação em UTI no estado da Paraíba. Foram utilizados dados registrados em um banco de dados da Secretaria Estadual de Saúde do Estado da Paraíba, que são as fichas de registro individuais dos casos de SRAG hospitalizados na rede do Sistema Único de Saúde do estado paraibano.

A Regressão Logística foi utilizada por Faria (2021) para avaliar a probabilidade de morte, as funções de sobrevivência foram calculadas utilizando o método de Kaplan-Meier e as razões de risco foram calculadas seguindo o modelo de riscos proporcionais de Cox. Os dados coletados incluíram informações sociodemográficas, clínicas e epidemiológicas dos pacientes. A amostra totalizou 2.084 observações, sendo 1.605 óbitos (77%) e 479 censuras (23%). A probabilidade de sobrevivência foi de 72,1% em 5 dias e 49,6% em 10 dias de internação segundo o método Kaplan-Meier. O tempo médio de internação em UTI até o óbito foi de 8,9 dias. Ademais, a análise revelou que diversos fatores influenciam significativamente a mortalidade de pacientes com SRAG por COVID-19 em UTIs.

Em outra área de estudo, por exemplo, a pesquisa de Mattos (2022) mostrou como a combinação das técnicas de aprendizado de máquina permitiram a criação de um sistema capaz de prever com alta precisão se uma transação de pagamento é fraudulenta, contribuindo para a segurança e confiabilidade das operações online.

Por fim, o estudo de Faria (2021) indica a importância de identificar características clínicas e sociodemográficas que impactam o prognóstico dos pacientes com SRAG por COVID-19. Fatores como idade avançada, necessidade de ventilação invasiva, baixa saturação de oxigênio e imunodepressão aumentam significativamente o risco de mortalidade. Por outro lado, residir em áreas periurbanas e não necessitar de ventilação invasiva são fatores que diminuem esse risco.

Os estudos demonstraram a evolução notável da eficácia dos métodos de aprendizado de máquina e de predição em comparação com abordagens manuais tradicionais, ressaltando a importância da utilização dessas tecnologias inovadoras na área da saúde, principalmente. Isso reforça a importância de algoritmos de aprendizado de máquina na medicina diagnóstica e a SRAG como objeto de estudo deste trabalho através de ML.

3. METODOLOGIA

O uso de aprendizado de máquina tem se destacado como uma solução promissora para a análise desses dados complexos e multivariados. Modelos preditivos como Regressão Logística, KNN, *Naive Bayes*, SVM, *Gradient Boosting*, Árvores de Decisão e Floresta Aleatória têm sido amplamente utilizados para ajudar na classificação e predição de doenças respiratórias. Esses modelos, quando adequadamente treinados e otimizados, podem identificar padrões ocultos nos dados e fornecer diagnósticos rápidos e precisos, o que é vital em emergências de saúde pública (Bughin et al., 2017).

A base de dados original utilizada neste trabalho compreendia um total de 190 colunas e 277.304 amostras. Observou-se que 84 colunas apresentavam mais de 80% dos dados faltantes, 0,3% das linhas da coluna objetivo, CLASSI_FIN, continham valores nulos, e foram identificados 196 registros duplicados. Além disso, não foram encontrados valores *outliers* significativos neste problema.

A base de dados apresentou uma distribuição significativa em relação à classe alvo, CLASSI_FIN, que representa a classificação final dos casos. Essa classificação é dividida em cinco categorias: (1) SRAG por influenza, (2) SRAG por outro vírus respiratório, (3) SRAG por outro agente etiológico, (4) SRAG não especificado e (5) SRAG por COVID-19.

A análise da Figura 1, que ilustra a divisão dos dados, revelou um desbalanceamento significativo. Dentro de uma base de dados com mais de 270 mil registros de casos, constatou-se que mais de 50% dos registros estavam classificados como 4 (SRAG não especificado). Outro ponto relevante é a classe de “SRAG por outro agente etiológico”, classificada como 3, que representava apenas 1,3% do total de casos mapeados.

