



FURG

UNIVERSIDADE FEDERAL DO RIO GRANDE - FURG

CENTRO DE CIÊNCIAS COMPUTACIONAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

CURSO DE MESTRADO EM ENGENHARIA DE COMPUTAÇÃO

Christian Giménez Barañano

**Aprendizado de Máquina para Predição de Sequelas de Pacientes Com
Diagnóstico de COVID-19**

Rio Grande, RS

2024

Christian Giménez Barañano

**Aprendizado de Máquina para Predição de Sequelas de Pacientes Com
Diagnóstico de COVID-19**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal do Rio Grande – FURG, como requisito para a obtenção do título de Mestre em Engenharia da Computação.

Orientador(a): Prof. Dr. Adriano Velasque Werhli

Rio Grande, RS

2024

Ficha Catalográfica

B227a Barañano, Christian Giménez.

Aprendizado de máquina para predição de sequelas de pacientes com diagnóstico de COVID-19 / Christian Giménez Barañano. – 2024.

99 f.

Dissertação (mestrado) – Universidade Federal do Rio Grande – FURG, Programa de Pós-Graduação em Computação, Rio Grande/RS, 2024.

Orientador: Dr. Adriano Velasque Werhli.

1. COVID-19 2. HU-FURG 3. Mineração de dados 4. Ciência de dados I. Werhli, Adriano Velasque II. Título.

CDU 004:616

Catálogo na Fonte: Bibliotecário José Paulo dos Santos CRB 10/2344

DISSERTAÇÃO DE MESTRADO

Aprendizado de Máquina para Predição de Sequelas de Pacientes Com Diagnóstico de COVID-19

CHRISTIAN GIMÉNEZ BARAÑANO

Banca examinadora:

Documento assinado digitalmente
 **KARINA DOS SANTOS MACHADO**
Data: 26/11/2024 19:32:52-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Karina dos Santos Machado

Documento assinado digitalmente
 **CARLOS HENRIQUE DA SILVEIRA**
Data: 22/11/2024 07:41:00-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Carlos Henrique da Silveira

Documento assinado digitalmente
 **ADRIANO VELASQUE WERHLI**
Data: 21/11/2024 11:56:27-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Adriano Velasque Werhli
Orientador

Dedico esta conquista aos meus pais por sempre acreditarem em mim. E à minha esposa, por seu apoio e paciência.

AGRADECIMENTOS

Agradeço primeiramente a Deus; sem Ele nada seria possível.

Aos colegas de mestrado Cristiano de Moura Borges e Karine de Oliveira Fonseca, pelas suas contribuições em vários momentos.

Aos professores do curso, Prof. Dr. Eduardo Nunes Borges, Prof. Dr. Vagner Santos da Rosa pelo auxílio e incentivo. Especialmente, agradeço ao Prof. Dr. Adriano Veslasque Werhli pela sua orientação e valiosas contribuições ao longo deste processo.

Aos membros da banca, Profa. Dra. Karina dos Santos Machado e Prof. Dr. Carlos Henrique da Silveira, pelas suas contribuições que enriqueceram este trabalho.

À secretária Caroline Drawanz Dias, sempre disposta a prestar auxílio e esclarecimentos quantos às regras e formalidades do curso.

À médica Msc.Heruza Einsfeild Zogbi, pela sua orientação no quesito médico do trabalho.

À Profa. Dra. Audrei Giménez Barañano pela consultoria em estrutura e confecção do trabalho.

E agradeço a todas as pessoas que de alguma forma fizeram parte desta etapa da minha vida.

RESUMO

A COVID-19, doença causada pelo vírus SARS-CoV-2, impactou o mundo com intensidade sem precedentes, levando a um alto número de óbitos e a uma crise econômica global, além de mergulhar os sistemas de saúde em um estado de emergência contínua. Em resposta, esforços de pesquisa emergiram mundialmente para investigar múltiplos aspectos da doença, incluindo sua origem, mecanismos de transmissão, sintomas, desenvolvimento de vacinas e potenciais sequelas a longo prazo. Este trabalho propôs uma investigação com foco regional, voltada para a análise dos dados de pacientes diagnosticados com COVID-19 que foram atendidos pelo Hospital Universitário da Universidade Federal do Rio Grande (HU-FURG). O estudo examinou o tempo de internação inicial, analisando dados como vacinação, comorbidades pré-existentes, medicamentos utilizados, métodos de tratamento, idade do paciente, exames realizados e tempo de internação desses pacientes. E também os dados relacionados ao seu retorno ao hospital, como exames realizados, no qual relataram sintomas persistentes e que foram diagnosticados como sequelas da COVID-19. Os dados desses dois momentos do paciente no HU-FURG (internação e retorno pós recuperação) foram compilados e analisados, utilizando técnicas de Aprendizado de Máquina. De todos os dados disponíveis, foram retirados todos os que foram considerados irrelevantes ou que poderiam influenciar de forma equivocada o resultado final. Campos sem dados informados, ou com registro de dados idênticos (ocorrência única) foram removidos. Outros, com dados muito diferenciados, foram convertidos em faixas (como no caso da idade, convertidos em faixas como criança, adulto, etc). Após o tratamento inicial desses dados, foram submetidos à algoritmos de predição, visando prever qual desses algoritmos seria o mais adequado à obtenção do resultado final, considerando também que o volume de dados obtido da base do HU-FURG é consideravelmente pequena. Para a obtenção dos resultados, foi utilizada a técnica de aprendizado supervisionado. Primeiramente, dividindo a base inicial em duas partições, uma de treino e a segunda de teste. Após, foi utilizado o cross-validation, onde a base original foi dividida em 10 partições, e empregado o método k-fold. A conclusão foi que o algoritmo Naive Bayes é o mais adequado, produzindo os melhores resultados em relação aos demais algoritmos testados para este cenário. Os resultados desse estudo pretendem contribuir para uma compreensão mais profunda dos efeitos prolongados da COVID-19 e auxiliar no planejamento de cuidados a longo prazo para os pacientes afetados.

Palavras-chave: COVID-19, HU-FURG, mineração de dados, ciência de dados

ABSTRACT

COVID-19, the disease caused by the SARS-CoV-2 virus, has impacted the world with unprecedented intensity, leading to a high number of deaths, a global economic crisis, and overwhelming health systems with a continuous state of emergency. In response, research efforts emerged worldwide to investigate multiple aspects of the disease, including its origin, transmission mechanisms, symptoms, vaccine development, and potential long-term sequelae. This study proposed a regional-focused investigation aimed at analyzing data from patients diagnosed with COVID-19 who were treated at the University Hospital of the Federal University of Rio Grande (HU-FURG). The study examined the initial hospitalization period by analyzing data such as vaccination status, pre-existing comorbidities, medications used, treatment methods, patient age, tests performed, and the length of hospital stay. It also considered data related to their return to the hospital, including tests conducted for patients reporting persistent symptoms, which were diagnosed as COVID-19 sequelae. The data from these two patient moments at HU-FURG (hospitalization and post-recovery return) were compiled and analyzed using Machine Learning techniques. From all the available data, irrelevant or potentially misleading data were excluded. Fields with missing information or identical records (single occurrences) were removed. Other data with significant variability were grouped into ranges (e.g., age categorized into groups such as child, adult, etc.). After this initial data preprocessing, predictive algorithms were applied to determine which algorithm would be the most suitable for obtaining the final results, especially considering the relatively small dataset obtained from the HU-FURG database. To achieve the results, supervised learning techniques were used. Initially, the original dataset was divided into two partitions: one for training and another for testing. Subsequently, cross-validation was applied, where the original dataset was divided into 10 partitions, employing the k-fold method. The conclusion was that the Naive Bayes algorithm proved to be the most suitable, delivering the best results compared to other tested algorithms for this scenario. The findings of this study aim to contribute to a deeper understanding of the prolonged effects of COVID-19 and to assist in planning long-term care for affected patients.

Keyword: COVID-19, HU-FURG, data mining, datascience

LISTA DE FIGURAS

Figura 1- Coronavírus, causador da COVID-19.....	23
Figura 2 - Ciclo de Transmissão do Vírus	23
Figura 3 - Evolução do COVID-19 no Brasil.....	24
Figura 4 - Sintomas Comuns	27
Figura 5 - Transmissão por Aerossol	28
Figura 7 - As cinco etapas da Descoberta de Conhecimento.....	31
Figura 8 - Exemplo de Árvore de Decisão	35
Figura 9 - Exemplo de Random Forest.....	37
Figura 10 - Evolução do XGBoost a partir da Árvore de Decisão	39
Figura 11 - Modelo Simples de Rede Neural	41
Figura 12 - Modelo de Rede Neural com duas camadas de processamento.....	41
Figura 13 - Modelo de Rede Neural Recorrente	42
Figura 14 - Fluxograma do projeto.....	46
Figura 15 – Comportamento das Médias das Métricas em relação aos Métodos.....	58
Figura 16 - Comparativos das Partições - 15 classes	60
Figura 17 - Comparativo das Partições - 3 classes	61
Figura 18 - Comparativo das Partições - 2 classes	61
Figura 19 - Banco de Dados COVID-19 HU-FURG	76
Figura 20 - Banco de Dados COVID-19 HU-FURG - Quadrante 1	77
Figura 21 - Banco de Dados COVID-19 HU-FURG - Quadrante 2	78
Figura 22 - Banco de Dados COVID-19 HU-FURG - Quadrante 3	79
Figura 23 - Tela principal do sistema Covid do HU-FURG.....	80
Figura 24 - Tela dados do paciente do sistema Ficha Covid-19 do HU-FURG	80
Figura 25 - Tela dados de comorbidades do paciente do sistema Covid-19 do HU-FURG.....	80
Figura 26 - Tela dados de medicações usadas no tratamento do paciente do sistema Covid-19 do HU-FURG	81
Figura 27 - Tela dados de sintomas referidos do paciente do sistema Covid-19 do HU-FURG	81
Figura 28 - Tela dados de sinais de gravidade do paciente do sistema Covid-19 do HU-FURG	81
Figura 29 - Tela dados de evidência de infecção do paciente do sistema Covid-19 do HU-FURG	81
Figura 30 - Tela dados de RX do tórax do paciente do sistema Covid-19 do HU-FURG	81
Figura 31 - Tela dados de tomografia do tórax do paciente do sistema Covid-19 do HU-FURG	82
Figura 32 - Tela dados de prescrição respiratória do paciente do sistema Covid-19 do HU-FURG	82
Figura 33 - Tela dados de oferta de O2 do paciente do sistema Covid-19 do HU-FURG	82
Figura 34 - Tela dados de dispositivos invasivos do paciente do sistema Covid-19 do HU-FURG	82
Figura 35 - Tela dados de RT-PCR do paciente do sistema Covid-19 do HU-FURG	82
Figura 36 - Tela dados de antígeno do paciente do sistema Covid-19 do HU-FURG.....	82
Figura 37 - Tela dados de TG IGG/IGM do paciente do sistema Covid-19 do HU-FURG	83
Figura 38 - Tela dados de ECG do paciente do sistema Covid-19 do HU-FURG	83
Figura 39 - Tela dados de cultura do paciente do sistema Covid-19 do HU-FURG	83
Figura 40 - Tela dados de evolução do paciente do sistema Covid-19 do HU-FURG.....	83
Figura 41 - Tela dados de vacinas do paciente do sistema Covid-19 do HU-FURG	83
Figura 42 - Tela dados de tromboembolismo do paciente do sistema Covid-19 do HU-FURG	

.....	83
Figura 43 - Tela dados de evolução do paciente do sistema Pós-Covid do HU-FURG	84
Figura 44 - Tela dados de diagnóstico do paciente do sistema Pós-Covid do HU-FURG	84
Figura 45 - Tela dados de tomografia do paciente do sistema Pós-Covid do HU-FURG	84
Figura 46 - Tela de Evolução pós-covid	85
Figura 47 - Tela de Diagnóstico pós-covid.....	85
Figura 48 - Resultado de Tomografia pós-covid	85

LISTA DE TABELAS

Tabela 1 – Variantes da COVID-19 e seus sintomas	26
Tabela 2 - Tempo do Coronavírus Ativo em Superfícies.....	29
Tabela 3 – Modelo de Matriz de Confusão – duas classes	42
Tabela 4 – Modelo de Matriz de Confusão - multiclasse	43
Tabela 5 - Distribuição dos Dados da Pesquisa	52
Tabela 6 - Balanceamento dos Dados Pesquisados - 2 classes.....	52
Tabela 7 - Balanceamento dos Dados Pesquisados - 3 classes.....	52
Tabela 8 - Balanceamento dos Dados Pesquisados - 15 classes.....	53
Tabela 9 - Comparativo de Resultados, com 15 classes, 3 classes e 2 classes. 74% dos registros como base de treino e 26% como base de teste	56
Tabela 10 - Médias das Médias das Métricas dos Algoritmos Utilizados na modalidade Cross-Validation - 15 classes	59
Tabela 11 - Médias das Médias das Métricas dos Algoritmos Utilizados na modalidade Cross-Validation - 3 classes	60
Tabela 12 - Médias das Médias das Métricas dos Algoritmos Utilizados na modalidade Cross-Validation - 2 classes	61
Tabela 13 - Comparativo das Médias das Médias das Métricas dos Algoritmos Utilizados na modalidade Cross-Validation, 15 classes.....	64
Tabela 14- Campos da tabela de cadastro de Materiais	86
Tabela 15 – Campos da tabela de cadastro de Comorbidades	86
Tabela 16 - Campos da tabela de cadastro de Vacinas.....	86
Tabela 17 - Campo da tabela base de cadastro de Pacientes	86
Tabela 18 - Campos da tabela de cadastro de Agente Infeccioso	87
Tabela 19 - Campos da tabela de cadastro de Amostras de Cultura	87
Tabela 20 - Campos da tabela de cadastro de Dispositivos Invasivos.....	87
Tabela 21 - Campos da tabela de cadastro de Focos	88
Tabela 22 - Campos da tabela de cadastro de Oferta de O2	88
Tabela 23 - Campos da tabela de cadastro de Exames de Raio-x/Tomografia.....	88
Tabela 24 - Campos da tabela de cadastro de Sinais de Gravidade.....	89
Tabela 25 - Campos da tabela de cadastro de Sintomas	89
Tabela 26 - Campos da tabela de cadastro de Situação de Exames.....	89
Tabela 27 - Campos da tabela de cadastro de Medicamentos	89
Tabela 28 - Campos da tabela de cadastro de Diagnósticos Pós-Covid	90
Tabela 29 - Campos da tabela de Consultas do paciente	90
Tabela 30 - Campos da tabela de Comorbidades.....	90
Tabela 31 - Campos da tabela da Ficha de Covid-19	90
Tabela 32 - Campos da tabela de Culturas do paciente	91
Tabela 33 - Campos da tabela de Dispositivos Invasivos.....	91
Tabela 34 - Campos da tabela de Evidências de Infecção	92
Tabela 35 - Campos da tabela de Evolução do paciente.....	92
Tabela 36 - Campos da tabela de infecção por H1N1	92
Tabela 37 - Campos da tabela de Medicamentos do paciente	93
Tabela 38 - Campos da tabela de Oferta de O2	93
Tabela 39 - Campos da tabela de Prescrição Respiratória.....	93
Tabela 40 - Campos da tabela de Exames RT-PCR / Antígeno	93
Tabela 41 - Campos da tabela de Exames de Raio-X / Tomografia Computadorizada.....	94
Tabela 42 - Campos da tabela de Saturação de O2.....	94
Tabela 43 - Campos da tabela de Sinais de Gravidade do paciente	95

Tabela 44 - Campos da tabela de Sintomas do paciente.....	95
Tabela 45 - Campos da tabela de Teste Rápido de IgG/IgM	95
Tabela 46 - Campos da tabela de Tromboembolismo do paciente	95
Tabela 47 - Campos da tabela de Vacinas do paciente	96
Tabela 48 - Campos da tabela de Gasometria do paciente	96
Tabela 49 - Campos da tabela de Internação do paciente.....	96
Tabela 50 - Campos da tabela de medicamentos de Evidência de Infecção.....	97
Tabela 51 - Campos da tabela de paciente.....	97
Tabela 52 - Campos da tabela de paciente covid.....	97

LISTA DE ABREVIATURAS E SIGLAS

ACC	Acurácia (do inglês <i>Accuracy</i>)
AD	Árvore de Decisão
ANN	Rede Neural Artificial (do inglês <i>Artificial Neural Network</i>)
CNN	Rede Neural Convolucional (do inglês <i>Convolutional Neural Network</i>)
CSG	Grupo de Estudo Coronaviridae do Comitê Internacional de Taxonomia de Vírus (do inglês <i>Coronaviridae Study Group of the International Committee on Taxonomy of Viruses</i>)
DM	Mineração de Dados (do inglês <i>Data Mining</i>)
ECG	Eletrocardiograma
ESPII	Emergência de Saúde Pública de Importância Internacional
FN	Falso Negativo
FP	Falso Positivo
IA	Inteligência Artificial (do inglês <i>AI – artificial intelligence</i>)
KDD	Descoberta de Conhecimentos em Bases de Dados (do inglês <i>Knowledge Discovery in Databases</i>)
NB	Naive Bayes
NHS	Sistema Nacional de Saúde Inglês
OMS	Organização Mundial de Saúde
OPAS	Organização Pan-Americana de Saúde
PRC	Precisão (do inglês <i>Precision</i>)
RCL	Recall
RF	Random Forest
RN	Rede Neural
RNN	Rede Neural Recorrente (do inglês <i>Recurrent Neural Network</i>)
RT-PCR	reação da transcriptase reversa seguida pela reação em cadeia da polimerase (do inglês “reverse-transcriptase-polymerase-chain-reaction”)
RX	Exame de raio-x
SARS-Cov2	Síndrome Respiratória Aguda Grave – coronavírus 2 (do inglês <i>Severe Acute Respiratory Syndrome-Coronavirus-2</i>)
SIM-P	Síndrome Inflamatória Multissistêmica Pediátrica
SpO2	Saturação de Oxigênio
TC do tórax	Tomografia Computadorizada do Tórax

TR IgG/IgM	Teste Rápido Imunoglobulina G / Teste Rápido Imunoglobulina M
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
XG	XGBoost

SUMÁRIO

1	INTRODUÇÃO.....	16
1.1	Justificativa.....	17
1.1.1	Motivação	19
1.2	Objetivo geral	20
1.3	Objetivos específicos	20
1.4	Aspectos Éticos.....	21
1.4.1	Processo e método de consentimento e/ou assentimento (forma de acesso ao paciente/ informações e documento de registro da participação do público na pesquisa	21
1.4.2	Riscos e Benefícios.....	21
1.4.3	Privacidade e Confidencialidade dos Dados	21
1.5	Contribuições da pesquisa para HU-Furg/Ebserh	21
2	REVISÃO DA LITERATURA – COVID-19.....	22
2.1	Histórico da COVID-19	22
2.1.1	Histórico da pandemia de COVID-19	22
2.1.2	Emergência de Saúde Pública de Importância Internacional	25
2.2	Folha informativa sobre COVID-19.....	25
2.3	Apresentação da Doença	26
2.4	Transmissão: Gotícula X Aerossol	27
2.5	Covid Longa	29
3	PESQUISAS RELACIONADAS.....	30
4	REVISÃO DA LITERATURA – ALGORITMOS.....	31
4.1	Seleção dos Dados.....	31
4.1.1	Preparação dos Dados.....	31
4.1.2	Pré-processamento e Limpeza dos Dados	32
4.1.3	Transformação dos Dados	32
4.2	Mineração dos Dados	32
4.3	Aprendizado de Máquina.....	33
4.3.1	Aprendizado de Máquina Supervisionado.....	34
4.4	Algoritmos de Classificação Utilizados	34
4.4.1	Árvore de Decisão	34
4.4.2	Random Forest.....	36
4.4.3	Naive Bayes.....	38
4.4.4	Xgboost.....	39
4.4.5	Rede Neural	40
4.5	Métricas de Avaliação.....	42

4.5.1	Matriz de Confusão	42
4.5.2	Acurácia.....	43
4.5.3	Precisão.....	43
4.5.4	Recall.....	44
4.5.5	F1	44
4.5.6	Métricas para Problemas Multiclasse	44
5	METODOLOGIA.....	46
5.1	Apresentação	46
5.2	Sistemas Fontes de Dados	48
5.3	Geração dos Modelos	51
6	RESULTADOS	52
6.1	Interpretação e Avaliação dos Resultados	52
6.1.1	Naive Bayes.....	53
6.1.2	Random Forest.....	54
6.1.3	Árvore de Decisões.....	54
6.1.4	XGBoost	55
6.1.5	Rede Neural	56
6.1.6	Divisão dos Dados em Partições	58
7	DISCUSSÃO	63
8	CONCLUSÃO.....	66
9	REFERÊNCIAS	68
	APÊNDICE A – Banco de Dados COVID-19 HU-FURG	76
	APÊNDICE B – Telas de controle de pacientes internados por COVID-19.....	80
	APÊNDICE C – Telas de controle de pacientes pós-covid	85
	APÊNDICE D – Dicionário de dados do sistema origem	86

1 INTRODUÇÃO

Não se sabe a data exata do surgimento dos vírus, mas sabe-se que já existiam quando os primeiros humanos começaram a povoar o planeta. De fato, os vírus vêm influenciando há milhões de anos a evolução da espécie humana (SILVA, DELATORRE, *et al.*, 2023). Mas, o mais importante, os vírus podem causar doenças, muitas delas fatais.

Há um número gigantesco de vírus, que podem ser divididos em famílias. Uma dessas famílias são os coronavírus. Essa família de vírus em particular é comum em muitas espécies de animais, como camelos, gado, gatos, morcegos e também o homem (MINISTÉRIO DA SAÚDE, 2023).

O SARS-Cov-2 pertence à família *Coronaviridae*, e é o causador da COVID-19 (UZUNIAN, 2020). O vírus foi inicialmente observado em dezembro de 2019 na cidade de Wuhan, na China. Os doentes tinham em comum o contato prévio com o mercado de Wuhan, conhecido por vender alimentos da cultura local, como animais considerados exóticos para ocidentais (TOZZI, LOURENÇO, *et al.*, 2024). Em questão de meses, o vírus havia se espalhado globalmente, atingiu mais de 200 países, apresentando taxas crescentes tanto no número de casos quanto no número de mortes (ALVES, 2020). O SARS-CoV-2 vitimou pessoas de todos os níveis sociais. Muitas pessoas acometidas por essa doença morreram (OMS afirma que COVID-19 é agora caracterizada como pandemia, 2020); muitas outras apresentaram sequelas (algumas delas, permanentes), agora identificadas como covid longa. A covid longa é uma nova entidade nosológica surgida na sequência da covid-19. Ela está impulsionando pesquisas sobre a sua prevalência, sintomas, diagnóstico, duração e métodos de tratamento (SEGATA e LÖWY, 2024).

A infecção pelo SARS-CoV-2 pode variar de casos assintomáticos e manifestações clínicas leves, até quadros moderados, graves e críticos, sendo necessária atenção especial aos sinais e sintomas que indicam piora do quadro clínico que exijam a hospitalização do paciente. (MINISTÉRIO DA SAÚDE, 2021)

Grande parte dos pacientes acometidos por essa doença conseguiram se recuperar. Mas um certo número desses pacientes voltaram às unidades de saúde, com reclamações de desconfortos. Os sintomas mais comumente relatados foram fadiga, dispneia, distúrbio do sono e dificuldade de concentração (32%, 25%, 24% e 22%, respectivamente, no acompanhamento de 3 a 12 meses) (ALKODAYMI, OMRANI, *et al.*, 2022). Esses novos sintomas apresentados pelos pacientes são relacionados à COVID-19, e denominados por pós-covid, ou covid longa. Essa condição é reconhecida pela OMS, que define a condição pós-

covid como a presença de sintomas persistentes por pelo menos três meses após a infecção inicial, sem outra explicação diagnóstica. Estes sintomas variam amplamente e afetam tanto crianças quanto adultos (OPAS - ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE, 2024).

Portanto, existe hoje um grande interesse em um diagnóstico preciso dessa nova condição, bem como um prognóstico da mesma, levando a um aumento no interesse do uso de métodos computacionais nesta tarefa devido ao grande volume de dados de pacientes, sendo as técnicas de inteligência computacional as mais promissoras.

A inteligência artificial é um ramo da ciência da computação que se dedica no desenvolvimento de sistemas e algoritmos que desempenhem funções que normalmente requerem inteligência humana. Como exemplo teríamos reconhecimento de padrões, tomada de decisão, aprendizado e outras (RUSSEL e NORVIG, 2003)

Nos últimos anos ocorreu um crescimento acelerado da presença da Inteligência Artificial (IA) no nosso cotidiano. Muitas vezes utilizamos modelos criados por algoritmos de IA (FACELI, LORENA, *et al.*, 2021). Pode ser empregada nos mais variados campos de atuação desde assistentes virtuais, reconhecimento facial, lavar roupas, controle de carros autônomos (CARVALHO, 2021); e medicina.

1.1 Justificativa

A partir de dezembro de 2019 o mundo se deparou com um novo vírus (apesar de pertencer a uma família de vírus já conhecida, os coronavírus), que se caracterizou por ter um alto nível de contágio entre humanos e também grande letalidade da doença causada (BRITO, BRAGA, *et al.*, 2020). Até o dia 03/11/2024 havia 776.798.109 casos confirmados e 7.074.387 óbitos (WORLD HEALTH ORGANIZATION, 2024). Uma doença nova, para a qual a humanidade não possuía imunidade, com taxas de transmissão elevadas. Pode saturar os serviços de saúde caso medidas não farmacológicas não sejam adotadas. (SILVA, MINAYO e GOMES, 2020)

No Brasil não foi diferente. Hospitais lotados, muitos pacientes sem ao menos conseguir tratamento foram fatos marcantes durante a pandemia. Relatos diários na imprensa traziam números atualizados de vítimas, e sempre um número superior ao do dia anterior. Em 05/12/2024 são 39.030.239 casos confirmados, com 714.306 mortes (PAINEL CORONAVÍRUS, 2024). E cerca de 50% das pessoas diagnosticadas com COVID-19 apresentaram sintomas pós-infecção (MIRANDA, GOMES, *et al.*, 2022).

O Rio Grande do Sul seguiu na mesma esteira. Em 31/12/2020, eram 494.323 casos confirmados; em 31/12/2021, 1.493.183 casos confirmados; em 31/12/2022, 2.928.671 casos confirmados (PAINEL CORONAVÍRUS RS, 2024). Em 05/12/2024, 3.151.125 casos confirmados (PAINEL CORONAVÍRUS, 2024). Mas, teve um agravante: o inverno. As doenças respiratórias são mais disseminadas por hábitos relacionados ao frio, como a maior permanência em espaços fechados e com pouca ventilação (SECOM - RS, 2024). Este fato impacta ainda mais a população, ocasionando ainda o aumento das filas para o atendimento nos hospitais.

A COVID-19 tornou-se na cidade de Rio Grande, assim como no resto do mundo, um grande problema: contaminou muitas pessoas de forma muito rápida, indiscriminadamente, e com várias vítimas fatais. Até o dia 10/12/2024 havia 3.153.249 casos confirmados e 43.095 óbitos (PAINEL CORONAVÍRUS RS, 2024). Algumas delas, por problemas decorrentes da COVID-19, mesmo após a fase aguda da doença. Esses problemas decorrentes da COVID-19 são agora denominados covid longa.

A infecção pelo SARS-CoV-2 pode variar de casos assintomáticos e manifestações clínicas leves, até quadros moderados, graves e críticos, sendo necessária atenção especial aos sinais e sintomas que indicam piora do quadro clínico que exijam a hospitalização do paciente. (MINISTÉRIO DA SAÚDE, 2021)

Existem várias pesquisas sobre covid longa. Pesquisa com brasileiros que tiveram COVID-19 demonstra que quase 60% deles são acometidos pela covid longa, tendo sintomas a partir de três meses após a fase aguda (MARQUES, 2023). Em outra pesquisa, o Ministério da Saúde vai visitar 33.250 pessoas que tiveram a doença e que residem em 133 municípios no Brasil, visando identificar as pessoas impactadas pela covid longa e criar políticas públicas de assistência (LABOISSIÈRE, 2024). Há o estudo FENIX da Universidade Federal de São Paulo (AZEVEDO, SANTOS, *et al.*, 2022) e o Levantamento de Sintomas Persistentes em Pacientes com Covid Longa (EICH, SCHNEIDER, *et al.*, 2023). A OPAS está desenvolvendo pesquisa sobre as sequelas surgidas após a infecção pelo vírus, visando melhorar o atendimento dos pacientes (OPAS - ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE, 2022); e há também a pesquisa sobre a influência da atividade física sobre pessoas com covid longa, para identificar como a atividade física beneficia as pessoas que ainda apresentam sintomas, para refletir sobre como a covid longa interfere no estado físico destas pessoas e compreender o impacto da atividade física na recuperação de sequelas (SILVA e BRAUER, 2022)

Com o intuito de evitar as complicações da covid longa, foi criado um sistema de prognóstico da doença, usando para tanto os dados disponíveis no sistema do HU-FURG. Tal sistema tem o potencial de diminuir os efeitos da covid longa, pois permitiria o início do tratamento imediatamente, aumentando assim a chance de cura.

1.1.1 Motivação

Muito já se falou sobre a COVID-19; busca-se saber onde começou, quais os problemas que o paciente apresenta quando infectado, tempo de internação e como combater essa doença. Tornou-se necessário também acompanhar esses pacientes após sua recuperação, pelo possível acometimento da covid longa.

