

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

ARATEUS CRYSHAM FRANCO MENESES

**Predição de Necessidade de UTI Hospitalar
para Pacientes Internados Utilizando
Métodos de Aprendizado de Máquina**

Monografia apresentada como requisito parcial
para a obtenção do grau de Bacharel em Ciência
da Computação

Orientador: Profa. Dra. Carla Maria Dal Sasso
Freitas

Co-orientador: Prof. Dr. Rodrigo Pires dos Santos

Porto Alegre
2021

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos André Bulhões

Vice-Reitora: Prof^a. Patricia Helena Pranke

Pró-Reitora de Graduação: Prof^a. Cíntia Inês Boll

Diretora do Instituto de Informática: Prof^a. Carla Maria Dal Sasso Freitas

Coordenador do Curso de Ciência de Computação: Prof. Rodrigo Machado

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

AGRADECIMENTOS

Gostaria de começar agradecendo a minha família, especialmente a meus pais pelo carinho, incentivo e suporte durante toda a minha jornada até o presente momento.

Um agradecimento especial a Professora Dra. Carla Maria Dal Sasso Freitas pela atenção, orientação e boa vontade durante o decorrer deste projeto.

Quero agradecer a Qualis e todo o seu experiente time de profissionais da saúde e da TI, em especial ao Prof. Dr. Rodrigo Pires dos Santos que permitiu este projeto acontecer. Pelas oportunidades, fornecimento dos dados e orientação durante o decorrer deste projeto, obrigado.

Agradeço ao André Pocos e Leonardo Sternberg que sempre me apoiaram e me ajudaram, a parceria de vocês sempre foi fundamental nos bons e maus momentos.

Agradeço a Brenda Maciel por toda a companhia e palavras de incentivo, me ajudaram a acreditar que poderia chegar neste momento.

Por fim, gostaria de agradecer ao James Lemmy por toda parceria e instrução durante esta caminhada, obrigado.

RESUMO

O setor da saúde vem testando a contribuição da Inteligência Artificial numa perspectiva de renovação dos processos de gestão hospitalar. Este estudo investigou a possibilidade de algoritmos de Aprendizado de Máquina auxiliarem na predição da admissão de pacientes em Unidade de Terapia Intensiva (UTI), através da análise de suas primeiras 24, 48 e 72 horas de internação. A base de dados utilizada foi de obtida um hospital real, contendo dados estruturados e não estruturados no período entre 01/06/2020 a 31/07/2021, durante pandemia da COVID-19. Os dados correspondem a internações, evoluções, prescrições de antimicrobianos, cirurgias, sinais vitais e exames dos pacientes internados. Os algoritmos utilizados no trabalho foram o *Random Forest* e o *Gradient Boost*. Os modelos que apresentaram as melhores métricas para cada período de internação foram o *Gradient Boost* com dados das primeiras 24 horas com métricas de AUROC de 92.7%, Precisão de 61%, Sensibilidade de 81.6%, Especificidade de 86.1%, Acurácia de 85.1% e F1-Score de 69.9%; o *Gradient Boost* com dados de 48 horas com métricas de AUROC de 91.3%, Precisão de 71.2%, Sensibilidade de 78.1%, Especificidade de 90.5%, Acurácia de 87.6% e F1-Score de 74.4%; e o *Random Forest* com dados das primeiras 72 horas com métricas de AUROC de 92.2%, Precisão de 70.1%, Sensibilidade de 81.4%, Especificidade de 88.7%, Acurácia de 86.9% e F1-Score de 75.4%; O estudo mostrou-se viável como instrumento preditivo para admissão em UTI hospitalar, à semelhança de outros estudos referidos no trabalho.

Palavras-chave: Inteligência Artificial. Aprendizado de máquina. Pré-processamento de dados. Unidade de Tratamento Intensivo.

Hospital ICU prediction using Machine Learning Methods

ABSTRACT

The health sector has been testing the contribution of Artificial Intelligence from the perspective of hospital management's process renovation. This study investigated the possibility of using Machine Learning Algorithms on helping patients admitted to the Intensive Care Unit (ICU), by analyzing their first 24, 48 and 72 hospitalization hours. The database used on this project was from a real hospital, having structured and unstructured data in the interval of 06/01/2020 to 07/31/2021, during the COVID-19 pandemic. The data was composed of patients' admissions, evolutions, antimicrobial prescriptions, surgeries, vital signs and exams. The used algorithms were the Random Forest and the Gradient Boost. The models which had the best performance metrics for each time interval were the Gradient Boost on the first 24 hours of data with 91.3% AUROC, 61% Precision, 81.6% Sensibility, 86.1% Specificity, 85.1% Accuracy and 69.9% F1-Score; the Gradient Boost on the first 48 hours of data with 91.3% AUROC, 71.2% Precision, 78.1% Sensibility, 90.5% Specificity, 87.6% Accuracy and 74.4% F1-Score; and the Random Forest on the first 72 hours of data with 92.2% AUROC, 70.1% Precision, 81.4% Sensibility, 88.7% Specificity, 86.9% Accuracy and 75.4% F1-Score. The study showed the viability of being used as a predictive instrument for ICU admission, similarly to other mentioned studies.

Keywords: Artificial Intelligence, Machine Learning, Data Preprocessing, Intensive Care Unit.

LISTA DE FIGURAS

Figura 2.1 Exemplo de Árvore de Decisão.	14
Figura 3.1 Tabela de resultados dos experimentos de Fernandes et al. (2020).....	19
Figura 3.2 Tabela de resultados dos experimentos de Araz, Olson e Ramirez-Nafarrate (2019).....	20
Figura 3.3 Tabela de resultados do experimento de Santos et al. (2021).....	21
Figura 3.4 Tabela de resultados dos experimentos de Grawe (2021).....	22
Figura 4.1 Pré processamento dos dados	26
Figura 6.1 Importância de variáveis no modelo Gradient Boost 24h.....	35
Figura 6.2 Importância de variáveis no modelo Gradient Boost 48h.....	35
Figura 6.3 Importância de variáveis no modelo Random Forest 72h	36
Figura 6.4 Métricas por grupos de idade no Gradient Boost 24h	37
Figura 6.5 Métricas por grupos de idade utilizando Gradient Boost 48h	37
Figura 6.6 Métricas por grupos de idade no Random Forest 72h	38
Figura 6.7 Métricas conforme gênero	38
Figura 6.8 Métricas para COVID-19 no algoritmo Gradient Boost 24h.....	39
Figura 6.9 Métricas para COVID-19 no uso do algoritmo Gradient Boost 48h	39
Figura 6.10 Métricas para COVID-19 no modelo Random Forest 72h	40
Figura 6.11 Matrizes de confusão decorrentes do modelo 24h.....	40
Figura 6.12 Matrizes de confusão decorrentes do modelo 48h.....	41
Figura 6.13 Matrizes de confusão decorrentes do modelo 72h.....	41

LISTA DE TABELAS

Tabela 3.1	Resumo dos trabalhos relacionados.....	23
Tabela 4.1	Modelo de Dados: 24, 48 e 72 horas de atendimento	25
Tabela 5.1	Resultados dos testes com Validação Cruzada	32
Tabela 5.2	Resultados dos experimentos em Validação.....	33

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
ANN	Artificial Neural Network
AUROC	Area Under Receiver Operating Characteristic
CSV	Comma-Separated Values
GB	Gradient Boost
IA	Inteligência Artificial
LR	Logistic Regression
LSTM	Long Short-Term Memory
ML	Machine Learning
MLP	Multilayer Perceptron
NB	Multinomial Naive Bayes
NLP	Natural Language Processing
RES	Registros Eletrônicos na Saúde
SGBD	Sistema Gerenciador de Banco de Dados
TCN	Temporal Convolutional Networks
TTS	Textual Time-Series
UTI	Unidade de Terapia Intensiva
SD	Structured Data
SPN	Sum-Product Network
STS	Structured Time-Series
SVM	Support Vector Machine
RF	Random Forest
UTI	Unidade de Tratamento Intensivo

sum-product network

SUMÁRIO

1 INTRODUÇÃO	10
1.1 Objetivo.....	11
1.2 Estrutura do texto	11
2 CONTEXTO DO TRABALHO E FUNDAMENTAÇÃO TEÓRICA	12
2.1 Gestão hospitalar	12
2.2 Métodos de Aprendizado de Máquina	13
2.2.1 Random Forest	13
2.2.2 Gradient Boost	13
2.3 Métricas para avaliação dos modelos.....	15
2.4 Threshold - Youden's index.....	16
3 TRABALHOS RELACIONADOS	18
3.1 Predição de pacientes candidatos a UTI hospitalar	18
3.2 Controle de infecção hospitalar usando Inteligência Artificial.....	20
3.3 Características e preditores de óbito para pacientes internados	21
3.4 Sumário	22
4 MATERIAIS E MÉTODOS	24
4.1 Descrição dos Conjuntos de Dados e Armazenamento	24
4.2 Pré-processamento dos dados	26
4.2.1 Limpeza dos dados.....	26
4.2.2 Transformações	27
4.3 Otimização de Hiperparâmetros	30
4.4 Tecnologias utilizadas	30
5 EXPERIMENTOS	31
5.1 Treinamento com Validação Cruzada	31
5.2 Treinamento e Validação	32
6 ANÁLISE DOS RESULTADOS OBTIDOS	34
7 CONCLUSÕES	42
REFERÊNCIAS	44
APÊNDICE A — GLOSSÁRIO	47
A.1 Atributos de acordo com os modelos adotados	47

1 INTRODUÇÃO

A admissão de pacientes em hospitais em todo o mundo tem gerado uma preocupação efetiva para os gestores quanto à capacidade de atendimento, em especial em áreas metropolitanas e de baixa renda como apontam Lambe et al. (2003), Cowan e Trzeciak (2005). Pacientes são mantidos por períodos mais longos na entrada em emergências por limitações das unidades de internação gerando superlotação (HOOT; ARONSKY, 2008; LEVIN et al., 2008) e aumentando as taxas de mortalidade (BERNSTEIN et al., 2009; POLEVOI; QUINN; KRAMER, 2005; SINGER et al., 2011; CHALFIN et al., 2007) como registra Araz, Olson e Ramirez-Nafarrate (2019).

Nesse sentido, Vaart, Vastag e Wijngaard (2011) sinalizam a importância da gestão do tempo no desempenho operacional dos Departamentos de Emergência (DE), considerando parâmetros tais como: eficácia do sistema de triagem, utilização de recursos e dinâmicas de lotação e de pessoal, a fim de ajustarem a capacidade das unidades de emergência e de internação.

Pacientes com doenças críticas identificadas após admissão em DE, podem ser transferidos para Centros de Terapia Intensiva (UTI) buscando melhorar a qualidade do tratamento intensivo e prevenir a mortalidade (CHOU et al., 2020). A estratificação precoce dos pacientes de acordo com a gravidade é considerada vital, uma vez que uma triagem inadequada pode levar ao aumento de morbidade, mortalidade e custos (FERNANDES et al., 2020).

Chalfin et al. (2007) descobriram que pacientes com processos de admissão mais longos em pronto-socorro tiveram tempo de permanência em UTI por mais tempo, o que, um modelo de predição de admissão preciso, poderia reduzir como também melhorar a qualidade do serviço, possibilitando aos administradores do hospital gerenciar um processo de admissão hospitalar mais eficaz, como apregoa Araz, Olson e Ramirez-Nafarrate (2019).

Dentro do contexto do COVID-19, a demanda por Centros de Terapia Intensiva (UTIs) hospitalares aumentou consideravelmente e se tornou o principal desafio dos sistemas de saúde de todo o mundo. Em Maio de 2021, nos Estados Unidos, foi constatado, segundo Figliozzi et al. (2020), mais de 32 milhões de casos confirmados, resultando em 5 milhões de internações hospitalares com aproximadamente 1.2 milhões de casos considerados graves, demonstrando que, em média, um terço dos pacientes estavam sendo hospitalizados necessitando de atendimento especial em UTI hospitalar.

A capacidade preditiva de modelos de Inteligência Artificial (IA) capazes de classificar pacientes com maior possibilidade de serem internados em UTI permite planejar o processo de admissão hospitalar com mais qualidade, reduzindo o tempo de permanência e melhor planejamento de recursos (FERNANDES et al., 2020; ARAZ; OLSON; RAMIREZ-NAFARRATE, 2019). De acordo com Iaccarino et al. (2020), entender preditores de Aprendizado de Máquina capazes de prever a admissão de pacientes que necessitam de UTIs pode ser de grande ajuda para o planejamento de intervenção e gerenciamento de doenças.

No presente trabalho é demonstrado o uso de técnicas de Aprendizado de Máquina (AM) com base em Registros Eletrônicos de Saúde (RES) visando a identificação de pacientes internados que necessitam de leitos de UTI hospitalar, bem como, a mensuração dos dados e métricas obtidas, melhorando a eficácia do sistema.

1.1 Objetivo

O presente estudo objetiva propor modelos que identifiquem pacientes que virão a necessitar cuidados intensivos com base nos critérios clínicos apresentados nas primeiras 24h, 48h e 72h de internação.

A pergunta norteadora da pesquisa recai sobre a importância do uso de técnicas de Aprendizado de Máquina para a identificação de pacientes que irão necessitar de UTI hospitalar e de índices favoráveis a tomada de decisão quanto ao gerenciamento de leitos e demais recursos, ou seja: O quanto os métodos de Aprendizado de Máquina apresentados podem auxiliar a predição da necessidade de UTI hospitalar para pacientes internados?

1.2 Estrutura do texto

O Capítulo 2 apresenta o contexto do trabalho e sua fundamentação teórica. O Capítulo 3 revisa brevemente outros trabalhos relacionados ao tema. O Capítulo 4 apresenta os processos de seleção de dados, de transformações e limpeza e as tecnologias utilizadas. No Capítulo 5 são detalhados os experimentos e os testes realizados. Já no Capítulo 6 é apresentada a análise dos experimentos. Por fim, no Capítulo 7, elabora-se a conclusão do estudo e aponta-se trabalhos futuros.

2 CONTEXTO DO TRABALHO E FUNDAMENTAÇÃO TEÓRICA

O presente trabalho estuda a predição da necessidade de UTI nas 24, 48 e 72 horas iniciais de atendimento, através do uso de modelos de IA.

Serão apresentados a seguir conceitos e fundamentos visando melhor compreensão sobre a pesquisa realizada.