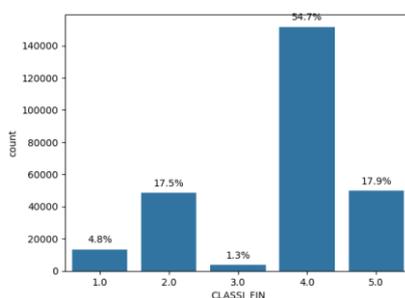


Figura 1. Percentagem de dados para cada classe
Fonte: elaborado pelos autores

Durante a fase de análise exploratória dos dados, foi identificado que a coluna CLASSE_FIN, utilizada como variável dependente (*target*), apresentava mais de 3,8% de valores nulos.

O tratamento dos dados visa garantir a qualidade das informações contidas na base de dados. Nesse contexto, foram realizadas as seguintes ações: remoção das colunas com mais de 80% dos dados nulos, exclusão dos 196 registros duplicados e eliminação de 52 colunas que apresentaram uma correlação extremamente baixa, com um valor de correlação entre $-0,12$ e $0,12$. Considerando que a presença de valores ausentes na coluna alvo pode comprometer a capacidade do modelo de aprendizado de máquina em aprender padrões e realizar previsões precisas, e que linhas duplicadas podem distorcer os resultados do modelo, optou-se pela remoção desses valores antes de avançar para a etapa de limpeza dos dados.

A remoção, assegurou que a base de dados estivesse completa e adequada para o treinamento do modelo, evitando problemas na avaliação das métricas de desempenho e nas previsões. No entanto, em outras colunas foi feito o tratamento dos dados nulos, colocando os mesmos com o valor de ignorado, dessa forma, foi possível manter a quantidade aceitável de linhas.

A partir dos dados mencionados, essa etapa do processo foi o ajuste dos valores nulos das colunas de sintomas e características conforme o dicionário de dados disponibilizado pelo DataSUS. Os valores foram padronizados como segue: 1 para confirmar, 2 para negar o sintoma ou característica, e 9 para ignorado. Assim, os valores nulos restantes nas colunas foram convertidos para 9, resultando em uma base de dados sem valores faltantes. Foi implementada a engenharia de atributos para a criação da coluna CORIZA. Observou-se em trabalhos correlatos que o sintoma de coriza é frequentemente utilizado na classificação de SRAG por ser um dos sintomas prevalentes citados nas fichas de notificação de SRAG e de Síndrome Gripal, doença que precede o comprometimento da função respiratória antes de se tornar crítico (Discacciati, 2022). Esse campo foi gerado a partir dos dados da coluna OUTRO_DES, a qual é uma coluna de texto livre e que apresenta outros sintomas que o paciente pode possuir, dentre esses sintomas diversos Coriza foi extraído e adicionado a uma coluna própria.

Essa extração foi feita de forma que independe se o texto estava em maiúsculo ou minúsculo. Também foi verificado a ocorrência de 26.832 vezes a ocorrência de Coriza. Em projetos de aprendizado de máquina, se faz necessário o tratamento de variáveis categóricas. Modelos de ML requerem dados numéricos para realizar cálculos matemáticos e identificar padrões. Como as variáveis categóricas, não possuem uma ordem ou valor numérico, foi utilizado neste trabalho o *OneHotEncoder*, o qual é uma função da biblioteca *scikit-learn* e tem como função transformar essas categorias em uma representação binária, garantindo que o modelo possa interpretá-las corretamente sem atribuir pesos incorretos às categorias. Ao utilizar o *OneHotEncoder*, cada categoria dentro de uma variável é convertida em uma nova coluna binária, onde o valor 1 indica a presença daquela categoria, e o valor 0 indica sua ausência.

Por fim, foi realizada a normalização de texto. Esse processo foi implementado para aprimorar a performance do modelo, uma vez que a normalização reduz a variabilidade e elimina inconsistências nos dados textuais. Ao uniformizar o formato dos dados, a normalização facilita a identificação de padrões pelo algoritmo de aprendizado de máquina, resultando em uma maior precisão e eficiência na modelagem.