A covid longa segue subestimada; não foi estabelecida uma definição sobre o que seria a covid longa. Estima-se que a prevalência varie de 10-70% até 24 meses pós-infecção por SARS-CoV-2. E mesmo 24 meses após, sintomas neuropsicológicos têm sido frequentemente relatados (RAMOS JR., 2024).

Foi necessário se proteger contra algo sobre o qual nada se sabia; e que se espalhou rápido. A vacina foi a grande barreira, que conseguiu conter os seus avanços. Desenvolvida em tempo recorde, utilizando as técnicas mais modernas disponíveis, e com várias frentes (vários países criaram suas próprias vacinas), foi o que finalmente conseguiu bloquear a contaminação em grande escala. O vírus pode não desaparecer, então o grande desafio é controlar sua disseminação. As vacinas são seguras, protegem coletivamente e protegem contra variantes (VICTOR, 2023).

A OMS recomendou várias vacinas contra a COVID-19 (produzidas pelos fabricantes Pfizer/BioNTech, AstraZeneca/Oxford, Janssen, Moderna, Sinopharm, Sinovac, Bharat, Novavax, Casino e Valneva), e elas são muito eficazes na prevenção de infecções graves, hospitalização e morte contra todas as variantes do vírus SARS-CoV-2, incluindo a ômicron e suas sub-linhagens (dados atualizados em 31/07/2023). Além disso, as vacinas são altamente eficazes na redução da transmissão do vírus, embora possam não impedir totalmente a infecção. (Perguntas frequentes: vacinas contra a COVID-19, 2023)

A vacina também reduz os sintomas em quem porventura veio a contrair a doença. E pessoas vacinadas têm menos risco de desenvolver a covid longa do que aqueles que não tomaram vacina (PORTAL DO BUTANTAN, 2023).

Após o início da vacinação, houve uma grande redução no número de casos graves e óbitos; mesmo a transmissão diminuiu consideravelmente. Entretanto, os sobreviventes da

COVID-19 podem apresentar por muito tempo sintomas persistentes relacionados a sequelas cardiopulmonares, neurológicas, psicológicas, entre outras. A persistência dos sintomas por mais de 12 semanas após a infecção aguda tem sido denominada síndrome pós-covid (ou covid longa) e tem grande impacto na qualidade de vida e no estado de saúde, ao influenciar as atividades da vida diária e o retorno ao trabalho negativamente, com consequências para a saúde mental desses pacientes. (AZEVEDO, SANTOS, *et al.*, 2022)

Há muitas pesquisas sobre a covid longa; sob contextos bastantes diversos. Há a necessidade de se saber melhor como os pacientes atendidos na cidade de Rio Grande estão evoluindo, considerando as condições sócio ambientais presentes aqui. Este trabalho tem este objetivo: analisados os dados dos pacientes da cidade de Rio Grande, mas especificamente os pacientes atendidos no HU-FURG. Nesta pesquisa, há a implementação e utilização de ferramentas que possibilitem reunir informações sobre a passagem desses pacientes no HU-FURG, permitindo que sejam propostos mapas e relatórios com predições de impactos possíveis de ocorrer nesses pacientes, a nível regional, será de grande relevância no entendimento da pandemia. E a proposição de tratamentos precoces trarão qualidade de vida aos pacientes.

1.2 Objetivo geral

Desenvolver um processo baseado em aprendizado de máquina, capaz de organizar os dados disponíveis no HU-FURG e fazer predições, permitindo aos profissionais de saúde, através das ferramentas que o sistema contém e outras que ainda serão desenvolvidas, de acordo com a necessidade (gráficos, relatórios estatísticos), identificar o perfil clínico e epidemiológico dos pacientes com síndrome pós-covid. Isso a fim de prever, utilizando as características dos pacientes, quais problemas terão maior probabilidade de ocorrer, permitindo um acompanhamento adequado de cada paciente.

1.3 Objetivos específicos

Como objetivos específicos são apresentados os seguintes tópicos:

- Desenvolver sistema de aquisição e tratamento de dados de pós-covid.
- Comparar de maneira sistemática algoritmos para a predição de evolução do pós-covid.
- Disponibilizar algoritmo para predição de evolução do pós-covid.

1.4 Aspectos Éticos

1.4.1 Processo e método de consentimento e/ou assentimento (forma de acesso ao paciente/ informações e documento de registro da participação do público na pesquisa

Conforme a LEI Nº 13.709, DE 14 DE AGOSTO DE 2018, art 7º (o tratamento de dados pessoais somente poderá ser realizado nas seguintes hipóteses) , item IV, para a realização de estudos por órgão de pesquisa, garantida, sempre que possível, a anonimização dos dados pessoais. (TEMER, JARDIM, *et al.*, 2018).

Contempla completamente este projeto.

1.4.2 Riscos e Benefícios

Não há risco direto para o paciente. O resultado será de predição, portanto será explicado ao paciente a respeito, sem prejuízo do mesmo.

Como benefício, o conhecimento de uma possibilidade permite manejo precoce e melhora da qualidade de vida do paciente.

1.4.3 Privacidade e Confidencialidade dos Dados

Todos os dados oriundos dos sistemas fontes de pesquisa serão anonimizados; dados como nome, CPF, prontuário, que poderiam identificar o paciente serão suprimidos; estes não são dados interessantes ou necessários para o desenvolvimento e aplicação do projeto.

1.5 Contribuições da pesquisa para HU-Furg/Ebserh

A presente pesquisa abre a possibilidade de outras pesquisas semelhantes, para a predição de outras doenças (diabetes, hipertensão), partindo da ideia central estabelecida.

2 REVISÃO DA LITERATURA – COVID-19

Neste capítulo serão apresentados estudos e pesquisas que estão sendo conduzidos sobre infecção e disseminação sobre o COVID-19, bem como os métodos não farmacológicos adotados para evitar a sua disseminação. Na falta de estudos mais locais, serão apresentados estudos mais gerais, a nível de Brasil e mundo.

Este trabalho foi realizado mediante revisão de literatura por meio de busca de artigos em bancos eletrônicos, tais quais PubMed, SciELO, Google Acadêmico e em endereços eletrônicos governamentais.

2.1 Histórico da COVID-19

2.1.1 Histórico da pandemia de COVID-19

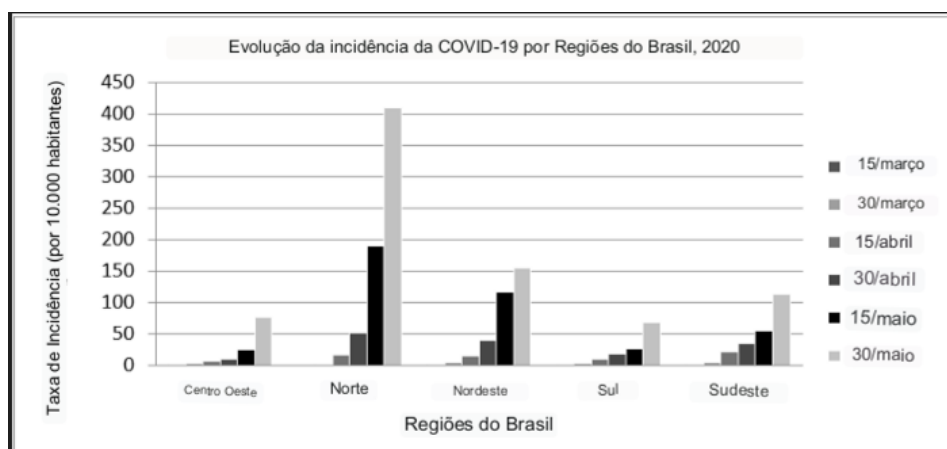
Em 31 de dezembro de 2019, a cidade de Wuhan (na China) teve vários casos de pneumonia (CRODA e GARCIA, 2020). A ligação com um novo coronavírus é estabelecida posteriormente.

Em janeiro de 2020, o CSG (Coronaviridae Study Group) reconhece este vírus como sendo da mesma família do vírus da síndrome respiratória aguda grave humana e de morcego (SARS-CoVs) da espécie Coronavírus relacionado à síndrome respiratória aguda grave, e o designa como SARS-CoV-2 (GORBALENYA, BAKER, *et al.*, 2020).

Já foram identificados sete coronavírus humanos (HCoV): HCoV-229E, HCoV-OC43, HCoV-NL63, HCoV-HKU1, SARS-COV (que causa síndrome respiratória aguda grave), MERS-COV e o, mais recente, que foi temporariamente nomeado 2019-nCoV e, em 11 de fevereiro de 2020, recebeu o nome de SARS-CoV-2). Esse último é responsável por causar a doença COVID-19 (OPAS - ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE)

Em março de 2020 a OMS declara a pandemia de COVID-19 (WORLD HEALTH ORGANIZATION, 2020). Na Figura 1, representação do vírus causador da COVID-19.

Figura 3 - Evolução do COVID-19 no Brasil



Fonte: (GOUVEIA, GOUVEIA, *et al.*, 2023) - adaptado

Em 2021 a variante Delta torna-se predominante em vários países. Alguns países que já estavam relaxando suas medidas protetivas viram o número de casos disparar novamente (PORTAL DO BUTANTÃ, 2021). Em 26 de novembro de 2021, Na África do Sul, foi detectada a variante Ômicron (PORTAL DO BUTANTAN, 2022); mais transmissível, porém menos letal.

Em 2022, novas vacinas e tratamentos são desenvolvidas para enfrentar as variantes do vírus. E, embora existam versões de vacinas com cobertura para as variantes, as vacinas baseadas na cepa original ainda trazem forte proteção contra a doença grave e morte. Ainda devem ser usadas para inicial o ciclo vacinal. (OPAS - ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE, 2022)

“Brasília, 5 de maio de 2023 – A Organização Mundial da Saúde (OMS) declarou nesta sexta-feira (5/05), em Genebra, na Suíça, o fim da Emergência de Saúde Pública de Importância Internacional (ESPII) referente à COVID-19.” (OPAS - ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE, 2023)

O vírus pode não desaparecer, então o grande desafio é controlar sua disseminação (VICTOR, 2023). A cada dia novos estudos são feitos sobre os mais variados aspectos da doença, seja prevenção seja em tratamento, visando a melhoria contínua da qualidade de vida da população, como o Impacto da COVID-19 grave na qualidade de vida (FONTES, COSTA, *et al.*, 2022) e Qualidade de Vida dos Profissionais de Saúde Pós-COVID-19 (PIRES, BOSCO, *et al.*, 2021) .

2.1.2 Emergência de Saúde Pública de Importância Internacional

Em 30 de janeiro de 2020, a Organização Mundial da Saúde declarou, em Genebra, na Suíça, que o surto do novo coronavírus (2019-nCoV) constitui uma Emergência de Saúde Pública de Importância Internacional (ESPII), com casos em 19 países, com transmissão entre humanos na China, Alemanha, Japão, Vietnã e Estados Unidos da América. (OPAS - ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE, 2020).

É a sexta vez na história que uma Emergência de Saúde Pública de Importância Internacional é declarada. As outras foram:

- **fevereiro de 2009:** pandemia de H1N1
- **2014:** surto de Ebola na África Ocidental
- **2016:** a febre do Chikungunya e vírus Zika e suas consequências graves
- **2018:** surto de ebola na República Democrática do Congo
- **2018:** epidemia de Sarampo, espalhando-se por vários países das Américas. (CARMO, 2020)

A responsabilidade de se determinar se um evento constitui uma Emergência de Saúde Pública de Importância Internacional cabe ao diretor-geral da OMS e requer a convocação de um comitê de especialistas – chamado de Comitê de Emergências do RSI (OPAS - ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE)

2.2 Folha informativa sobre COVID-19

Os sintomas principais da COVID-19 são febre, tosse, dispneia (falta de ar), mialgia (dor muscular), confusão mental e cefaleia (dor de cabeça). Há outros sintomas não tão comuns que são: dor de garganta, rinorreia (coriza), dor torácica, diarreia, náuseas e vômitos (LIMA, 2020).

Surge a variante Ômicron. Em 26 de novembro de 2021, a OMS designou a variante da COVID-19 B.1.1.529 como uma variante de preocupação. A variante se espalha rapidamente pelo mundo, sendo que em janeiro (2022) já era a cepa predominante no planeta (PORTAL DO BUTANTAN, 2022). As outras variantes de preocupação ainda estão em circulação e são: Alfa, Beta, Gama e Delta.

Os procedimentos para prevenção continuam os mesmos: manter ambientes ventilados, evitando contaminar outras pessoas e, por esse meio permitir o surgimento de mais variantes, higiene das mãos, distanciamento social e uso de máscaras.

2.3 Apresentação da Doença

Os casos da variante ômicron se propagaram rapidamente e em escala global. Os sintomas mais comuns são febre, coriza, dor de garganta e dor no corpo, nada semelhantes à perda de paladar, de olfato e tosse seca comuns às outras variantes (PORTAL DO BUTANTAN, 2022).

Os sintomas são muito parecidos com os sintomas da gripe, fazendo com que muitos profissionais não solicitem o teste de COVID-19. Entretanto, testar é fundamental pois, com a doença identificada, os protocolos são aplicados, evitando a sua disseminação e o surgimento de novas variantes. Apesar de ser mais branda, a variante ômicron ainda apresenta riscos à saúde e à vida.

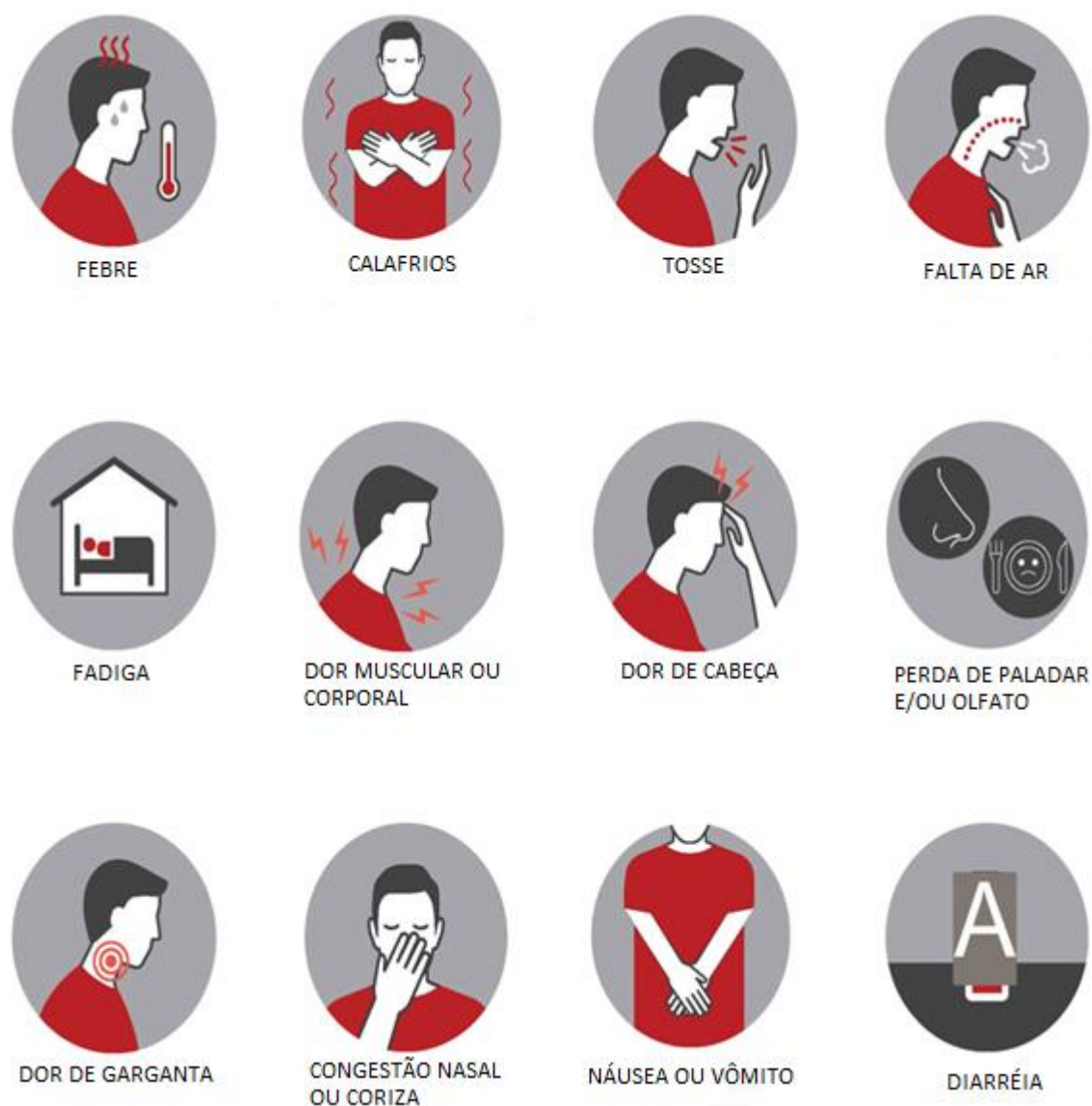
Na Figura 4 estão alguns dos sintomas mais comuns da COVID-19 E na Tabela 1 estão listadas as variantes do vírus e os seus principais sintomas.

Tabela 1 – Variantes da COVID-19 e seus sintomas

Variante	Sintomas mais comuns
Omicron	cansaço extremo, dores pelo corpo, dor de cabeça e dor de garganta
Delta	coriza, dor de cabeça, espirros, dor de garganta, tosse persistente e febre
Gama	febre, tosse, dor de garganta, falta de ar, diarreia, vômito, dor no corpo, cansaço e fadiga
Alfa	perda ou alteração do olfato, perda ou alteração do paladar, febre, tosse persistente, calafrios, perda de apetite e dores musculares
Beta	febre, tosse, dor de garganta, falta de ar, diarreia, vômito, dor no corpo, cansaço e fadiga

Fonte: (INSTITUTO BUTANTÃ, 2021)

Figura 4 - Sintomas Comuns

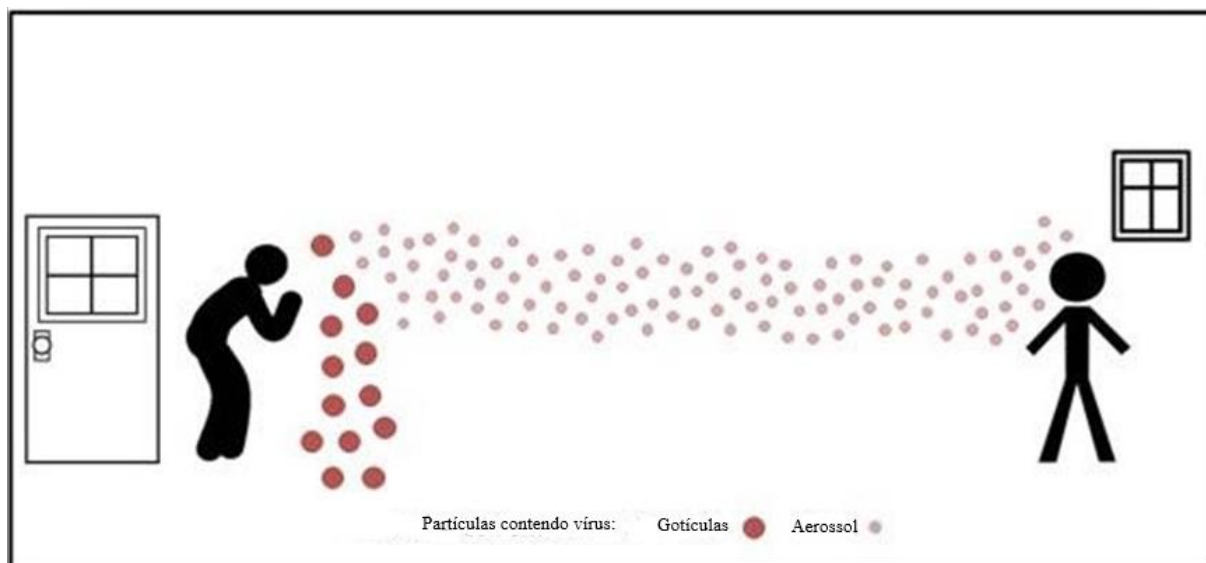


Fonte: (THE OHIO STATE UNIVERSITY - WEXNER MEDICAL CENTER, 2024) - adaptado

2.4 Transmissão: Gotícula X Aerossol

Além das gotículas maiores, o vírus pode ser transmitido por gotículas e aerossóis (gotas ainda menores), principalmente em ambientes fechados, que permite a concentração do vírus. (MORAWSKA e CAO, 2020). A Figura 5 demonstra esse tipo de transmissão.

Figura 5 - Transmissão por Aerossol



Fonte: (HOWARD, 2020) - adaptado

Através de experimentos com imagens para visualizar exalações respiratórias, gotículas respiratórias podem ser aerossolizadas ou transportadas em uma nuvem de gás e ter trajetórias que vão além de dois metros a partir da fala (pessoa que fala mais alto transmite por uma maior distância, tosse ou espirro. (ASADI, BOUVIER, *et al.*, 2020)

Dessa forma, é possível destacar que a transmissão do novo coronavírus costuma ocorrer pelo ar ou por contato pessoal próximo e/ou objetos que contenham essas secreções contaminadas, como:

- Gotículas de saliva;
- Espirro;
- Tosse;
- Contato pessoal próximo, como toque ou aperto de mão;
- Contato com objetos ou superfícies contaminadas, seguido de contato com a boca, nariz ou olhos;
- Compartilhamento de copos, bebedouro com jato direto na boca, rodas coletivas para consumo do chimarrão (erva-mate regada com água quente) ou do tereré (erva-mate regada com água gelada), rodas coletivas para consumo de narguilé, entre outras ações que requeiram contato direto com gotículas de saliva, que podem estar contaminadas. (FIOCRUZ CAMPUS VIRTUAL, 2020)

A Tabela 2 exibe o tempo de permanência do vírus ativo em algumas superfícies,

Tabela 2 - Tempo do Coronavírus Ativo em Superfícies

Superfície	Tempo
Aço Inoxidável	72 horas
Plástico	72 horas
Papelão	24 horas
Cobre	4 horas
Aerossol / Poeira	40 minutos a 2 horas e meia

Fonte. (DOREMALEN, BUSHMAKER, *et al.*, 2020)

2.5 Covid Longa

Conhecida também como síndrome pós-covid A OMS a reconhece a covid longa oficialmente em agosto de 2020, definindo-a como condição pós-COVID-19 em pessoas com infecção por SARS-CoV-2 provável ou confirmada, habitualmente após três meses, com sinais/sintomas que duram pelo menos dois meses e que não são explicados por outro diagnóstico (RAMOS JR., 2024).

Com contornos ainda vagos, ela está impulsionando pesquisas sobre a sua prevalência, sintomas, diagnóstico, duração e métodos de tratamento (SEGATA e LÖWY, 2024).

Há prevalência de covid longa em pessoas do sexo feminino, maior faixa etária, com uma relevância de dados para persistência dos sintomas, com muitos ou sintomas específicos no início do quadro como fator de risco, cuja manifestações clínicas mais frequentes foram fadiga, dispneia e quadro neuropsicológico, com progressão variável e impacto significativo na saúde e vida (NUNES, ALVES, *et al.*, 2022).

A covid longa é uma pandemia à espera de diagnóstico. E, além disso, não há tratamento. (SEGATA e LÖWY, 2024). É ofertada uma combinação de tratamentos, sendo o farmacológico direcionado ao alívio dos sintomas e os tratamentos não farmacológicos aqueles que abrangem a reabilitação para melhora das funções comprometidas, apoio psicoterapêutico e social (CURVELO, RIBEIRO e UEHARA, 2024).

3 PESQUISAS RELACIONADAS

Há outras pesquisas ao redor do mundo, tratando do mesmo assunto. Nos Estados Unidos, usando o repositório de prontuários eletrônicos de saúde do National COVID Cohort Collaborative (N3C), foram desenvolvidos modelos de aprendizado de máquina XGBoost para identificar pacientes em potencial com COVID longa (PFAFF, GIRVIN, *et al.*, 2022).

Na Inglaterra, foram usados oito conjuntos de dados vinculados do National Health Service (NHS), para pessoas vivas em 23 de janeiro de 2020. Dados sobre testes de COVID-19, vacinação, registros de cuidados primários e secundários e registros de óbitos foram coletados até 30 de novembro de 2021, visando entender o impacto populacional da covid longa. (THYGESEN, TOMLINSON, *et al.*, 2022).

Na Alemanha, foram usados dados que compreendem históricos médicos de pacientes de atenção primária, para prever condições pós-COVID-19 em pacientes após o diagnóstico de COVID-19 e para avaliar os fatores relevantes associados a essas condições usando métodos de aprendizado de máquina (KESSLER, PHILIPP, *et al.*, 2023).

Em Punta Arenas, Chile, foi conduzido utilizando modelagem matemática e aprendizado de máquina como ferramentas potentes para orientar e otimizar as diferentes medidas de controle da COVID-19. E também utiliza aprendizado de máquina para investigar a COVID longa, identificando padrões e relações de sintomas, prevendo indicadores de risco e permitindo a avaliação precoce de covid longa (VARÓN, GONZÁLEZ-PUELMA, *et al.*, 2023).

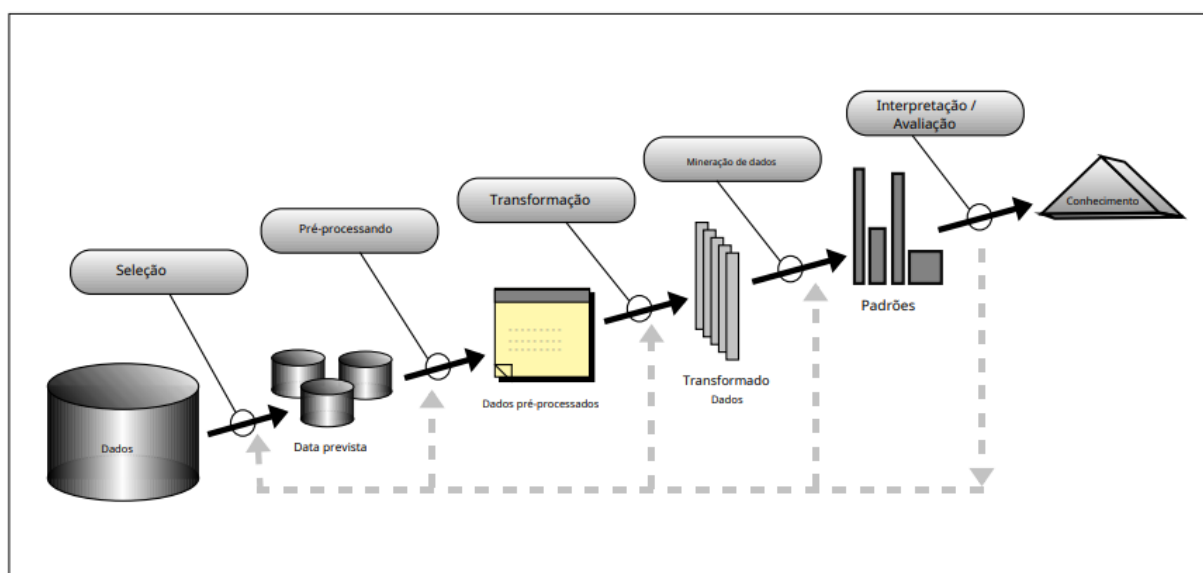
Na Universidade de Fortaleza, uma equipe de pesquisadores desenvolveu pesquisa semelhante à abordada neste trabalho, ao se utilizar de inteligência artificial (IA) para prever sequelas da COVID-19. Mas, neste caso, é um modelo híbrido. Os resultados são obtidos por meio de IA são complementares aos resultados obtidos de forma tradicional (diagnóstico médico) (ANDRADE, PINHEIRO, *et al.*, 2023).

Pode ser estabelecida relação entre os estudos citados e o estudo conduzido neste trabalho, pois todos analisam dados de pacientes que tiveram de COVID-19 e sintomas classificados como covid longa. Os dados analisados são em sua maioria comuns entre os estudos, como dados de vacinação, sintomas relatados e outros. Contudo, o estudo deste trabalho se utiliza de dados da região abrangida pelo HU-FURG, o que torna seus resultados mais precisos regionalmente.

4 REVISÃO DA LITERATURA – ALGORITMOS

A Descoberta de Conhecimentos em Bases de Dados (do inglês *Knowledge Discovery in Databases KDD*) é um conjunto de técnicas e ferramentas utilizadas para transformar os dados contidos em bases de dados regulares de estabelecimentos (neste caso, um hospital) em conhecimento. Para essa transformação são necessárias cinco etapas: seleção dos dados; pré-processamento e limpeza dos dados; transformação dos dados; mineração de dados (*Data Mining*); e interpretação e avaliação dos resultados, (STEINER, SOMA, *et al.*, 2006) e que serão utilizadas neste trabalho. A Figura 6 exibe graficamente as cinco etapas.

Figura 6 - As cinco etapas da Descoberta de Conhecimento



Fonte: Da mineração de dados à descoberta de conhecimento em Bancos de dados (do inglês) *Data Mining to Knowledge Discovery in Databases* (FAYYAD, PIATETSKY-SHAPIO e SMYTH, 1996) – Traduzido para o português

4.1 Seleção dos Dados

4.1.1 Preparação dos Dados

A preparação dos dados é uma etapa muito importante para se obter bons resultados com o modelo preditivo. (OLIVEIRA, 2020)

Também conhecida como “Redução de Dados”, esta é a primeira etapa no processo de descoberta de informação. Nesta etapa é definido o conjunto de dados a ser utilizados e, por isso, possui papel fundamental no resultado final, pois contém todos os atributos e instâncias possíveis que se pretende analisar. Os dados podem ser selecionados das mais diversas fontes,

tais como: banco de dados relacional, arquivo texto legado, Data Warehouses (Armazéns de Dados), planilhas, dentre outros, o que torna este processo bastante complexo (ROSA, 2017).