2.1 Gestão hospitalar

A administração de saúde (*Health Administration*) abrange a liderança e a gestão de hospitais, redes hospitalares e sistemas de saúde, sendo o gerenciamento de todas as funções não clínicas envolvidas em uma unidade de saúde, tais como: contratação de pessoal, orçamento e finanças até o planejamento estratégico de longo prazo para garantir o sucesso do provedor ou sistema de saúde (Healthcare Administration EDU, 2021).

A administração de saúde no contexto hospitalar, de acordo com a Herd et al. (2016) e com o *National Center for Health Leadership*¹ (NCHL), requer um gestor hospitalar multidisciplinar, com conhecimento de contabilidade, orientado à realização de tarefas, adequado para lideranças de equipes, colaborativo, ter habilidades de comunicação, habilidades com finanças, habilidade de influenciar, persuadir e convencer outros a tomar decisões; promover ideias inovadoras, ter senso organizacional, autoconfiança, orientação estratégica, talento para desenvolver e liderar times.

Ainda, o estudo aponta dois tipos de gestores hospitalares, o generalista e o especialista. O Generalista é responsável pelo gerenciamento e auxílio a outros profissionais, observando o contexto organizacional de forma macro. O Especialista é responsável pela eficiência e efetividade de operações específicas, como análises de políticas, finanças, contabilidade, recursos monetários, marketing, recursos humanos e gerenciamento operacional de pacientes e leitos hospitalares (HERD et al., 2016).

Mudanças no campo da saúde, conforme aponta o estudo de Patterson et al. (2016), incluem a necessidade de aprendizagens multifuncionais e organizacionais para líderes de saúde. Araz, Olson e Ramirez-Nafarrate (2019), por sua vez, ressaltam que administradores de hospitais precisam estar cientes da dinâmica do fluxo de pacientes para ajustar à capacidade das unidades de emergência e de UTI, sugerindo o uso de algoritmos de IA para auxiliar neste processo, o que valida a motivação do presente trabalho.

¹<https://www.nchl.org/>

2.2 Métodos de Aprendizado de Máquina

Métodos de Aprendizado de Máquina tem demonstrado resultados expressivos para a identificação de padrões e classificação em diferentes processos (FERNANDES et al., 2020). Na área da saúde, algoritmos cujos resultados podem ser representados por grafos ou outra estrutura de representação - modelos considerados não caixa-preta - e que indiquem a importância dos atributos necessários para tomadas de decisão têm tido forte preferência por conta da validação clínica e estatística (SWE, 2019).

No presente trabalho, foram utilizados os métodos Random Forest (HO, 1995) e Gradient Boost (FRIEDMAN, 2001), os quais são brevemente revisados a seguir.

2.2.1 Random Forest

O *Random Forest* (RF) (HO, 1995) é um algoritmo baseado em árvores de decisão (*Decision Trees*) (UTGOFF, 1989). Árvore de decisão é um modelo baseado numa estrutura em árvore bastante utilizado para classificação. Cada nodo desta árvore representa uma característica ou pergunta, as folhas representam classes, e os ramos correspondem a múltiplas escolhas, como mostra a Figura 2.1.

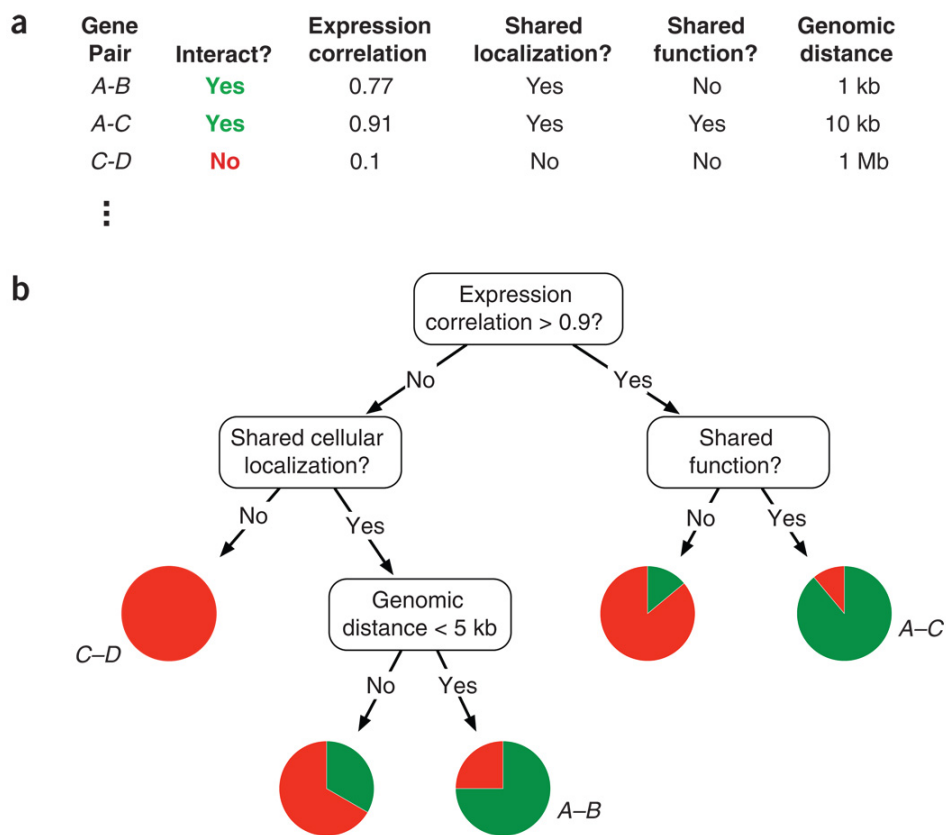
No *Random Forest*, cada árvore representa um modelo preditivo de uma amostra de dados retirados de um conjunto de dados de treinamento, utilizando-se (ou não) o método de *bootstrap*². Então, a partir desta amostra, gera-se a *Decision Tree* deste conjunto de dados, com a escolha aleatória para os atributos. Executa-se este processo N vezes, onde N é a quantidade de árvores a serem criadas, estabelecida pelo usuário. Com uma “floresta” de *Decision Trees* aleatórias, realiza-se a votação entre todas as árvores a fim de se definir o valor preditivo para cada dado que busca ser testado, reduzindo os riscos de *overfitting* e facilitando a determinação da importância dos atributos.

2.2.2 Gradient Boost

O *Gradient Boost* é uma técnica comumente usada em Aprendizado de Máquina para modelos de regressão e classificação. O seu modelo preditivo final é gerado a partir

²Seleção aleatória de uma seleção de linhas e colunas, sendo que as linhas podem se repetir, já que a seleção aleatória não exclui da lista de opções de escolha uma linha já escolhida anteriormente para a criação deste conjunto de dados

Figura 2.1 – Exemplo de Árvore de Decisão.



Fonte: (UTGOFF, 1989)

de um conjunto de modelos fracos (pequenas árvores de decisão). Criado com o objetivo de superar o *Random Forest*, o *Gradient Boost* constrói seu modelo em estágios (típico de métodos de *Boosting*) e os generaliza, desta forma permitindo a otimização de uma função de perda diferenciável arbitrária (FRIEDMAN, 2001).

Este método funciona pela criação de *stumps* (árvores de decisão simples, considerados classificadores fracos), com exceção da “primeira árvore” (de tamanho 1) que é feita a partir da *Logistic Function*. Quando o método de *Gradient Boost* é aplicado em uma situação de classificação, para definir-se o valor preditivo de uma instância, é feito um cálculo que busca definir os valores de saída das folhas da árvore que está sendo gerada e, a partir desses valores, é calculado valor preditivo e respectiva probabilidade, conforme as equações a seguir.

$$ValorDeSaida_i = \frac{\sum Residuo_i}{\sum (ProbabilidadeAnterior_i * (1 - ProbabilidadeAnterior_i))}$$

$$PredicaoDeChances_i = \sum (TaxaDeAprendizagem * ValorDeSaida_i)$$

$$Probabilidade_i = \frac{e^{PredicaoDeChances_i}}{1 + e^{PredicaoDeChances_i}}$$

2.3 Métricas para avaliação dos modelos

Nesta subseção são listadas as métricas utilizadas neste trabalho para avaliar os modelos e analisar as predições realizadas pelos mesmos. As predições são caracterizadas como:

- **Verdadeiras Positivas:** tem condição positiva e o teste é positivo.
- **Verdadeiras Negativas:** tem condição negativa e o teste é negativo.
- **Falsas Positivas:** tem condição negativa e o teste é positivo.
- **Falsas Negativas:** tem condição positiva e o teste é negativo.

A partir destas categorias estabelecidas, temos as métricas que avaliam a performance dos modelos:

- **Precisão:** ou valor preditivo positivo, é a relação entre os elementos positivos preditos corretamente divididos pelos preditos como positivo.

$$\frac{VP}{VP + FP}$$

- **Sensibilidade:** ou taxa de verdadeiros positivos; representa a taxa de acerto para predizer corretamente a classe positiva.

$$\frac{VP}{VP + FN}$$

- **Especificidade:** ou taxa de verdadeiros negativos; refere-se a taxa de acerto para predizer corretamente a classe negativa.

$$\frac{VN}{VN + FP}$$

- **Acurácia:** é a relação de acertos positivos e negativos dividido pela quantidade

total de predições.

$$\frac{VP + VN}{VP + VN + FP + FN}$$

- **F1-Score:** é a média harmônica entre Precisão e Sensibilidade.

$$\frac{2 * Precisao * Sensibilidade}{(Precisao + Sensibilidade)}$$

- **AUROC:** a Área Sob a Curva Característica de Operação do Receptor é a métrica que define como os modelos são capazes de distinguir as classes. Através de uma representação gráfica a curva ROC ilustra o desempenho do modelo por exibir a compensação entre a Taxa de Verdadeiros Positivos e a Taxa de Falsos Positivos de um sistema de classe binária à medida que o seu limiar (*threshold*) varia. A área sob a curva representa a precisão geral do teste, com um valor próximo a 1.0 indicando a alta Sensibilidade e a Especificidade. A linha pontilhada do gráfico representa a discriminação com Área Sob a Curva ROC (AUC) de 0.5 (50%), isto é, o teste não é melhor do que jogar uma moeda (LALKHEN; MCCLUSKEY, 2008).

Foram, também, analisadas as importâncias dos atributos para a predição nos modelos. O cálculo da importância dos atributos é baseado na média decrescente em impureza.

2.4 Threshold - Youden's index

O limiar (*threshold*) pode ser entendido como um ponto limítrofe que serve de parâmetro, delimitando uma condição em uma saída contínua. Provost (2000) alerta que, no entanto, se utilizado em distribuições de classes desbalanceadas, usando os classificadores produzidos por algoritmos de Aprendizado de Máquina padrão sem ajustar o limite de saída, pode muito bem induzir a um erro crítico. Nesse caso, entende o autor, que a escolha de outro método se torna o mais indicado.

Nesse sentido, Bekkar, Djemaa e Alitouche (2013) apresentam o Índice de Youden como sendo convencionalmente usado para estimar testes de diagnóstico, melhorando a eficiência dos algoritmos quando usados na área clínica.

Youden (1950) sugeriu uma maneira de reduzir o rendimento de uma prova de diagnóstico onde o valor pode ser de 0 a 1, o que ficou denominado de Índice de Youden. Quando o teste de diagnóstico dá a mesma proporção de resultados positivos para um

grupo com ou sem enfermidade, tem o valor de 0, isto é, o teste é inútil. Já o valor 1 indica que não há falsos positivos ou falsos negativos, ou seja, o teste é perfeito.

O Índice de Youden é comumente usado em conjunção com a análise da Curva ROC (*Receiver Operating Characteristic*). É definido para todos os pontos da curva e o valor máximo do índice pode ser usado como critério para seleção do ponto de corte ideal onde um teste resulta em um valor numérico, ao invés de um resultado dicotômico. O índice é representado graficamente pela altura acima da linha de chance (linha de sorte), o que é equivalente a área sob a curva subentendida por um único ponto de operação (LALKHEN; MCCLUSKEY, 2008).

A fórmula que define o cálculo do índice é definido a partir da Sensibilidade e da Especificidade, correspondendo à taxa de verdadeiros positivos (TPR) subtraindo-se a taxa de falsos positivos (FPR), como apresentado nas fórmulas a seguir.

$$J = \text{Sensibilidade} + \text{Especificidade} - 1$$

ou

$$J = TPR - FPR$$

3 TRABALHOS RELACIONADOS

Diversas pesquisas no contexto das ciências médicas tem demonstrado que a detecção precoce de doenças baseada em pontuações clínicas utilizando métodos de aprendizado de máquina é benéfica para abreviar o tempo e melhorar a qualidade do tratamento, bem como, produzir a redução de custos, como já referido na Seção 2.2.

Nesta seção, são revisados alguns estudos que serviram de subsídio para o trabalho ora apresentado, pelo uso de recursos de aprendizado de máquina, mesmo que não tratando diretamente com predição de internação em UTIs (ARAZ; OLSON; RAMIREZ-NAFARRATE, 2019; FERNANDES et al., 2020; IZQUIERDO et al., 2020; SANTOS et al., 2021; BERENGUER et al., 2020; GRAWE, 2021).

3.1 Predição de pacientes candidatos a UTI hospitalar

Fernandes et al. (2020) apresenta experiência em dois hospitais, a saber o *Hospital Beatriz Ângelo* (HBA), em Portugal, e o *Beth Israel Deaconess Medical Center* (BIDMC), nos EUA, com uso de Regressão Logística (*Logistic Regression*), *Random Forest Regression* e *Random Undersampling Boosting* para prever o risco de admissão em UTI em momento de triagem hospitalar. Foram utilizadas todas as variáveis clínicas e a queixa principal, com exceção de prioridade de triagem. Em ambos os hospitais, pelos modelos de *Logistic Regression*, receberam alta classificação para a predição de admissão na UTI os dados de frequência cardíaca, oximetria de pulso, frequência respiratória e pressão arterial sistólica.

A pesquisa registra ainda que os desempenhos do modelo desenvolvido para o BIDMC, com variáveis clínicas e queixa principal com calibração isotônica, e do multimodelo para o HBA, combinando o modelo desenvolvido com variáveis clínicas e a queixa principal com um modelo desenvolvido apenas com variáveis clínicas, atingiram boas performances e demonstraram bons equilíbrios entre Sensibilidade e Especificidade, conforme a Figura 3.1.