Para garantir que todas as variáveis estivessem na mesma escala, os dados foram padronizados, ajustando-os de modo que possuam média zero e desvio padrão igual a um. Entretanto, algumas variáveis, como idade, apresentam valores com maior amplitude numérica. Antes de aplicar as técnicas de ML, foi necessário dividir os dados em duas partes: o conjunto de treino e o conjunto de teste. Essa separação foi feita para validar se o modelo é capaz de atingir resultados aceitáveis. Os dados foram divididos em 75% para treino, correspondendo a 106.029 linhas e 56 colunas para os valores de X (variáveis independentes), e 25% para teste, totalizando 35.343 linhas e 56 colunas para os valores de X. Adicionalmente, os valores y (variável de destino) para treino contaram com 106.029 linhas e 1 coluna, enquanto os valores y para teste tiveram 35.343 linhas e 1 coluna. Essa divisão foi realizada de forma estratificada, preservando a proporção das classes da variável de destino nos dois conjuntos.

Ademais, foi optado para o uso do método de *GridSearch*, ou busca em grade. Ao definir um conjunto de valores para cada hiperparâmetro, a busca em Grade avalia todas as combinações possíveis utilizando validação cruzada para determinar a configuração que maximiza uma métrica de desempenho. No caso desse trabalho, foi optado por utilizar a F-medida para essa avaliação. Inicialmente, todos os modelos foram executados com seus parâmetros padrões. Posteriormente, aplicou-se a busca em grade para otimizar os hiperparâmetros de cada modelo, com o objetivo de melhorar o desempenho nas previsões.

A seguir é apresentado na Tabela 1 os parâmetros padrões de cada algoritmo de aprendizado de máquina supervisionado, bem como o intervalo de busca e o melhor valor encontrado para cada um dos parâmetros buscados.

Tabela 1. Parâmetros utilizados com e sem a busca em grade

Método	Parâmetros	Valores padrão	Intervalo procurado	Melhor valor
Regressão Logística	C	1.0	0,01; 0.1; 1;10	10
	Penalty	L2	L1; L2	L2
	solver	lbfgs	Liblinear; lbfgs	lbfgs
KNN	n_neighbors	5	3; 5; 7; 9	3
	weights	uniform	Uniform; distance	distance
	p	2	1; 2	1
Naive Bayes	var_smoothing	1e-09	1e-9; 1e-8; 1e-7; 1e-6	1e-09
SVM	kernel	rbf	Linear; rbf	rbf
	c	1	0,1; 1; 10	10
	gamma	scale	Scale; auto	auto
	max_iter	7000	7000	7000
Gradient Boosting	n_estimators	100	100; 200; 300	300
	learning_rate	0.1	0,01; 0,05; 0,1	0,1
	max_features	None	None; sqrt; log2	sqrt
	max_depth	3	3; 5	3
Árvores de Decisão	max_depth	None	3; 5; 10	3
	min_samples_split	2	2; 4; 10	2
Floresta Aleatória	n_estimators	100	100; 200	200
	min_samples_split	2	10; 20	10
	min_samples_leaf	1	5; 10	5
	max_depth	None	3; 5; 10	10
	max_features	sqrt	Sqrt; log2	sqrt

Fonte: elaborado pelos autores

4. RESULTADOS

Neste capítulo, são apresentados os resultados obtidos com a aplicação dos modelos de classificação selecionados para este trabalho. Os resultados das métricas de ROC AUC e F-medida, utilizando a busca em grade em todos os modelos são apresentados na Tabela 2.