4.1.2 Pré-processamento e Limpeza dos Dados

A limpeza de dados é um processo que visa identificar, corrigir e remover inconsistências, erros e ruídos presentes em conjuntos de dados. Essa técnica envolve a aplicação de etapas como detecção de valores ausentes, tratamento de dados inconsistentes, remoção de outliers e normalização de formatos (IWATA e ITO, 2023).

Também é verificado se é possível diminuir atributos e instâncias, visando melhorar o desempenho dos algoritmos de análise (GOLDSCHMIDT e PASSOS, 2005). Dados irrelevantes para a pesquisa, como atributos com somente um valor ou vazios, são descartados.

4.1.3 Transformação dos Dados

Nesta etapa os conjuntos brutos de dados são convertidos em uma forma padrão de uso. (GOLDSCHMIDT e PASSOS, 2005)

É implementada através de processamento dos dados, visando organizá-los para auxiliar o trabalho sucedido pelas fases posteriores do processo KDD. Não existe um critério único de transformação dos dados e diversas técnicas podem ser usadas de acordo com os objetivos pretendidos (ROSA, 2017). Como técnicas há a normalização de dados, conversões de valores simbólicos para valores numéricos, discretização e composição de atributos (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996).

Neste trabalho foi utilizada a discretização no campo Idade, que foi convertido para faixas, demonstradas a seguir:

- 0 a 12: criança;
- 13 a 17: adolescente;
- 18 a 25: jovem;
- 26 a 40: adulto;
- 41 a 60: meia_idade;
- mais de 61: terceira_idade.

4.2 Mineração dos Dados

Não se deve confundir DM (mineração de dados – do inglês *data mining*) e KDD (descoberta de conhecimentos em bases de dados (do inglês *Knowledge Discovery in Databases*), considerando que DM é uma parte do processo de KDD, que é todo o processo até que se chegue ao resultado de um padrão de comportamento das variáveis. (TAN, STEINBACH e KUMAR, 2009)

A mineração de dados é o processo de explorar dados à procura de padrões consistentes, como regras de associação, correlações, agrupamentos ou sequências temporais, para detectar relacionamentos sistemáticos entre variáveis, detectando assim novos subconjuntos de dados (O'BRIEN, 2005). É uma técnica que faz parte de uma das etapas da descoberta de conhecimento em banco de dados. Ela é capaz de revelar, automaticamente, o conhecimento que está implícito em grandes quantidades de informações armazenadas nos bancos de dados de uma organização (CARDOSO e MACHADO, 2008).

“Durante a etapa de Mineração de Dados, é realizada a busca efetiva por conhecimentos úteis no contexto da aplicação de KDD.” (GOLDSCHMIDT e PASSOS, 2005)

A característica principal da mineração de dados é o algoritmo, que será capaz de extrair de maneira competente, o conhecimento implícito e útil de uma base de dados, podendo auxiliar na predição de um conhecimento futuro. (ROSA, 2017)

4.3 Aprendizado de Máquina

O aprendizado de máquina é um campo da inteligência artificial, que surgiu dentro da ciência da computação (ROZA, 2016), e se baseia na ideia de que sistemas computacionais possam aprender a partir de dados e padrões, retornando o que seria uma decisão, sem ou com pouca interferência humana. Como exemplo há os assistentes de voz (como a Alexia, da Amazon ou o Google Assistente), que reconhecem a fala humana (PEREIRA e CÂMARA, 2021).

É uma área de IA cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática. Um sistema de aprendizado é um programa de computador que toma decisões baseado em experiências acumuladas através da solução bem-sucedida de problemas anteriores (MONARD e BARANAUSKAS, 2003).

Existem três tipos principais de Aprendizado de Máquina: Supervisionado, Não Supervisionado e Aprendizado por Reforço (LUDERMIR, 2021).

4.3.1 Aprendizado de Máquina Supervisionado

O aprendizado supervisionado se caracteriza pelo fato de que, para cada exemplo apresentado ao algoritmo de aprendizado é necessário apresentar a resposta desejada (recebe um conjunto de entradas e suas correspondentes saídas) (LUDERMIR, 2021). Esta é a forma de seu treinamento. Possui duas classes de algoritmos: classificação e regressão.

A classificação tem por objetivo prever uma classe ou rótulo associado com uma variável de entrada contendo determinados atributos. O algoritmo é treinado com um conjunto de dados com classes conhecidas, podendo ser somente duas classes ou várias classes (FONTANA, 2020). É o processo que localiza um determinado grupo de funções que descrevem e diferenciam classes ou conceitos, com o intuito de utilizar o modelo localizado para prever a classe de objetos que ainda não foi classificado (MAIA, GOMES e CHAGAS, 2017).

A regressão faz também previsões, mas tem como resultado valores contínuos. Um exemplo de um problema de regressão seria a previsão do rendimento em um processo de fabricação química no qual as entradas consistem nas concentrações de reagentes, temperatura e pressão (BISHOP, 2006). A saída (rendimento) poderia ser qualquer valor numérico.

Os algoritmos de classificação possuem rótulos de classe discretos, ou seja, seu objetivo é prever uma classe ou rótulo (FONTANA, 2020), e por esse motivo são utilizados neste estudo. Os escolhidos foram Árvore de Decisão, Naive Bayes e Random Forest.

As métricas obtidas que serão avaliadas serão Acurácia, Precisão Recall e F1.

4.4 Algoritmos de Classificação Utilizados

4.4.1 Árvore de Decisão

O funcionamento é baseado no conceito dividir para conquistar, de forma que os dados são divididos em subconjuntos, de acordo com suas características (GARCIA, 2003), ou seja, divide um conjunto de dados em subconjuntos menores com base em critérios de decisão, criando uma estrutura de árvore para representar as decisões a serem tomadas. Esse processo é repetido para cada subconjunto de forma recursiva, chamada de particionamento recursivo. O processo recursivo é encerrado quando a divisão não agrega mais valor às previsões

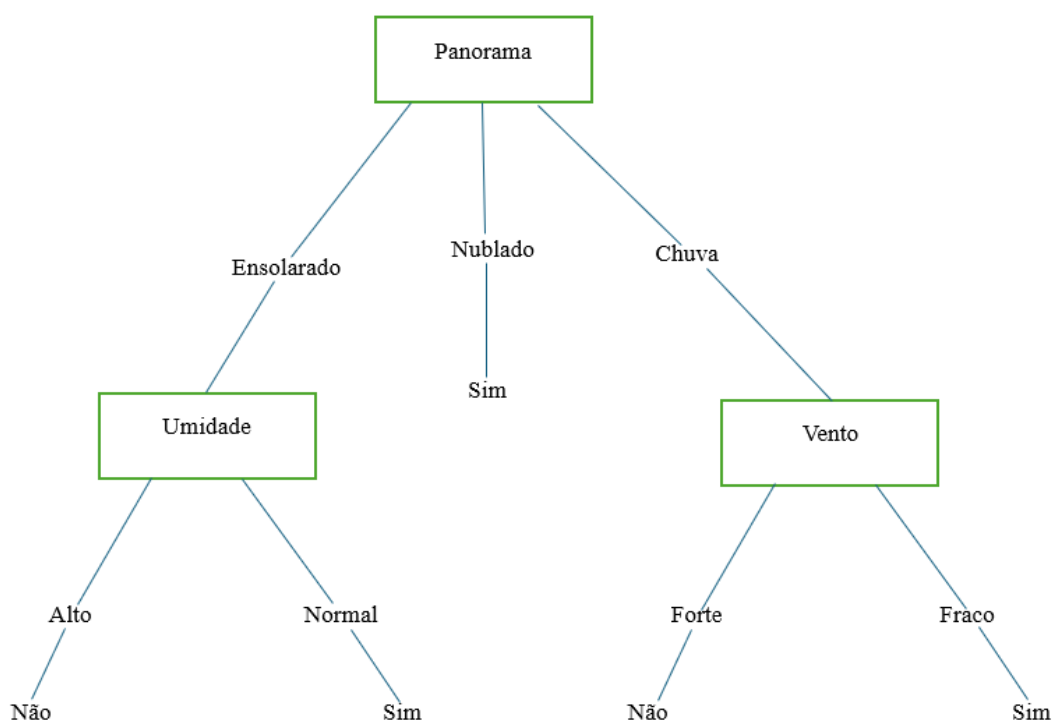
(QUINLAN, 1986). Este é o funcionamento do algoritmo guloso, que é a estratégia mais comum para aprender árvores de decisão (ROKACH e MAIMON, 2005).

A classificação ocorre quando os caminhos descritos pelos nós (os subconjuntos de dados são chamados de nodos (nós) são percorridos, até se chegar ao nó final do caminho percorrido. Este nó é chamado de folha; o nó inicial é chamado de raiz.

Os resultados são apresentados hierarquicamente, ou seja, com prioridades. E podem ser usadas em conjunto com a tecnologia de indução de regras. O primeiro nó é o atributo mais importante; os demais atributos são exibidos nos nós subsequentes (CREPALDI, AVILA, *et al.*, 2007). A Figura 7 exhibe exemplo de Árvore de Decisão de altura 5, onde o elemento Panorama representa o nodo raiz, enquanto os elementos Sim e Não representam as folhas. Visa identificar se o fim de semana seguinte seria propício para viajar para uma cidade com praia. O Sim indica que a viagem que o destino seria agradável no fim de semana; o Não indica que seria melhor evitar viajar!

Figura 7 - Exemplo de Árvore de Decisão

Visa identificar se o fim de semana seguinte seria propício para viajar para uma cidade com praia.



Fonte: Adaptado de Machine Learning (MITCHELL, 1997)

É necessário calcular a entropia dos dados, para identificar quais dados tem maior ganho de informação. O dado com maior ganho de informação será o nó raiz; os nós subsequentes serão obtidos a partir do nó-raiz, através do decréscimo do ganho de informação

até se chegar ao nó-folha, que será correspondente às classes, considerando-se algoritmo de classificação (QUINLAN, 2021).

Há dois tipos de árvore de decisão: árvore de classificação e árvore de regressão. A árvore de classificação é quando o resultado previsto é uma classe discreta, enquanto que uma árvore de regressão é quando o resultado previsto é um modelo de regressão, como o tempo de um paciente em um hospital (BREIMAN, FRIEDMAN, *et al.*, 1984).

Existem muitos de algoritmos de árvore de decisão; alguns exemplos são ID3, C4.5 (sucessor do ID3), CART (árvore de classificação e regressão), CHAID e MARS. Cada um se utiliza de algumas métricas para seu processamento, como:

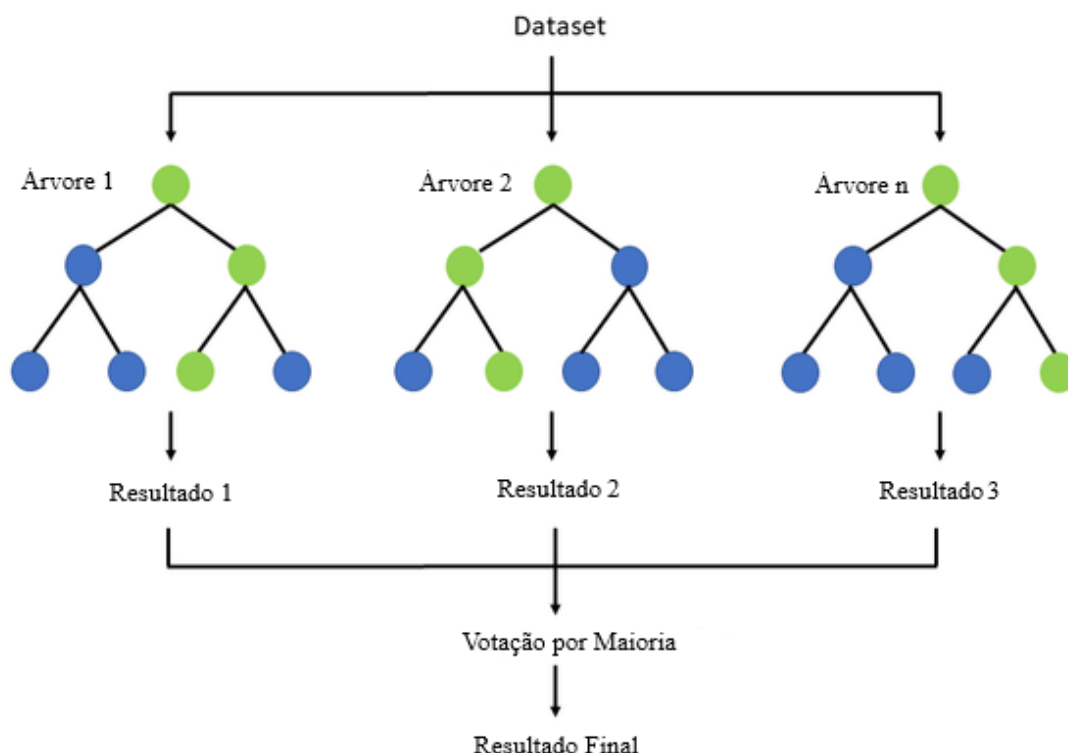
- **Estimativa de Correção Positiva** – Identifica até que ponto os verdadeiros positivos superam os falsos positivos.
- **Impureza de Gini** – utilizado pelo algoritmo CART para árvores de classificação (BREIMAN, FRIEDMAN, *et al.*, 1984). O algoritmo procura dividir os dados em grupos cada vez mais homogêneos em termos de rótulos de classe.
- **Redução de Variação** – também utilizado pelo algoritmo CART, mas para árvores de regressão. É empregada nos casos em que a variável alvo é contínua. Para o uso de outras métricas, a variável alvo precisaria primeiro ser discretizada (BREIMAN, FRIEDMAN, *et al.*, 1984).
- **Medida de Bondade** – busca o equilíbrio entre a capacidade de uma divisão candidata de criar filhos puros com sua capacidade de criar filhos de tamanhos iguais. O processo é repetido para cada nó impuro até que a árvore esteja completa (LAROSE e LAROSE, 2014).
- **Ganho de Informação** - Para calcular o ganho de informação, é necessário analisar um atributo. Se os registros da base para cada lado são homogêneos ou próximo disso, há então um alto ganho de informação. Usado pelos algoritmos de geração ID3, C4.5 e C5.0 (WITTEN, FRANK e HALL, 2011).

4.4.2 Random Forest

Random Forest é um método de aprendizado de máquina que se baseia em uma coleção de árvores de decisão. Cada árvore de decisão no Random Forest considera um subconjunto aleatório ao formar perguntas e só tem acesso a um conjunto aleatório de pontos de dados de treinamento (KOEHRSEN, 2017).

Ao final, combina os resultados das árvores para a obtenção do resultado final (LIMA e AMORIM, 2020). É bastante utilizado para problemas de classificação e regressão. A Figura 8 apresenta um exemplo simples de Random Forest.

Figura 8 - Exemplo de Random Forest



Fonte: Adaptado de Random Forest Simple Explanation (KOEHRSEN, 2017)

Algumas propriedades do Random Forest são:

- **importância variável** – pode ser usada para classificar a importância das variáveis em um problema de regressão ou classificação (BREIMAN, 2001).
- **importância da permutação** – primeiro passo é ajustar a random forest aos dados, para determinar a importância da variável no conjunto de dados. Durante o processo, os *out-of-bag error* são registrados e é calculada a média do random forest.

O Random Forest faz parte dos métodos Ensemble. Estes métodos são construídos da mesma forma que algoritmos mais básicos, mas possuem uma característica principal que é a combinação de diferentes modelos para a obtenção de um único resultado. A ideia é treinar vários preditores e então agregar as várias previsões em apenas uma previsão. Quanto

maiores as diferenças entre os preditores, melhor. Ou seja, preditores independentes e com erros não correlacionados são o melhor cenário (POHLENZ e JÚNIOR, 2020).

O Random Forest é amplamente adotado pois:

- tem suas origens na forma mais básica e inicial de um algoritmo de suporte a decisão (as árvores de decisão), fica muito fácil entender seu funcionamento;
- Random Forests resolvem problemas tanto de regressão quanto de classificação;
- apresenta bons resultados em diversos tipos de problema;
- tem bom desempenho;
- possui implementações nas mais diversas bibliotecas de inteligência artificial disponíveis atualmente (PESSANHA, 2019).

4.4.3 Naive Bayes

O algoritmo Naive Bayes refere-se à construção de um modelo probabilístico Bayesiano que atribui uma probabilidade de classe posterior a uma instância. O classificador Bayes simples e ingênuo usa essas probabilidades para atribuir uma instância a uma classe (BERRAR, 2018). Ingênuo, neste caso, quer dizer que os atributos são independentes entre si. É assumido também que todos os atributos são igualmente importantes. Em casos que não seja dessa forma, o algoritmo deixa de ser interessante.

Não é um algoritmo único; mas uma coleção de algoritmos de classificação baseados no Teorema de Bayes. O Teorema de Bayes é representado pela equação:

$$P(A|B) = P(B|A) \times P(A) / P(B)$$

Onde:

$P(B|A)$ significa a probabilidade de B acontecer já que o evento A se confirmou

$P(A)$ é a probabilidade de A acontecer

$P(B)$ é a probabilidade de B acontecer (GUSMÃO, 2022)

O treinamento de máxima verossimilhança pode ser feito avaliando uma expressão de forma fechada (RUSSEL e NORVIG, 2003). Isso também quer dizer que se pode trabalhar com Naive Bayes sem aceitar a probabilidade bayesiana ou usar quaisquer métodos bayesianos.

Uma vantagem do Naive Bayes é que ele requer apenas uma pequena quantidade de dados de treinamento para estimar os parâmetros necessários para a classificação (FAN, 2018),

4.4.4 Xgboost

Caracteriza-se por ter alta eficiência e desempenho, tanto se utilizado em regressão quanto em classificação. Baseia-se no conceito de gradient boosting, que combina modelos de aprendizado fracos em um único modelo de aprendizagem forte, de maneira iterativa (CHEN e GUESTRIN, 2016). A ideia se originou na observação de que o *boosting* pode ser interpretado como um algoritmo de otimização em uma função de custo adequada (BREIMAN, 1997). A

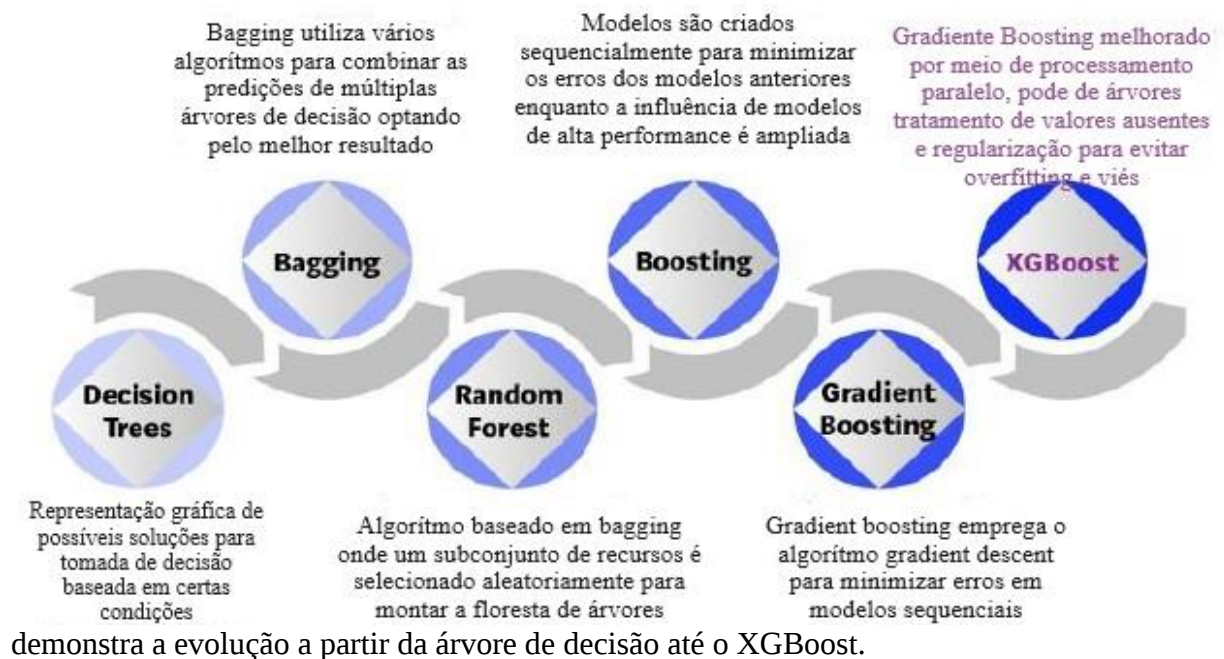
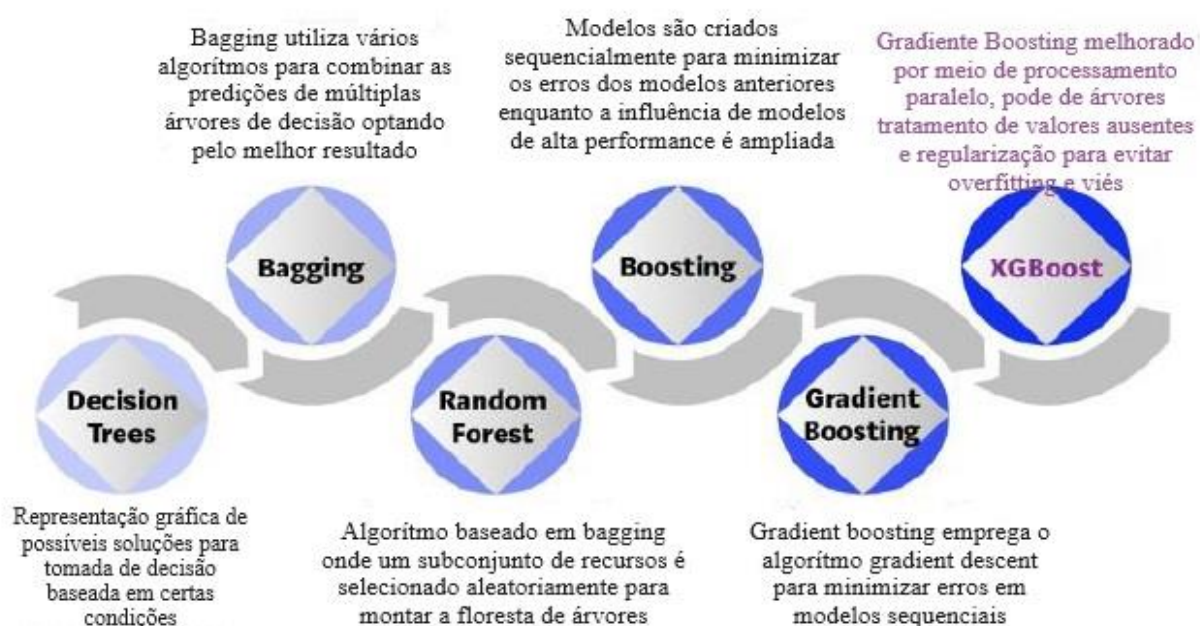


Figura 9 - Evolução do XGBoost a partir da Árvore de Decisão



Fonte: (KRISHNAN, NEYAZ e QINGZHONG, 2021) – adaptado

Entre os recursos importantes XGBoost que o tornam diferente de outros algoritmos de gradient boosting estão os seguintes:

- **Eficiência e Velocidade:** Suporta paralelismo em múltiplas CPUs e GPUs, além de poder ser distribuído em clusters; o processamento paralelo é utilizado na construção de árvores. Desta forma, o treinamento e a previsão são significativamente acelerados, especialmente em grandes conjuntos de dados (CHEN e GUESTRIN, 2016).
- **Regularização:** Inclui termos de regularização L1 (lasso) e L2 (ridge) para prevenir overfitting, o que permite ajustes em modelos mais robustos e pode resultar em melhor desempenho do que outras implementações de boosting (CHEN e GUESTRIN, 2016).
- **Manejo de Dados Faltantes:** XGBoost pode lidar automaticamente com valores ausentes nos dados de entrada. Ao invés de descartá-los, o XGBoost os trata de maneira inteligente, identificando o melhor caminho para cada ponto de dados (CHEN e GUESTRIN, 2016).
- **Flexibilidade:** Oferece uma ampla gama de funções de custo e métricas de avaliação, permitindo um ajuste fino dos modelos (FRIEDMAN, 2001).

Há biblioteca que funciona em Windows, Linux e MacOs e possui implementação em várias linguagens, incluindo o R.

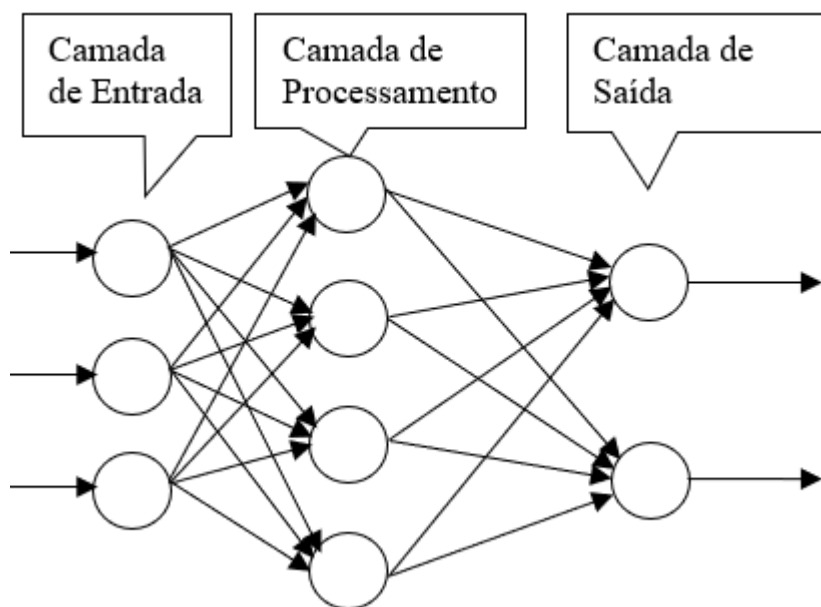
4.4.5 Rede Neural

É um campo da inteligência artificial, que busca implementar modelos matemáticos que se assemelhem às estruturas neurais biológicas (FERNEDA, 2006). Busca na estrutura cerebral inspiração para sua concepção. O processamento da informação em ANN é feito nos neurônios artificiais (LUDERMIR, 2021).

Há alguns tipos de Redes Neurais, que são:

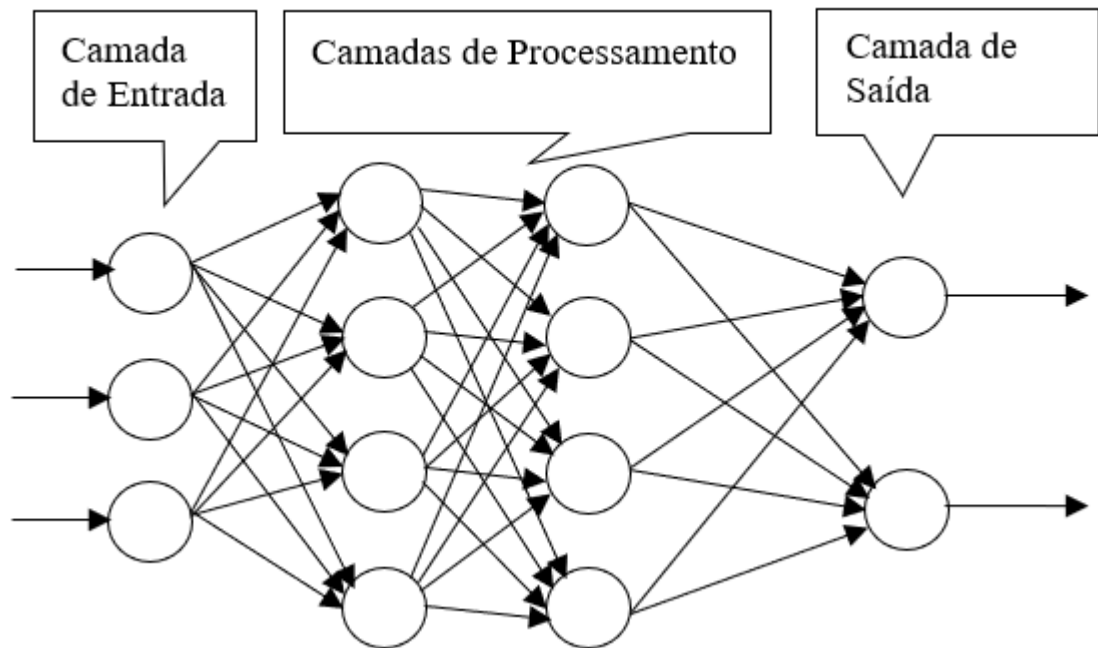
- **Redes Neurais Artificiais (ANN):** O tipo mais básico de rede neural, usado principalmente para classificação e regressão. O modelo mais simples de rede neural é o Perceptron, definido em 1957 por Frank Rosembat. É uma estrutura de camada única, contendo neurônios e regra de aprendizado (LUDERMIR, 2021). Com o modelo multicamadas, se caracteriza o deep learning. A Figura 10 exibe um modelo de rede neural simples, e a Figura 11 exibe modelo de rede neural com duas camadas, já caracterizando *deep learning* (aprendizado profundo, quando a rede possui mais de uma camada oculta (FALCÃO, MOREIRA, *et al.*, 2014)) .