O estudo conclui que uma superestimativa das probabilidades de admissão à UTI para o grupo de pacientes de alto risco pode ser melhor do que uma subestimação do risco.

Araz, Olson e Ramirez-Nafarrate (2019) apresenta uma alternativa de análise para admissão de pacientes em momento de triagem através do emprego de modelos de predição baseados em mineração de dados. De 2006 até 2009, incluindo um período de

Figura 3.1 – Tabela de resultados dos experimentos de Fernandes et al. (2020)

	BIDMC model with no priority		HBA multi-model with no priority	
	No calibration	Isotonic calibration	No calibration	Isotonic calibration
AUROC	0.91 [0.90-0.92]	0.91 [0.90-0.92]	0.85 [0.83-0.86]	0.85 [0.83-0.86]
AUPRC	0.32 [0.28-0.35]	0.30 [0.27-0.33]	0.06 [0.05-0.07]	0.06 [0.05-0.07]
AP	0.31 [0.28-0.35]	0.30 [0.27-0.33]	0.06 [0.05-0.07]	0.06 [0.05-0.07]
Specificity	0.86 [0.86-0.87]	0.86 [0.86-0.86]	0.81 [0.80-0.81]	0.85 [0.84-0.85]
Recall	0.81 [0.79-0.84]	0.82 [0.80-0.84]	0.73 [0.7-0.76]	0.68 [0.65-0.72]
SMR	5.44 [5.1-5.78]	5.62 [5.28-5.97]	26.27 [24.27-28.5]	25.66 [23.68-27.84]
Threshold	0.035	0.030	0.009	0.007
N-gram range	Combination of unigrams and bigrams		M1: unigrams	
Words from vocabulary	9500		M1: 29000	
Warm start	No		Yes	
Regularization constant	1		1	

M1 corresponds to model which uses chief complaint for prediction. In brackets is the result for 100 bootstrapping iterations in 95% confidence intervals. The models selected have recall highlighted in bold. BIDMC—Beth Israel Deaconess Medical Center. HBA—Hospital Beatriz Ângelo. AUROC—area under the ROC curve. AUPRC—area under the precision recall curve. AP—Average precision. SMR—Standardized Mortality Ratio.

Fonte: Fernandes et al. (2020)

epidemia *influenza* em 2009 (DUGAS et al., 2012), foram obtidos dados de um hospital real, compostos por dados de triagem como as informações de idade, gênero, modo de chegada, nível de gravidade, temporada de gripe, turno de chegada ao Departamento de Emergência no dia (DE) e indicador de caso de doença semelhante à *influenza* (ILI).

O estudo propõe o uso dos algoritmos de *Logistic Regression*, *Decision Tree*, *Artificial Neural Network*, *Support Vector Machines* e *Random Forest* para identificar a admissão hospitalar com o emprego de informações clínicas mínimas.

Os experimentos foram divididos em quatro modelos diferentes: um modelo com dados apenas de 2006, outro modelo com dados de 2009, outro com dados de 2006 e 2007 e, por fim, um modelo definitivo com todos os dados entre 2006 e 2009. Em todas as etapas dos testes o algoritmo *XGBoost* apresentou as melhores métricas de AUROC (*area under the receiver operating characteristic curve*) e Acurácia. O melhor resultado de AUROC foi obtido pelo modelo com dados de 2006 a 2009, com o valor de 86.3%, enquanto o melhor valor de Acurácia foi apresentado pelo modelo com dados de 2006, com o valor 82.1%. Todos os resultados podem ser verificados na Figura 3.2.

Em outro estudo, utilizando informações de texto livre de *Electronic Health Records* de um hospital da Espanha, Izquierdo et al. (2020) usa Processamento de Linguagem Natural (PLN) e *Word Embeddings* para transformar dados de textos não-estruturados em informações estruturadas, que foram processadas com o uso de *Decision Trees* a fim de identificar fatores associados a pacientes com COVID-19 em UTI hospitalar e que tiveram necessidade de internar em uma UTI. O estudo atinge métricas finais de AUROC 84%, Acurácia 85% e Sensibilidade de 57%, identificando uma combinação de fatores

Figura 3.2 – Tabela de resultados dos experimentos de Araz, Olson e Ramirez-Nafarrate (2019)

Model 2006 Data N = 43,703				N = 87,900 2006–2007 Data				
	AUC	*Model Accuracy	Positive Predictive Value (TP/(TP + FP) (%))	Models	AUC	% Change	*Model Accuracy	% Change
Logistic Regression	0.840	0.820	95.5	Logistic Regression	0.830	-1.19	0.810	-1.22
ANN	0.850	0.816	94.0	ANN	0.830	-2.35	0.800	-1.96
SVM	0.840	0.816	95.2	SVM	0.800	-4.76	0.780	-4.41
Decision Tree	0.820	0.819	95.2	Decision Tree	0.810	-1.22	0.800	-2.32
Random Forest	0.790	0.818	95.7	Random Forest	0.770	-2.53	0.800	-2.20
XGradient Boosting (LR)	0.860	0.821	94.8	XGBoost (LR)	0.830	-3.48	0.810	-1.33
Model 2009 Data N = 47,002				N = 118,005 2006–2009 Data				
	AUC	*Model Accuracy	Positive Predictive Value (TP/(TP + FP) (%))	Models	AUC	% Change	*Model Accuracy	% Change
Logistic Regression	0.824	0.801	81.6	Logistic Regression	0.832	-0.95	0.803	-2.07
ANN	0.834	0.804	82.3	ANN	0.835	-1.76	0.805	-1.35
SVM	0.742	0.806	82.0	SVM	0.790	-5.95	0.783	-4.04
Decision Tree	0.786	0.793	82.2	Decision Tree	0.810	-1.22	0.802	-2.08
Random Forest	0.785	0.808	81.6	Random Forest	0.770	-2.53	0.805	-1.59
XGradient Boosting (LR)	0.843	0.810	83.3	XGBoost (LR)	0.863	+0.35	0.817	-0.04

Fonte: Araz, Olson e Ramirez-Nafarrate (2019)

clínicos atrelados à classe positiva: idade, febre e taquipneia.

3.2 Controle de infecção hospitalar usando Inteligência Artificial

A pesquisa apresentada por Santos et al. (2021) descreve a utilização de inteligência artificial para a vigilância do controle de infecção hospitalar. O estudo contemplou 18 meses de informação de prontuário eletrônico de todos os pacientes com 18 anos ou mais, sendo 12 meses para treinamento (1º de janeiro de 2017 a 31 de dezembro de 2017) e 6 meses para validação (1 de janeiro de 2018 a 3 de junho de 2018). Os dados analisados incluíram resultados laboratoriais, sinais vitais, prescrições de medicamentos, imagens e evoluções durante a internação. O conjunto de dados de treinamento constituiu-se de 84,655 (69.2%) pacientes-dia e o de validação de 37,606 (30.8%) pacientes-dia.

Conforme pode ser observado na Figura 3.3, os modelos vencedores para a classificação das infecções variaram entre o *Random Forest* (RF) e o *Multinomial Naive Bayes* (MNB). Para o preditor de Pneumonia, o melhor resultado foi atingido através do modelo de RF com AUROC de 93.47%. Referente à Pneumonia associada à Ventilação Mecânica, o melhor resultado foi atingido pelo RF com AUROC de 95.67%. Na Infecção de Sítio Cirúrgico, o melhor modelo foi atingido pelo algoritmo MNB com AUROC de 85.74%. Para a Infecção do Trato Urinário, o algoritmo vencedor foi o RF com AUROC de 88.87%. Já, referente à Infecção de Corrente Sanguínea, o melhor modelo foi o MNB com AUROC de 88.28% e para a Traqueobronquite, o modelo vencedor foi o RF com AUROC de 98.62%.

Figura 3.3 – Tabela de resultados do experimento de Santos et al. (2021)

	Algorithm	Dataset	AUROC (SD)	NPV (SD)	PPV (SD)	Sensitivity (SD)	Specificity (SD)	Accuracy (SD)
Global	MLP	Categorical/Numerical	90.27% (0.15)	99.66% (0.01)	8.92% (0.07)	88.57% (0.51)	78.86% (0.17)	79.0% (0.16)
Pneumonia	RF	Categorical/Numerical	93.47% (0.13)	99.99% (0.0)	1.59% (0.05)	97.31% (0.77)	80.66% (0.65)	80.7% (0.64)
Ventilator-associated Pneumonia	RF	Categorical/Numerical	95.67% (0.15)	99.99% (0.0)	2.62% (0.02)	95.69% (0.48)	89.02% (0.05)	89.0% (0.05)
Surgical Site Infection	MNB	Categorical/Numerical	85.74% (0.52)	99.99% (0.0)	0.44% (0.01)	95.09% (1.35)	70.03% (1.06)	70.0% (1.06)
Urinary Tract Infection	RF	Categorical/Numerical	88.87% (0.33)	99.85% (0.03)	3.42% (0.49)	84.0% (4.3)	81.49% (4.04)	81.5% (3.98)
Bloodstream Infection	MNB	Categorical/Numerical	88.28% (0.19)	99.9% (0.01)	3.26% (0.03)	90.85% (0.67)	77.91% (0.12)	78.0% (0.12)
Tracheobronch-itis	RF	Text Bag-of-Words (TF-IDF)	98.62% (0.06)	99.99% (0.0)	5.26% (0.30)	97.22% (0.93)	96.63% (0.18)	96.6% (0.18)

MLP – Multilayer Perceptron; RF – Random Forest; MNB - Multinomial Naive Bayes; AUROC - Area Under the Receiver Operating Characteristic Curve; NPV – Negative Predictive Value; PPV- Positive Predictive Value; SD – Standard Deviation. TF-IDF – Term Frequency-Inverse Document Frequency.

Fonte: Santos et al. (2021)

O modelo final utilizou uma rede neural *Multilayer Perceptron* (MLP) alcançando uma AUROC de 90,27%; Especificidade de 78,86%; Sensibilidade de 88,57%; detectando corretamente 67 entre 73 pacientes com Associação a Infecções Hospitalares (AIHs), demonstrando a viabilidade do método. As infecções respiratórias tiveram os melhores resultados (AUROC 93,47%).

O algoritmo de inteligência artificial mostrou-se capaz de identificar a maioria das AIHs demonstrando potencial para economizar tempo, promover vigilância precisa em todo o ambiente hospitalar e melhorar o desempenho de prevenção de infecção.

3.3 Características e preditores de óbito para pacientes internados

O estudo retrospectivo observacional de Berenguer et al. (2020) apresenta coleta de dados demográficos, clínicos, laboratoriais, de tratamento e complicações de pacientes que tiveram COVID-19 positivo no período entre 17 de março de 2020 e 17 de abril de 2020, em 127 centros espanhóis, realizando análises a partir destas informações a fim de definir as características que pudessem ter relação com o alto risco de morte através da utilização de regressão de Cox univariada e multivariável (*Proportional hazards model*).

Os dados registrados descrevem os sintomas mais comuns apresentados tais como: febre, tosse, mal-estar e dispneia. Durante a internação dos 4035 pacientes, 31.5% dos pacientes desenvolveu síndrome do desconforto respiratório aguda, 18.5% internações em UTI e 15.5% foram submetidos à ventilação mecânica. Tiveram óbito 1131 (28%) pacientes, sendo que mortalidade aumentou com a idade (85.6% ocorrendo em maiores de 65 anos). O risco aumentado de morte foi associado a dezessete fatores, sendo os mais representativos: idade avançada (20.49%), cirrose hepática (1.77%), baixa saturação de oxigênio ajustada para idade (2.62%), concentrações mais altas de proteína criativa (2.87%) e menor taxa de filtração glomerular estimada (3.85%).

Um trabalho recente no Programa de Pós-Graduação em Computação da UFRGS (GRAWÉ, 2021) explorou a predição de mortalidade, usando apenas a janela das primeiras 48 horas de estadia hospitalar dos pacientes. Os experimentos usaram um conjunto de dados da base MIMIC-III³, com mais de 20mil estadias em UTIs.

Foram utilizados algoritmos de Aprendizado de Máquina e *Deep Learning* cujos resultados levaram a treinamentos de modelos de *ensemble* (WOLPERT, 1992; TING; WITTEN, 1999; POLIKAR, 2006). Estes modelos foram agrupados, resultando, então, nos modelos definitivos. Os experimentos finais mostraram que o uso de *ensemble* melhorou a performance final dos modelos, como constatado nas linhas 19 e 20 da Figura 3.4.

Figura 3.4 – Tabela de resultados dos experimentos de Grawe (2021)

	Algorithm	Balancing	Data Types	Precision	Recall	W F-Score	U F-Score	TP Rate	AUROC
1	SVM	Cost-Sensitive	STS + TTS	0.870	0.761	0.794	0.655	0.764	0.847
2	LR	Cost-Sensitive	STS + TTS	0.871	0.755	0.790	0.652	0.775	0.848
3	SVM	Cost-Sensitive	SD + STS	0.875	0.622	0.678	0.558	0.889	0.842
4	LR	Cost-Sensitive	SD + STS	0.869	0.751	0.786	0.647	0.771	0.841
5	SVM	Cost-Sensitive	SD + TTS	0.861	0.504	0.554	0.462	0.892	0.776
6	LR	Cost-Sensitive	SD + TTS	0.851	0.703	0.748	0.599	0.717	0.776
7	SVM	Cost-Sensitive	SD + STS + TTS	0.878	0.661	0.712	0.588	0.885	0.855
8	LR	Cost-Sensitive	SD + STS + TTS	0.874	0.760	0.793	0.657	0.786	0.855
9	NB	Undersampling	STS + TTS	0.872	0.752	0.788	0.651	0.788	0.843
10	SVM	Undersampling	STS + TTS	0.874	0.739	0.778	0.643	0.813	0.844
11	LR	Undersampling	STS + TTS	0.873	0.755	0.790	0.654	0.788	0.844
12	NB	Undersampling	SD + STS	0.871	0.759	0.793	0.656	0.775	0.840
13	SVM	Undersampling	SD + STS	0.875	0.666	0.718	0.591	0.869	0.841
14	LR	Undersampling	SD + STS	0.872	0.761	0.795	0.657	0.775	0.841
15	NB	Undersampling	SD + TTS	0.849	0.627	0.685	0.548	0.771	0.756
16	SVM	Undersampling	SD + TTS	0.864	0.466	0.522	0.436	0.911	0.759
17	LR	Undersampling	SD + TTS	0.850	0.650	0.704	0.564	0.759	0.761
18	NB	Undersampling	SD + STS + TTS	0.873	0.754	0.790	0.653	0.790	0.850
19	SVM	Undersampling	SD + STS + TTS	0.881	0.573	0.615	0.518	0.927	0.852
20	LR	Undersampling	SD + STS + TTS	0.876	0.761	0.795	0.661	0.800	0.853

Fonte: Grawe (2021)

3.4 Sumário

Os trabalhos citados apresentam aspectos e parâmetros que influenciaram o presente projeto. A Tabela 3.1 descreve as características gerais de cada um dos trabalhos aqui apresentados.