Tabela 2. Resultado geral com busca em grade. Fonte: elaborado pelos autores

Métodos	ROC AUC		F-medida	
	Treino	Teste	Treino	Teste
Regressão Logística	93,60%	93,12%	90,36%	90,20%
KNN	100,00%	81,72%	99,95%	82,06%
Naive Bayes	81,27%	80,83%	72,94%	72,78%
SVM	96,21%	93,16%	92,17%	90,95%
Gradient Boosting	99,07%	95,93%	95,67%	93,54%
Árvores de Decisão	92,34%	92,04%	92,09%	91,99%
Floresta Aleatória	97,19%	95,27%	92,97%	92,67%

Com base nos dados apresentados na Tabela 2, os resultados indicam que o modelo de *Gradient Boosting* obteve o melhor desempenho geral, com altos valores de ROC AUC e F-medida, tanto nos dados de treino quanto nos dados de teste. Da mesma forma, os algoritmos de SVM, Floresta Aleatória e Árvores de Decisão apresentaram bons resultados, com uma baixa discrepância entre as métricas de treino e teste, demonstrando uma boa capacidade de generalização. No entanto, o KNN, apesar de atingir 100% de ROC AUC nos dados de treino, apresentou uma queda acentuada nos testes, sugerindo um problema de *overfitting*.

Além disso, o *Naive Bayes* mostrou os valores mais baixos em ambas as métricas, refletindo um desempenho inferior em comparação com os demais modelos. A Regressão Logística teve um desempenho consistente, com valores próximos de 90% tanto para ROC AUC quanto para F-medida, mesmo sem grandes variações entre treino e teste.

Em particular, o uso da busca em grade foi essencial para melhorar o desempenho dos modelos de Árvores de Decisão e Floresta Aleatória, que inicialmente apresentaram sinais de *overfitting*. A busca em grade, ao ajustar o parâmetro *max_depth*, permitiu limitar a profundidade das árvores, evitando o excesso de ajuste aos dados de treino e melhorando a performance nos testes. Na Tabela 3, são apresentados os resultados do tempo de execução dos modelos, com e sem o uso de busca em grade

Tabela 3. Custo computacional de cada modelo

Algoritmos	Tempo de execução	Tempo utilizando busca em grade
Regressão Logística	6,3s	2min 31,9s
KNN	1min 4,8s	25min 4s
Naive Bayes	1,1s	6,3s
SVM	49min 40,5s	3 h 52 min e 24,9 s
Gradient Boosting	2min 47,3s	1 h 2 min e 49 s
Árvores de Decisão	2s	15s
Florestas Aleatórias	21,7s	3min 22,7s

Fonte: elaborado pelos autores

A análise do custo computacional dos modelos, com parâmetros padrões e Busca em Grade, mostra que o SVM apresentou o maior custo, especialmente ao usar a busca em grade. O *Gradient Boosting* também teve um custo elevado, mas inferior ao SVM. *Naive Bayes* e Árvore de Decisão foram os mais eficientes em termos de tempo de execução, enquanto a Regressão Logística e o KNN tiveram custos intermediários, com aumento no KNN ao aplicar busca em grade. Esses resultados indicam a importância de considerar o custo computacional ao escolher o algoritmo, especialmente em problemas médicos e SRAG.

Com base na análise anterior, são abordados os quatro modelos mais relevantes de diferentes famílias de algoritmos: SVM, *Gradient Boosting*, KNN e *Naive Bayes*. Esses algoritmos representam abordagens distintas em seus métodos de previsão, permitindo uma compreensão mais profunda de suas características no contexto do problema do SRAG. A seguir, na Figura 2, são apresentados os gráficos da Curva ROC, que avaliam o desempenho em teste desses quatro modelos.

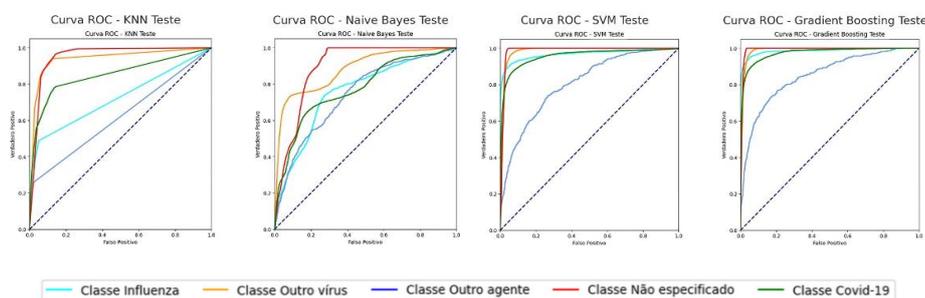


Figura 2. Curva ROC dos quatro modelos interessantes
Fonte: elaborado pelos autores