Figura 10 - Modelo Simples de Rede Neural



Fonte: o próprio autor

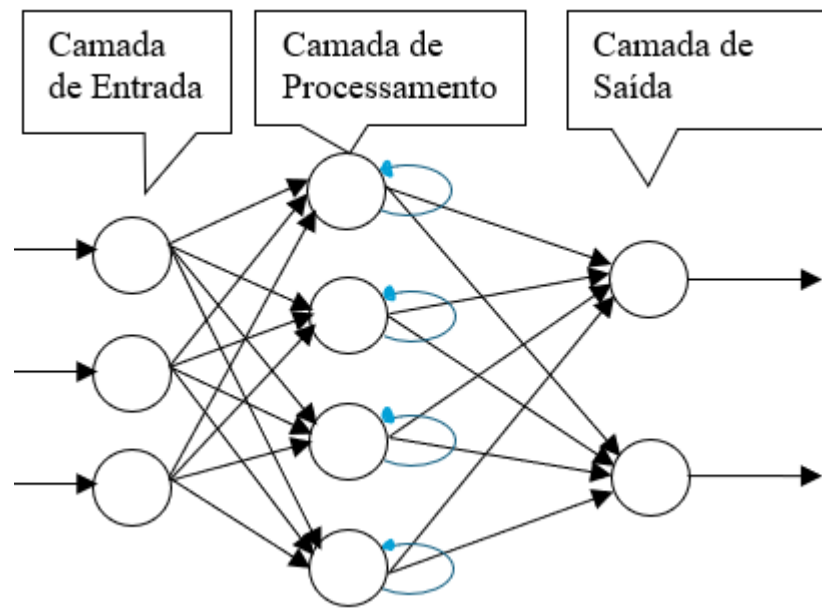
Figura 11 - Modelo de Rede Neural com duas camadas de processamento



Fonte: o próprio autor

- **Redes Neurais Convolucionais (CNN):** Utilizadas principalmente em reconhecimento de imagens, explorando a hierarquia espacial dos dados. É um algoritmo de aprendizado profundo. Uma camada de convolução é composta por diversos neurônios, cada um responsável por aplicar um filtro em um pedaço específico da imagem (VARGAS, PAES e VASCONCELOS, 2016);
- **Redes Neurais Recorrentes (RNN):** Usadas para dados sequenciais, como séries temporais ou processamento de linguagem natural (DEEN, 2016). Enquanto uma rede neural multicamadas pode mapear apenas entradas com saídas, as RNNs guardam um histórico das entradas para cada saída (SILVA, 2022). A Figura 12 exibe modelo de rede neural recorrente.

Figura 12 - Modelo de Rede Neural Recorrente



Fonte: o próprio autor

4.5 Métricas de Avaliação

As Métricas de Avaliação são utilizadas para identificar se o modelo é adequado para o uso à qual está sendo proposto; o quão bem está se comportando em relação às predições (DUARTE, 2023).

4.5.1 Matriz de Confusão

Necessária para o entendimento do funcionamento das métricas, a matriz de confusão indica os erros e acertos do modelo.

A Tabela 3 exibe um modelo de Matriz de Confusão de duas classes.

Tabela 3 – Modelo de Matriz de Confusão – duas classes

		Predição	
		Sim	Não
Real	Sim	Verdadeiro Positivo	Falso Negativo
	Não	Falso Positivo	Verdadeiro Negativo

Fonte: o próprio autor

Onde **Verdadeiro Positivo** indica a correta identificação da classe como positiva (correta); **Falso Positivo**, quando o modelo prevê como positivo uma classe negativa; **Falso**

Negativo, quando o modelo prevê como negativa uma classe positiva; e **Verdadeiro Negativo**, quando o modelo prevê como negativa uma classe negativa. (RODRIGUES, 2019).

A Tabela 4 exibe um modelo de Matriz de Confusão multiclasse. Neste exemplo, com quatro classes, considerando que a classe 1 está sendo analisada.

Tabela 4 – Modelo de Matriz de Confusão - multiclasse

		Predição			
		1	2	3	4
Real	1	Verdadeiro Positivo	Falso Negativo	Falso Negativo	Falso Negativo
	2	Falso Positivo	Verdadeiro Negativo	Verdadeiro Negativo	Verdadeiro Negativo
	3	Falso Positivo	Verdadeiro Negativo	Verdadeiro Negativo	Verdadeiro Negativo
	4	Falso Positivo	Verdadeiro Negativo	Verdadeiro Negativo	Verdadeiro Negativo

Fonte: o próprio autor

Ao entender a matriz de confusão, podemos então prosseguir para as Métricas de Avaliação.

4.5.2 Acurácia

A acurácia (ACC) é a métrica mais geral. Indica, dentre todas as classificações, quantas foram corretas (RODRIGUES, 2019). Considerando VP - Verdadeiro Positivo, VN - Verdadeiro Negativo, FP - Falso Positivo e FN - Falso Negativo, a fórmula da acurácia (equação 1111) é:

$$ACC = \frac{VP + VN}{VP + VN + FP + FN} \quad (1)$$

4.5.3 Precisão

Identifica, dentre todas as classificações, quais estão corretamente positivas (RODRIGUES, 2019). Na equação 2, a fórmula da Precisão (PRC).

$$PRC = \frac{VP}{VP + FP} \quad (2)$$

4.5.4 Recall

Dentre todas as situações de classe Positivo como valor esperado, quantas estão corretas (RODRIGUES, 2019). Ou seja, dentre todas as classes positivas, quantas o modelo acertou. Na equação 3, a fórmula do Recall (RCL).

$$RCL = \frac{VP}{VP + FN} \quad (3)$$

4.5.5 F1

É a média harmônica entre precisão e recall. Na equação 4, a fórmula da F1 (RODRIGUES, 2019).

$$F1 = 2 * \frac{PRC * RCL}{PRC + RCL} \quad (4)$$

Utilizando os dados de retorno diretamente da matriz de confusão, temos a equação 5.

$$F1 = \frac{VP}{VP + \frac{1}{2} * (FP + FN)} \quad (5)$$

4.5.6 Métricas para Problemas Multiclasse

- **Macro:** é a média aritmética. Todas as classes contribuem igualmente para a média final (KHALUSOVA, 2022). A fórmula está descrita na equação 6.

$$Macro = \frac{\Sigma \text{Pontuação}}{\text{Número de Classes}} \quad (6)$$

- **Micro:** todas as amostras contribuem igualmente para a métrica média final (KHALUSOVA, 2022). Na equação 7, a fórmula de cálculo da Precisão. Na equação 8, a fórmula de cálculo da F1.

$$PRC \text{ Micro} = \frac{\Sigma VP}{\Sigma VP + \Sigma FP} \quad (7)$$

$$F1 \text{ Micro} = \frac{\Sigma VP}{\Sigma VP + \frac{1}{2} * (\Sigma FP + \Sigma FN)} \quad (8)$$

- **Macro Weighted:** é a média ponderada. A contribuição de cada classe para a média é ponderada pelo seu tamanho (KHALUSOVA, 2022), como descrito na equação 9.

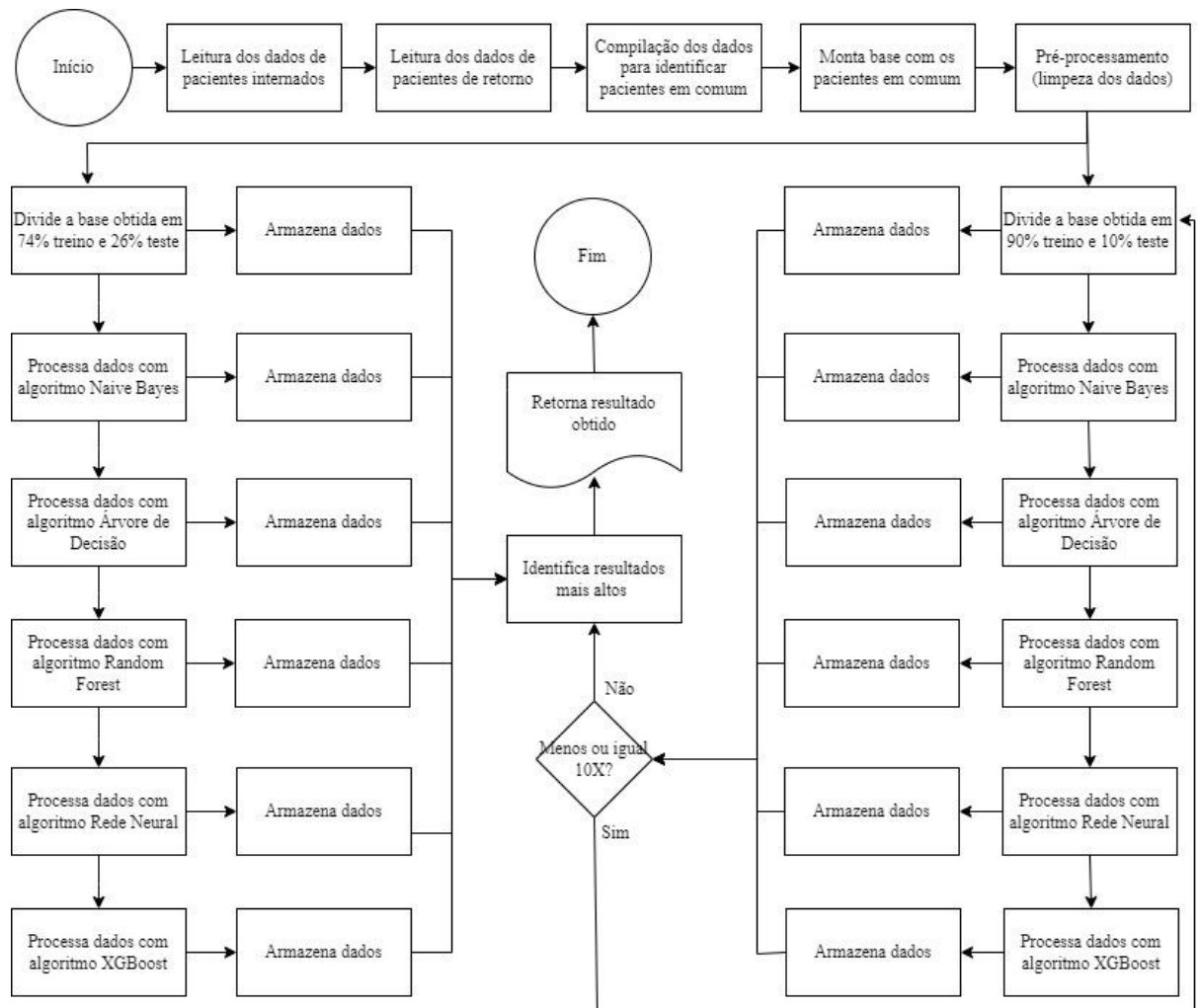
$$Macro \text{ Weighted} = \frac{\Sigma \text{Pontuação} * \text{Peso}}{\Sigma \text{Peso}} \quad (9)$$

5 METODOLOGIA

5.1 Apresentação

A Figura 13 exibe fluxograma do projeto, desde a captação dos dados origem até a devolução do resultado.

Figura 13 - Fluxograma do projeto



Fonte: o próprio autor

O levantamento dos dados foi feito a partir de dois subsistemas que já estão em uso no HU-FURG, desenvolvidos durante a pandemia, quando surgiu a necessidade de um controle mais criterioso dos dados. O primeiro deles contém os dados dos pacientes com COVID-19, que foram internados no HU-FURG, onde constam dados de:

- **Comorbidades** – contém as comorbidades preexistentes do paciente, conforme Figura 25, no apêndice B;

- **Medicamentos utilizados** – contém os medicamentos utilizados no paciente durante o tratamento. Separados em medicamentos convencionais e medicamentos de tratamento antirretroviral (TARV), conforme Figura 26, no apêndice B;
- **Sintomas referidos** – sintomas relatados pelo paciente e constatados pelo médico, bem como a data que começaram a se manifestar, conforme Figura 27, no apêndice B;
- **Sinais de gravidade** - são registrados os sinais de gravidade apresentados pelo paciente, bem como a data que começaram, conforme Figura 28, no apêndice B;
- **Evidência de Infecção** - dados de evidência de infecção do paciente. Esses dados são tipo, foco, agente e medicamentos usado no combate à Infecção, conforme Figura 29, no apêndice B;
- **RX do tórax** - dados de exames de raio-x do tórax do paciente, contendo a data de realização e o resultado, conforme Figura 30, no apêndice B;
- **TC do tórax** - dados de exames de tomografia computadorizada do paciente, contendo a data de realização e o resultado, conforme Figura 31, no apêndice B;
- **Prescrição respiratória** - dados do tratamento da função respiratória do paciente, contendo a medicação administrada, data de início e data de fim da administração, conforme Figura 32, no apêndice B;
- **Oferta de O2** - dados de oferta de O2 ao paciente, contendo a via de administração, data de início, data de fim e o volume administrado, conforme Figura 33, no apêndice B;
- **Dispositivos invasivos** – dados de uso de dispositivos invasivos, contendo o dispositivo utilizado, data de início e data de fim, conforme Figura 34, no apêndice B;
- **RT-PCR para SARS-CoV-2** – contém os dados dos exames RT-PCR do paciente, contendo a situação do PCR, a amostra utilizada e a data de realização, conforme Figura 35, no apêndice B;
- **Teste rápido de antígeno para SARS-CoV-2** – dados do exame de antígenos, contendo a situação do PCR, a amostra coletada e data de realização, conforme Figura 36, no apêndice B;
- **TR IGG/IGM (Teste Rápido Imunoglobulina G / Teste Rápido Imunoglobulina M) para SARS-CoV-2** – dados do exame de TG IGG/IGM,

contendo a situação do TR e data de realização, conforme Figura 37, no apêndice B;

- **ECG (eletrocardiograma)** – dados dos exames de ECG, contendo o resultado e data de realização, conforme Figura 38, no apêndice B;
- **Cultura** – dados das culturas, contendo o micro-organismo, a amostra e a data da coleta, conforme Figura 39, no apêndice B;
- **Evolução** - dados da evolução, contendo a data da evolução e a descrição da evolução, conforme Figura 40, no apêndice B;
- **Vacinas** - as vacinas que foram administradas no paciente, contendo o nome da vacina, qual a dose aplicada e a data de aplicação, conforme Figura 41, no apêndice B;
- **Tromboembolismo** – dados dos tromboembolismos, contendo a data da ocorrência e o evento propriamente dito, conforme Figura 42, no apêndice B.

5.2 Sistemas Fontes de Dados

Há dois subsistemas em uso no HU-FURG para registro dos dados relativos à COVID-19: o sistema de Fichas covid, onde estão registrados os pacientes que foram internados por terem contraído essa doença, e o sistema de Pós-covid, onde estão registrados os dados dos pacientes que foram internados no HU-FURG e que retornaram com sintomas, diagnosticados como sequelas decorrentes da COVID-19. Os dados estão armazenados em tabelas do banco de dados Postgresql, utilizado pelo HU-FURG para manter sua base de dados. A extração destes dados é feita via programa PHP, desenvolvido pelo autor do projeto, que gera saída no formato de arquivo .csv a ser lido pela linguagem de processamento R. No sistema onde constam os dados dos pacientes que foram internados, constam 300 registros. Os dados são compostos pelos dados pessoais do paciente, comorbidades, medicamentos utilizados (convencional e anti-retroviral), sintomas referidos pelo paciente (sintoma e data de início), sinais de gravidade (sinal e data de início), evidência de infecção (tipo, foco, agente, e drogas utilizadas, contendo sua dosagem, data de início e data de fim de aplicação), raio X de tórax (contendo alteração e data), tomografia computadorizada de tórax (alteração e data), prescrição respiratória (droga, data de início e data de fim de uso), oferta de O2 (oferta (como cateter nasal), data de início e data de fim de uso e volume administrado), dispositivos invasivos (dispositivo, data de início e data de fim), RT-PCR (situação do PCR, amostra e

data), teste de antígeno (situação do PCR, amostra e data), TR IGG/IGM (situação TR e data), ECG (alteração e data), cultura (microorganismo, amostra e data de coleta), evolução (data e descrição), vacinas (vacina, dose e data de aplicação), e tromboembolismo (data e evento). No sistema de controle de retorno, há 400 registros, contendo os dados pessoais do paciente, evolução (data e descrição), diagnóstico (diagnóstico padronizado, data e descrição do diagnóstico) e tomografia (data e descrição). Ao cruzar os dados dos dois sistemas, visando identificar os pacientes que foram internados e retornaram com sintomas, foram obtidos 130 registros, com 570 colunas para verificação. Os registros das tabelas precisaram ser convertidos em colunas, possibilitando seu processamento pelo R. Para exemplificar, as vacinas, que são registros dos pacientes e caracterizados por doses, foram convertidas em colunas (covid-19 dose 1; outra coluna covid-19 dose 2; outra coluna gripe). E cada coluna recebe os valores Verdadeiro ou Falso, caso tenha sido ou não administrada àquele paciente.

Os resultados esperados são a predição.

No R, os dados precisam ser tratados após a importação (pré-processamento), começando pela conversão de alguns dados, para se adaptarem ao projeto. A transformação foi:

- Conversão dos dados do diagnóstico para dados categóricos;
- Conversão das datas, do tipo caracter para o tipo data;
- Discretização da idade dos pacientes, de números para faixas de idade.

Após a adaptação dos dados, foi necessário limpá-los. Para este projeto, a limpeza dos dados consiste de:

- exclusão de colunas com valores nulos em todas as instâncias;
- exclusão de colunas com ocorrências em um mínimo de instâncias;
- exclusão de colunas irrelevantes para a obtenção de resultado, como o caso de nº de internação e o nº da ficha, que são valores únicos;
- exclusão de instância de paciente falecido, pois também está fora do que é proposto nesta pesquisa;
- exclusão de instância de paciente com coluna alvo Depressão, pois é ocorrência única. Para o processamento utilizando 3 ou 2 classes, Depressão se converte para True e esta instância não é excluída;
- exclusão de instância de paciente com coluna alvo Insônia, pelo mesmo motivo da anterior. Para o processamento utilizando 3 ou 2 classes, Insônia se converte para True e esta instância não é excluída;

- exclusão de colunas com somente um resultado, pois não impactam na obtenção do resultado.

Após o tratamento dos dados, restam 127 registros (ou 129, quando na modalidade de 2 ou 3 classes. Diarréia e Insônia contemplam um registro cada e são eliminadas na limpeza de dados de 15 classes. Mas na modalidade de 2 ou 3 classes, são convertidos em Positivo) com 203 colunas para serem processados.

Para avaliação de performance dos algoritmos e relevância dos resultados, os resultados esperados foram divididos em três modalidades: duas classes (Positivo ou Negativo), três classes (Positivo, Negativo e Astenia). A Astenia foi a sequela com mais ocorrências de todas dentre os casos pesquisados. E quinze classes (Agravamento de Doenças Pré-existent, Alopécia, Amnésia, Ansiedade, Astenia, Cefaléia, Depressão, Diarréia, Dificuldade Cognitiva, Disgeusia, Dispnéia, Dores Abdominais, Edema, Falso, Insônia, Mialgia e Tosse). Depressão e Insônia também foram observados, porém em somente um dos pacientes pesquisados cada, e foram eliminados na fase de limpeza de dados.

A seguir, base é dividida em duas partes: a base de treino com 74% dos dados originais (quinze classes), e a base de teste, com os 26% de dados restantes.

De modo a aprofundar a pesquisa, foi executado um segundo processamento, onde foi aplicada a técnica de avaliação Cross Validation, utilizando o método K-fold. A opção foi dividir todos os 127 registros disponíveis em 10 partições, utilizando uma das partições como teste e as demais como treino a cada repetição. Foram processadas em nove repetições para cada partição, utilizando cada um dos cinco algoritmos selecionados. As sete primeiras partições contendo 13 registros cada e as partições restantes, 12 registros cada. Nesta segunda forma de processamento, foram utilizadas as três modalidades de pesquisa, com duas, três e quinze classes.

A Figura 22 exibe a tela principal do sistema de Fichas covid, onde constam dados de identificação do paciente. Este programa permite a manutenção das fichas, onde poderão ser incluídos novos paciente que sejam internados, bem como atualizar seus dados.

A Figura 24 exibe dados detalhados de paciente que está sendo incluído no sistema – uma ficha de controle está sendo gerada. Os dados relativos à internação são carregados para conferência, diretamente do sistema de controle de internação. Os dados relativos ao paciente são carregados do cadastro de pacientes. São preenchidos nesta tela os demais dados necessários ao controle da COVID-19.

O segundo sistema contém os dados de pós-covid – pacientes que procuraram o HU após a fase aguda da COVID-19, com algum problema decorrente da doença. Neste sistema

fica registrada a evolução; o motivo pelo qual o paciente procurou o HU-FURG novamente, o parecer médico sobre a situação do paciente e o tratamento prescrito ao paciente.

Através da compilação dos dados dos dois subsistemas, espera-se prever possíveis recorrências dos pacientes que foram tratados no HU-FURG. Essa previsão ficará gravada no sistema e disponível aos profissionais de saúde para consulta, juntamente com a data em que foi obtida.

A Figura 45 exibe a tela principal onde são registrados os dados da evolução do paciente, contendo a data do registro, a evolução e o nome do profissional que registrou a evolução no sistema.

A Figura 46 exibe a tela de registro de diagnóstico, contendo a data do registro e a descrição.

A Figura 47 exibe a tela de registro de tomografias, contendo a data do registro e a descrição do resultado.

Para a compilação desses dados será utilizada a técnica de *machine learning* (aprendizado de máquina).

5.3 Geração dos Modelos

Para ser obtido o melhor modelo de previsão, foram construídos cinco modelos, e seus retornos foram comparados, para identificar qual o mais adequado para esta utilização no HU-FURG. Os algoritmos utilizados para a confecção dos modelos foram Árvore de Decisão, Random Forest, Naive Bayes, XGBoost e Rede Neural.

6 RESULTADOS

6.1 Interpretação e Avaliação dos Resultados

Nesta última etapa de KDD, também conhecida como pós-processamento, os dados obtidos precisam ser interpretados e avaliados. Todos os envolvidos devem avaliar criteriosamente os dados obtidos, para que, desta forma, o modelo seja interpretado e o resultado extraído.

“É nesta etapa que o especialista em KDD e o especialista no domínio da aplicação avaliam os resultados obtidos e definem novas alternativas de investigação dos dados.” (GOLDSCHMIDT e PASSOS, 2005, p. 55)

Para análise foram usadas as métricas Acurácia, Precisão, Recall e F1.

Tabela 5 - Distribuição dos Dados da Pesquisa

Dataset	Instâncias	Atributos	Treino	Teste	Balanceamento
	127	203	74%	23%	15 classes
	127	203	CrossValidation Fold = 10		15 classes
	129	203	CrossValidation Fold = 10		3 classes
	129	203	CrossValidation Fold = 10		2 classes

Fonte: o próprio autor

Tabela 6 - Balanceamento dos Dados Pesquisados - 2 classes

Classes	Ocorrências
Positivo	112
Negativo	18
Total	130

Fonte: o próprio autor

Tabela 7 - Balanceamento dos Dados Pesquisados - 3 classes

Classes	Ocorrências
Positivo	112
Negativo	18
Astenia	23
Total	130

Fonte: o próprio autor

Tabela 8 - Balanceamento dos Dados Pesquisados - 15 classes

Classes	Ocorrências
Agravamento de Doenças Pré existentes	4
Alopecia	5
Amnesia	5
Ansiedade	11
Astenia	23
Cefaleia	3
Depressao	1
Diarreia	6
Dificuldade Cognitiva	13
Disgeusia	2
Dispneia	22
Dor Abdominal	2
Edema	4
Negativo	18
Insonia	1
Mialgia	6
Tosse	4
Total	130

Fonte: o próprio autor

6.1.1 Naive Bayes

O resultado obtido pelo processamento do algoritmo Naive Bayes foi acurácia de 11,54%, utilizando a base de treino e gerando a matriz de confusão pela base de teste. Esta base está com todas as 15 classes (Agravamento de Doenças Pré-existentes, Alopecia, Amnésia, Ansiedade, Astenia, Cefaléia, Depressão, Diarréia, Dificuldade Cognitiva, Disgeusia, Dispneia, Dores Abdominais, Edema, Falso, Insônia, Mialgia e Tosse); acurácia baixa. O NIR calculado está dentro do intervalo de confiança, indicando que o modelo é ruim, com os parâmetros padrão, descritos a seguir:

- **formula:** a coluna classe é indicada, bem como é descrito que todas as colunas da base serão utilizadas na pesquisa para obter a coluna classe;
- **data:** especifica a base de treino como fonte de dados.
- **laplace:** não utilizado. Default 0;
- **na.action:** não utilizado. Padrão: os NAs (nulos) encontrados não são contados.

Ao alterar o parâmetro *laplace* para 1, a acurácia passou para 26,92%. O parâmetro *laplace*, quando aumentado, evita a probabilidade 0; basicamente, adiciona um pequeno

número a cada uma das contagens na tabela de frequências (RUSSEL e NORVIG, 2003) (neste caso, 1). Contudo, afeta a precisão do modelo com amostras muito pequenas. Alterar o parâmetro *na.action* (alterado para *na.omit* – NAs são rejeitados) foi inócuo.

Quando utilizada a base com 3 classes, positivo, negativo e Astenia (fadiga), a acurácia foi de 54,84%.

Ao utilizar a base com 2 classes, positivo e negativo, a acurácia de 81,25%. Precisão, Recall e F1 todos retornaram 25%. Acurácia um pouco mais elevada, mas os demais indicadores muito baixos.

6.1.2 Random Forest

O resultado obtido pelo processamento do algoritmo Random Forest, a acurácia foi de 26,92%, utilizando a base de treino e gerando a matriz de confusão pela base de teste. Os parâmetros utilizados foram:

- **formula:** a coluna classe é indicada, bem como é descrito que todas as colunas da base serão utilizadas na pesquisa para obter a coluna classe;
- **data:** especifica a base de treino como fonte de dados.
- **importance:** informada como TRUE, indica que a importância dos preditores deve ser avaliada. Sem a informação deste parâmetro, a acurácia muda para 23,08%.

Esta base está utilizando todas as 15 classes. Acurácia baixa. O NIR está dentro do intervalo de confiança, indicando que o modelo é ruim.

Quando utilizada a base com 3 classes (positivo, negativo e Astenia) a acurácia foi de 70,97%.

Ao utilizar a base com 2 classes (positivo e negativo) a acurácia foi de 87,50%. Mas todos os registros retornaram como verdadeiro, tornando irrelevantes os indicadores Precisão, Recall e F1. Acurácia já bem mais elevada, quando comparada ao processamento com 3 classes e com todas as classes.

6.1.3 Árvore de Decisões

Para o algoritmo Árvore de Decisão, a acurácia foi de 7,69%, utilizando a base de treino e gerando a matriz de confusão pela base de teste, e a coluna classe com 15 valores distintos. Acurácia baixa. O NIR está dentro do intervalo de confiança, indicando que o

modelo é ruim. Os parâmetros de controle utilizados foram `minsplit=5`, `minbucket=1`, `maxdepth=10`, que são:

- **Minsplit** é o número mínimo de observações que devem existir em um nó para que uma divisão seja tentada;
- **Minbucket** é o número mínimo de observações em qualquer nó terminal (folha). Se apenas um `minbucket` ou `minsplit` for especificado, o código definirá `minsplit` como `minbucket*3` ou `minbucket` como `minsplit/3`, conforme apropriado;
- **Maxdepth** é profundidade máxima de qualquer nó da árvore final, com o nó raiz contado como profundidade 0. Valores maiores que 30 `rpart` darão resultados absurdos em máquinas de 32 bits. (`rpart.control`: Control for `Rpart` Fits)

Os parâmetros de controle foram modificados para `minsplit=30`, `minbucket=10`, `maxdepth=30`. Após a modificação dos parâmetros, a acurácia retornou 23,08%.

Quando utilizada a base com 3 classes, positivo, negativo e Astenia, a acurácia foi de 58,06% com os parâmetros originais e 70,97% com os parâmetros modificados.

Ao utilizar a base com 2 classes, positivo e negativo, a acurácia foi de 87,50%, com os parâmetros padrão e repetindo o resultado para os parâmetros modificados. Acurácia já bem mais elevada, quando comparada ao processamento com 3 classes e com todas as classes.

6.1.4 XGBoost

Para o algoritmo XGBoost com todas as classes, não foi possível estabelecer uma base de teste para uso, pois são 130 os registros disponíveis; todas as etapas retornaram sem resultado. Os parâmetros informados fora:

- **data**: matriz `xgboost` gerada a partir da base de treino;
- **max.depth**: profundidade máxima de árvores de decisão. Neste caso foi utilizado como 3. Valor maior ou menor resultou em resultados menores quando utilizado para 2 ou 3 classes;
- **nrounds**: número de iterações. Após testes sucessivos (com análise dos dados de retorno) 499 foi o número mais adequado, visando um resultado com percentual mais alto e com melhor desempenho;
- **verbose**: utilizado como 0 para o R não retornar mensagens de progresso.

Já para classe com 3 valores distintos (positivo, negativo e Astenia), a acurácia de 58,06%, utilizando a base de treino e gerando a matriz de confusão pela base de testes. Para classes de valores binários, a acurácia foi de 84,38%.

6.1.5 Rede Neural

Com o algoritmo de Rede Neural e todas as classes, não foi possível estabelecer uma base de teste para uso, pois são muito poucos os registros disponíveis. Para a geração do modelo foram utilizados os parâmetros a seguir:

- **formula:** as colunas classe foram indicadas (cada valor de classe gerou coluna nova, para satisfazer a condição da rede neural de processar dados numéricos), bem como é descrito que todas as colunas da base serão utilizadas na pesquisa para obter a coluna classe;
- **data:** matriz gerada a partir da base de treino;
- **hidden:** com o valor 2, indica uma camada oculta com dois neurônios;
- **act.fct:** com o valor *logistic*, indica que o valor retornado será 0 ou 1;
- **linear.output:** com o valor *false*, indica que o valor retornado será passado pela função de ativação (indicada pelo parâmetro *act.fct*).