³<https://physionet.org/content/mimiciii/1.4/>

Tabela 3.1 – Resumo dos trabalhos relacionados

Autor(es)	Tema	Modelo de dados	Métodos	Conclusões
Fernandes et al. (2020)	Preditor de admissão em UTI de pacientes vindos da Emergência.	Dados coletados de departamentos de emergência (triagem) de 2 hospitais para pacientes com mais de 18 anos.	PLN, LR, RUSBoost, AdaBoost e Random Forest.	O multimodelo mostrou melhor calibração para definir as probabilidades de admissão de UTI.
Araz, Olson e Ramirez-Nafarrate (2019)	Preditor de admissão hospitalar a partir da emergência usando informações de triagem.	Dados coletados de departamentos de emergência (triagem).	LR, ANN, SVM, Decision Tree, RF e XGBoost.	O uso de data mining baseado em modelos preditivos foram capazes de identificar pacientes que precisaram ser admitidos.
Santos et al. (2021)	Preditor de infecções hospitalares.	Dados-dia da coleta de resultados laboratoriais, sinais vitais, prescrições de antibióticos, laudos de imagem e notas de evolução.	MLP, RF e MNB.	Os algoritmos identificaram infecções hospitalares com sensibilidade e acurácia aceitáveis.
Berenguer et al. (2020)	Características e preditores de morte em pacientes com COVID-19.	Registros eletrônicos em saúde de pacientes com COVID-19.	Clustering.	O modelo enuncia características determinantes correlacionadas a morte para pacientes com COVID-19.
Grawe (2021)	Preditor de mortalidade com métodos de Ensemble.	MIMIC-III: Pacientes, estadia em UTI, notas médicas, medidas fisiológicas.	SVM, LR, NB, LSTM e TCN.	Métodos de Ensemble melhoraram e otimizaram as métricas finais dos modelos.

Fonte: O autor

4 MATERIAIS E MÉTODOS

Neste capítulo será descrito o conjunto de dados utilizado (Seção 4.1), todas as transformações e métodos aplicados (Seção 4.2) e tecnologias utilizadas (Seção 4.4).

4.1 Descrição dos Conjuntos de Dados e Armazenamento

A base de dados foi obtida de um hospital real em um período de Pandemia do COVID-19, compreendido entre 01/06/2020 e 31/07/2021, onde o intervalo de 01/06/2021 até 31/03/2021 foi utilizado para treinamento e o de 01/04/2021 até 31/07/2021, para validação. Os dados foram enviados pelo hospital em planilhas tipo *Comma-separated Values* (CSV) e armazenados em um banco de dados relacional PostgreSQL ⁴ (Seção 4.4).

Foram analisados e tratados dados referentes às internações, evoluções, laudos de exames de imagem, prescrição de medicamentos, laudos de cirurgias, sinais vitais e exames de todos os pacientes que tiveram atendimento durante o período relatado no estudo, considerando as suas primeiras 24, 48 e 72 horas de internação.

A coleta de exames e sinais vitais (HGT, pressão arterial sistólica, pressão arterial diastólica, frequência respiratória, frequência cardíaca, oximetria, temperatura, medições de dor, diurese, altura e peso) apresentou os seguintes dados: número do prontuário do paciente, tipo do registro da coleta, data e hora da coleta, valor de medida do que foi coletado, a unidade de medida e o perfil de quem executou a coleta.

Os exames laboratoriais (escarro, abscesso, cateter, líquido, tecido, secreção, fragmento, coleção, ferida, hemocultura positiva, hemocultura negativa, bacterioscopia positiva, bacterioscopia negativa, urocultura positiva e urocultura negativa) informaram: número de prontuário do paciente, nome do exame laboratorial coletado, código do exame, nome do campo do exame, data e hora do pedido do exame, data e hora da entrega do exame, e resultado/valor de entrega do exame.

Os laudos de exames de imagem (tomografia, raio-X, ecografia) trouxeram o número de registro do prontuário do paciente, o nome do exame, o código do exame, data e hora do pedido do exame, data e hora da entrega do exame e resultado do laudo do exame de imagem coletado.

As informações relacionadas às prescrições de medicamentos continham o registro de prontuário do paciente, nome do medicamento, código do medicamento, data e

⁴<https://www.postgresql.org/>

hora da prescrição, data e hora de vigência para o início do uso do medicamento, data e hora para o fim do uso do medicamento, dose, sua unidade de medida, frequência e via.

Nas evoluções dos pacientes teve-se o respectivo código de registro do prontuário, data e hora da evolução, unidade hospitalar onde foi feita a evolução, o perfil de quem a fez e o texto da evolução, o qual já foi recebido anonimizado e descaracterizado da fonte.

As descrições cirúrgicas registraram: o número do prontuário do paciente, código de atendimento da cirurgia, perfil de quem a executou, data e hora do início da cirurgia, data e hora do fim da cirurgia e texto com as informações referentes à cirurgia propriamente dita.

As informações dos pacientes apresentaram o número de prontuário, sexo e data de nascimento, de onde foi extraída a idade do paciente considerando a data do atendimento.

As informações relacionadas às internações, por sua vez, forneceram o registro do prontuário do paciente, o código do atendimento, data e hora do atendimento, unidade hospitalar onde o paciente foi atendido, data e hora da alta e tipo da alta, quando registrada.

Com estes dados, após serem preparados, conforme explicado na Seção 4.2, os modelos utilizados para treinamento e validação possuíam as características exibidas na Tabela 4.1, contendo uma proporção aproximada de 70% para treinamento e 30% para validação em todos os conjuntos de dados, com valores desbalanceados para as classes: classe positiva correspondente nos conjuntos de dados de validação de 21.12%, 27.25% e 32.94% para os respectivos modelos de 24, 48 e 72 horas.

Tabela 4.1 – Modelo de Dados: 24, 48 e 72 horas de atendimento

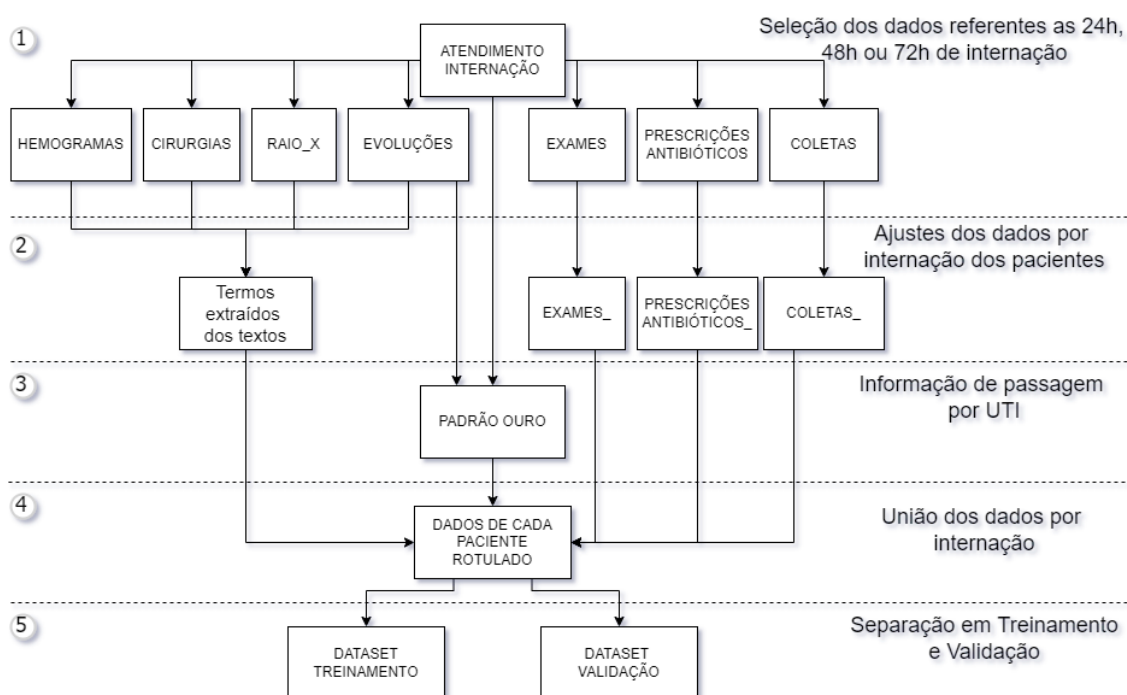
Característica	Treinamento 24h	Validação 24h	Treinamento 48h	Validação 48h	Treinamento 72h	Validação 72h
Atendimentos	6583 (69.9%)	2836 (30.1%)	4828 (69.3%)	2136 (30.7%)	3783 (69.2%)	1688 (30.8%)
Idade						
18-20	83 (1.26%)	57 (2.01%)	42 (0.87%)	37 (1.73%)	16 (0.42%)	22 (1,30%)
21-30	902 (13.70%)	393 (13.86%)	437 (9.05%)	242 (11.33%)	220 (5.82%)	135 (8,00%)
31-40	1038 (15.77%)	424 (14.95%)	568 (11.76%)	239 (11.19%)	347 (9.17%)	161 (9,54%)
41-50	639 (9.71%)	296 (10.44%)	466 (9.65%)	213 (9.97%)	352 (9.30%)	172 (10,19%)
51-60	771 (11.71%)	381 (13.43%)	619 (12.82%)	325 (15.22%)	532 (14.06%)	271 (16,05%)
61-70	1123 (17.06%)	483 (17.03%)	932 (19.30%)	388 (18.16%)	779 (20.59%)	335 (19,85%)
71-80	1080 (16.41%)	433 (15.27%)	913 (18.91%)	365 (17.09%)	781 (20.64%)	314 (18,60%)
81-90	717 (10.89%)	285 (10.05%)	635 (13.15%)	250 (11.70%)	555 (14.67%)	212 (12,56%)
91+	230 (3.49%)	84 (2.96%)	216 (4.47%)	77 (3.60%)	201 (5.31%)	66 (3,91%)
Sexo						
F	3789 (57.56%)	1643 (57.93%)	2550 (52.82%)	1151 (53.89%)	1862 (49.22%)	852 (50.47%)
M	2794 (42.44%)	1193 (42.07%)	2278 (47.18%)	985 (46,11%)	1921 (50.78%)	836 (49.53%)
Alvo						
0	5472 (83.12%)	2237 (78.88%)	3753 (77.73%)	1554 (72.75%)	2737 (72.35%)	1132 (67.06%)
1	1111 (16.88%)	599 (21.12%)	1075 (22.27%)	582 (27.25%)	1046 (27.65%)	556 (32.94%)

4.2 Pré-processamento dos dados

Esta seção apresenta o processamento de dados utilizado para adequar o conjunto de dados para fins de aplicação de algoritmos de Aprendizado de Máquina. As transformações dos dados foram realizadas através do *software* de processamento de dados Dataiku (2021) (versão 4.5).

Na Figura 4.1 é possível conferir as etapas do processamento.

Figura 4.1 – Pré processamento dos dados



Fonte: O autor.

4.2.1 Limpeza dos dados

Foram removidos todos os dados que não continham informação de Unidade Hospitalar (pois este dado é o que delimita se um paciente teve passagem por UTI ou não) e todos os dados que não faziam parte do período citado.

No que concerne aos exames foram selecionados os 21 exames com maior incidência no período relatado, sendo estes: Detecção por PCR de Coronavírus 2019 (Covid19), creatinina, gasometria arterial, sódio, potássio, EQU, ureia, tempo de protrombina, plaquetas contagem, Ktíp, proteína C reativa, bilirrubinas total e frações, lactato

arterial (ácido láctico), Tgp(Alt), Tgo(Ast), eritrograma, Hiv 1 e 2, magnésio, glicose, gasometria venosa e contagem de leucócitos.

No que se refere aos resultados em texto de alguns exames foi executado um método de padronização textual. O mesmo procedimento foi aplicado na coluna referente ao texto das cirurgias, exame de imagem e evoluções.

Na tabela de sinais vitais foram anuladas todas as células onde não existia um valor numérico válido representado.

Referente aos laudos de exame de imagem, foram selecionadas as informações relacionadas aos exames de tórax e abdômen, visto que são os exames de imagem com maior quantidade de ocorrências no período do estudo.

No que se refere aos medicamentos, foram mantidos os dados associados à administração dos seguintes remédios: Levofloxacina, Teicoplanina, Amicacina, Ganciclovir, Oseltamivir, Metronidazol, Mebendazol, Azitromicina, Amoxicilina, Rifampicina, Aciclovir, Cefalexina, Ertapenem, Amoxicilina + Clavulanato, Sulfametoxazol + Trimetoprim, Gentamicina, Ampicilina, Polimixina B, Cefuroxima, Anfotericina B, Albendazol, Piperaciclina + Tazobactam, Clindamicina, Micafungina, Claritromicina, Vancomicina, Ivermectina, Eritromicina, Nistatina, Ciprofloxacina, Sulfadiazina, Penicilina + Benzatina, Cefoxitina, Cloranfenicol, Fluconazol, Ceftazidima, Itraconazol, Meropenem, Cefazolina, Cefepime, Tigeciclina, Voriconazol, Ampicilina + Sulbactam, Doxiciclina, Ceftriaxona, Nitrofurantoina, Norfloxacina, Pirimetamina, Linezolida, Tetraciclina, Oxacilina, Neomicina; bem como as vias de aplicação: Endovenosa, Oral, Enteral, Otológica, Nasoenteral, Vaginal, Intramuscular, Tópico, Oftálmica, Intraperitoneal e Gastrostomia, totalizando 52 medicamentos e 11 vias.

Todos estes filtros e descrições estão representados pela Etapa 1 da Figura 4.1.