Os gráficos mostram que os métodos analisados tiveram resultados expressivos, com o *Gradient Boosting* e o SVM se destacando. O *Gradient Boosting* apresentou o melhor desempenho, com AUC ROC próxima de 1 para a maioria das classes, enquanto o SVM também obteve resultados muito positivos. O KNN teve um desempenho inferior, evidenciando *overfitting*, já que os resultados de teste foram piores que os resultados de treino. O *Naive Bayes* Gaussiano, embora simples, alcançou uma AUC de até 89% para algumas classes. Esses resultados confirmam a superioridade de modelos mais complexos, como o *Gradient Boosting* e o SVM, no problema de classificação de SRAG.

Com base nos dados, observa-se que o *Gradient Boosting* foi o modelo mais adequado para o problema de SRAG, apresentando os melhores resultados em precisão e um custo computacional balanceado. Em predições de problemas de saúde, a precisão diagnóstica é primordial, e o *Gradient Boosting* destacou-se ao obter ótimos acertos nas diferentes classes analisadas, tornando-se o método ideal para esta aplicação.

5. CONCLUSÃO

A SRAG representa um desafio crítico para a saúde pública, especialmente em momentos de pandemias, como a recente crise da COVID-19. A SRAG abrange uma gama de doenças respiratórias que, se não diagnosticadas e tratadas rapidamente, podem levar a complicações graves e óbitos. Um dos principais desafios é a necessidade de classificar rapidamente os casos e identificar a etiologia correta, uma vez que a SRAG pode ser causada por diferentes agentes.

No presente estudo, foi analisado e comparado o uso de algoritmos de aprendizado de máquina para a predição de casos de SRAG com base em dados do SIVEP-Gripe, além de destacar a importância da aplicação de técnicas avançadas de aprendizado de máquina na saúde pública brasileira. A contribuição do estudo está na aplicação prática de modelos preditivos e na validação de suas capacidades em um contexto real de saúde pública. Os principais modelos preditivos, incluindo Regressão Logística, KNN, *Naive Bayes*, SVM, *Gradient Boosting*, Árvores de Decisão e Floresta Aleatória, foram aplicados e comparados.

O *Gradient Boosting* destacou-se como o modelo mais adequado para a classificação dos casos, apresentando os melhores resultados em termos de precisão e um custo computacional equilibrado. Foi demonstrado que a precisão diagnóstica, fundamental no contexto de predição de problemas de saúde, pode ser maximizada com o uso desse modelo, que obteve altos índices de acerto nas diferentes classes de SRAG analisadas.

Além disso, o estudo reforçou a importância de métodos avançados de aprendizado de máquina na análise de grandes volumes de dados epidemiológicos, especialmente em momentos de crise sanitária, como a pandemia de COVID-19. A utilização desses modelos permitiu a identificação de padrões complexos nos dados, o que é decisivo para o controle e prevenção de doenças respiratórias graves. Assim, a pesquisa contribuiu para o campo da saúde pública ao validar o uso de *Machine Learning* como uma ferramenta eficaz no enfrentamento de crises de saúde, especialmente na predição de casos de SRAG.