Ao utilizar a classe com 3 valores distintos, a acurácia foi de 66,67%, utilizando a base de treino e gerando a matriz de confusão pela base de teste. Acurácia mais elevada, quando comparada ao processamento com 3 classes.

Já ao utilizar a classe com valores binários, a acurácia foi de 86,84%, utilizando a base de treino e gerando a matriz de confusão pela base de teste. Acurácia mais elevada, quando comparada ao processamento com 3 classes.

A Tabela 9 traz um comparativo entre os resultados dos processamentos dos algoritmos utilizados, bem como o formato da pesquisa integral (quinze classes), com três classes e binário. A Figura 14 apresenta os mesmos dados em um gráfico de barras horizontal, que ilustra as médias dos valores calculados para as métricas em relação aos algoritmos.

Tabela 9 - Comparativo de Resultados, com 15 classes, 3 classes e 2 classes. 74% dos registros como base de treino e 26% como base de teste

	NB	NB	NB	RF	RF	RF	AD	AD	AD	XG	XG	XG	RN	RN	RN
Classes	15	3	2	15	3	2	15	3	2	15	3	2	15	3	2
Accuracy	11,54	54,84	81,25	23,08	70,97	87,50	23,08	70,97	87,50	..	58,06	84,38	..	16,67	86,84
Recall – macro	5,00	32,58	57,14	10,42	33,33	50,00	12,92	33,33	50,00	..	44,39	48,21	..	57,14	86,84

Recall – micro	11,54	54,84	81,25	23,08	70,97	87,50	23,08	70,97	87,50	..	58,06	84,38	..	16,67	86,84
Recall - macro weighted	11,54	54,84	81,25	23,08	70,97	87,50	23,08	70,97	87,50	..	58,06	84,38	..	16,67	86,84
Precision – macro	5,95	29,86	57,14	14,57	70,97	87,50	21,95	70,97	87,50	..	47,15	43,55	..	34,67	50,00
Precision – micro	11,54	54,84	81,25	23,08	70,97	87,50	23,08	70,97	87,50	..	58,06	84,38	..	16,67	86,84
Precision - macro weighted	11,90	51,95	81,25	21,87	70,97	87,50	26,79	70,97	87,50	..	65,70	76,21	..	97,34	100,00
F_meas – macro	7,58	31,11	57,14	18,30	83,02	93,33	23,10	83,02	93,33	..	44,76	45,76	..	16,35	92,96
F_meas – micro	11,54	54,84	81,25	23,08	70,97	87,50	23,08	70,97	87,50	..	58,06	84,38	..	16,67	86,84
F_meas - macro weighted	13,26	53,33	81,25	27,46	83,02	93,33	27,24	83,02	93,33	..	61,25	80,08	..	24,52	92,96

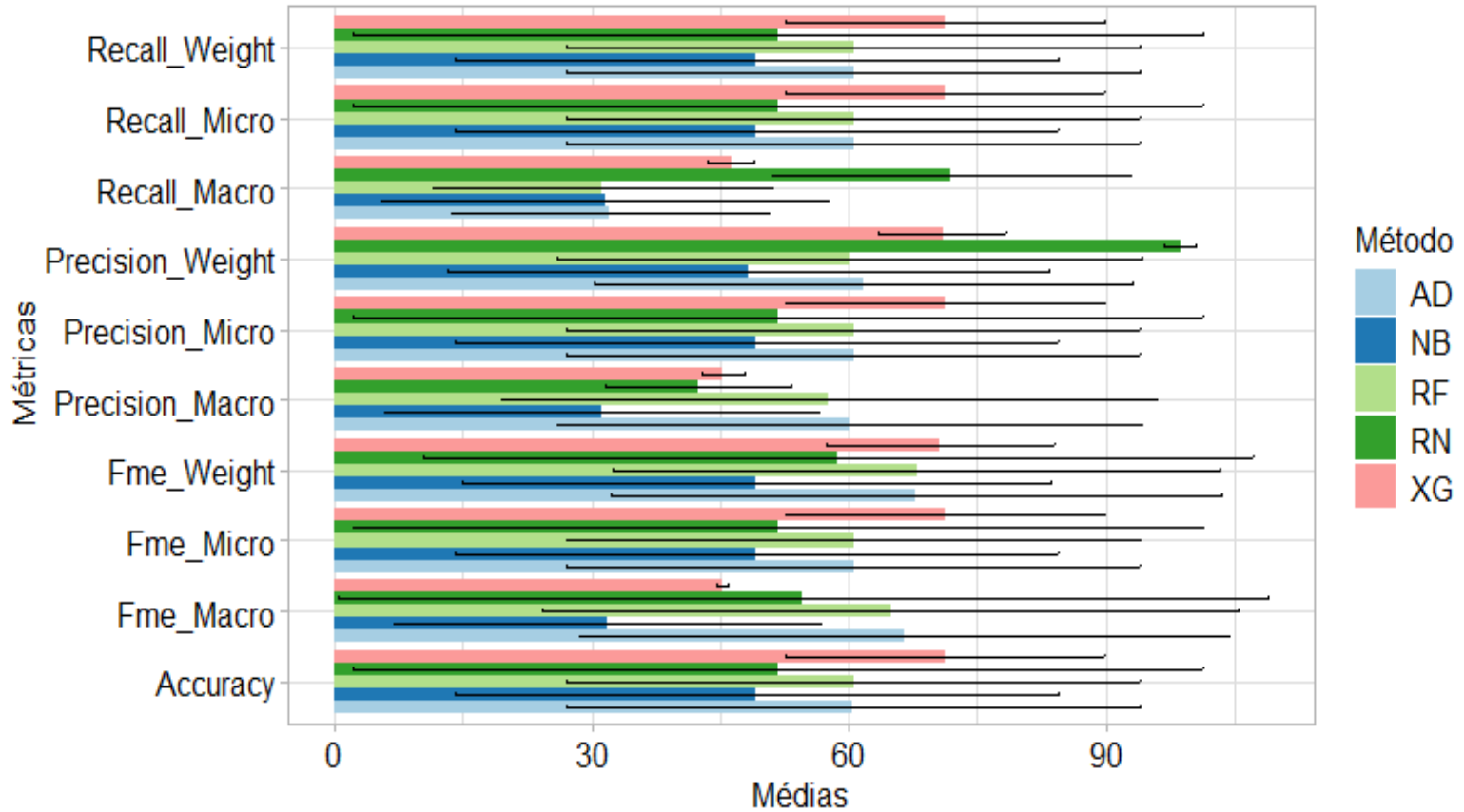
Legenda:

AD	Árvore de Decisão
NB	Naive Bayes
RF	Random Forest
XG	XGBoost
RN	Rede Neural

Fonte: o próprio autor

Figura 14 – Comportamento das Médias das Métricas em relação aos Métodos

As barras pretas indicam o desvio padrão das médias. Legenda: AD – Árvore de Decisão, NB – Naive Bayes, RF – Random Forest, XG – XGBoost e RN – Rede Neural.



Fonte: o próprio autor

6.1.6 Divisão dos Dados em Partições

Para melhor análise, foi utilizada a técnica de Cross Validation, onde os dados foram divididos em 10 partições e analisados separadamente. A divisão foi feita da seguinte forma:

- as 7 primeiras partições com 13 registros e as 3 últimas com 12 registros para análise do banco de dados com 15 classes;
- todas as partições com 13 registros para análise do banco de dados com 2 ou 3 classes.

Na Tabela 10 constam os resultados do processamento das partições de dados, com 15 classes.

A Figura 15 apresenta os mesmos dados em um gráfico de barras horizontal, que ilustra as médias dos valores calculados para as métricas em relação aos algoritmos.

Na Tabela 11 constam os resultados do processamento das partições de dados, com 3 classes.

A Figura 16 apresenta os mesmos dados em um gráfico de barras horizontal, que ilustra as médias dos valores calculados para as métricas em relação aos algoritmos.

Na Tabela 12 constam os resultados do processamento das partições de dados, com duas classes.

A Figura 17 apresenta os mesmos dados em um gráfico de barras horizontal, que ilustra as médias dos valores calculados para as métricas em relação aos algoritmos.

Tabela 10 - Médias das Médias das Métricas dos Algoritmos Utilizados na modalidade Cross-Validation - 15 classes

Médias	Accuracy	Recall Macro	Recall Micro	Recall Weight	Precision Macro	Precision Micro	Precision Weight	Fme Macro	Fme Micro	Fme Weight
AD	12,63	12,11	12,63	12,63	13,65	12,63	28,55	26,99	12,63	26,18
NB	18,97	15,53	18,97	18,97	18,04	18,97	32,63	27,80	18,97	29,34
RF	15,83	11,51	15,83	15,83	20,01	15,83	28,58	19,79	15,83	23,98
XG	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
RN	3,27	1,88	3,27	3,27	3,33	3,27	18,18	3,93	3,27	5,52

Legenda: AD Árvore de Decisão

NB Naive Bayes

RF Random Forest

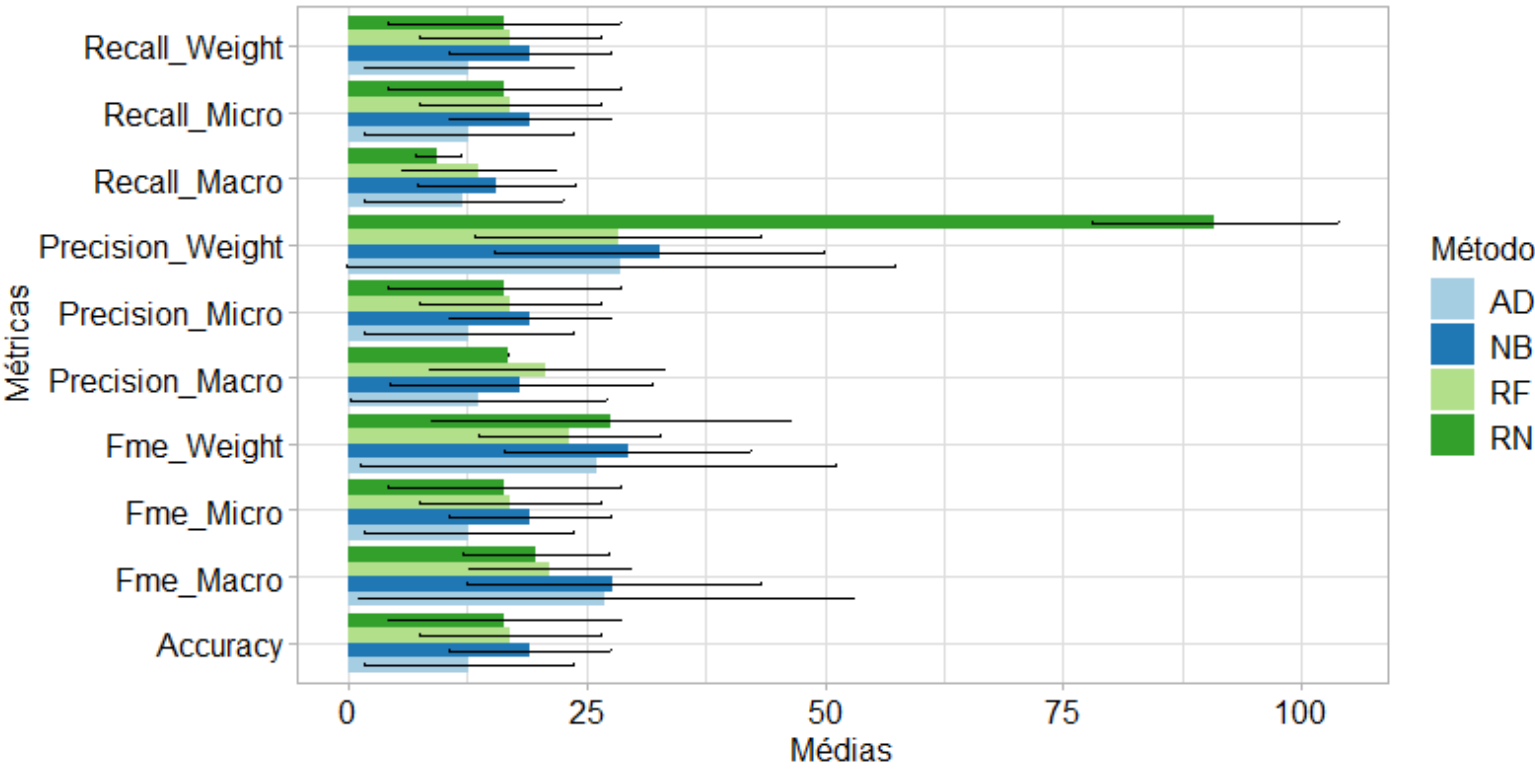
XG XGBoost

RN Rede Neural

Fonte: o próprio autor

Figura 15 - Comparativos das Partições - 15 classes

As barras pretas indicam o desvio padrão das médias. Legenda: AD – Árvore de Decisão, NB – Naive Bayes, RF – Random Forest e RN – Rede Neural.



Fonte: o próprio autor

Tabela 11 - Médias das Médias das Métricas dos Algoritmos Utilizados na modalidade Cross-Validation - 3 classes

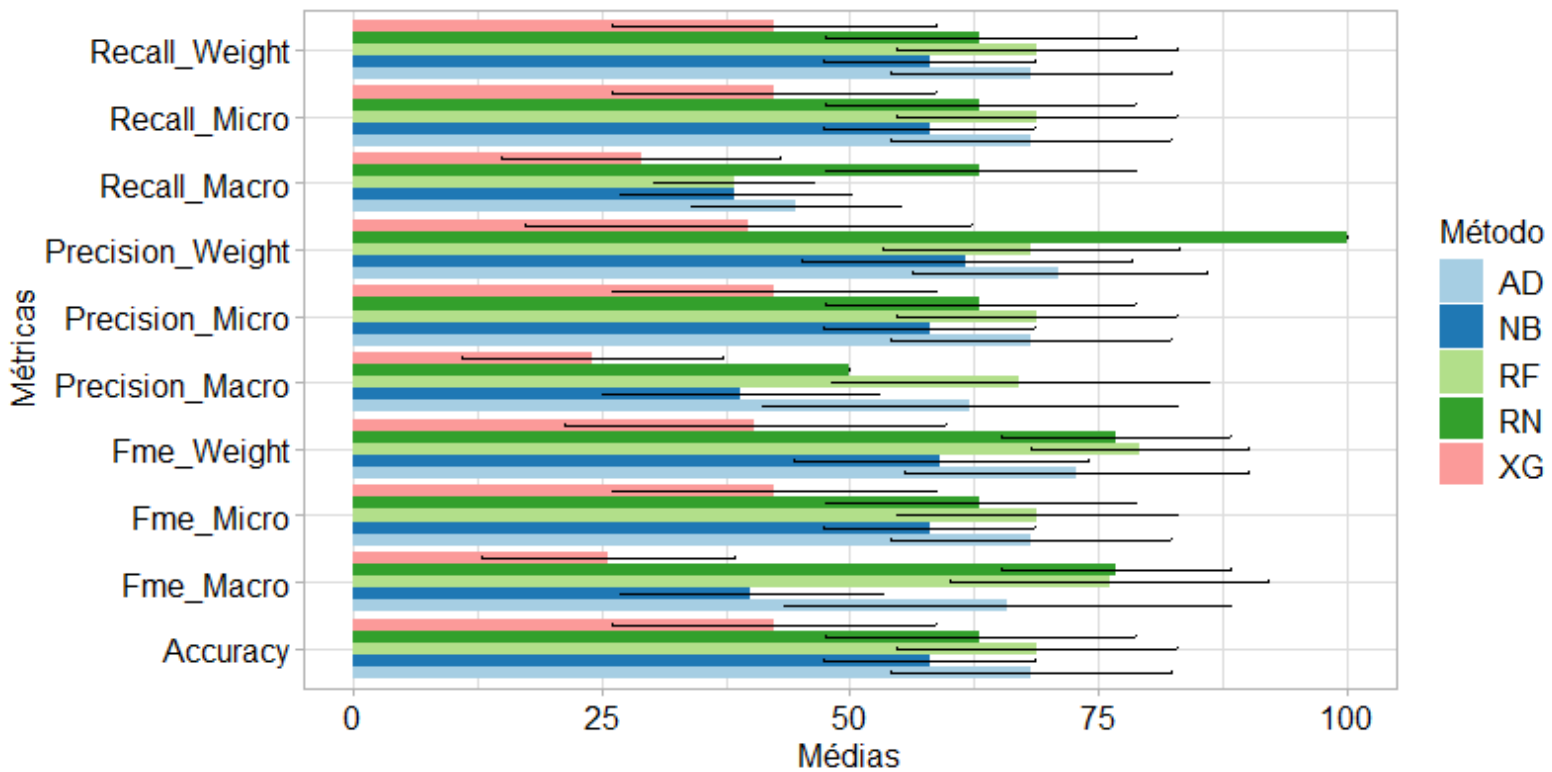
Médias	Accuracy	Recall Macro	Recall Micro	Recall Weight	Precision Macro	Precision Micro	Precision Weight	Fme Macro	Fme Micro	Fme Weight
AD	68,27	44,50	68,27	68,27	62,04	68,27	71,11	65,88	68,27	72,78
NB	58,08	38,43	58,08	58,08	38,99	58,08	61,77	40,04	58,08	59,19
RF	68,85	38,33	68,85	68,85	67,08	68,85	68,29	76,11	68,85	79,12
XG	8,46	5,79	8,46	8,46	4,81	8,46	7,97	5,13	8,46	8,09
RN	63,09	63,09	63,09	63,09	50,00	63,09	100,00	76,75	63,09	76,75

Legenda: AD Árvore de Decisão
NB Naive Bayes
RF Random Forest
XG XGBoost
RN Rede Neural

Fonte: o próprio autor

Figura 16 - Comparativo das Partições - 3 classes

As barras pretas indicam o desvio padrão das médias. Legenda: AD – Árvore de Decisão, NB – Naive Bayes, RF – Random Forest, XG – XGBoost e RN – Rede Neural.



Fonte: o próprio autor

Tabela 12 - Médias das Médias das Métricas dos Algoritmos Utilizados na modalidade Cross-Validation - 2 classes

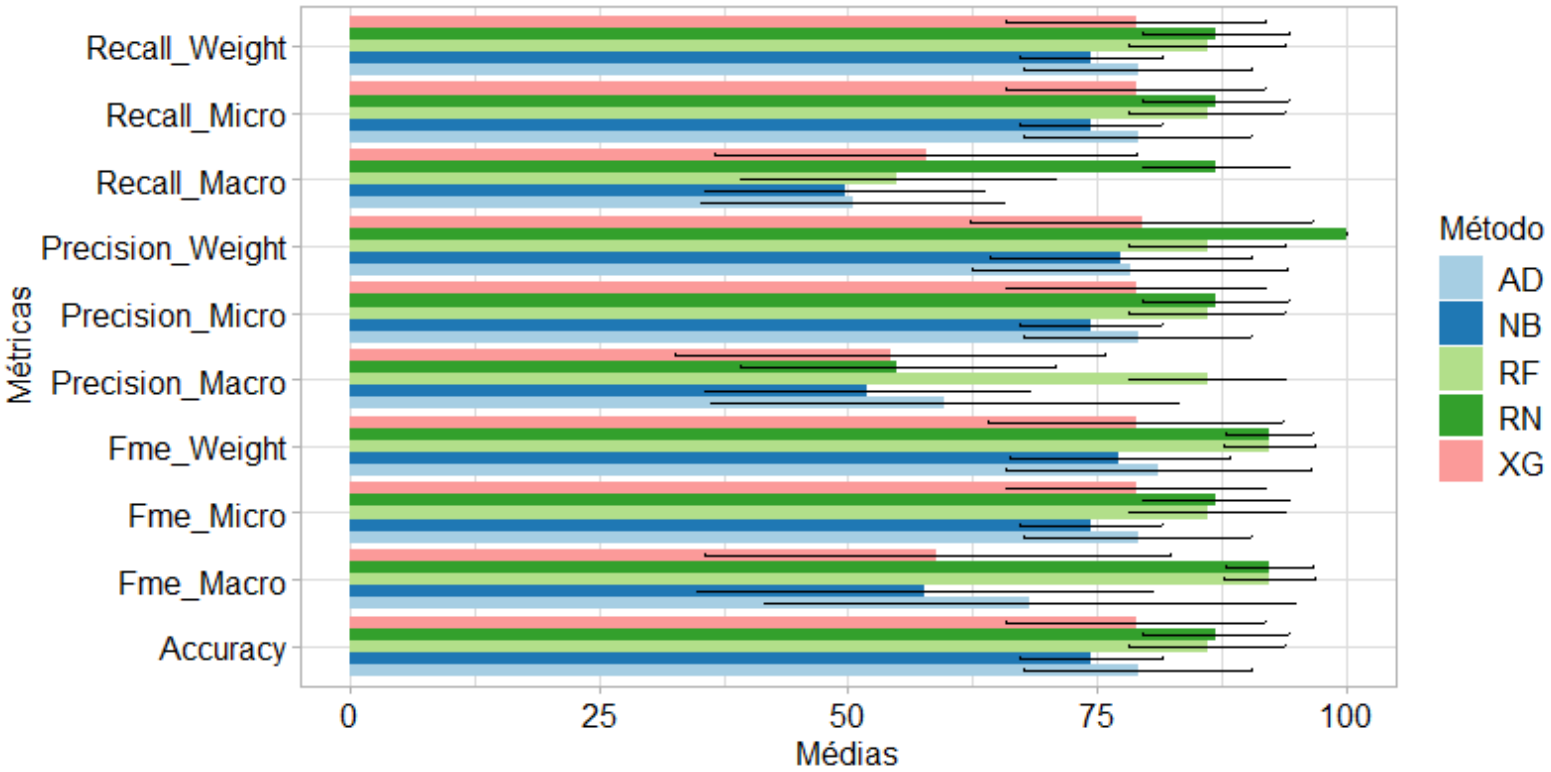
Médias	Accuracy	Recall Macro	Recall Micro	Recall Weight	Precision Macro	Precision Micro	Precision Weight	Fme Macro	Fme Micro
AD	79,17	50,48	79,17	79,17	59,70	79,17	78,36	68,16	79,17
NB	74,42	49,65	74,42	74,42	51,95	74,42	77,42	57,67	74,42
RF	86,09	55,00	86,09	86,09	86,09	86,09	86,09	92,35	86,09
XG	55,26	40,49	55,26	55,26	37,99	55,26	55,67	41,27	55,26
RN	86,95	86,95	86,95	86,95	55,00	86,95	100,00	92,34	86,95

Legenda: AD Árvore de Decisão
 NB Naive Bayes
 RF Random Forest
 XG XGBoost
 RN Rede Neural

Fonte: o próprio autor

Figura 17 - Comparativo das Partições - 2 classes

As barras pretas indicam o desvio padrão das médias. Legenda: AD – Árvore de Decisão, NB – Naive Bayes, RF – Random Forest, XG – XGBoost e RN – Rede Neural.



Fonte: o próprio autor

7 DISCUSSÃO

Para a análise dos resultados obtidos com diferentes algoritmos de aprendizado de máquina neste estudo, foram utilizadas comparações entre gráficos (Figura 14 à Figura 17) gerados a partir do uso da técnica de validação cruzada (cross-validation). Esta técnica é amplamente empregada em tarefas de classificação, uma vez que garante maior confiabilidade na avaliação do desempenho dos algoritmos, reduzindo o viés causado pela divisão dos dados em conjuntos de treino e teste.

Os resultados destacam que, entre os algoritmos avaliados, o Árvore de Decisão apresentou resultados mistos. Embora tenha se destacado em algumas comparações, apresentando bom desempenho, em outros cenários o seu resultado foi insatisfatório, chegando a apresentar acurácia nula. Isso demonstra que o algoritmo pode ser sensível à variabilidade dos dados ou à complexidade do problema, o que é uma característica conhecida deste tipo de abordagem, que pode sofrer com overfitting ou underfitting, dependendo do balanceamento das classes e da complexidade dos dados.

O algoritmo XGBoost, reconhecido pela sua eficiência e robustez em diversas tarefas de classificação, apresentou uma limitação importante: nas partições 1, 7 e 9 não foi possível gerar resultados. Essa limitação pode estar relacionada à forma como os dados foram estruturados ou à configuração dos hiperparâmetros, já que o XGBoost tende a ser sensível ao ajuste fino dos parâmetros, como a taxa de aprendizado, a profundidade das árvores e o número de iterações. A ausência de resultados em algumas partições levou à desconsideração deste modelo para o contexto geral do trabalho, sugerindo que sua aplicação pode não ser adequada para o conjunto de dados específico utilizado.

Ao analisar o desempenho do processamento com classe binária, a Rede Neural emergiu como o algoritmo com os melhores resultados. Isso é condizente com a literatura, já que redes neurais, especialmente com arquiteturas adequadas, são poderosas em capturar padrões não lineares e complexos presentes nos dados, especialmente em problemas binários, onde a separação entre as classes pode ser mais facilmente aprendida. No entanto, ao aumentar a complexidade da tarefa, passando de 2 para 3 classes, a Rede Neural perde a liderança para o Random Forest, indicando que este algoritmo consegue lidar melhor com a separação de múltiplas classes. O Random Forest, com sua abordagem baseada em múltiplas árvores de decisão, tende a ser mais robusto em tarefas com maior complexidade de classes, oferecendo uma boa generalização, especialmente em cenários com mais de duas classes.

Quando a tarefa foi expandida para 15 classes, o desempenho tanto da Rede Neural quanto do Random Forest se deteriorou em várias partições. Em alguns casos, a Rede Neural nem sequer conseguiu calcular resultados, o que pode indicar um aumento significativo da dificuldade do problema, aliado à necessidade de mais dados. O mesmo problema foi identificado para o Random Forest, sugerindo que a complexidade da classificação em 15 classes pode ter exigido uma maior capacidade de generalização do que os modelos poderiam proporcionar sem sofrer com problemas de overfitting ou underfitting. Somente a Árvore de Decisão e o Naive Bayes geraram dados em todos os classificadores.

Diante desse cenário, o algoritmo Naive Bayes foi o que se destacou no contexto da classificação com 15 classes. Na Tabela 13 estão exibidas as médias das médias das métricas dos algoritmos utilizados na modalidade Cross-Validation, 15 classes.

Tabela 13 - Comparativo das Médias das Médias das Métricas dos Algoritmos Utilizados na modalidade Cross-Validation, 15 classes

Médias	Accuracy	Recall Macro	Recall Micro	Recall Weight	Precision Macro	Precision Micro	Precision Weight	Fme Macro	Fme Micro	Fme Weight
AD	12,63	12,11	12,63	12,63	13,65	12,63	28,55	26,99	12,63	26,18
NB	18,97	15,53	18,97	18,97	18,04	18,97	32,63	27,80	18,97	29,34
RF	15,83	11,51	15,83	15,83	20,01	15,83	28,58	19,79	15,83	23,98
XG	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
RN	3,27	1,88	3,27	3,27	3,33	3,27	18,18	3,93	3,27	5,52

Legenda: AD Árvore de Decisão
 NB Naive Bayes
 RF Random Forest
 XG XGBoost
 RN Rede Neural

Fonte: o próprio autor

Apesar de ser um classificador baseado em uma premissa forte (a independência entre as variáveis preditoras), o Naive Bayes é conhecido por sua simplicidade e eficiência em cenários com muitas classes. Seu desempenho consistente e equilibrado, sem a ocorrência de resultados nulos e sem a oscilação extrema entre máximos e mínimos, torna-o uma escolha pragmática para este estudo. Embora o Naive Bayes não tenha alcançado os melhores resultados em termos de acurácia máxima, ele também não apresentou os piores, sendo capaz de fornecer uma classificação estável e confiável para todas as classes, o que é um fator crítico no cenário da classificação multiclasse.

Essa característica de robustez do Naive Bayes torna-se ainda mais relevante quando se considera a necessidade prática dos médicos do HU-FURG, que preferem trabalhar com modelos capazes de lidar eficientemente com a classificação de 15 classes. Assim, mesmo que algoritmos mais complexos, como redes neurais e Random Forest, possam oferecer melhores resultados em cenários mais simples (binários ou com poucas classes), a escolha do Naive Bayes se justifica pela sua estabilidade e adequação ao contexto do problema. Este equilíbrio entre performance e praticidade é essencial em contextos aplicados, especialmente em áreas críticas como a saúde, onde a confiabilidade do modelo é muitas vezes mais importante do que a maximização de uma métrica isolada.

Portanto, a escolha do Naive Bayes como o algoritmo final deste estudo não apenas reflete sua adequação técnica, mas também atende às necessidades práticas do cenário de aplicação, demonstrando um alinhamento entre os resultados da pesquisa e as demandas dos profissionais de saúde que utilizarão o modelo.

8 CONCLUSÃO

A presente dissertação abordou algoritmos preditivos para identificar sequelas decorrentes da covid longa, utilizando uma base de dados composta por informações de pacientes tratados no HU-FURG. Este estudo teve como objetivo principal contribuir para o conhecimento local sobre as sequelas associadas à COVID-19, a partir de dados específicos da região, como comorbidades, histórico de uso de medicamentos, e persistência de sintomas pós-alta hospitalar.

Diferentemente de muitos estudos mais amplos sobre os efeitos da COVID-19, que frequentemente analisam populações diversas em contextos generalizados, o foco deste trabalho recaiu sobre a população atendida pelo HU-FURG. Essa escolha se justifica pela necessidade de considerar as particularidades locais da população do sul do Brasil, mais precisamente a região atendida pelo HU-FURG, onde fatores socioeconômicos, ambientais e de infraestrutura de saúde podem diferir substancialmente de outras regiões do país e do mundo. Ao centrar a análise em uma realidade específica, foi possível gerar resultados mais contextualizados, oferecendo um suporte mais robusto para a tomada de decisões médicas na própria instituição hospitalar e para a população local.