4.2.2 Transformações

Para o Padrão Ouro (Etapa 3 da Figura 4.1) definiu-se que, caso o paciente tivesse passagem por UTI hospitalar no intervalo de sua internação após o período que concerne o seu atendimento e o respectivo intervalo de 24, 48 ou 72 horas, seria considerado positivo para fins do estudo e, caso contrário, seria negativo. Isto é, hospitalizações de pacientes que ingressaram direto em UTI hospitalar ou foram admitidos em uma entre a sua admissão e os intervalos citados, foram considerados negativos para o estudo.

Referente às informações relacionadas às prescrições de medicamentos, executou-

se uma padronização sobre a data e a hora de prescrição do remédio no formato ‘dd/MM/yyyy HH:mm’. Removeu-se, a seguir, as colunas de data e hora referentes a de vigência do medicamento. Executou-se uma padronização nos nomes dos remédios para ter-se os nomes em um formato único. Após isso, criou-se colunas “dummy” a partir da coluna com o nome dos remédios (colunas “pivotadas” a partir da coluna medicamento, contendo o algarismo 1 no caso de haver o nome do medicamento ou algarismo 0 caso não apresente o nome). O mesmo critério foi adotado para o caso das vias de aplicação. Então, considerando o registro de cada paciente e as respectivas datas de prescrição dos medicamentos, gerou-se um agrupamento contendo a contagem de todos os medicamentos e vias pivotados. Uniu-se estas informações referentes aos remédios e vias à respectiva internação do paciente, mantendo os medicamentos e vias prescritos no intervalo das primeiras 24h, 48h ou 72h, conforme cada um dos modelos testados. Depois foi feito o agrupamento destas informações, fazendo-se contagens de cada medicamento e via no período por código de atendimento do paciente.

Nos exames começou-se por fazer um agrupamento das informações referentes aos “resultados de exame”, conforme o código do paciente, nome do exame, tipo do exame, data e hora do pedido e entrega do exame. Executou-se a padronização sobre a data e hora do pedido e entrega do exame e removeu-se todas as linhas em que as colunas de resultado do exame, data e hora de entrega ou pedido de exame tiverem valores vazios. Na sequência, aplicaram-se ajustes de normalização nos resultados dos exames. Concatenou-se a coluna do nome exame com a do item do exame, pivotando esta nova coluna concatenada com o valor do resultado. Executou-se uma limpeza nas colunas EQU - Eritrocitos, EQU - Volume, EQU - Leucocitos, Tgp(Alt) - Tgp, Gasometria Arterial - Fio2, Kttp - Kttp, Kttp - Tempo De Tromboplastina, EQU - Ph, EQU - Densidade, Glicose - Glicose, Eritrograma - Rdw, Tempo De Protrombina - Atividade, Gasometria Arterial - So2, garantindo a existência apenas de valores numéricos. Após foi executada a união destas informações conforme estavam nos intervalos das primeiras 24, 48 ou 72 horas. Fez-se um agrupamento destas informações por atendimento do paciente onde, nas informações numéricas teve-se os valores máximos, mínimos, médias e contagens, obtendo-se, então, o agregado das informações textuais nas informações categóricas.

Referente aos exames de laboratório fez-se busca de informações (termos) tais como: escarro, abscesso, catéter, líquido pulmonar, secreção, fragmento, coleção, ferida, se o exame demonstra hemocultura positiva, sem crescimento bacteriológico, bacterioscopia positiva, bacterioscopia negativa, urocultura positiva, urocultura negativa, exame

bacteriológico positivo e negativo, mantendo apenas os dados referentes ao período estabelecido de internação, agrupando as informações de texto pelo código de atendimento do paciente.

Nas coletas realizadas pela equipe de enfermagem, constituída por enfermeiros e líderes de enfermagem, teve-se a quantidade de informações totais coletadas no período, de onde foram extraídas métricas de máximo, mínimo, primeiro, último, média, desvio padrão e contagens das seguintes informações: HGT (exame de glicemia), pressão arterial sistólica, pressão arterial diastólica, frequência respiratória, temperatura, níveis de dor, oximetria, frequência cardíaca, diurese, altura e peso. Assim como também foi inserido um atributo booleano que identifica como febril a temperatura maior que 38°C.

Nos textos das evoluções, laudos de exames de imagem e laudos de cirurgias foram agrupados os textos e analisada a quantidade de ocorrências de cada laudo ou evolução registrado no período. Buscou-se identificar, a partir desses textos agrupados alguns termos e comorbidades que foram indicativos diretamente relacionados a pacientes com critério grave utilizando-se expressões regulares, conforme relatado por Izquierdo et al. (2020), ao considerar este conjunto de sintomas e comorbidades como mais relevantes, a saber: hipertensão, diabetes, obesidade, carcinoma, bronquite, cirrose, tosse, febre, dispneia, estertor respiratório, disenteria, mialgia e cefaleia.

Estas transformações foram realizadas nos conjuntos de dados (Etapa 2 da Figura 4.1), fazendo união entre as colunas por internação do paciente (Etapa 4 da Figura 4.1). Resultaram conjuntos de dados com 424 atributos que foram utilizados nos modelos de dados finais, conforme o período de 24h, 48h ou 72h, sendo denominados, então, como: ‘dataset_label_24h’, ‘dataset_label_48h’ e ‘dataset_label_72h’, respectivamente.

A partir destes conjuntos de dados, então, foram separados os conjuntos para os treinamentos e validações (Etapa 5 da Figura 4.1).

Por fim, no que se refere aos dados faltantes, foi delimitado o valor constante de -999.0 para as células numéricas vazias, visto que é um valor cujo intervalo numérico não é apresentado em qualquer uma das *características* do modelo. Também, para os dados categóricos, foi feito o processo *Label Encoder* da biblioteca Scikit Learn⁵, onde as informações textuais categóricas foram convertidas em números.

⁵<https://scikit-learn.org/>

4.3 Otimização de Hiperparâmetros

Otimização de Hiperparâmetros é a etapa do processo de aprendizado de máquina que tem como objetivo otimizar a escolha dos parâmetros a fim de identificar a melhor combinação de valores possíveis que podem ser utilizados como parâmetro de entrada para o algoritmo que será utilizado (HO, 1995).

O método *Random Search* foi utilizado nesse processo. Através de um *grid* de parâmetros, o método faz combinações aleatórias para testar os modelos. A combinação que tiver a melhor entrada conforme a métrica especificada (AUROC, F1-Score, Acurácia, Precisão, etc) é a escolhida para os parâmetros do modelo, conforme assinala Bergstra e Bengio (2012).

4.4 Tecnologias utilizadas

Para gerenciar o armazenamento de todos os dados deste projeto foi escolhido o sistema gerenciador de banco de dados (SGBD) relacional PostgreSQL (versão 9).

Em código aberto (*open source*), o PostgreSQL se utiliza da linguagem SQL para fazer uma série de combinações de comandos para armazenar e manusear grandes cargas de dados, garantindo integridade e tolerância a falhas para os mais diversos projetos.

Para a realização dos experimentos deste projeto, foi utilizada a linguagem de programação Python⁶ (versão 3.9). A biblioteca Scikit Learn (versão 0.24.2) foi a escolhida para o processamento dos modelos de Inteligência Artificial, ajustes de hiperparâmetros (*RandomizedSearchCV*) e coleta das métricas. A biblioteca Pandas⁷ (versão 1.3.3) foi utilizada para a manipulação dos conjuntos de dados.

⁶<https://www.python.org/>

⁷<https://pandas.pydata.org/>

5 EXPERIMENTOS

Nesta seção estão descritos os experimentos realizados no estudo e os resultados das métricas de cada modelo, a fim de se definir seu poder preditivo.

Para início da análise, o primeiro passo foi a definição dos melhores parâmetros para cada modelo. Foi utilizado o método *Random Search* para definir qual a combinação vencedora de parâmetros que maximiza o valor de AUROC (WANG et al., 2015). A partir de um *grid* de possibilidades para cada modelo com intervalos já pré estabelecidos, todos os métodos foram submetidos a dez mil combinações geradas de forma aleatória com Validação Cruzada (*Cross Validation*) igual a cinco, gerando um total de cinquenta mil *fits* para cada modelo.

Referente ao método *Random Forest* foi estabelecido o *grid* com os valores de 50, 75, 100, 150, 250, 500 para a quantidade de árvores geradas. Para a profundidade limite de cada árvore teve-se valores entre sete e 26. A quantidade mínima de ramificações e folhas ficou estabelecida em um intervalo entre um e dez. As opções de critério de construção das árvores foram *Gini* e *Entropy*. Os pesos de classe estabelecidos foram: ‘0: 0.1’, ‘0: 0.01’, ‘0: 0.001’, ‘0: 0.05’, ‘0: 0.005’ e ‘balanceado’.

No método de *Gradient Boost* teve-se os números de estimadores definidos em um intervalo de cem a mil. Na taxa de aprendizagem e na subamostra teve-se o intervalo numérico entre zero e um. A profundidade máxima estabelecida ficou entre dois a vinte. A quantidade mínima de ramificações e folhas ficou estabelecida em um intervalo entre um e dezoito. O critério de escolha para o máximo de atributos foi ‘log2’ ou ‘raiz quadrada’.

5.1 Treinamento com Validação Cruzada

O processo de treinamento foi realizado utilizando-se, então, a escolha vencedora dos parâmetros com Validação Cruzada igual a dez, em um total de seis modelos.

Com base nos experimentos realizados a partir dos métodos de treinamento com Validação Cruzada igual a dez, utilizando o método de Youden (1950), que maximiza a relação entre Sensibilidade e Especificidade, para escolha de *threshold*, a partir dos melhores parâmetros obtidos com o método de *Random Search* foram obtidas as médias dos resultados mostrados na Tabela 5.1.

Como pode-se observar, todos os modelos tiveram qualidade para definir e classificar a classe positiva, especialmente o modelo *Gradient Boost* com o conjunto de dados

Tabela 5.1 – Resultados dos testes com Validação Cruzada

	Modelo	Precisão	Sensibilidade	Especificidade	Acurácia	F1-Score	AUROC
1	RF-24h	0.608	0.787	0.892	0.874	0.682	0.914
2	GB-24h	0.588	0.795	0.882	0.867	0.671	0.915
3	RF-48h	0.653	0.800	0.885	0.867	0.712	0.914
4	GB-48h	0.696	0.802	0.910	0.888	0.743	0.917
5	RF-72h	0.710	0.790	0.906	0.881	0.745	0.910
6	GB-72h	0.676	0.779	0.889	0.865	0.719	0.894

Fonte: O autor.

das 48h de atendimento (linha 4), com valor de 80.2%, o qual também obteve as melhores métricas de Especificidade, Acurácia e AUROC, com 91%, 88.8% e 91.7% respectivamente. O modelo *Random Forest* com o conjunto de dados das primeiras 72h (linha 5) obteve a melhor Precisão com um valor de 71% e a melhor métrica F1-score com um valor de 74.5%.

Os modelos de 48 e 72 horas (linhas 3, 4, 5 e 6) mostraram uma melhora significativa nos valores de Precisão e, conseqüentemente, F1-Score. O aumento de informações de internação para os pacientes auxiliou o modelo, de forma geral, a acertar mais a classe negativa. Os modelos de 48 horas (linhas 3 e 4) tiveram uma ligeira melhora na Sensibilidade, demonstrando que existiu um balanceamento favorável entre a quantidade de dados para aprendizado com relação à classe positiva.

Também notou-se que o modelo *Gradient Boost* em 72 horas (linha 6) teve perda de efetividade, visto que, em relação à AUROC e à Sensibilidade, teve um desempenho levemente inferior se comparado aos outros modelos.

Os modelos de 48h e 72h apresentaram um aumento de Precisão e F1-Score, demonstrando um comportamento semelhante se comparado aos modelos de 24h.

5.2 Treinamento e Validação

Na sequência, analisou-se os modelos considerando treinamento e validação (Tabela 5.2). O treinamento foi realizado utilizando-se todo o conjunto de dados de treinamento previamente definido. A análise posterior de predição foi realizada com um conjunto de dados diferente daquele da etapa de treinamento.

Percebe-se que houve um aumento da AUROC de todos os modelos, se comparada a etapa de Treinamento e Teste, com exceção do modelo *Gradient Boost* em 48 horas (linha 4) que teve uma pequena piora de efetividade, principalmente pela perda de

Tabela 5.2 – Resultados dos experimentos em Validação.

	Modelo	Precisão	Sensibilidade	Especificidade	Acurácia	F1-Score	AUROC
1	RF-24h	0.573	0.861	0.828	0.835	0.688	0.922
2	GB-24h	0.610	0.816	0.861	0.851	0.699	0.927
3	RF-48h	0.670	0.799	0.882	0.863	0.729	0.924
4	GB-48h	0.712	0.781	0.905	0.876	0.744	0.913
5	RF-72h	0.701	0.814	0.887	0.869	0.754	0.922
6	GB-72h	0.656	0.790	0.865	0.847	0.717	0.898

Fonte: O autor.

Sensibilidade e Especificidade.

A melhor AUROC foi obtida pelo modelo *Random Forest* em 48 horas (linha 3) 92.4%, enquanto as melhores métricas de Especificidade e Acurácia ocorreram no modelo *Gradient Boost* 48 horas (linha 4) 90.5% e 87.6%, respectivamente. O maior valor para Sensibilidade foi atingido pelo modelo *Random Forest* 24 horas (linha 1), com um valor de 86.1% porém, este modelo obteve a pior métrica para Precisão com 57.3%. Os maiores valores de Precisão e F1-Score foram atingidos pelo modelo *Random Forest* de 72 horas (linha 5), com valores de 70.1% e 75.4%, respectivamente. Observou-se também que o modelo *Gradient Boost* de 72 horas (linha 6) obteve uma perda de efetividade em sua AUROC.

Com os resultados dessa etapa, foi feita a seleção dos melhores modelos com base no maior valor de F1-Score conforme o período, os quais estão analisados no Capítulo 6. Os modelos considerados vencedores foram: GB-24h (Linha 2), GB-48h (Linha 4) e RF-72h (Linha 5).

6 ANÁLISE DOS RESULTADOS OBTIDOS

A partir da seleção dos modelos vencedores de cada período (GB-24h, GB-48h e RF-72h), inicialmente, foi analisada a importância dos atributos conforme reportada pelos modelos.