REFERÊNCIAS

- Altino, Rita de Cássia et al. A importância da notificação compulsória frente à Síndrome Respiratória Aguda Grave (SRAG) e Covid-19. *Rev. Salusvita (Online)*, p. 627-649, 2020.
- Bastos, Leonardo Soares et al. COVID-19 e hospitalizações por SRAG no Brasil: uma comparação até a 12ª semana epidemiológica de 2020. *Cadernos de Saúde Pública*, v. 36, p. e00070120, 2020.
- Braga, A. V.; Lins, A. F.; Soares, L. S.; Fleury, L. G.; Carvalho, J. C.; Prado, R. S. do. Machine learning: O Uso da Inteligência Artificial na Medicina/ Machine learning: The Use of Artificial Intelligence in Medicine. *Brazilian Journal of Development*, [S. l.], v. 5, n. 9, p. 16407–16413, 2019. DOI: 10.34117/bjdv5n9-190. Disponível em: <https://ojs.brazilianjournals.com.br/ojs/index.php/BRJD/article/view/3437>. Acesso em: 17 mar. 2024.
- Brasil, M. da S. Ficha de registro individual-casos de síndrome respiratória aguda grave hospitalizados. instruções para preenchimento. SIVEP-Gripe. Ministério da Saúde, 2020.
- Bughin, J. et al. McKinsey Global Institute: Artificial Intelligence The Next Digital Frontier? Discussion Paper. p.11-39. 2017. Disponível em: <https://www.mckinsey.com/~media/mckinsey/industries/advanced%20electronics/our%20insights/how%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/mgi-artificial-intelligence-discussion-paper.ashx>. Acesso em: 19 mar. 2024.

- Croda, J. H. R.; Garcia, L. P. Resposta imediata da Vigilância em Saúde à epidemia da COVID-19. *Epidemiologia e Serviços de Saúde*, Brasília, DF, v. 29, n. 1, p. e2020002, 2020. EpubMar. 23, 2020. DOI: <https://doi.org/10.5123/s1679-49742020000100021>. Disponível em: <https://www.scielo.br/j/ress/a/zMMJJZ86vnrBdqPKtfsPL5w/?lang=pt>. Acesso em: 16 mar. 2024.
- Discacciati, Renata Gabrielli Pereira. Perfil epidemiológico da síndrome respiratória aguda grave (SRAG) e síndrome inflamatória multissistêmica pediátrica (SIM-P) associada à COVID-19 em indivíduos de 0 a 11 anos no estado do Pará: janeiro de 2020 a junho de 2022. 2022. 85 f. Trabalho de Conclusão de Curso (Bacharelado em Medicina) - Faculdade de Medicina, Instituto de Ciências da Saúde, Universidade Federal do Pará, Belém, 2022. Disponível em: <https://bdm.ufpa.br:8443/jspui/handle/prefix/6118>. Acesso em: 20 mar. 2024.
- Faria, Aline Roseane Queiroz de Paiva. Análise de sobrevivência e fatores prognósticos associados à mortalidade em pacientes com SRAG por Covid-19 hospitalizados em UTI na Paraíba. 2021. 174 f. Tese apresentada ao Programa de Pós-Graduação em Modelos de Decisão e Saúde – Nível Doutorado, Centro de Ciências Exatas e da Natureza, Departamento de Estatística, Universidade Federal da Paraíba, João Pessoa, 2021. Disponível em: <https://repositorio.ufpb.br/jspui/handle/123456789/22652>. Acesso em: 19 mar. 2024.
- Lorena, A. C.; DE CARVALHO, A. C. P. L. F. Uma Introdução às Support Vector Machines. *Revista de Informática Teórica e Aplicada*, [S. l.], v. 14, n. 2, p. 43–67, 2007. DOI: 10.22456/2175-2745.5690. Disponível em: https://seer.ufrgs.br/index.php/rita/article/view/rita_v14_n2_p43-67. Acesso em: 24 apr. 2024.
- Magalhães, Maria Iranilda Silva; VARGAS, Conrado Vilela; BOMFIM, Vitoria Vilas Boas da Silva; FERREIRA, Tainan Gomes; BEHRENS, Priscila de Almeida Castro. IMPACTO DA INTELIGÊNCIA ARTIFICIAL NO DIAGNÓSTICO MÉDICO: DESAFIOS E OPORTUNIDADES. *Revista Ibero-Americana de Humanidades, Ciências e Educação*, [S. l.], v. 10, n. 1, p. 1477–1485, 2024. DOI: 10.51891/rease.v10i1.12819. Disponível em: <https://periodicorease.pro.br/rease/article/view/12819>. Acesso em: 17 mar. 2024.
- Mattos, Lucas de. Aplicação De Técnicas De Machine Learning No Apoio À Detecção De Fraudes Em Pagamentos Online. 2022.