O trabalho incluiu, além da análise dos dados clínicos, uma revisão sobre a COVID-19, suas causas, efeitos e a progressão da pandemia, além de um estudo sobre as abordagens mais adequadas para o tratamento desses dados. Com base nessa revisão, optou-se pela utilização de técnicas de aprendizado de máquina (machine learning) para o desenvolvimento dos algoritmos preditivos, sendo que diferentes métodos de classificação foram testados e comparados. Os algoritmos foram avaliados quanto à sua capacidade de prever corretamente a ocorrência de sequelas associadas à covid longa, considerando o conjunto de variáveis disponíveis nos dados dos pacientes.

Os resultados obtidos demonstraram que os algoritmos de aprendizado de máquina são ferramentas promissoras para a predição de sequelas em pacientes com histórico de COVID-19, proporcionando aos profissionais de saúde uma forma de identificar precocemente os indivíduos em maior risco de desenvolver complicações de longo prazo. Entre os algoritmos testados, alguns apresentaram melhor desempenho em contextos específicos de predição, enquanto outros se mostraram menos eficazes ou mais sensíveis às variáveis do conjunto de dados. Esta variabilidade reforça a importância de continuar explorando e ajustando esses modelos, de modo a obter resultados cada vez mais precisos e confiáveis.

A escolha do Naive Bayes como o algoritmo mais adequado para este estudo, especialmente no contexto da predição de sequelas em 15 classes de complicações, destacou-se pelo seu equilíbrio entre simplicidade e eficiência. Embora outras técnicas, como redes neurais e o Random Forest, tenham mostrado desempenho superior em cenários menos complexos, o Naive Bayes se mostrou mais robusto no cenário de múltiplas classes, fornecendo previsões consistentes e estáveis.

Em termos práticos, os resultados deste estudo possuem implicações diretas no campo da saúde pública local. A possibilidade de prever quais pacientes têm maior propensão a desenvolver sequelas da COVID-19 permite que o HU-FURG e outros hospitais da região implementem estratégias mais eficazes de acompanhamento e tratamento pós-alta, possibilitando intervenções mais precisas e, potencialmente, uma melhoria na qualidade de vida dos pacientes. Além disso, a abordagem baseada em aprendizado de máquina proporciona um avanço significativo no uso de tecnologias para a personalização da medicina, algo que pode ser ampliado e replicado em outras instituições e para outras condições de saúde.

Dessa forma, este trabalho contribui não apenas para o avanço da ciência de dados aplicada à saúde, mas também para o desenvolvimento de estratégias práticas e locais que beneficiam a população atendida pelo HU-FURG. Os resultados evidenciam a importância de continuar investindo na coleta e análise de dados regionais, bem como no aprimoramento das técnicas de machine learning, a fim de aumentar a precisão e a confiabilidade das previsões e, consequentemente, melhorar o atendimento médico e o prognóstico dos pacientes.

9 REFERÊNCIAS

- AGÊNCIA Nacional de Vigilância Sanitária - Anvisa. **Anvisa aprova por unanimidade uso emergencial das vacinas**, 2021. Disponível em: <<https://www.gov.br/anvisa/pt-br/assuntos/noticias-anvisa/2021/anvisa-aprova-por-unanimidade-uso-emergencial-das-vacinas#:~:text=A%20Diretoria%20Colegiada%20da%20Anvisa,Institute%20of%20India%20C%20em%20parceria>>. Acesso em: 12 out. 2023.
- ALKODAYMI, M. S. et al. Prevalence of post-acute COVID-19 syndrome symptoms at different follow-up periods: a systematic review and meta-analysis. **Pubmed**, 2022. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/35124265/>>. Acesso em: 09 dez. 2024.
- ALVES, J. E. D. O avanço da pandemia de Covid-19 no mundo e no Brasil no mês de março. **EcoDebate**, 2020. Disponível em: <<https://www.ihu.unisinos.br/categorias/597706-o-avanco-da-pandemia-de-covid-19-no-mundo-e-no-brasil-no-mes-de-marco>>.
- ANDRADE, E. C. D. et al. Modelo híbrido para identificação precoce de sequelas pós-Covid-19. **Journal of Ambient Intelligence & Humanized Computing**, 06 fev. 2023.
- ASADI, S. et al. The coronavirus pandemic and aerosols: Does COVID-19 transmit via expiratory particles? **Taylor & Francis Online**, 03 abr. 2020. 635-638. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/02786826.2020.1749229>>.
- AZEVEDO, H. M. J. D. et al. Persistência de sintomas e retorno ao trabalho após hospitalização por COVID-19. **Jornal Brasileiro de Pneumologia**, São Paulo, 19 out. 2022. Disponível em: <<https://www.scielo.br/j/jbpneu/a/Dcy8wVnjrMYq4C58mVRPBQn/?format=pdf&lang=pt>>.
- BERRAR, D. Bayes' Theorem and Naive Bayes Classifier. In: RANGANATHAN, S.; NAKAI, K.; SCHONBACH, C. **Encyclopedia of Bioinformatics and Computational Biology**. [S.l.]: Elsevier, v. 1, 2018. p. 403-412.
- BISHOP, C. M. **Pattern Recognition and Machine learning**. Ney York: Springer Science+Business Media, LLC, 2006.
- BREIMAN, L. **Arcing the Edge - Technical Report 486**, Statistics Department. University of California. Berkeley, p. 14. 1997.
- BREIMAN, L. Random Forests. In: BREIMAN, L. **Machine Learning**. 3. ed. Berkeley: Kluwer Academic Publisher, v. 113, 2001. p. 5-32.
- BREIMAN, L. et al. **Classification and Regression Trees**. 1. ed. New York: Chapman and Hall/CRC, 1984.
- BRITO, S. B. P. et al. Pandemia da COVID-19: o maior desafio do século XXI. **Vigilância Sanitária em Debate**, 28 abr. 2020. 54-63. Disponível em: <<https://www.redalyc.org/journal/5705/570567430007/570567430007.pdf>>.
- CARDOSO, N. P.; MACHADO, R. T. M. Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. **Revista de Administração Pública**, Rio de Janeiro, 05-06 2008. 495-528.
- CARMO, E. H. Emergências de saúde pública: breve histórico, conceitos e aplicações. **Saúde Debate**, Rio de Janeiro, jul. 2020. 9 -19.
- CARVALHO, A. C. P. D. L. F. D. Inteligência Artificial: riscos, benefícios e uso responsável. **Estudos Avançados**, São Paulo, 01-04 2021. 35. Disponível em: <<https://www.scielo.br/j/ea/a/ZnKyrclLVqzhZbXGgXTwDtn/#>>. Acesso em: 07 abr. 2024.
- CHEN, T. Q.; GUESTRIN, C. XGBoost: A Scalable Tree Boosting System, jun. 2016.
- CREPALDI, P. et al. UM ESTUDO SOBRE A ÁRVORE DE DECISÃO E SUA IMPORTÂNCIA NA HABILIDADE DE APRENDIZADO. **Revista eletrônica - Faculdade Integrada Inesul**, Londrina, 2007. Disponível em: <https://www.inesul.edu.br/revista/arquivos/arq-idvol_15_1320100263.pdf>.
- CRODA, J. H. R.; GARCIA, L. P. Resposta imediata da Vigilância em Saúde à epidemia da

- COVID-19. **Scielo Saúde Pública**, 2020. Disponível em: <<https://www.scielo.org/article/ress/2020.v29n1/e2020002/pt/#>>. Acesso em: 07 abr. 2024.
- CURVELO, R. D.; RIBEIRO, A. C.; UEHARA, S. C. D. S. A. Assistência à saúde de pacientes com COVID longa. **Revista da Escola de Enfermagem da USP**, São Carlos - SP, 23 ago. 2024. 2.
- DEEN, F. V. The Neural Network Zoo. **The Asimov Institute**, 2016. Disponível em: <<https://www.asimovinstitute.org/neural-network-zoo/>>. Acesso em: 20 set. 2024.
- DOREMALEN, N. V. et al. Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. **The New England Journal of Medicine**, 382, n. 16, 17 mar. 2020. Disponível em: <<https://portal.fiocruz.br/pergunta/quanto-tempo-o-coronavirus-permanece-ativo-em-diferentes-superficies#:~:text=Um%20estudo%20publicado%20no%20New,reduzindo%20o%20risco%20de%20contamina%C3%A7%C3%A3o.>>. Acesso em: 28 out. 2024.
- DUARTE, R. Métricas de Avaliação em Modelos de Classificação em Machine Learning. **Sigmoidal**, 2023. Disponível em: <<https://sigmoidal.ai/metricas-de-avaliacao-em-modelos-de-classificacao-em-machine-learning/#:~:text=M%C3%A9tricas%20de%20Avalia%C3%A7%C3%A3o%20em%20Modelos%20de%20Classifica%C3%A7%C3%A3o%20em%20Machine%20Learning,-por%20Rafael%20Duarte&text=A%20avalia%C3%A>>. Acesso em: 3 mar. 2024.
- EICH, S. L. et al. Levantamento de Sintomas Persistentes em Pacientes com Covid. **10º Congresso Internacional em Saúde**, 16 a 19 maio 2023.
- FACELI, K. et al. **Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina**. 2. ed. Rio de Janeiro: LTC, 2021. ISBN 9788521618805.
- FALCÃO, J. et al. Redes Neurais Deep Learning com Tensorflow. **RE3C - Revista Eletrônica Científica de Ciência da Computação**, Alfenas, 14, 2014. 5.
- FAN, S. Understanding the mathematics behind Naive Bayes. **Shuzhan Fan**, 2018. Disponível em: <<https://shuzhanfan.github.io/2018/06/understanding-mathematics-behind-naive-bayes/>>. Acesso em: 19 mar. 2024.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **American Association for Artificial Intelligence**, p. 54, 1996. Disponível em: <<https://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>>. Acesso em: 18 set. 2023.
- FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Scielo Brasil**, Universidade de São Paulo, Faculdade de Filosofia Ciências e Letras de Ribeirão Preto, São Paulo, Brazil, abr. 2006. Disponível em: <<https://www.scielo.br/j/ci/a/SQ9myjZWLxnyXfstXMgCdCH/#>>. Acesso em: 20 set. 2024.
- FIOCRUZ CAMPUS VIRTUAL. Curso Autoinstrucional para o Enfrentamento da Covid-19 no Sistema Prisional. **Fiocruz Campus Virtual**, 2020. Disponível em: <<https://mooc.campusvirtual.fiocruz.br/rea/covid-prisional/unidade1-p3.html>>. Acesso em: 16 out. 2023.
- FONTANA, É. **Introdução aos Algoritmos de Aprendizagem Supervisionada**. Universidade Federal do Paraná. Curitiba, p. 6. 2020.
- FONTES, L. C. D. S. F. et al. Impacto da COVID-19 grave na qualidade de vida relacionada com a saúde e a incapacidade: uma perspectiva de follow-up a curto-prazo. **Scielo Brasil**, 2022. Disponível em: <<https://www.scielo.br/j/rbti/a/pCzcbtTHRG8FxN9vCQGcXjP/#>>. Acesso em: 19 out. 2024.
- FRIEDMAN, J. H. Greedy Function Approximation: A Gradient Boosting Machine. **Annals of Statistics**, 19 abr. 2001.
- GARCIA, S. C. **O Uso de Árvores de Decisão na Descoberta de Conhecimento na Área da Saúde**. Universidade Federal do Rio Grande do Sul. Porto Alegre, p. 14. 2003.

- GOLDSCHMIDT, R.; PASSOS, E. **Data Mining um Guia Prático**. Rio de Janeiro: Elsevier Editora Ltda, v. I, 2005.
- GORBALENYA, A. E. et al. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. **PubMed**, 2020. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/32123347/>>. Acesso em: 10 out. 2023.
- GOUVEIA, G. P. D. M. et al. Historical series of cases of COVID-19 in brazilian regions in the year of 2020. **J. Health Biol Sci.** **2020; 8(1)**, 27 set. 2023. 1-7.
- GUSMÃO, A. Rockcontent. **O que é Naive Bayes e como funciona esse algoritmo de classificação**, 2022. Disponível em: <<https://rockcontent.com/br/blog/naive-bayes/>>. Acesso em: 03 nov. 2023.
- HOWARD, J. COVID-19 and the Workplace. **National Institute for Occupational Safety and Health**, 2020. Disponível em: <https://glcohs.uic.edu/wp-content/uploads/sites/463/2020/05/Dr.-HOWARD.COVID_.19.NSC_.30.April_.2020.pdf>. Acesso em: 23 nov. 2024.
- HU, Q.; HE, ; ZHANG,. Community Transmission via Indirect Media-To-Person Route: A Missing Link in the Rapid Spread of COVID-19. **Frontiers in Public Health**, 2021. Disponível em: <<https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2021.687937/full>>. Acesso em: 23 nov. 2024.
- INSTITUTO BUTANTÃ. Conheça os sintomas mais comuns da ômicron e de outras variantes da Covid-19. **Portal do Butantã**, 2021. Disponível em: <<https://butantan.gov.br/noticias/conheca-os-sintomas-mais-comuns-da-omicron-e-de-outras-variantes-da-covid-19>>. Acesso em: 28 out. 2024.
- IWATA, C. K.; ITO,. Limpeza de Dados em Big data: Uma Revisão Bibliométrica. **Simpósio dos Programas de Mestrado Profissional**, São Paulo, 21 nov. 2023. 1262.
- KESSLER, R. et al. Predictive Attributes for Developing Long COVID-A Study Using Machine Learning and Real-World Data from Primary Care Physicians in Germany. **National Library of Medicine**, 17 maio 2023. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/37240616/#:~:text=The%20most%20common%20predictive%20features,as%20well%20as%20cough%20preparations.>>. Acesso em: 31 dez. 2024.
- KHALUSOVA, M. **MACHINE LEARNING MODEL EVALUATION METRICS PART 2: MULTI-CLASS CLASSIFICATION**, 2022. Disponível em: <<https://www.mariakhalusova.com/posts/2019-04-17-ml-model-evaluation-metrics-p2/>>. Acesso em: 02 mar. 2024.
- KOEHRSEN, W. Random Forest Simple Explanation. **Medium**, 2017. Disponível em: <<https://williamkoehrsen.medium.com/random-forest-simple-explanation-377895a60d2d>>. Acesso em: 09 dez. 2023.
- KRISHNAN, S.; NEYAZ, ; QINGZHONG, L. IoT Network Attack Detection using Supervised Machine Learning. **International Journal of Artificial Intelligence and Expert Systems**, 23 jun. 2021. 16.
- LABOISSIÈRE, P. Pesquisa avalia sequelas da covid-19 na população brasileira. **Agência Brasil**, 2024. Disponível em: <<https://agenciabrasil.ebc.com.br/saude/noticia/2024-03/pesquisa-avalia-sequelas-da-covid-19-na-populacao-brasileira>>. Acesso em: 22 abr. 2024.
- LAROSE, D. T.; LAROSE, C. D. **Discovering Knowledge in Data: An Introduction to Data Mining**. 2ª. ed. [S.l.]: [s.n.], 2014.
- LEVIN, M. J. et al. Intramuscular AZD7442 (Tixagevimab-Cilgavimab) for Prevention of Covid-19. **National Library of Medicine**, Massachusetts, 09 jun. 2022. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/35443106/>>.
- LIMA, C. M. A. D. O. Informações sobre o novo coronavírus (COVID-19). **SciELO Brasil**, 2020. Disponível em: <<https://www.scielo.br/j/rb/a/MsJJz6qXfjjpkXg6qVj4Hfj/?lang=pt>>. Acesso em: 12 dez. 2024.

- LIMA, M.; AMORIM, F. Random Forest. **LAMFO**, 2020. Disponível em: <<https://lamfo-unb.github.io/2020/07/08/Random-Forest/>>. Acesso em: 31 out. 2023.
- LUDERMIR, T. B. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. **Scielo Brasil**, 2021. Disponível em: <<https://doi.org/10.1590/s0103-4014.2021.35101.007>>. Acesso em: 28 out. 2023.
- MAIA, C. M.; GOMES, J. C. M.; CHAGAS, D. Estudo Sobre o Uso de Árvores de Decisão na Área da Saúde. **Encontro de Computação do Oeste Potiguar**, Angicos, RN, 2017. 24. Disponível em: <<https://periodicos.ufersa.edu.br/ecop>>.
- MARQUES, F. Pesquisa analisa aspectos da Covid-19 longa no Brasil. **FioCruz**, 2023. Disponível em: <<https://portal.fiocruz.br/noticia/pesquisa-analisa-aspectos-da-covid-19-longa-no-brasil>>. Acesso em: 02 out. 2023.
- MINISTÉRIO DA SAÚDE. Sintomas. **Ministério da Saúde**, 08 abr. 2021. Disponível em: <<https://www.gov.br/saude/pt-br/coronavirus/sintomas#:~:text=Caracterizado%20a%20partir%20da%20presen%C3%A7a,%20C%20fadiga%20e%20Fou%20cefaleia.>>.
- MINISTÉRIO DA SAÚDE. Coronavírus. **Ministério da Saúde**, 2023. Disponível em: <<https://www.gov.br/saude/pt-br/coronavirus>>. Acesso em: 02 out. 2023.
- MIRANDA, D. A. P. D. et al. Long COVID-19 syndrome: a 14-months longitudinal study during the two first epidemic peaks in Southeast Brazil. **Transactions of the Royal Society of Tropical Medicine & Hygiene**, Oxford, 116, 06 maio 2022. 1007-1014. Disponível em: <[https://portal.fiocruz.br/noticia/pesquisa-da-fiocruz-avalia-sindrome-da-covid-longa#:~:text=A%20pesquisa%20acompanhou%20por%2014,OMS\)%20classifica%20como%20Covid%20longa.](https://portal.fiocruz.br/noticia/pesquisa-da-fiocruz-avalia-sindrome-da-covid-longa#:~:text=A%20pesquisa%20acompanhou%20por%2014,OMS)%20classifica%20como%20Covid%20longa.)>. Acesso em: 14 abr. 2024.
- MITCHELL, T. M. **Machine Learning**. New York: McGraw-Hill, 1997.
- MONARD, M. C.; BARANAUSKAS, J. A. **Conceitos sobre Aprendizado de Máquina**. Universidade de São Paulo. São Paulo, p. 89-114. 2003.
- MORAWSKA, L.; CAO, J. Airborne transmission of SARS-CoV-2: The world should face the reality. **Environment International**, 139, jun. 2020. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S016041202031254X>>.
- NUNES, M. D. C. et al. Síndrome da COVID longa: uma revisão integrativa. **Research, Society and Development**, 15 out. 2022.
- O'BRIEN, J. A. **Sistemas de Informação e as decisões gerenciais na era da internet**. 2ª. ed. São Paulo: Saraiva, 2005. 143 p. ISBN 9788502044074. Acesso em: 28 abr. 2024.
- OLIVEIRA, M. M. D. O Uso de Aprendizado de Máquina para Identificar Alunos em Risco de Evasão na Educação a Distância, Rio Grande, 2020. Disponível em: <<https://repositorio.furg.br/bitstream/handle/1/9858/0000013522.pdf?sequence=1&isAllowed=y>>. Acesso em: 24 Julho 2023.
- OMS afirma que COVID-19 é agora caracterizada como pandemia. **OPAS - Organização Pan-Americana da Saúde**, 2020. Disponível em: <<https://www.paho.org/pt/news/11-3-2020-who-characterizes-covid-19-pandemic>>. Acesso em: 14 abr. 2024.
- OPAS - ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE. OMS declara emergência de saúde pública de importância internacional por surto de novo coronavírus. **OPAS - Organização Pan-Americana de Saúde**, 2020. Disponível em: <<https://www.paho.org/pt/news/30-1-2020-who-declares-public-health-emergency-novel-coronavirus>>. Acesso em: 13 out. 2023.
- OPAS - ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE. Declarações sobre vacinas contra a COVID-19 atuais e atualizadas. **OPAS - Organização Pan-Americana de Saúde**, 2022. Disponível em: <<https://www.paho.org/pt/noticias/17-6-2022-declaracoes-sobre-vacinas-contracovid-19-atuais-e-atualizadas>>. Acesso em: 13 out. 2023.
- OPAS - ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE. Rede Colaborativa formada pela

OPAS desenvolve pesquisa para conhecer melhor perfil de pessoas internadas e com sequelas de COVID-19 no Brasil. **OPAS - Organização Pan-Americana de Saúde**, 2022. Disponível em: <<https://www.paho.org/pt/noticias/27-10-2022-rede-colaborativa-formada-pela-opas-desenvolve-pesquisa-para-conhecer-melhor>>. Acesso em: 23 out. 2023.

OPAS - ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE. OMS declara fim da Emergência de Saúde Pública de Importância Internacional referente à COVID-19. **OPAS - Organização Pan-Americana de Saúde**, 2023. Disponível em: <<https://www.paho.org/pt/noticias/5-5-2023-oms-declara-fim-da-emergencia-saude-publica-importancia-internacional-referente>>. Acesso em: 13 out. 2023.

OPAS - ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE. OPAS e OMS debatem COVID longa e impactos das mudanças climáticas na saúde, em sessão do G20 no Brasil. **OPAS - Organização Pan-americana de Saúde**, 2024. Disponível em: <<https://www.paho.org/pt/noticias/6-6-2024-opas-e-oms-debatem-covid-longa-e-impactos-das-mudancas-climaticas-na-saude-em#:~:text=A%20OMS%20define%20a%20condi%C3%A7%C3%A3o,afetam%20tanto%20crian%C3%A7as%20quanto%20adultos.>>. Acesso em: 09 dez. 2024.

OPAS - ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE. Histórico da pandemia de COVID-19. **OPAS - Organização Pan-Americana de Saúde**. Disponível em: <<https://www.paho.org/pt/covid19/historico-da-pandemia-covid-19>>. Acesso em: 10 out. 2023.

PAINEL CORONAVÍRUS. Painel Coronavírus. **CORONAVÍRUS BRASIL**, 2024. Disponível em: <<https://covid.saude.gov.br/>>. Acesso em: 09 dez. 2024.

PAINEL CORONAVÍRUS RS. Painel Coronavírus RS. **Coronavirus - COVID-19**, 2024. Disponível em: <<https://ti.saude.rs.gov.br/covid19/>>. Acesso em: 22 abr. 2024.

PEREIRA, J. C.; CÂMARA, C. E. ALEXA: UMA ANÁLISE DE INTERFACE BASEADA EM VOZ. **Revista Ubiquidade**, Jundiaí - SP, 01 - 07 2021. 17.

PERGUNTAS frequentes: vacinas contra a COVID-19. **Organização Pan-Americana da Saúde**, 2023. Disponível em: <<https://www.paho.org/pt/vacinas-contracovid-19/perguntas-frequentes-vacinas-contracovid-19#:~:text=As%20vacinas%20contra%20a%20COVID%2D19%20recomendadas%20pela%20OMS%20s%C3%A3o,omicron%20e%20suas%20sub%2Dlinhagens.>>. Acesso em: 20 set. 2023.

PESSANHA, C. Random Forest: como funciona um dos algoritmos mais populares de ML. **Medium**, 2019. Disponível em: <<https://medium.com/cinthiabpessanha/random-forest-como-funciona-um-dos-algoritmos-mais-populares-de-ml-cc1b8a58b3b4>>. Acesso em: 20 mar. 2024.

PFAFF, E. R. et al. Identifying who has long COVID in the USA: a machine learning approach using N3C data. **The Lancet Digital Health**, 4, 16 maio 2022. Disponível em: <[https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(22\)00048-6/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(22)00048-6/fulltext)>. Acesso em: 31 dez. 2024.

PIRES, B. M. F. B. et al. QUALIDADE DE VIDA DOS PROFISSIONAIS DE SAÚDE PÓS-COVID-19: UM ESTUDO TRANSVERSAL. **Cogitare Enfermagem**, Curitiba, 2021. Disponível em: <<https://www.scielo.br/j/cenf/a/VBpnyMhyPTdgSqCsgfyyJD/>>.

POHLENZ, V.; JÚNIOR, E. C. V. Decision Trees, Random Forests e Ensemble. **GEAM - Grupo de Estudos em Aprendizagem de Máquina**, 2020. Disponível em: <<chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://geam.paginas.ufsc.br/files/2020/02/decision-tree-ensemble.pdf>>. Acesso em: 20 mar. 2024.

PORTAL DO BUTANTÃ. Por que a variante delta passou a preocupar países já avançados na vacinação? **Portal do Butantã**, 2021. Disponível em: <<https://butantan.gov.br/noticias/por-que-a-variante-delta-passou-a-preocupar-paises-ja-avancados-na-vacinacao-entenda-o-que>>.

- aconteceu-nos-estados-unidos-em-israel-e-no-reino-unido>. Acesso em: 12 out. 2023.
- PORTAL DO BUTANTAN. Saiba quais são os primeiros sintomas da ômicron e quando eles surgem. **Portal do Butantan**, 2022. Disponível em: <<https://butantan.gov.br/noticias/saiba-quais-sao-os-primeiros-sintomas-da-omicron-e-quando-eles-surgem>>. Acesso em: 15 out. 2023.
- PORTAL DO BUTANTAN. Seis fatos sobre a ômicron, a variante mais transmissível da Covid-19. **Portal do Butantan**, 2022. Disponível em: <<https://butantan.gov.br/noticias/seis-fatos-sobre-a-omicron-a-variante-mais-transmissivel-da-covid-19#:~:text=A%20variante%20%C3%B4micron%20do%20v%C3%ADrus,em%2026%2F11%2F2021.>>>. Acesso em: 12 out. 2023.
- PORTAL DO BUTANTAN. Vacinados têm menos risco de desenvolver Covid longa e dose de reforço pode ampliar a proteção, afirma estudo. **Portal do Butantan**, 2023. Disponível em: <<https://butantan.gov.br/noticias/vacinados-tem-menos-risco-de-desenvolver-covid-longa-e-dose-de-reforco-pode-ampliar-a-protecao-afirma-estudo>>. Acesso em: 12 dez. 2024.
- QUINLAN, J. **Induction of Decision Trees**. Boston: Kluwer Academic Publisher, 1986.
- QUINLAN, J. Decision trees and Decision-making. **IEEE Transactions on Systems, Man, and Cybernetics**, mar. 2021. 339-346.
- RAMOS JR., A. N. Desafios da COVID longa no Brasil: uma agenda inacabada para o Sistema Único de Saúde. **Scielo Brasil**, 2024. Disponível em: <<https://www.scielo.br/j/csp/a/xvVfjNQxX4c5MqmRbZDxSRz/?lang=pt>>. Acesso em: 11 dez. 2024.
- RITTER, C. Working Together to Tackle Coronavirus Disease (COVID-19). **WashU Medicine**, 2020. Disponível em: <<https://nephrology.wustl.edu/working-together-to-tackle-coronavirus-disease-covid-19/>>. Acesso em: 01 out. 2023.
- RODRIGUES, V. Métricas de Avaliação: acurácia, precisão, recall... quais as diferenças? **Medium**, 2019. Disponível em: <<https://vitorborbarodrigues.medium.com/m%C3%A9tricas-de-avalia%C3%A7%C3%A3o-acur%C3%A1cia-precis%C3%A3o-recall-quais-as-diferen%C3%A7as-c8f05e0a513c>>. Acesso em: 23 fev. 2024.
- ROKACH, L.; MAIMON, O. Top-Down Induction of Decision Trees Classifiers—A survey. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, v. 35, p. 476-487, 2005.
- ROSA, C. R. M. **Uma metodologia para a descoberta de conhecimento em bases de dados visando a classificação de padrões**. Pontifícia Universidade Católica do Paraná. Curitiba, p. 162. 2017.
- ROZA, S. D. **Aprendizagem de máquina para apoio à tomada de**. Universidade Federal de Santa Catarina. Florianópolis, p. 16. 2016.
- RPART.CONTROL: Control for Rpart Fits. **RDocumentation**. Disponível em: <<https://www.rdocumentation.org/packages/rpart/versions/4.1.19/topics/rpart.control>>. Acesso em: 15 nov. 2023.
- RUSSEL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 2ª. ed. [S.l.]: Prentice Hall, 2003.
- SECOM - RS. Governo do Estado Rio Grande do Sul. **Chegada do inverno reforça necessidade de vacinação contra gripe e covid-19**, 2024. Disponível em: <<https://www.estado.rs.gov.br/chegada-do-inverno-reforca-necessidade-de-vacinacao-contragripe-e-covid-19#:~:text=A%20dissemina%C3%A7%C3%A3o%20de%20v%C3%ADrus%20respirat%C3%B3rios,o%20refor%C3%A7o%20nas%20medidas%20preventivas.>>>. Acesso em: 09 dez. 2024.
- SEGATA, J.; LÖWY,. Covid longa, a pandemia que não terminou. **Horizontes Antropológicos**, Porto Alegre, 09-12 2024. Disponível em:

- <<https://www.scielo.br/j/ha/a/nvbMkWCWBZ9sh5QHwcpjJjc/>>. Acesso em: 14 dez. 2024.
- SILVA, A. A. M.; MINAYO, M. C. D. S.; GOMES, R. Epidemiologia, ciências sociais e políticas de saúde no enfrentamento da COVID-19. **SciELO Brasil**, Junho 2020. Disponível em: <<https://www.scielo.br/j/csc/a/TbBjtnfFVgwHynj9F6SmksB/?lang=pt#>>. Acesso em: 06 dez. 2022.
- SILVA, A. D. S. D. et al. **Tópicos em Virologia**. Rio de Janeiro: Editora FIOCRUZ, 2023.
- SILVA, G. S. D. Avaliação de arquiteturas de Redes Neurais Recorrentes. **UNIVERSIDADE FEDERAL DO AMAPÁ**, 2022. Disponível em: <https://riu.ufam.edu.br/bitstream/prefix/6321/6/TCC_GuilhermeSilva.pdf>. Acesso em: 21 set. 2024.
- SILVA, N.; BRAUER, A. ATIVIDADE FÍSICA E REABILITAÇÃO DA COVID LONGA. **Anais do EVINCI - UniBrasil**, 2022. Disponível em: <<https://portaldeperiodicos.unibrasil.com.br/index.php/anaisevinci/article/view/6522>>. Acesso em: 23 out. 2023.
- SOUZA, A. S. R. et al. Aspectos gerais da pandemia de COVID-19. **Revista Brasileira de Saúde Materno Infantil**, Recife, 09 dez. 2020. Disponível em: <<https://www.scielo.br/j/rbsmi/a/8phGbzmbSsynCQRWjpXJL9m/?lang=pt#ModalTutorss1>>. Acesso em: 20 set. 2023.
- STEINER, M. T. et al. Abordagem de um problema médico por meio do processo de KDD com ênfase à análise exploratória dos dados. **SciELO Brasil**, 2006. Disponível em: <<https://www.scielo.br/j/gp/a/gC9RkgLD8B8FffPsNhBYWkB/#:~:text=O%20processo%20KDD%20refere%2Dse,e%20Data%20Mining%20como%20sin%C3%B4nimos.>>. Acesso em: 11 set. 2023.
- TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introdução ao Data Mining - Mineração de Dados**. Português. ed. Rio de Janeiro: Ciência Moderna, 2009.
- TEMER, M. et al. Planalto. **Presidência da República**, 2018. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm>. Acesso em: 03 fev. 2023.
- THE OHIO STATE UNIVERSITY - WEXNER MEDICAL CENTER. COVID-19: Symptoms, Prevention and What you can do. **The Ohio State University - Wexner Medical Center**, 2024. Disponível em: <<https://wexnermedical.osu.edu/infectious-diseases/covid-19/symptoms-and-prevention>>. Acesso em: 23 nov. 2024.
- THYGESEN, J. et al. COVID-19 trajectories among 57 million adults in England: a cohort study using electronic health records. **The Lancet Digital Health**, 4, 08 jun. 2022. Disponível em: <[https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(22\)00091-7/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(22)00091-7/fulltext)>. Acesso em: 31 dez. 2024.
- TOZZI, M. et al. Notas Recomendação COVID-19. **Secretaria de Estado de Saúde de Minas Gerais**, 2024. Disponível em: <<https://coronavirus.saude.mg.gov.br/blog/27-como-surgiu-o-coronavirus>>. Acesso em: 30 mar. 2024.
- UZUNIAN, A. Coronavírus SARS-CoV-2 e Covid-19. **SciELO Brasil**, 2020. Disponível em: <<https://www.scielo.br/j/jbpml/a/Hj6QN7mmmKC4Q9SNNt7xRh/?lang=pt#>>. Acesso em: 29 mar. 2024.
- VARGAS, A. C. G.; PAES, A.; VASCONCELOS, C. N. Um estudo sobre Redes Neurais Convolucionais e sua aplicação em detecção de pedestres, 2016. Disponível em: <<http://gibis.unifesp.br/sibgrapi16/eproceedings/wuw/7.pdf>>. Acesso em: 20 set. 2024.
- VARÓN, L. S. et al. The role of machine learning in health policies during the COVID-19 pandemic and in long COVID management. **Frontiers in Public Health**, Punta Arenas, 11, 10 abr. 2023. Disponível em: <<https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2023.1140353/full>>. Acesso em: 31 dez. 2023.
- VICTOR, N. VACINA PROTEGE. **Ministério da Saúde**, 2023. Disponível em:

<<https://www.gov.br/saude/pt-br/assuntos/noticias/2023/marco/cinco-motivos-para-voce-se-vacinar-contr-a-covid-19>>. Acesso em: 28 abr. 2024.