Nas Figuras 6.1, 6.2 e 6.3 percebeu-se que todos os atributos têm validação médica ou estatística e são possíveis variáveis que indicam se o paciente pode necessitar de UTI hospitalar. Também é possível notar que os atributos de temperatura, contagens de evolução, laudo de imagem e sinais vitais, frequência respiratória, HGT, pressão arterial diastólica e pressão arterial sistólica aparecem nos modelos como informações de grande importância para a tomada da decisão, com a informação de temperatura (*Temp_last*) aparecendo no *top* dois para os três modelos, indicando uma correlação elevada com a classificação preditiva final dos modelos.

Outro atributo que demonstrou relevância foi a informação de idade. Com isto, nas Figuras 6.4, 6.5 e 6.6, teve-se a avaliação da performance preditiva deste modelo com cada modelo agregando pacientes em grupos de idade entre [18, 20), [20, 30), [30, 40), [40, 50), [50, 60), [60, 70), [70, 80), [80, 90), 90+. É possível notar que, com o aumento da idade, as métricas perdem performance, tanto para os modelos de *Gradient Boost* 24h (GB-24h), *Gradient Boost* 48h (GB-48h) e *Random Forest* 72h (RF-72h). As métricas mais penalizadas para os modelos de GB-24h e GB-48h são a Precisão e o F1-Score, enquanto para o modelo de RF-72h é a Sensibilidade.

Foram também analisadas as métricas conforme gênero. A Figura 6.7 mostra que os modelos tem maior Especificidade e Acurácia para o sexo feminino, se comparado com o sexo masculino. Notou-se que todas as métricas melhoraram à medida que o intervalo de informações aumentou (24h, 48h e 72h), com exceção da AUROC no modelo *Gradient Boost* com 24h para o sexo feminino, que atingiu um valor de 94%. O modelo *Gradient Boost* com 48h para o sexo feminino obteve, por sua vez, a menor métrica de Sensibilidade com 75%.

Na sequência, foi analisado o comportamento do modelo com pacientes que realizaram ou não o exame de PCR para teste de COVID-19 no período da internação, em função do estudo ter ocorrido durante período de pandemia do novo Coronavírus. As Figuras 6.8, 6.9 e 6.10 mostram que os modelos foram capazes de prever a classe positiva com precisão. Na Figuras 6.8 notou-se que os pacientes que tiveram COVID-19 detectado através do exame PCR, apresentaram alta performance em suas métricas, tanto para defi-

Figura 6.1 – Importância de variáveis no modelo Gradient Boost 24h

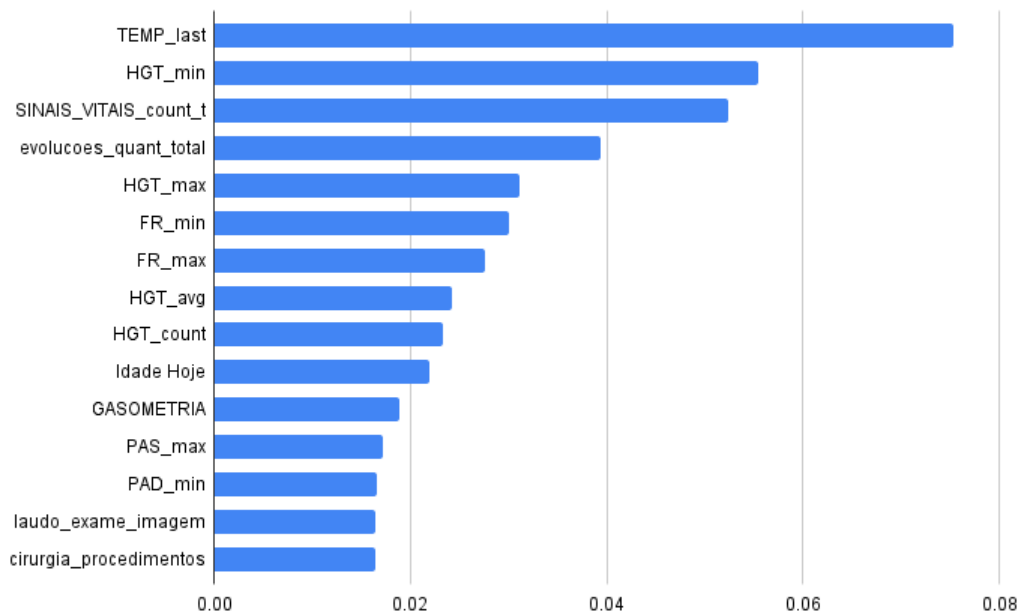


Figura 6.2 – Importância de variáveis no modelo Gradient Boost 48h

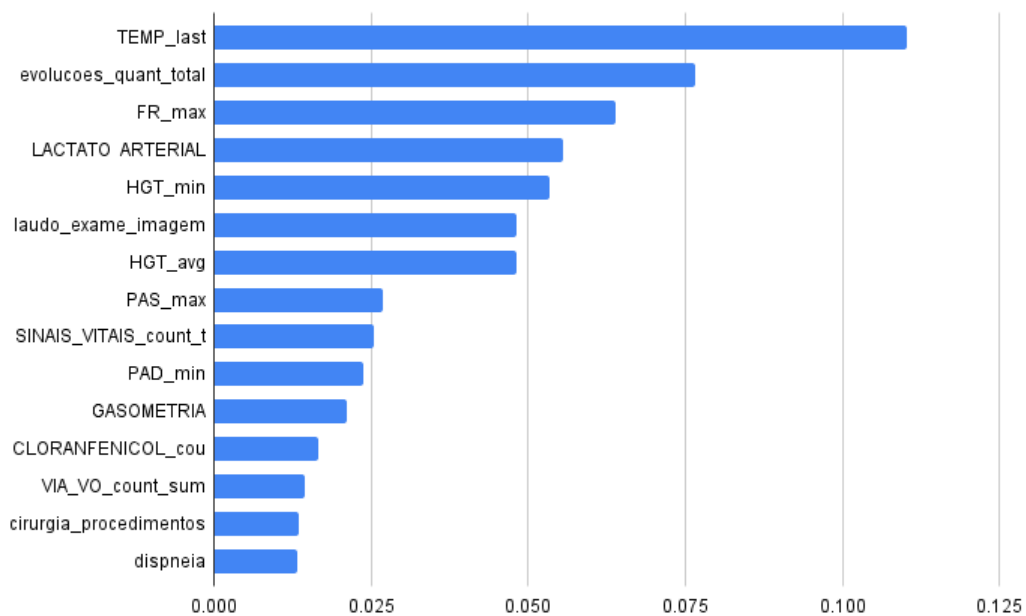
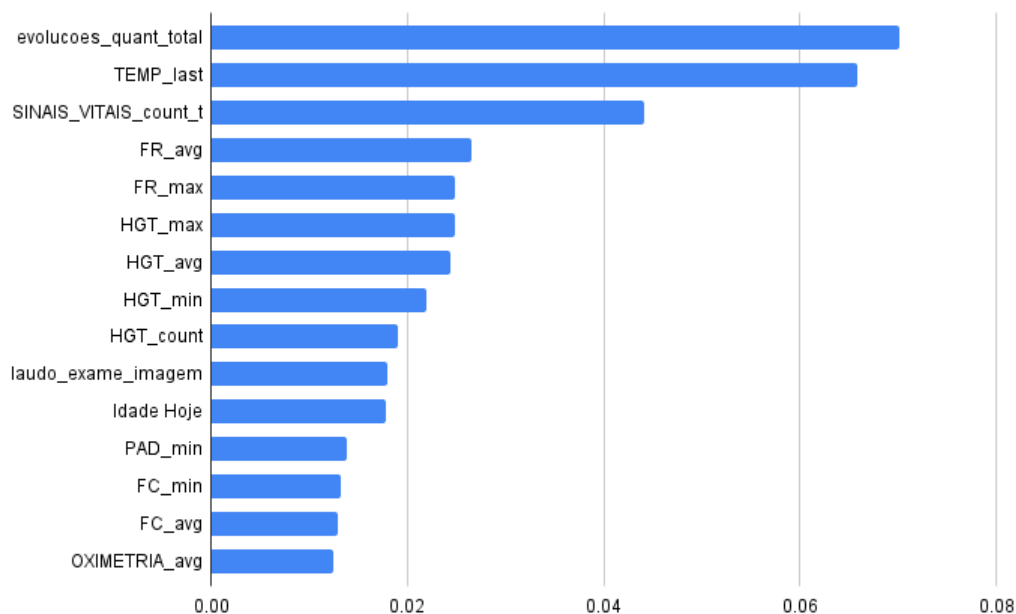


Figura 6.3 – Importância de variáveis no modelo Random Forest 72h



nição da classe positiva quanto negativa, enquanto os que não tiveram o exame do PCR positivo apresentaram uma redução das métricas de Precisão, Sensibilidade e F1-Score.

As Figuras 6.9 e 6.10, tiveram um comportamento em suas métricas inferior ao apresentado na Figura 6.8 (modelo GB-24h) para pacientes com COVID-19 detectado. Também pode-se notar que este atributo é uma informação que reduz a capacidade do modelo em prever a classe negativa, indicando que depende de outras informações para melhorar a Especificidade. Comparando as Figuras 6.9 e 6.10, percebeu-se que o modelo de GB-48h (Figura 6.9) tem melhores resultados para as métricas de AUROC, Precisão, Especificidade, Acurácia e F1-Score para pacientes com COVID-19 detectado. Já as métricas do RF-72h (Figura 6.10) apresentaram uma ligeira melhora em AUROC, Sensibilidade, Especificidade e F1-Score para pacientes não detectados.

Por fim, as Figuras 6.11b, 6.12b e 6.13a apresentam as matrizes de confusão dos modelos vencedores. Para o modelo GB-24h (Figura 6.11b) os verdadeiros positivos foram 489, verdadeiros negativos foram 1927, falsos positivos foram 312 e falsos negativos foram 110. No que se refere ao modelo GB-48h (Figura 6.12b), os verdadeiros positivos foram 385, verdadeiros negativos foram 1487, falsos positivos foram 156 e falsos negativos foram 108, enquanto para o modelo RF-72h (Figura 6.13a), os verdadeiros positivos foram 338, verdadeiros negativos foram 1129, falsos positivos foram 144 e falsos negativos foram 77.

Figura 6.4 – Métricas por grupos de idade no Gradient Boost 24h

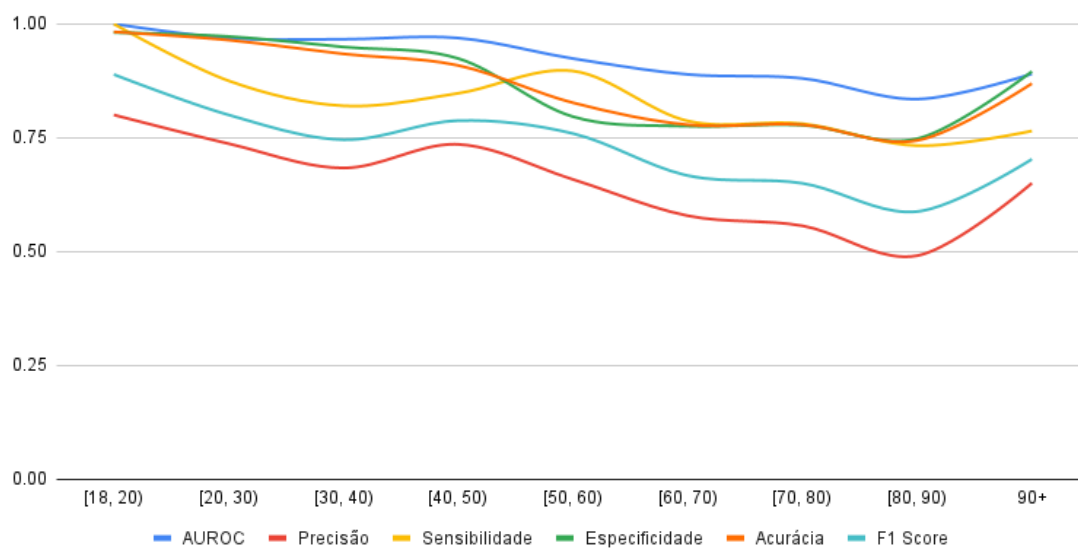


Figura 6.5 – Métricas por grupos de idade utilizando Gradient Boost 48h

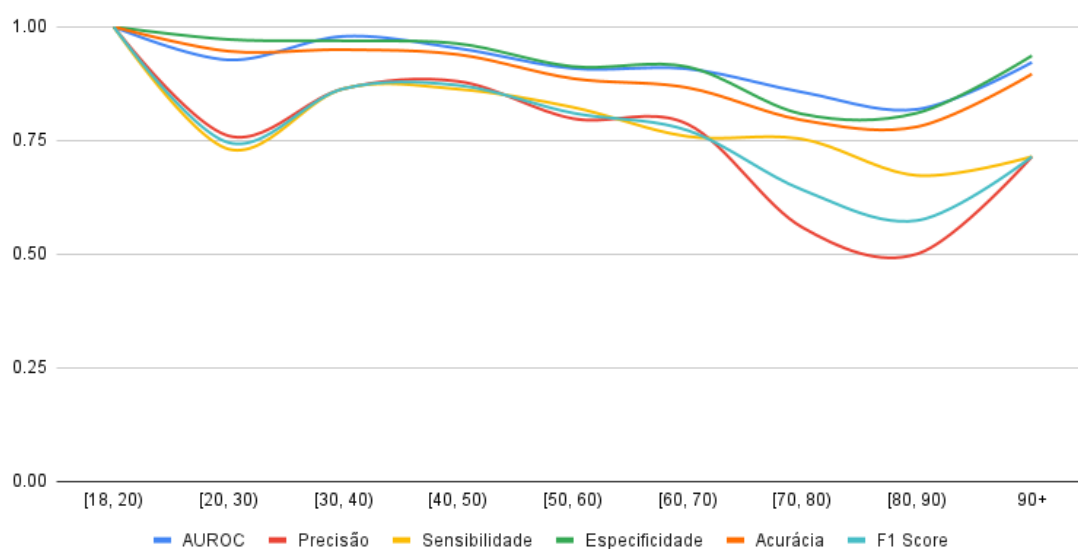


Figura 6.6 – Métricas por grupos de idade no Random Forest 72h

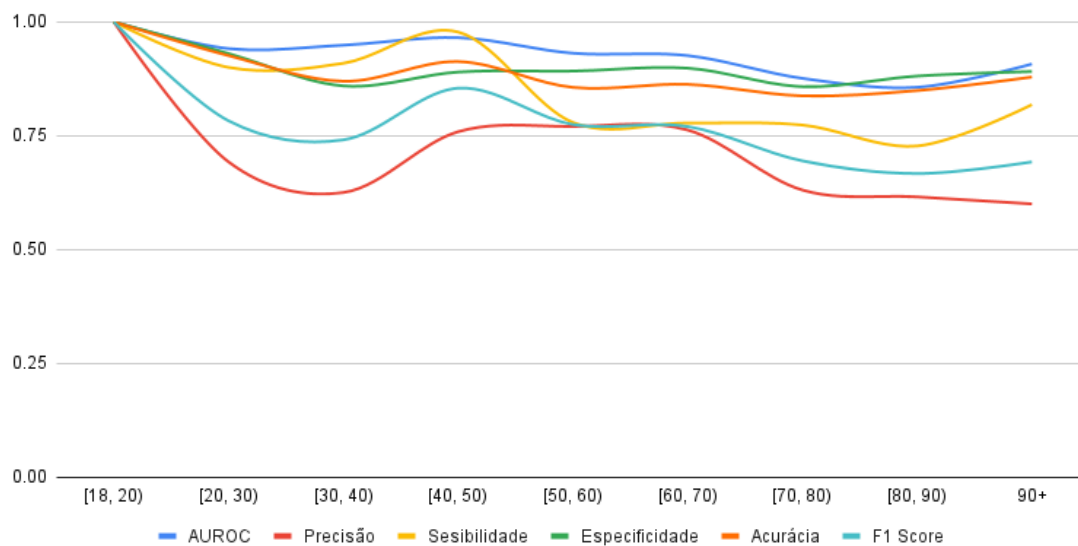


Figura 6.7 – Métricas conforme gênero

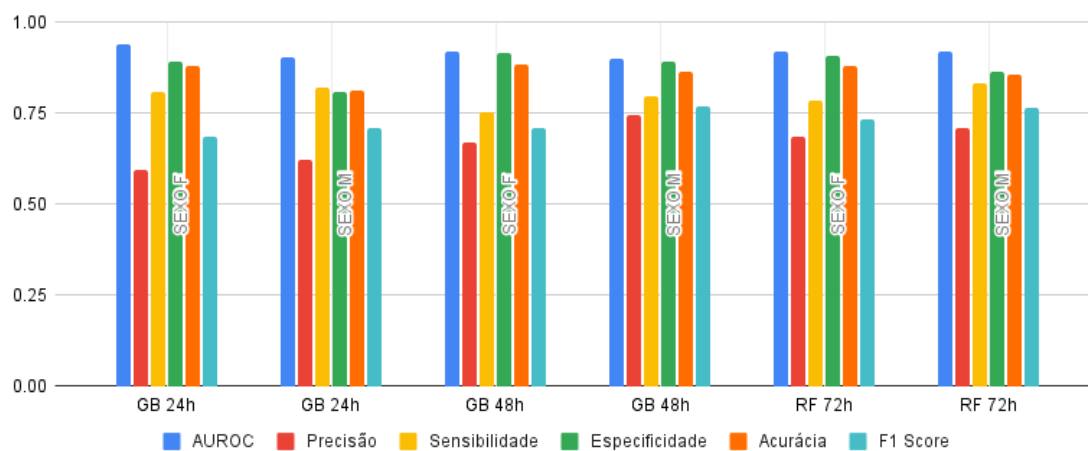


Figura 6.8 – Métricas para COVID-19 no algoritmo Gradient Boost 24h

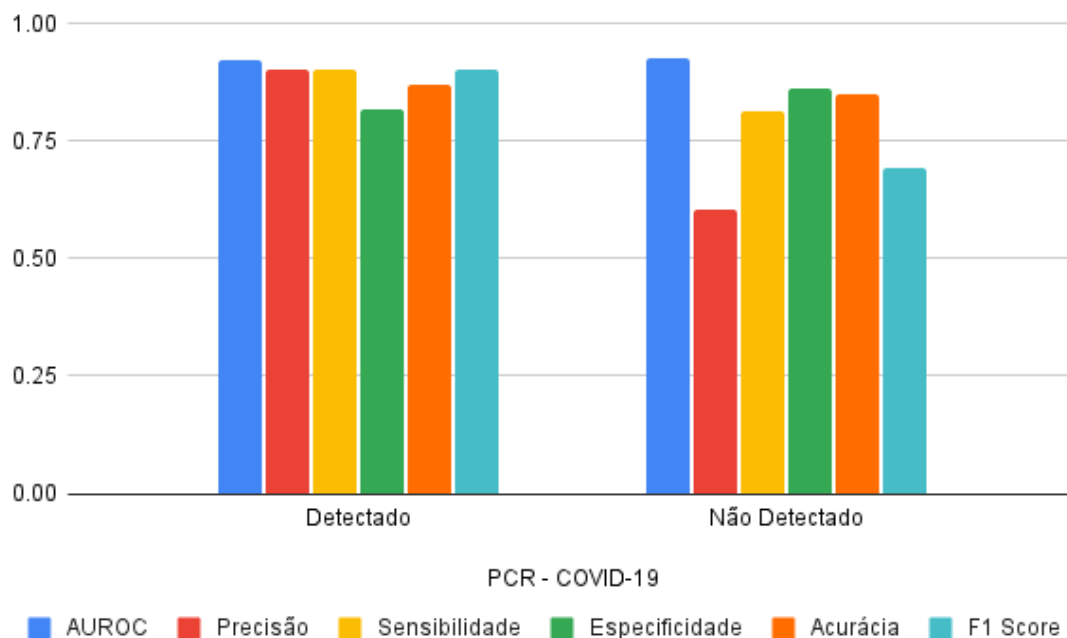


Figura 6.9 – Métricas para COVID-19 no uso do algoritmo Gradient Boost 48h

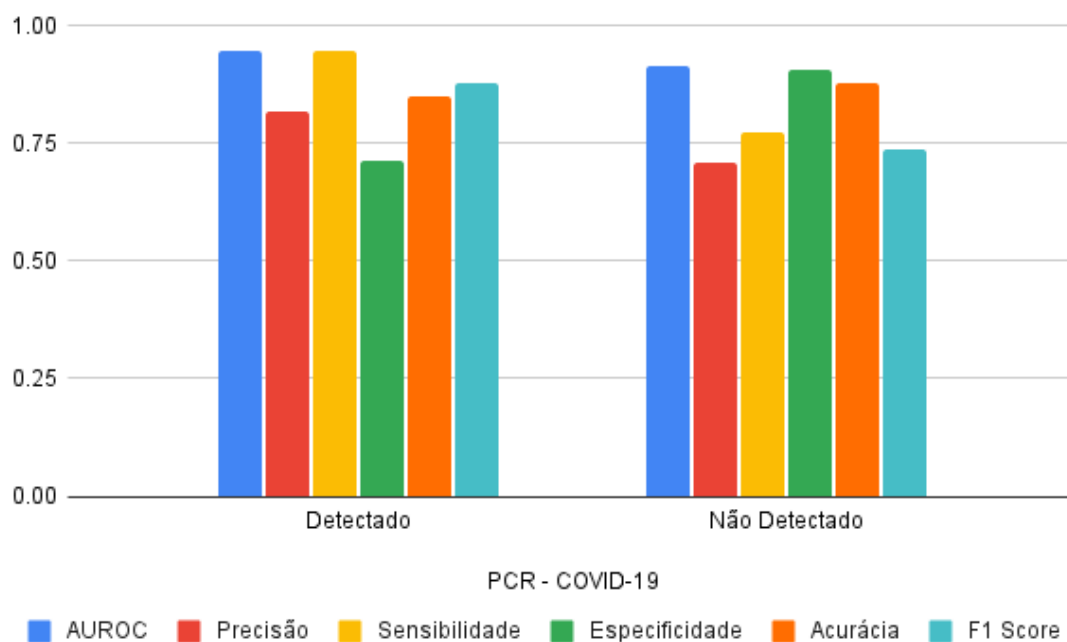


Figura 6.10 – Métricas para COVID-19 no modelo Random Forest 72h

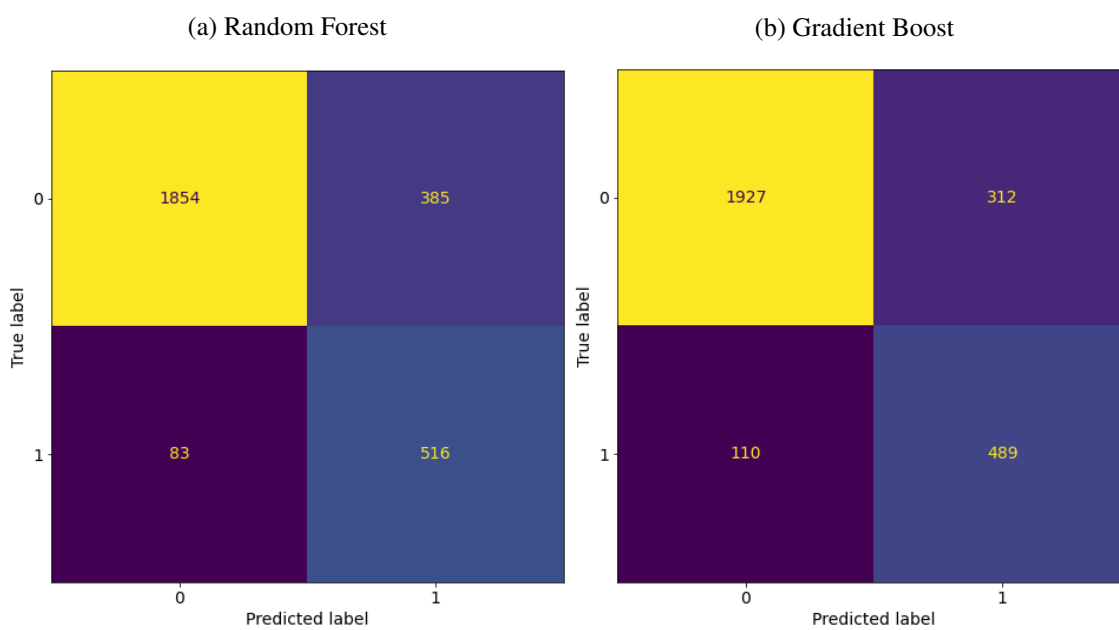
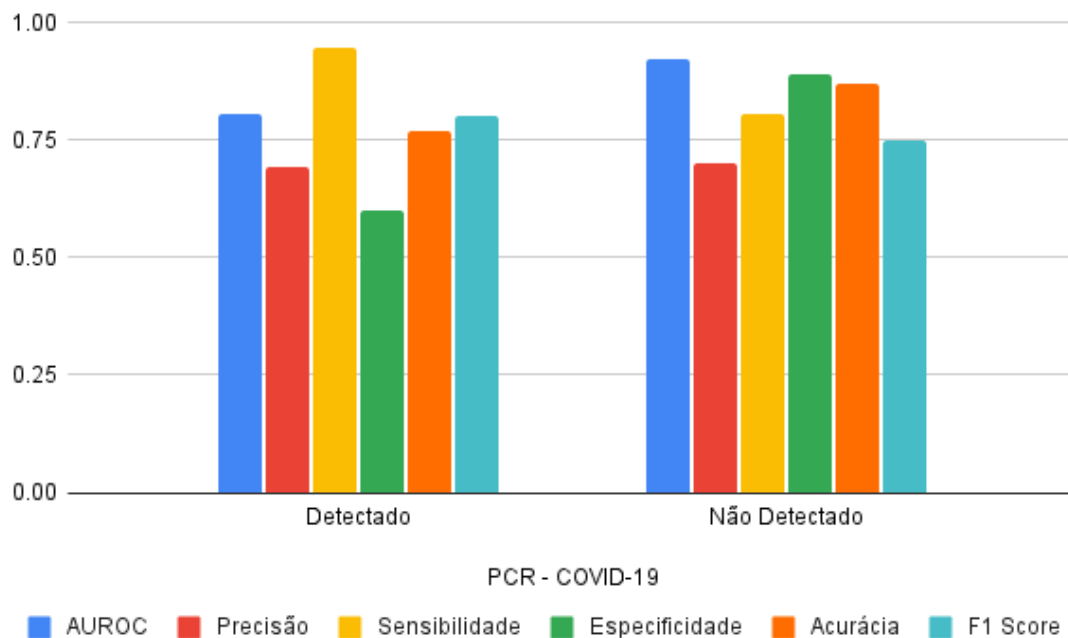