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Practical Machine Learning Tools and Techniques**. 3. ed. [S.l.]: Morgan Kaufmann, 2011.

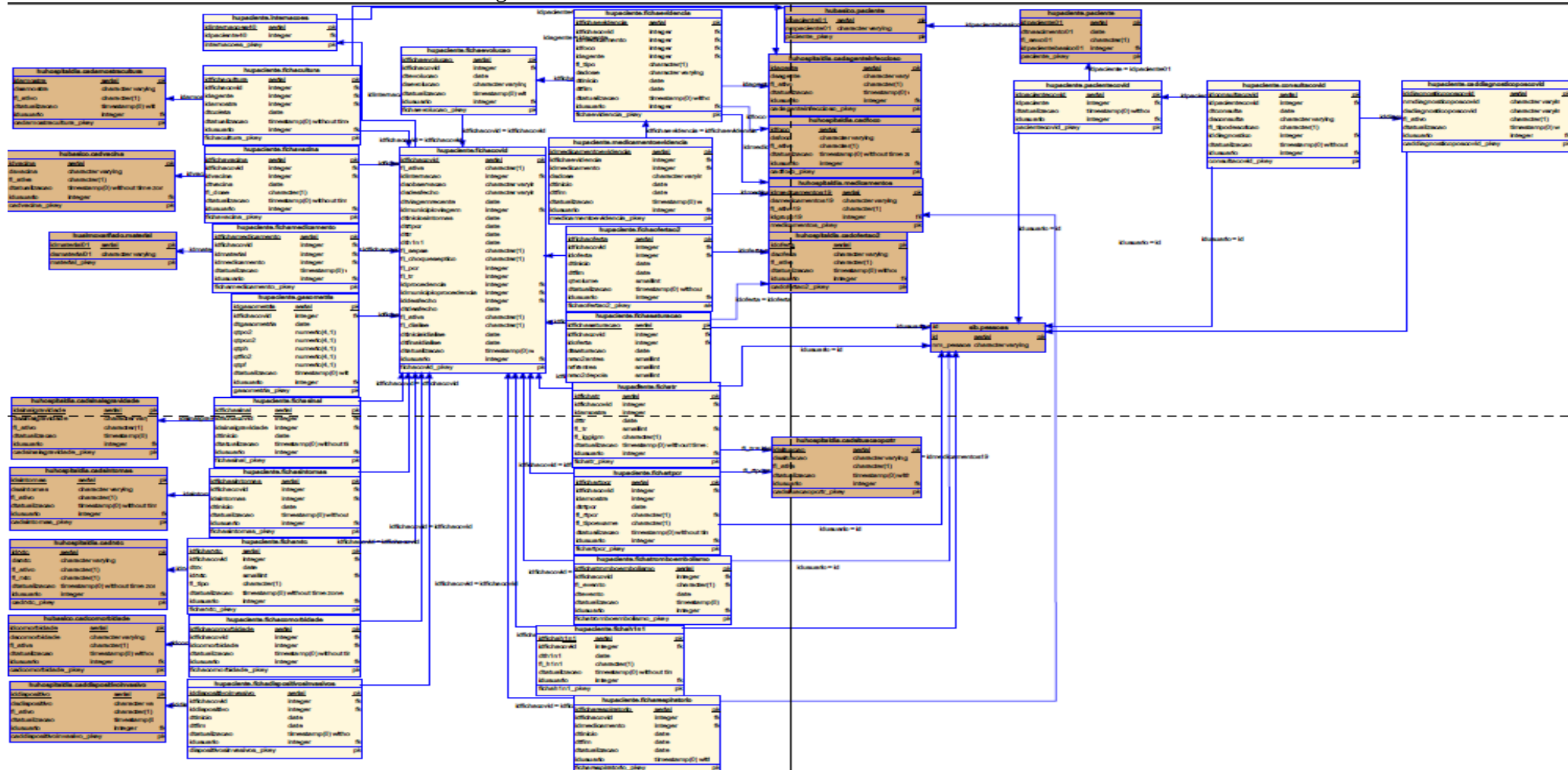
WORLD HEALTH ORGANIZATION. WHO Director-General's opening remarks at the media briefing on COVID-19. **World Health Organization**, 2020. Disponível em:

<<https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>>. Acesso em: 10 out. 2023.

WORLD HEALTH ORGANIZATION. WHO COVID-19 dashboard. **World Health Organization**, 2024. Disponível em: <<https://data.who.int/dashboards/covid19/cases?n=c>>. Acesso em: 10 dez. 2024.

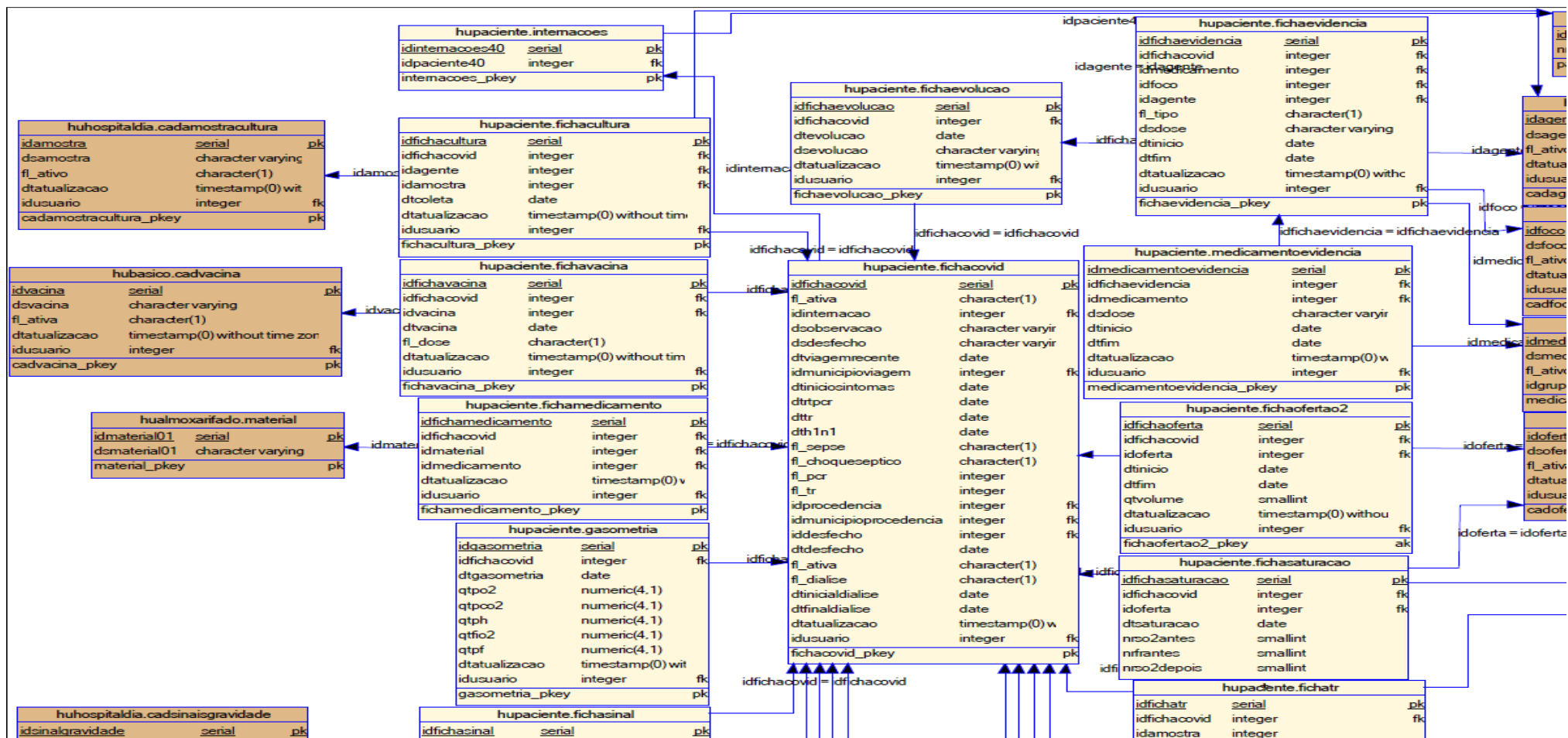
APÊNDICE A – Banco de Dados COVID-19 HU-FURG

Figura 18 - Banco de Dados COVID-19 HU-FURG



Fonte: o próprio autor

Figura 19 - Banco de Dados COVID-19 HU-FURG - Quadrante 1



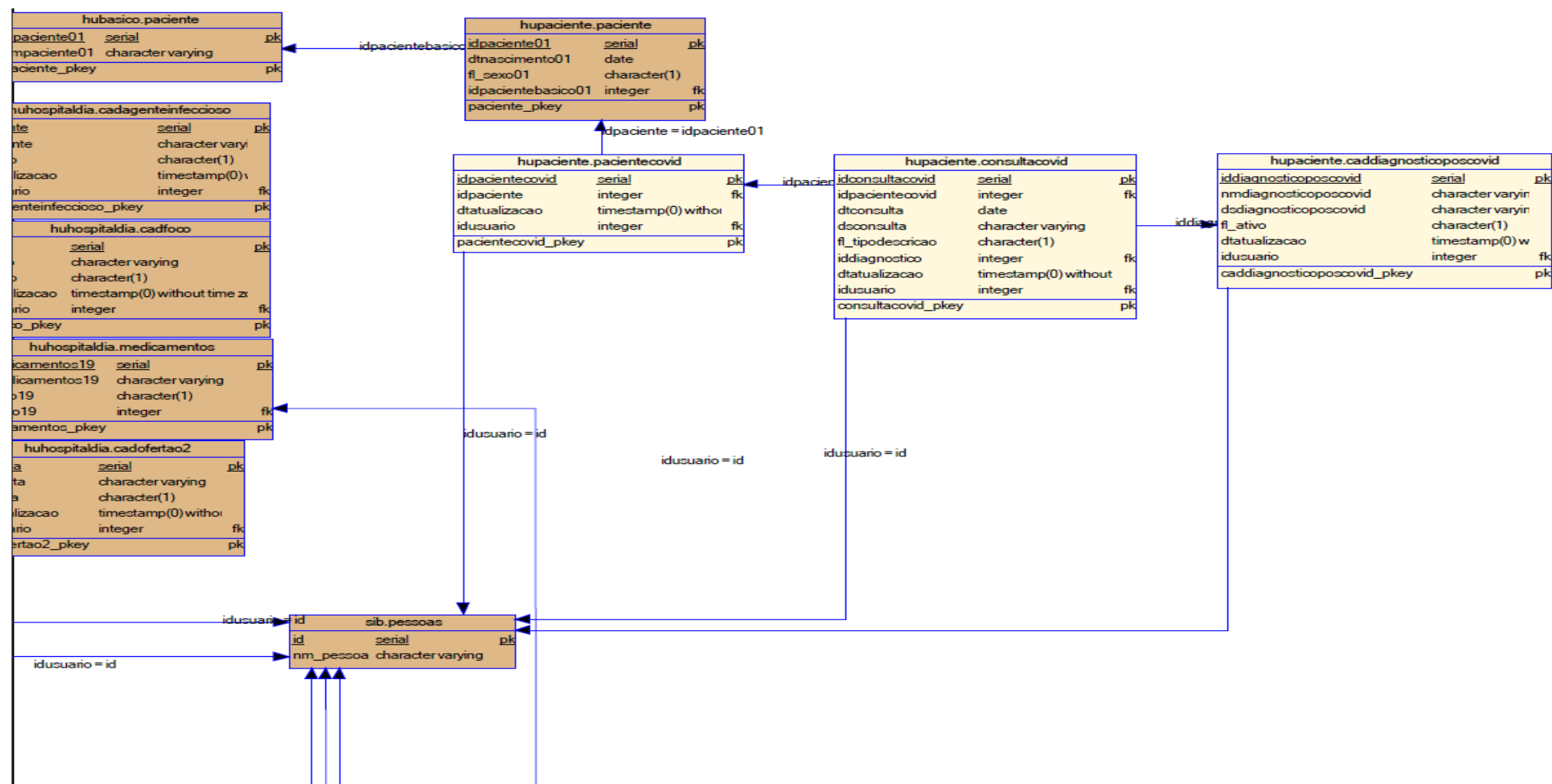
Fonte: o próprio

atomas	integer	fk			
--------	---------	----	--	--	--



anco de dados.

Figura 21 - Banco de Dados COVID-19 HU-FURG - Quadrante 3



Fonte: o próprio autor

APÊNDICE B – Telas de controle de pacientes internados por COVID-19

Figura 22 - Tela principal do sistema Covid do HU-FURG

Internação	Data de internação	Nº Prontuário	Nº Registro	Nome paciente	Data de Nascimento	Cartão SUS	
☐	2249028	01/01/2020	XXXXXX	XXXXXX	JOHN XXXXX XXXXX XXXXX	17/06/1983	709 XXXX XXXX XXXX
☐	2284334	03/05/2020	XXXXXX	XXXXXX	VAGNER XXXXXXXX XXXX	30/01/1997	704 XXXX XXXX XXXX
☐	2285665	13/05/2020	XXXXXX	XXXXXX	FELIPE XXXXXXXX XXXXXXXX	14/05/1980	707 XXXX XXXX XXXX
☐	2287089	25/05/2020	XXXXXX	XXXXXX	CLENI XXXXXXXX XXXXX	27/07/1949	700 XXXX XXXX XXXX
☐	2287259	27/05/2020	XXXXXX	XXXXXX	ARACI XX XXXXX	06/12/1946	704 XXXX XXXX XXXX
☐	2288691	06/06/2020	XXXXXX	XXXXXX	PAULO XXXXXXXX XXXXXXXX XXXXXXXX	11/05/1980	707 XXXX XXXX XXXX
☐	2288727	06/06/2020	XXXXXX	XXXXXX	LIDIANE XXXXX XXXXXXXX	24/03/1994	706 XXXX XXXX XXXX
☐	2288779	07/06/2020	XXXXXX	XXXXXX	JOSE XXXXXXXX XXX XXXXXXXX	14/05/1937	700 XXXX XXXX XXXX
☐	2288962	08/06/2020	XXXXXX	XXXXXX	ZAILOR XXXXXXXX	10/11/1945	700 XXXX XXXX XXXX
☐	2289884	15/06/2020	XXXXXX	XXXXXX	ROSA XXXXXXXX XXXXX XXXXXXXX	30/03/1958	704 XXXX XXXX XXXX
☐	2290306	17/06/2020	XXXXXX	XXXXXX	NERVA XXXXXXXX XXXXXXXX	12/02/1950	706 XXXX XXXX XXXX
☐	2290781	19/06/2020	XXXXXX	XXXXXX	ROZA XX XXXX XXXXXXXX	30/08/1920	702 XXXX XXXX XXXX

Fonte: o próprio autor

Figura 23 - Tela dados do paciente do sistema Ficha Covid-19 do HU-FURG

Fonte: o próprio autor

Figura 24 - Tela dados de comorbidades do paciente do sistema Covid-19 do HU-FURG

Fonte: o próprio autor

Figura 25 - Tela dados de medicações usadas no tratamento do paciente do sistema Covid-19 do HU-FURG

Comorbidades	Medicações em uso	Sintomas referidos	Sinais de Gravidade	Evidência de infecção	Rx Tórax	TC Tórax	Prescrição respiratória	Oferta O2	Dispositivos Invasivos	RT-PCR	Antígeno	TR IGG/IGM	ECG	Cultura	Evolução
Vacinas	Tromboembolismo														

Medicações em uso HDia040.002

+ - =

Medicamento	Medicamento TARV - TB
☐ LOSARTANA POTÁSSICA 25MG LOSARTANA POTÁSSICA, 25 MG, COMPRIMIDO, APRESENTAÇÃO: COMPRIMIDO, REVESTIDO E UNITARIZADO, CÓD INST: 63834, CATMAT: 270786, CÓD EBSERH: EBF01140	
☐ HIDROCLOROTIAZIDA 25MG HIDROCLOROTIAZIDA, 25 MG, COMPRIMIDO, APRESENTAÇÃO: COMPRIMIDO, CÓD INST: 63738, CATMAT: 267674, CÓD EBSERH: EBF01132	
☐	Rifampicina + Isoniazida + Etambutol - RHZE (4 em 1)
☐	DOLUTEGRAVIR

Fonte: o próprio autor

Figura 26 - Tela dados de sintomas referidos do paciente do sistema Covid-19 do HU-FURG

Comorbidades	Medicações em uso	Sintomas referidos	Sinais de Gravidade	Evidência de infecção	Rx Tórax	TC Tórax	Prescrição respiratória	Oferta O2	Dispositivos Invasivos	RT-PCR	Antígeno	TR IGG/IGM	ECG	Cultura	Evolução
Vacinas	Tromboembolismo														

Sintomas referidos HDia040.003

+ - =

Sintomas	Data início
☐ Febre (>= 37,8)	01/05/2020
☐ Cefaleia	
☐ Prostração	

Fonte: o próprio autor

Figura 27 - Tela dados de sinais de gravidade do paciente do sistema Covid-19 do HU-FURG

Comorbidades	Medicações em uso	Sintomas referidos	Sinais de Gravidade	Evidência de infecção	Rx Tórax	TC Tórax	Prescrição respiratória	Oferta O2	Dispositivos Invasivos	RT-PCR	Antígeno	TR IGG/IGM	ECG	Cultura	Evolução
Vacinas	Tromboembolismo														

Sinais de Gravidade HDia040.004

+ - =

Sinal	Data início
☐ SpO2 <95% em ar ambiente	03/05/2020
☐ Movimento paradoxal do abdome (em crianças)	01/05/2020

Fonte: o próprio autor

Figura 28 - Tela dados de evidência de infecção do paciente do sistema Covid-19 do HU-FURG

Comorbidades	Medicações em uso	Sintomas referidos	Sinais de Gravidade	Evidência de infecção	Rx Tórax	TC Tórax	Prescrição respiratória	Oferta O2	Dispositivos Invasivos	RT-PCR	Antígeno	TR IGG/IGM	ECG	Cultura	Evolução
Vacinas	Tromboembolismo														

Evidência de infecção HDia040.005

+ - =

Tipo	Foco	Agente	Medicamentos
☐ Comunitária	Abdominal	Acinetobacter baumannii	
☐ Comunitária	Abdominal	Acinetobacter baumannii	
☐ Comunitária	Abdominal	Acinetobacter baumannii	Antfotericina B
☐ Comunitária	Pulmonar	Streptococcus pneumoniae	Ceftriaxone Dose: 1g
☐ Comunitária	Trato respiratorio superior		Amoxicilina + clavulanato Dose: 1g 8/8 D.L.: 06/07/2020
☐ Comunitária	Pulmonar		Ceftriaxone
☐ Comunitária	Trato respiratorio superior		Vancomicina Dose: 1g de 12/12 e 1 g de 8/8 D.L.: 01/05/2020 D.L.F.: 15/05/2021 Meropenem Dose: 1g de 12/12 e 1 g de 8/8 D.L.: 01/05/2020 D.L.F.: 15/05/2021
☐ Hospitalar	Trato urinário baixo	Candida spp	Meropenem
☐ Hospitalar	Corrente Sanguinea	Staphylococcus aureus	Vancomicina Dose: 1g de 12/12h D.L.: 01/05/2020
☐ Hospitalar	Trato urinário baixo	Neisseria meningitidis	Vancomicina D.L.: 01/05/2020
☐ Hospitalar	Genal	Acinetobacter baumannii	Aminapinas

Fonte: o próprio autor

Figura 29 - Tela dados de RX do tórax do paciente do sistema Covid-19 do HU-FURG

Comorbidades	Medicações em uso	Sintomas referidos	Sinais de Gravidade	Evidência de infecção	Rx Tórax	TC Tórax	Prescrição respiratória	Oferta O2	Dispositivos Invasivos	RT-PCR	Antígeno	TR IGG/IGM	ECG	Cultura	Evolução
Vacinas	Tromboembolismo														

Rx Tórax HDia040.006

+ - =

Data	Alteração
☐ 06/05/2020	Opacidade com halo
☐ 07/05/2020	Infiltrado D
☐ 07/05/2020	Vidro fosco
☐ 06/07/2020	Vidro fosco

Fonte: o próprio autor

Figura 30 - Tela dados de tomografia do tórax do paciente do sistema Covid-19 do HU-FURG

Comorbidades	Medicações em uso	Sintomas referidos	Sinais de Gravidade	Evidência de infecção	Rx Tórax	TC Tórax	Prescrição respiratória	Oferta O2	Dispositivos Invasivos	RT-PCR	Antígeno	TR IGG/IGM	ECG	Cultura	Evolução
Vacinas	Tromboembolismo														

TC Tórax HDia040.007

+ - =

	Data	Alteração
☐	08/05/2020	Infiltrado D
☐	08/05/2020	Última imagem

Fonte: o próprio autor

Figura 31 - Tela dados de prescrição respiratória do paciente do sistema Covid-19 do HU-FURG

Comorbidades	Medicações em uso	Sintomas referidos	Sinais de Gravidade	Evidência de infecção	Rx Tórax	TC Tórax	Prescrição respiratória	Oferta O2	Dispositivos Invasivos	RT-PCR	Antígeno	TR IGG/IGM	ECG	Cultura	Evolução
Vacinas	Tromboembolismo														

Prescrição respiratória HDia040.008

+ - =

	Drugs	Data início	Data fim
☐	Oxetamiv	01/05/2020	06/05/2020
☐	Ivermectina	01/05/2020	
☐	Heparina plena	01/05/2020	03/05/2020
☐	Vasopressor		

Fonte: o próprio autor

Figura 32 - Tela dados de oferta de O2 do paciente do sistema Covid-19 do HU-FURG

Comorbidades	Medicações em uso	Sintomas referidos	Sinais de Gravidade	Evidência de infecção	Rx Tórax	TC Tórax	Prescrição respiratória	Oferta O2	Dispositivos Invasivos	RT-PCR	Antígeno	TR IGG/IGM	ECG	Cultura	Evolução
Vacinas	Tromboembolismo														

Oferta O2 HDia040.009

+ - =

	Oferta	Data início	Data fim	Volume (L/min)
☐	Cafeteir nasal	01/05/2020	03/05/2020	
☐	VM - ventilação	03/05/2020	11/05/2020	

Fonte: o próprio autor

Figura 33 - Tela dados de dispositivos invasivos do paciente do sistema Covid-19 do HU-FURG

Comorbidades	Medicações em uso	Sintomas referidos	Sinais de Gravidade	Evidência de infecção	Rx Tórax	TC Tórax	Prescrição respiratória	Oferta O2	Dispositivos Invasivos	RT-PCR	Antígeno	TR IGG/IGM	ECG	Cultura	Evolução
Vacinas	Tromboembolismo														

Dispositivos Invasivos HDia040.010

+ - =

	Dispositivo	Data início	Data fim
☐	Acesso venoso central	01/05/2020	13/05/2020
☐	Acesso venoso central	14/05/2020	20/05/2020

Fonte: o próprio autor

Figura 34 - Tela dados de RT-PCR do paciente do sistema Covid-19 do HU-FURG

Comorbidades	Medicações em uso	Sintomas referidos	Sinais de Gravidade	Evidência de infecção	Rx Tórax	TC Tórax	Prescrição respiratória	Oferta O2	Dispositivos Invasivos	RT-PCR	Antígeno	TR IGG/IGM	ECG	Cultura	Evolução
Vacinas	Tromboembolismo														

RT-PCR HDia040.011

+ - =

	Situação do PCR	Amostra	Data
☐	Positivo	Mini BAL	01/05/2020

Fonte: o próprio autor

Figura 35 - Tela dados de antígeno do paciente do sistema Covid-19 do HU-FURG

Comorbidades	Medicações em uso	Sintomas referidos	Sinais de Gravidade	Evidência de infecção	Rx Tórax	TC Tórax	Prescrição respiratória	Oferta O2	Dispositivos Invasivos	RT-PCR	Antígeno	TR IGG/IGM	ECG	Cultura	Evolução
Vacinas	Tromboembolismo														

Antígeno HDia040.020

+ - =

	Situação do PCR	Amostra	Data
☐	Positivo	Swab nasofarínge	01/07/2021

Fonte: o próprio autor

Figura 36 - Tela dados de TG IGG/IGM do paciente do sistema Covid-19 do HU-FURG

Comorbidades	Medicações em uso	Sintomas referidos	Sinais de Gravidade	Evidência de infecção	Rx Tórax	TC Tórax	Prescrição respiratória	Oferta O2	Dispositivos Invasivos	RT-PCR	Antígeno	TR IGG/IGM	ECG	Cultura	Evolução
Vacinas		Tromboembolismo													

TR IGG/IGM HDia040 012

+ -

Situação TR	Data
☒ Positivo	13/05/2020
☐ Em andamento	07/05/2020

Fonte: o próprio autor

Figura 37 - Tela dados de ECG do paciente do sistema Covid-19 do HU-FURG

Comorbidades	Medicações em uso	Sintomas referidos	Sinais de Gravidade	Evidência de infecção	Rx Tórax	TC Tórax	Prescrição respiratória	Oferta O2	Dispositivos Invasivos	RT-PCR	Antígeno	TR IGG/IGM	ECG	Cultura	Evolução
Vacinas		Tromboembolismo													

ECG HDia040 013

+ -

Data	Alteração
☒ 01/05/2020	Prolongamento do intervalo QT
☐	

Fonte: o próprio autor

Figura 38 - Tela dados de cultura do paciente do sistema Covid-19 do HU-FURG

Comorbidades	Medicações em uso	Sintomas referidos	Sinais de Gravidade	Evidência de infecção	Rx Tórax	TC Tórax	Prescrição respiratória	Oferta O2	Dispositivos Invasivos	RT-PCR	Antígeno	TR IGG/IGM	ECG	Cultura	Evolução
Vacinas		Tromboembolismo													

Cultura HDia040 014

+ -

Microorganismo	Amostra	Data coleta
☒ Aspergillus flavus	Sangue	14/05/2020
☐ Candida parapsilosis	Urina	01/05/2020

Fonte: o próprio autor

Figura 39 - Tela dados de evolução do paciente do sistema Covid-19 do HU-FURG

Comorbidades	Medicações em uso	Sintomas referidos	Sinais de Gravidade	Evidência de infecção	Rx Tórax	TC Tórax	Prescrição respiratória	Oferta O2	Dispositivos Invasivos	RT-PCR	Antígeno	TR IGG/IGM	ECG	Cultura	Evolução
Vacinas		Tromboembolismo													

Evolução HDia040 018

+ -

Data	Descrição
☒ 21/05/2021	teste T1 2

Fonte: o próprio autor.

Figura 40 - Tela dados de vacinas do paciente do sistema Covid-19 do HU-FURG

Comorbidades	Medicações em uso	Sintomas referidos	Sinais de Gravidade	Evidência de infecção	Rx Tórax	TC Tórax	Prescrição respiratória	Oferta O2	Dispositivos Invasivos	RT-PCR	Antígeno	TR IGG/IGM	ECG	Cultura	Evolução
Vacinas		Tromboembolismo													

Vacinas HDia040 019

+ -

Vacina	Dose	Data
☒ AstraZeneca	2ª dose	21/07/2021
☐ AstraZeneca	1ª dose	19/08/2021

Fonte: o próprio autor

Figura 41 - Tela dados de tromboembolismo do paciente do sistema Covid-19 do HU-FURG

Comorbidades	Medicações em uso	Sintomas referidos	Sinais de Gravidade	Evidência de infecção	Rx Tórax	TC Tórax	Prescrição respiratória	Oferta O2	Dispositivos Invasivos	RT-PCR	Antígeno	TR IGG/IGM	ECG	Cultura	Evolução
Vacinas		Tromboembolismo													

Tromboembolismo HDia040 022

+ -

Data	Evento
☒ 02/08/2021	TEP - tromboembolismo pulmonar
☐ 03/08/2021	TVP - trombose venosa profunda a E

Fonte: o próprio autor

Figura 42 - Tela dados de evolução do paciente do sistema Pós-Covid do HU-FURG

Evolução		
Evolução		
HURegistro057.001		
Data	Descrição	Atualizado por
05/09/2021	Realizo contato em 05/09, conversei com a paciente e oriento sobre consulta em 08/09 as 13:30 no ambulatório de pós covid	IASSANA SOUZA DOS SANTOS

Fonte: o próprio autor

Figura 43 - Tela dados de diagnóstico do paciente do sistema Pós-Covid do HU-FURG

Evolução		Diagnóstico	Tomografia		
Diagnóstico - Inserir		HURegistro057.002			
Data	05/12/2022				
Descrição					
Inserir	Voltar				

Fonte: o próprio autor

Figura 44 - Tela dados de tomografia do paciente do sistema Pós-Covid do HU-FURG

Evolução		Diagnóstico	Tomografia
Tomografia - Inserir		HURegistro057.003	
Data	08/12/2022		
Descrição			

Fonte: o próprio autor

APÊNDICE C – Telas de controle de pacientes pós-covid

Figura 45 - Tela de Evolução pós-covid

Data	Descrição	Atualizado por
08/09/2021	Realizo contato em 05/09, conversei com a paciente e oriento sobre consulta em 08/09 as 13:30 no ambulatório de pós covid	IASSANA SOUZA DOS SANTOS

Fonte: o próprio autor

Figura 46 - Tela de Diagnóstico pós-covid

Diagnóstico - Inserir

Data: 05/12/2022

Descrição:

Inserir Voltar

Fonte: o próprio autor

Figura 47 - Resultado de Tomografia pós-covid

Tomografia - Inserir

Data: 08/12/2022

Descrição:

Inserir Voltar

Fonte: o próprio autor

APÊNDICE D – Dicionário de dados do sistema origem

A Tabela 14 traz a descrição parcial do cadastro de Materiais. Somente os campos relevantes ao levantamento de dados tratado neste trabalho.

Tabela 14- Campos da tabela de cadastro de Materiais

Tabela: Material

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idmaterial01	1	Integer	32		Identificador.
dsmaterial01	3	character varying timestamp without			Descrição do material.
dtinsercao01	4	time zone			Data de inclusão no cadastro.
idusuario01	5	Integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 15 traz a descrição do cadastro de Comorbidades, as quais podem os pacientes estar acometidos.

Tabela 15 – Campos da tabela de cadastro de Comorbidades

Tabela: Cadcomorbidade

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idcomorbidade	1	integer	32		Identificador.
dscorbidade	2	character varying			Descrição da comorbidade.
fl_ativa	3	character timestamp without			Ativa (S/N) no cadastro.
dtatualizacao	4	time zone			Data e horário da última atualização.
Idusuario	5	integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 16 traz a descrição do cadastro de Vacinas.

Tabela 16 - Campos da tabela de cadastro de Vacinas

Tabela: Cadvacina

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
Idvacina	1	Integer	32		Identificador.
Dsvacina	2	character varying			Descrição da vacina.
fl_ativa	3	Character timestamp without			Ativa (S/N) no cadastro.
Dtactualizacao	4	time zone			Data e horário da última atualização.
Idusuario	5	Integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 17 traz a descrição do cadastro de Pacientes.

Tabela 17 - Campo da tabela base de cadastro de Pacientes

Tabela: hubasico.paciente

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idpaciente01	1	Integer	32		Identificador.
nmpaciente01	2	character varying			Nome do Paciente.
nrregistro01	3	Integer	32		Nº de Registro Interno.
nrprontuario01	4	Integer	32		Nº do Prontuário.
fl_sexo01	5	Character			Sexo – Masculino ou Feminino.
dtnascimento01	6	Date			Data de Nascimento.

nmmae01	7	character varying		Nome da Mãe
nmpai01	8	character varying		Nome do Pai
		timestamp without		
dtatualizacao01	9	time zone		Data de atualização do cadastro.
dsprofissao01	10	character varying		Profissão.
nrcartaosus01	11	character varying		Nº do Cartão SUS
dsnaturalidade01	12	character varying		Naturalidade.
dsendereco01	13	character varying		Endereço.
dsbairro01	14	character varying		Bairro.
dscidade01	15	character varying		Cidade.
dstelefone01	16	character varying		Nº de telefone
nrcep01	17	Integer	32	CEP de residência
dsuf01	18	Character		Sigla do Estado
nrcpf01	19	Character		CPF
nrci01	20	Character		Nº da Carteira de Identidade
nmsocialpaciente01	21	character varying		Nome Social
dtobito01	22	Date		Data do Óbito
idescolaridade01	23	Integer	32	Identificador de Escolaridade
idraca01	24	Integer	32	Identificador de Raça
fl_ativo01	31	Character		Ativo (S/N) no cadastro.
dsemail01	32	character varying		E-mail

Fonte: o próprio autor

A Tabela 18 traz a descrição do cadastro de Agente Infeccioso.

Tabela 18 - Campos da tabela de cadastro de Agente Infeccioso

Tabela: Cadagenteinfeccioso

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
Idagente	1	Integer	32		Identificador.
Dsagente	2	character varying			Descrição do agente infeccioso.
		timestamp without			
dtatualizacao	3	time zone			Data e horário da última atualização.
Idusuario	4	Integer	32		Usuário que fez a última atualização.
fl_ativo	5	Character			Ativo (S/N) no cadastro.

Fonte: o próprio autor

A Tabela 19 traz a descrição do cadastro de Amostras de Cultura.

Tabela 19 - Campos da tabela de cadastro de Amostras de Cultura

Tabela: Cadamostracultura

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idamostra	1	Integer	32		Identificador.
dsamostra	2	character varying			Descrição da amostra.
		timestamp without			
dtatualizacao	3	time zone			Data e horário da última atualização.
Idusuario	4	Integer	32		Usuário que fez a última atualização.
fl_ativo	5	Character			Ativo (S/N) no cadastro.

Fonte: o próprio autor

A Tabela 20 traz a descrição do cadastro de Dispositivos Invasivos.

Tabela 20 - Campos da tabela de cadastro de Dispositivos Invasivos

Tabela: Caddispositivoinvasivo

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
--------------	-------------	-------------	-------------	-------------	------------------

iddispositivo	1	Integer	32	Identificador.
dsdispositivo	2	character varying		Descrição do dispositivo.
fl_ativo	3	Character timestamp without		Ativa (S/N) no cadastro.
dtatualizacao	4	time zone		Data e horário da última atualização.
idusuario	5	Integer	32	Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 21 traz a descrição do cadastro de Focos.

Tabela 21 - Campos da tabela de cadastro de Focos

Tabela: Cadfoco

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
Idfoco	1	Integer	32		Identificador.
Dsfoco	2	character varying			Descrição do foco.
fl_ativo	3	Character timestamp without			Ativo (S/N) no cadastro.
dtatualizacao	4	time zone			Data e horário da última atualização.
idusuario	5	Integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 22 traz a descrição do cadastro de Oferta de O2.

Tabela 22 - Campos da tabela de cadastro de Oferta de O2

Tabela: Cadofertao2

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
Idoferta	1	Integer	32		Identificador.
Dsoferta	2	character varying			Descrição da oferta.
fl_ativa	3	Character timestamp without			Ativo (S/N) no cadastro.
dtatualizacao	4	time zone			Data e horário da última atualização.
Idusuario	5	Integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 23 traz a descrição do cadastro de Exames de Raio-X e Tomografia Computadorizada. Ambos são gravados no mesmo cadastro, diferenciados pelo identificador *fl_rxtc*.

Tabela 23 - Campos da tabela de cadastro de Exames de Raio-x/Tomografia

Tabela: Cadrxtc

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
Idrxtc	1	Integer	32		Identificador.
Dsrxtc	2	character varying			Descrição do exame.
fl_ativo	3	character varying			Ativo (S/N) no cadastro.
fl_rxtc	4	Character timestamp without			R - Raio-x; T - Tomografia computadorizada; A – ambos
dtatualizacao	5	time zone			Data e horário da última atualização.
Idusuario	6	Integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 24 traz a descrição do cadastro de Sinais de Gravidade.

Tabela 24 - Campos da tabela de cadastro de Sinais de Gravidade

Tabela: Cadsinaisgravidade

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idsinalgravidade	1	Integer	32		Identificador.
dssinalgravidade	2	character varying			Descrição do sinal de gravidade.
fl_ativo	3	Character timestamp without			Ativo (S/N) no cadastro.
dtatualizacao	4	time zone			Data e horário da última atualização.
idusuario	5	Integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 25 traz a descrição do cadastro de Sintomas.

Tabela 25 - Campos da tabela de cadastro de Sintomas

Tabela: Cadsintomas

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idsintomas	1	Integer	32		Identificador.
dssintomas	2	character varying			Descrição do sintoma.
fl_ativo	3	character varying timestamp without			Ativo (S/N) no cadastro.
dtatualizacao	4	time zone			Data e horário da última atualização.
idusuario	5	Integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 26 traz a descrição do cadastro de Situação de Exames.

Tabela 26 - Campos da tabela de cadastro de Situação de Exames

Tabela: Cadsituacaopcrtr

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idsituacao	1	Integer	32		Identificador.
dssituacao	2	character varying			Descrição da situação.
fl_ativa	3	Character timestamp without			Ativo (S/N) no cadastro.
DtAtualizacao	4	time zone			Data e horário da última atualização.
Idusuario	5	Integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 27 traz a descrição do cadastro de Medicamentos.

Tabela 27 - Campos da tabela de cadastro de Medicamentos

Tabela: Medicamentos

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idmedicamentos19	1	Integer	32		Identificador.
dsmedicamentos19	2	Character			Descrição do medicamento.
fl_ativo19	3	Character			Ativo (S/N) no cadastro.
idgrupo19	4	Integer	32		Identificador do grupo de medicamentos.

Fonte: o próprio autor

A Tabela 28 traz a descrição do cadastro de Diagnósticos Pós-Covid.

Tabela 28 - Campos da tabela de cadastro de Diagnósticos Pós-Covid

Tabela: Caddiagnosticoposccovid

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
Iddiagnosticoposccovid	1	integer	32		Identificador.
Nmdiagnosticoposccovid	2	character varying			Nome do diagnóstico.
Dsdiagnosticoposccovid	3	character varying			Descrição do diagnóstico.
fl_ativo	4	character timestamp without			Ativo (S/N) no cadastro.
DtAtualizacao	5	time zone			Data e horário da última atualização.
Idusuario	6	integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 29 traz a descrição da tabela de Consultas do paciente.

Tabela 29 - Campos da tabela de Consultas do paciente

Tabela: Consultacovid

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idconsultacovid	1	integer	32		Identificador.
idpacientecovid	2	integer	32		Identificador do paciente.
dtconsulta	3	date			Data da consulta.
dsconsulta	4	character varying			Descrição da consulta.
fl_tipodescricao	5	character timestamp without			E - Evolucao; D - Diagnostico; T – Tomografia
dtAtualizacao	6	time zone			Data e horário da última atualização.
idusuario	7	integer	32		Usuário que fez a última atualização.
iddiagnostico	8	integer	32		Identificador do diagnóstico padrão.

Fonte: o próprio autor

A Tabela 30 traz a descrição do cadastro de Comorbidades.

Tabela 30 - Campos da tabela de Comorbidades

Tabela: Fichacomorbidade

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idfichacomorbidade	1	integer	32		Identificador.
idfichacovid	2	integer	32		Identificador da ficha covid.
idcomorbidade	3	integer timestamp without	32		Identificador da comorbidade.
dtAtualizacao	4	time zone			Data e horário da última atualização.
Idusuario	5	integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 31 traz a descrição da Ficha de Covid-19.

Tabela 31 - Campos da tabela da Ficha de Covid-19

Tabela: Fichacovid

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idfichacovid	1	integer	32		Identificador.
idinternacao	2	integer	32		Identificador da internação.
dsobservacao	3	character varying			Observação
dsdesfecho	5	character varying			Descrição do desfecho.
dtviagemrecente	6	date			Data de viagem recente do paciente.

idmunicipioviagem	7	integer	32	Identificador do município destino da viagem.
dtiniciosintomas	8	date		Data de início dos sintomas.
dtrtpcr	9	date		Data do exame RT-PCR.
dttr	10	date		
dth1n1	11	date		Data de contaminação com H1N1.
		timestamp without		
dtatualizacao	12	time zone		Data e horário da última atualização.
Idusuario	13	integer	32	Usuário que fez a última atualização.
fl_sepse	14	character		Teve Sepse? (S/N)
fl_choqueseptico	15	character		Teve choque séptico (S/N)
fl_viaaereadificil	17	character		Via aérea difícil?
fl_pcr	18	integer	32	Não utilizado
fl_tr	19	integer	32	Não utilizado
idprocedencia	20	integer	32	Identificado de procedência
				Identificador do município de procedência.
idmunicipioprocendencia	21	integer	32	Identificador do município de procedência.
iddesfecho	22	integer	32	Identificador do desfecho.
dtdesfecho	23	date		Data do desfecho.
fl_ativa	24	character		Ativo (S/N) no cadastro.
fl_dialise	25	character		Fez diálise (S/N)
dtinicialdialise	26	date		Data de início das diálises.
dtfinaldialise	27	date		Data de fim das diálises.

Fonte: o próprio autor

A Tabela 32 traz a descrição da tabela de Culturas do paciente.

Tabela 32 - Campos da tabela de Culturas do paciente

Tabela: Fichacultura

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idfichacultura	1	Integer	32		Identificador.
idfichacovid	2	Integer	32		Identificador da ficha covid.
idagente	3	Integer	32		Identificador do agente.
idamostra	4	Integer	32		Identificador da amostra.
dtcoleta	5	Date			Data da coleta.
		timestamp without			
dtatualizacao	6	time zone			Data e horário da última atualização.
idusuario	7	Integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 33 traz a descrição da tabela dos Dispositivos Invasivos utilizados no tratamento do paciente.

Tabela 33 - Campos da tabela de Dispositivos Invasivos

Tabela: Fichadispositivosinvasivos

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
iddispositivoinvasivo	1	integer	32		Identificador.
idfichacovid	2	integer	32		Identificador da ficha covid.
iddispositivo	3	integer	32		Identificador do dispositivo.
dtinicio	4	date			Data de início de utilização.
dtfim	5	date			Data da fim de utilização.
		timestamp without			
dtatualizacao	6	time zone			Data e horário da última atualização.
idusuario	7	integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 34 traz a descrição da tabela de Evidências de Infecção e o tratamento para combatê-la.

Tabela 34 - Campos da tabela de Evidências de Infecção

Tabela: Fichaevidencia

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idfichaevidencia	1	integer	32		Identificador.
Idfichacovid	2	integer	32		Identificador da ficha covid.
idmedicamento	3	integer	32		Identificador do medicamento.
Idfoco	4	integer	32		Identificador do foco.
Idagente	5	integer	32		Identificador do agente.
fl_tipo	6	character			B - bacteriana; H – hospitalar
Dsdose	7	character varying			Descrição da dose.
Dtinicio	8	date			Data de início.
Dtfim	9	date			Data de fim.
		timestamp without time			Data e horário da última atualização.
Dtactualizacao	10	zone			Usuário que fez a última atualização.
Idusuario	11	integer	32		

Fonte: o próprio autor

A Tabela 35 traz a descrição da tabela de Evolução do paciente.

Tabela 35 - Campos da tabela de Evolução do paciente

Tabela: Fichaevolucão

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idfichaevolucão	1	Integer	32		Identificador.
idfichacovid	2	Integer	32		Identificador da ficha covid.
dtevolucão	4	Date			Data da evolução.
dsevolucão	5	character varying timestamp without time			Descrição da evolução.
dtactualizacao	6	zone			Data e horário da última atualização. Usuário que fez a última atualização.
idusuario	7	Integer	32		

Fonte: o próprio autor

A Tabela 36 traz a descrição da tabela de Infecção por H1N1.

Tabela 36 - Campos da tabela de infecção por H1N1

Tabela: Fichah1n1

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idfichah1n1	1	Integer	32		Identificador.
idfichacovid	2	Integer	32		Identificador da ficha covid.
dth1n1	3	Date timestamp without time zone			Data de infecção por H1N1.
dtactualizacao	4	time zone			Data e horário da última atualização.
idusuario	5	Integer	32		Usuário que fez a última atualização.
fl_h1n1	6	Character			Identificador de H1N1.

Fonte: o próprio autor

A Tabela 37 traz a descrição da tabela de Medicamentos utilizados no tratamento do paciente.

Tabela 37 - Campos da tabela de Medicamentos do paciente

Tabela: Fichamedicamento

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
Idfichamedicamento	1	Integer	32		Identificador.
Idfichacovid	2	Integer	32		Identificador da ficha covid.
Idmaterial	3	Integer	32		Identificador de material.
dtatualizacao	4	timestamp without time zone			Data e horário da última atualização.
Idusuario	5	Integer	32		Usuário que fez a última atualização.
idmedicamento	6	Integer	32		Identificador de medicamento.

Fonte: o próprio autor

A Tabela 38 traz a descrição da tabela de Oferta de O2 ao paciente.

Tabela 38 - Campos da tabela de Oferta de O2

Tabela: Fichaofer tao2

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idfichaofer ta	1	Integer	32		Identificador.
idfichacovid	2	Integer	32		Identificador da ficha covid.
idoferta	3	Integer	32		Identificador de oferta.
dtinicio	4	Date			Data de início de utilização.
dtfim	5	Date			Data de fim de utilização.
dtatualizacao	6	timestamp without time zone			Data e horário da última atualização.
idusuario	7	Integer	32		Usuário que fez a última atualização.
qtvolume	8	Smallint	16		Quantidade utilizada.

Fonte: o próprio autor

A Tabela 39 traz a descrição da tabela de prescrição de medicamentos para o tratamento respiratório.

Tabela 39 - Campos da tabela de Prescrição Respiratória

Tabela: Ficharespiratorio

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idficharespiratorio	1	Integer	32		Identificador.
idfichacovid	2	Integer	32		Identificador da ficha covid.
idmedicamento	3	Integer	32		Identificador do medicamento.
Dtinicio	4	Date			Data de início de utilização.
Dtfim	5	Date			Data de fim de utilização.
dtatualizacao	6	timestamp without time zone			Data e horário da última atualização.
Idusuario	7	Integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 40 traz a descrição da tabela de Exames RT-PCR ou Antígeno e seu resultado. A identificação do exame se dá pelo identificador *fl_tipoexame*.

Tabela 40 - Campos da tabela de Exames RT-PCR / Antígeno

Tabela: Fichartpcr

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
--------------	-------------	-------------	-------------	-------------	------------------

Idfichartpcr	1	Integer	32	Identificador.
Idfichacovid	2	Integer	32	Identificador da ficha covid.
				1 - Swab nasofaringe; 2 - Escarro; 3 - LBA; 4 - Aspirado traqueal; 5 - Mini BAL; 6 - Outro; 7 - Aspirado de nasofaringe.
Idamostra	3	Integer	32	Data do exame.
Dtrtpcr	4	Date timestamp without time zone		Data e horário da última atualização.
Dtactualizacao	5	Integer	32	Usuário que fez a última atualização.
Idusuario	6	Smallint	16	Identificador de situação.
fl_rtpcr	7	Character		R - RT-PCR; A - Antígeno.
fl_tipoexame	8			

Fonte: o próprio autor

A Tabela 41 traz a descrição da tabela de Exames Raio-X ou Tomografia Computadorizada. A identificação do exame se dá pelo identificador *fl_tipo*.

Tabela 41 - Campos da tabela de Exames de Raio-X / Tomografia Computadorizada

Tabela: Ficharxtc

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
Idficharxtc	1	Integer	32		Identificador.
Idfichacovid	2	Integer	32		Identificador da ficha covid.
Dtrx	3	Date			Data do exame.
Idrxtc	4	Smallint	16		Identificador do exame.
fl_tipo	5	Character			R - Raio-x; T - Tomografia computadorizada.
		timestamp without time zone			
Dtactualizacao	6	Integer	32		Data e horário da última atualização.
Idusuario	7	Integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 42 traz a descrição da tabela de Saturação de O2.

Tabela 42 - Campos da tabela de Saturação de O2

Tabela: Fichasaturacao

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idfichasaturacao	1	Integer	32		Identificador.
idfichacovid	2	Integer	32		Identificador da ficha covid.
Idoferta	3	Integer	32		Identificador de oferta de O2.
dtsaturacao	4	Date			Data de verificação
nrso2antes	5	Smallint	16		Saturação de O2 antes.
nrfrantes	6	Smallint	16		
nrso2depois	7	Smallint	16		Saturação de O2 depois.
nrfrdepois	8	Smallint	16		
		timestamp without time zone			
dtatualizacao	9	Integer	32		Data e horário da última atualização.
Idusuario	10	Integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 43 traz a descrição da tabela de Sinais de Gravidade do paciente.

Tabela 43 - Campos da tabela de Sinais de Gravidade do paciente

Tabela: Fichasinal

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idfichasinal	1	Integer	32		Identificador.
idfichacovid	2	Integer	32		Identificador da ficha covid.
idsinalgravidade	3	Integer	32		Identificador de sinal de gravidade.
dtatualizacao	4	timestamp without time zone			Data e horário da última atualização.
idusuario	5	Integer	32		Usuário que fez a última atualização.
dtinicio	6	Date			Data de início

Fonte: o próprio autor

A Tabela 44 traz a descrição da tabela de Sintomas do paciente.

Tabela 44 - Campos da tabela de Sintomas do paciente

Tabela: Fichasintomas

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idfichasintomas	1	Integer	32		Identificador.
idfichacovid	2	Integer	32		Identificador da ficha covid.
idsintomas	3	Integer	32		Identificador do sintoma.
dtatualizacao	4	timestamp without time zone			Data e horário da última atualização.
idusuario	5	Integer	32		Usuário que fez a última atualização.
dtinicio	6	Date			Data de início

Fonte: o próprio autor

A Tabela 45 traz a descrição da tabela de Teste Rápido IgG/IgM e seu resultado.

Tabela 45 - Campos da tabela de Teste Rápido de IgG/IgM

Tabela: Fichatr

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
Idfichatr	1	Integer	32		Identificador.
Idfichacovid	2	Integer	32		Identificador da ficha covid.
Idamostra	3	Integer	32		Identificador da amostra.
Dttr	4	Date			Data do teste
fl_tr	5	Smallint	16		Identificador da Situação.
Dtactualizacao	6	timestamp without time zone			Data e horário da última atualização.
Idusuario	7	Integer	32		Usuário que fez a última atualização.
fl_iggigm	8	Character			+ - Positivo; - - Negativo

Fonte: o próprio autor

A Tabela 46 traz a descrição da tabela de Tromboembolismo do paciente.

Tabela 46 - Campos da tabela de Tromboembolismo do paciente

Tabela: Fichatromboembolismo

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idfichatromboembolismo	1	Integer	32		Identificador.
Idfichacovid	2	Integer	32		Identificador da ficha covid.
fl_evento	3	Character			1 - TEP- tromboembolismo pulmonar 2 - TVP - trombose venosa profunda a D 3 - TVP - trombose venosa profunda a E 4 - Sistema nervoso central
Dtevento	4	Date			Data.
dtatualizacao	5	timestamp without time zone			Data e horário da última atualização.

		time zone		
Idusuario	6	Integer	32	Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 47 traz a descrição da tabela de Vacinas que o paciente tomou, contendo a vacina, a data de aplicação e a dose aplicada.

Tabela 47 - Campos da tabela de Vacinas do paciente

Tabela: Fichavacina

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idfichavacina	1	Integer	32		Identificador.
idfichacovid	2	Integer	32		Identificador da ficha covid.
ldvacina	3	Integer	32		Identificador da vacina.
Dtvacina	4	Date			Data da vacinação.
fl_dose	5	Character			1 - 1ª dose; 2 - 2ª dose; 3 - 3ª dose; R - Reforço.
dtatualizacao	6	timestamp without time zone			Data e horário da última atualização.
Idusuario	7	Integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 48 traz a descrição da tabela de Gasometria do paciente.

Tabela 48 - Campos da tabela de Gasometria do paciente

Tabela: Gasometria

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idgasometria	1	Integer	32		Identificador.
idfichacovid	2	Integer	32		Identificador da ficha covid.
dtgasometria	3	Date			
qtpo2	4	Numeric	4	1	
qtpco2	5	Numeric	4	1	
Qtph	6	Numeric	4	1	
qt fio2	7	Numeric	4	1	
Qtpf	8	Numeric	4	1	
dtatualizacao	9	timestamp without time zone			Data e horário da última atualização.
Idusuario	10	Integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor

A Tabela 49 traz a descrição da tabela da Internação do paciente.

Tabela 49 - Campos da tabela de Internação do paciente

Tabela: Internações

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idinternacoes40	1	Integer	32		Identificador.
idpaciente40	2	Integer	32		Identificador do paciente.
dtinternacao40	3	Date			Data da internação.
dtalta40	4	Date			Data de alta.
idlocal40	5	Integer	32		Identificador do local de internação.
idleito40	6	Integer	32		Identificador do leito.
dtatualizacao40	7	timestamp without time zone			Última atualização.

Fonte: o próprio autor

A Tabela 50 traz a descrição da tabela de medicamentos usados no tratamento de infecções, contendo o medicamento, data inicial e final de utilização e a dosagem utilizada.

Tabela 50 - Campos da tabela de medicamentos de Evidência de Infecção

Tabela: Medicamentoevidencia

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idmedicamentoevidencia	1	Integer	32		Identificador.
idfichaevidencia	2	Integer	32		Identificador da ficha covid.
idmedicamento	3	Integer	32		Identificador do medicamento.
		timestamp without			
Dtactualizacao	4	time zone			Data e horário da última atualização.
Idusuario	5	Integer	32		Usuário que fez a última atualização.
Dsdose	6	character varying			Dosagem utilizada.
Dtinicio	7	Date			Data de início de uso.
Dtfim	8	Date			Data de fim de uso.

Fonte: o próprio autor

A Tabela 51 traz a descrição da tabela de dados do paciente.

Tabela 51 - Campos da tabela de paciente

Tabela: Paciente

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
idpaciente01	1	Integer	32		Identificador.
nmpaciente01	2	character varying			Nome do paciente.
dtnascimento01	3	Date			Data de nascimento.
nrprontuario01	4	Integer	32		Número do prontuário
nrregistro01	5	Integer	32		Número de registro.
nmmae01	6	character varying			Nome da mãe.
fl_correto01	7	Character			Indicador de dados corretos (S/N)
nmsocialpaciente01	8	character varying			Nome social.
fl_sexo01	9	Character			Sexo.
nrcartaosus01	10	character varying			Número do cartão SUS
idpacientebasico01	17	Integer	32		Identificador do cadastro geral.
fl_prontuariosame	18	Character			Possui prontuário no SAME?
nrprontuariodv01	19	Integer	32		Dígito verificador do prontuário.

Fonte: o próprio autor

A Tabela 52 traz a descrição da tabela de dados do paciente covid; especialização da tabela de paciente.

Tabela 52 - Campos da tabela de paciente covid

Tabela: Pacientecovid

Campo	Seq.	Tipo	Tam.	Dec.	Descrição
Idpacientecovid	1	Integer	32		Identificador.
Idpaciente	2	Integer	32		Identificador do paciente.
		timestamp without			
Dtactualizacao	3	time zone			Data e horário da última atualização.
Idusuario	4	Integer	32		Usuário que fez a última atualização.

Fonte: o próprio autor