Figura 6.11 – Matrizes de confusão decorrentes do modelo 24h

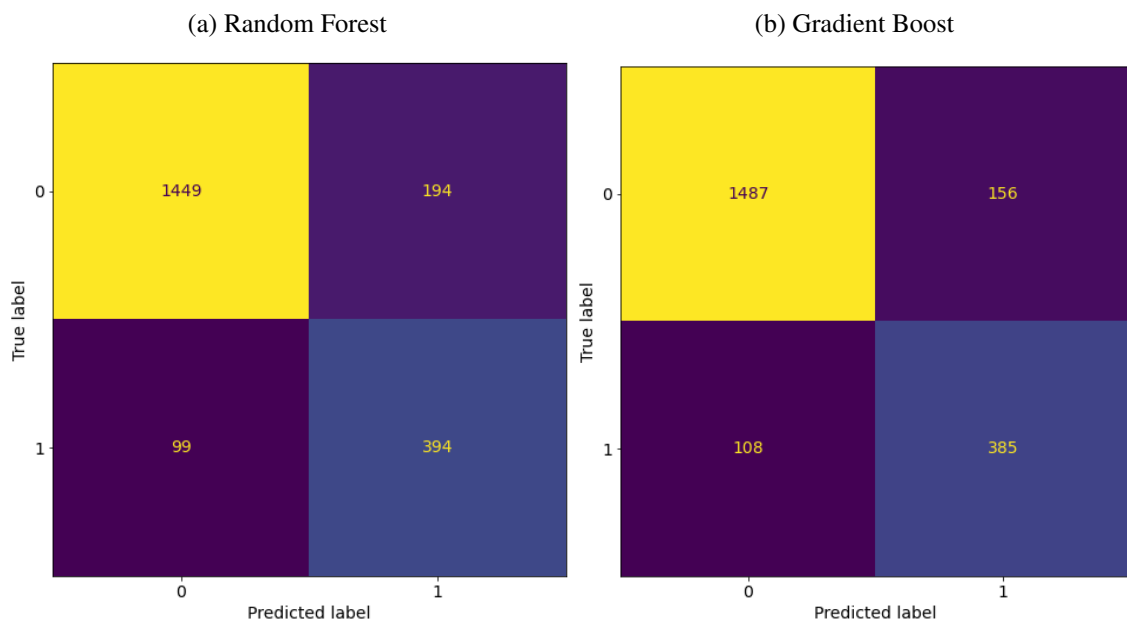


Figura 6.12 – Matrizes de confusão decorrentes do modelo 48h

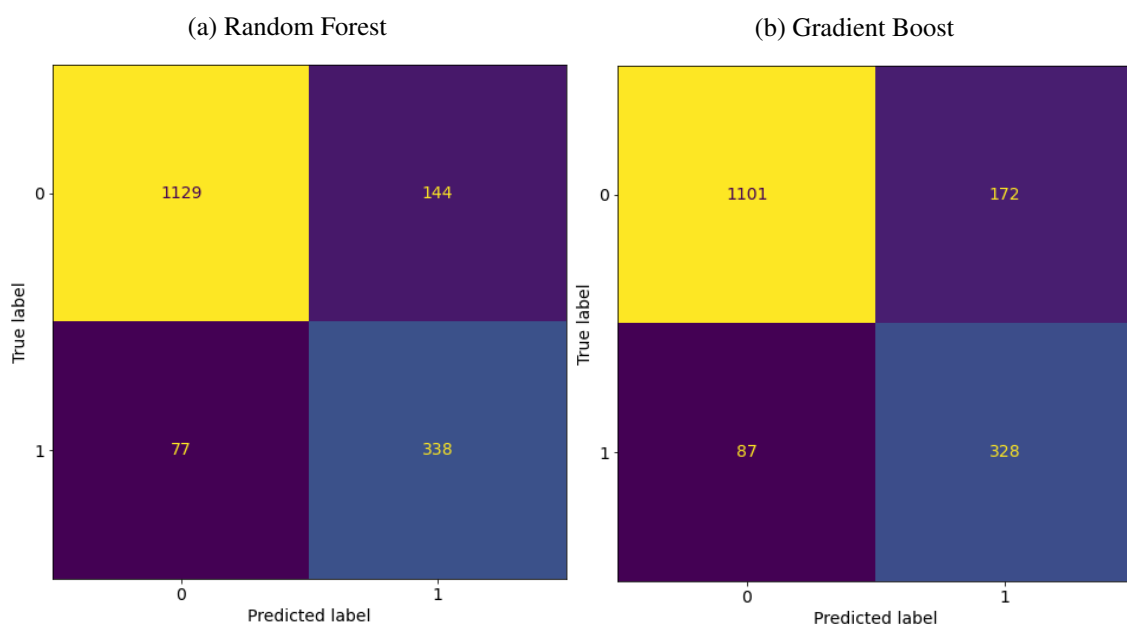


Figura 6.13 – Matrizes de confusão decorrentes do modelo 72h

7 CONCLUSÕES

Este trabalho testou métodos de Aprendizado de Máquina em dados estruturados e não estruturados de uma base de dados hospitalar. Com informações hospitalares de pacientes internados em um intervalo de treze meses durante um período da pandemia COVID-19, os métodos demonstraram qualidade em definir quais pacientes necessitaram de atendimento especial em UTI hospitalar, mesmo esta classe sendo desbalanceada - classe positiva correspondente a 21,12%, 27,25% e 32,94% das instâncias dos conjuntos de dados de 24 horas, 48 horas e 72 horas, respectivamente.

Os modelos de dados foram criados a partir de diferentes estruturas do hospital, contendo informações de prescrições de medicamentos, exames hospitalares, laudos de exame de imagem, sinais vitais, evoluções hospitalares e dados pessoais do paciente. Com transformações e extrações de novas métricas para a criação de novos atributos para o modelo, a metodologia aplicada visou seguir conceitos clínicos e metodologias da estatística para elaboração de modelos que pudessem ser processados por diversos métodos de Aprendizado de Máquina.

A partir da avaliação dos resultados, tanto nos testes com Validação Cruzada e em Teste e Validação, foi possível concluir que os modelos foram consistentes e eficazes em definir tanto a classe positiva quanto a negativa. Nos testes com Validação Cruzada, o modelo *Random Forest* com 72h de informações obteve as melhores métricas de precisão (71%) e F1-Score (74.5%) enquanto o modelo *Gradient Boost* com 48 horas de informação demonstrou as melhores métricas de Sensibilidade (80.2%), Especificidade (91%), Acurácia (88.8%) e AUROC (91.7%). Em validação, as maiores métricas de Precisão, Especificidade e Acurácia ficaram com o modelo de *Gradient Boost* em 48h de informações, com os valores de 71.2%, 90.5% e 87.6%, respectivamente. A melhor métrica de Sensibilidade foi constatada no modelo de *Random Forest* em 24h (86.1%) e o maior valor para AUROC foi obtido pelo modelo *Gradient Boost* de 24h (92.7%). Estes modelos são considerados “não caixa preta”, pelo fato de exibirem as árvores das quais se geraram as probabilidades e listagem da importância dos atributos. Pode-se perceber assim que os modelos seguiram critérios com embasamento clínico para a tomada de decisão. Esta decisão pode ser revisada e comprovada por profissionais da saúde a qualquer momento.

Diferentemente dos estudos realizados por Araz, Olson e Ramirez-Nafarrate (2019) e Fernandes et al. (2020), onde foram utilizados dados de Triagem para prever a necessidade de leitos dos pacientes, este estudo buscou identificar pacientes que pudessem vir

a ter necessidade de leitos de UTI hospitalar, através das análises de conjuntos de dados baseados em intervalos de internação prévia dos pacientes (24 horas, 48 horas e 72 horas). Enquanto Araz, Olson e Ramirez-Nafarrate (2019) atingiu valores de AUROC de 86.3% e Acurácia de 82.1% com o algoritmo *XGBoost*, Fernandes et al. (2020) atingiu os valores de AUROC e Acurácia de 91% e 86%, respectivamente, com o algoritmo de *Logistic Regression*. No contexto da COVID-19, Izquierdo et al. (2020) apresenta uma proposta com o uso de PLN e *Word Embeddings* em dados de texto para identificar fatores relacionados a pacientes com COVID-19 em CTI hospitalar atingindo métricas de AUROC 84%, Acurácia 85% e Sensibilidade de 57% com o algoritmo *Decision Trees*, ainda identificando uma combinação de fatores clínicos atrelados a classe positiva: idade, febre e taquipneia.

Possíveis melhorias para este trabalho podem focar em informações que auxiliassem na melhoria da Precisão dos modelos. Incluir outros atributos, que possibilitem enunciar outras comorbidades, ou que possam ter correlação maior com o valor preditivo (como sepsis, por exemplo), algoritmos de comitês de classificação (*ensemble methods*), critérios para a escolha de atributos (como qui-quadrado, F-measure, *Variance Inflation Factor*, etc), técnicas de subamostragem, outros métodos para busca de hiperparâmetros, uso de redes neurais ou outras técnicas de Aprendizado de Máquina e refazer a análise com base na métrica de F2-Score (ou $F\beta$ - Score), em função das classes desbalanceadas que constituem os conjuntos de dados.

Este trabalho experimental permite a reflexão sobre a viabilidade de aplicação de algoritmos de Inteligência Artificial na área da saúde, os quais, a partir de melhorias e mudanças com validação clínica, estatística e ética, possam se tornar um auxiliar de gestores hospitalares na agilização da oferta de tratamento, minimização de custos e, espera-se, melhoria das chances de sobrevivência do paciente.

REFERÊNCIAS

- ARAZ, O. M.; OLSON, D.; RAMIREZ-NAFARRATE, A. **Predictive analytics for hospital admissions from the emergency department using triage information.** 2019. 199–207 p.
- BEKKAR, M.; DJEMAA, H. K.; ALITOUCHE, T. A. Evaluation measures for models assessment over imbalanced data sets. **J Inf Eng Appl**, v. 3, n. 10, 2013. Available from Internet: <https://www.researchgate.net/publication/292718336_Evaluation_measures_for_models_assessment_over_imbalanced_data_sets>. Accessed 19 Sep. 2021.
- BERENGUER, J. et al. Characteristics and predictors of death among 4035 consecutively hospitalized patients with COVID-19 in Spain. **Clin. Microbiol. Infect.**, v. 26, n. 11, p. 1525–1536, nov. 2020. Available from Internet: <<https://onlinelibrary.wiley.com/doi/10.1111/eci.13362>>. Accessed 3 Sep. 2021.
- BERGSTRA, J.; BENGIO, Y. **Random search for hyper-parameter optimization.** 2012. <<https://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a>>. Accessed 21 Oct. 2021.
- BERNSTEIN, S. L. et al. **The Effect of Emergency Department Crowding on Clinically Oriented Outcomes.** 2009. 1–10 p.
- CHALFIN, D. B. et al. Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit. **Crit. Care Med.**, v. 35, n. 6, p. 1477–1483, jun. 2007.
- CHOU, C.-A. et al. Mixed-integer optimization approach to learning association rules for unplanned ICU transfer. **Artif. Intell. Med.**, v. 103, p. 101806, mar. 2020.
- COWAN, R. M.; TRZECIAK, S. Clinical review: Emergency department overcrowding and the potential impact on the critically ill. **Crit. Care**, v. 9, n. 3, p. 291–295, jun. 2005. Available from Internet: <<https://ccforum.biomedcentral.com/articles/10.1186/cc2981>>. Accessed 02 Nov. 2021.
- DATAIKU. **Dataiku.** 2021. <<https://www.dataiku.com/>>. Accessed 06 Jun. 2021.
- DUGAS, A. F. et al. Google flu trends: correlation with emergency department influenza rates and crowding metrics. **Clin. Infect. Dis.**, v. 54, n. 4, p. 463–469, fev. 2012.
- FERNANDES, M. et al. Predicting intensive care unit admission among patients presenting to the emergency department using machine learning and natural language processing. **PLoS One**, v. 15, n. 3, p. e0229331, mar. 2020. Available from Internet: <<https://pubmed.ncbi.nlm.nih.gov/32126097/>>. Accessed 8 Sep. 2021.
- FIGLIOZZI, S. et al. Predictors of adverse prognosis in COVID-19: A systematic review and meta-analysis. **Eur. J. Clin. Invest.**, v. 50, n. 10, p. e13362, out. 2020. Available from Internet: <<https://onlinelibrary.wiley.com/doi/10.1111/eci.13362>>. Accessed 3 Sep. 2021.

FRIEDMAN, J. H. Greedy function approximation: A gradient boosting machine. **Ann. Stat.**, Institute of Mathematical Statistics, v. 29, n. 5, p. 1189–1232, 2001. Available from Internet: <<https://statweb.stanford.edu/jhf/ftp/trebst.pdf>>. Accessed 01 Nov. 2021.

GRAWE, M. F. **Heterogeneous Ensemble Models for In-Hospital Mortality Prediction**. 2021. 53 p. PPGC da UFRGS (dissertação de mestrado).

Healthcare Administration EDU. **What is Health and Medical Administration?** 2021. <<https://www.healthcareadministrationedu.org/what-is-health-and-medical-administration/>>. Accessed 02 Nov. 2021.

HERD, A. M. et al. Finding what works: Leadership competencies for the changing healthcare environment. **Educ. Leadersh.**, v. 15, n. 4, 2016. Available from Internet: <https://journalofleadershiped.org/wp-content/uploads/2019/02/15_4_Adams_Pope.pdf>. Accessed 22 Oct. 2021.

HO, T. K. Random decision forests. In: **Proceedings of 3rd International Conference on Document Analysis and Recognition**. [S.l.: s.n.], 1995. v. 1, p. 278–282 vol.1.

HOOT, N. R.; ARONSKY, D. Systematic review of emergency department crowding: causes, effects, and solutions. **Ann. Emerg. Med.**, v. 52, n. 2, p. 126–136, ago. 2008.

IACCARINO, G. et al. Gender differences in predictors of intensive care units admission among COVID-19 patients: The results of the SARS-RAS study of the Italian Society of Hypertension. **PLoS One**, v. 15, n. 10, p. e0237297, out. 2020. Available from Internet: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0237297>>. Accessed 3 Sep. 2021.

IZQUIERDO, J. L. et al. Clinical characteristics and prognostic factors for intensive care unit admission of patients with COVID-19: Retrospective study using machine learning and natural language processing. **J. Med. Internet Res.**, v. 22, n. 10, p. e21801, out. 2020. Available from Internet: <<http://dx.doi.org/10.2196/21801>>. Accessed 7 Sep. 2021.

LALKHEN, A. G.; MCCLUSKEY, A. Clinical tests: sensitivity and specificity. **Contin Educ Anaesth Crit Care Pain**, Oxford Academic, v. 8, n. 6, p. 221–223, dez. 2008. Available from Internet: <<https://www.sciencedirect.com/science/article/pii/S1743181617304572?via%3Dihub>>. Accessed 01 Nov. 2021.

LAMBE, S. et al. Waiting times in California's emergency departments. **Ann. Emerg. Med.**, v. 41, n. 1, p. 35–44, jan. 2003. Available from Internet: <[https://www.annemergmed.com/article/S0196-0644\(02\)84925-3/fulltext](https://www.annemergmed.com/article/S0196-0644(02)84925-3/fulltext)>. Accessed 02 Nov. 2021.

LEVIN, S. R. et al. Optimizing cardiology capacity to reduce emergency department boarding: a systems engineering approach. **Am. Heart J.**, v. 156, n. 6, p. 1202–1209, dez. 2008.

PATTERSON, T. E. et al. Addressing the leadership gap in healthcare: What's needed when it comes to leader talent. **Centre for Creative Leadership**, 2016.

POLEVOI, S. K.; QUINN, J. V.; KRAMER, N. R. Factors associated with patients who leave without being seen. **Acad. Emerg. Med.**, v. 12, n. 3, p. 232–236, mar. 2005.

POLIKAR, R. Ensemble based systems in decision making. **IEEE Circuits and Systems Magazine**, v. 6, n. 3, p. 21–45, 2006.

PostgreSQL. **PostgreSQL**. 2021. <<https://www.postgresql.org/>>. Accessed 19 Sep. 2021.

PROVOST, F. Machine learning from imbalanced data sets 101. In: **Proceedings of the AAAI'2000 workshop on imbalanced data sets**. [S.l.: s.n.], 2000. v. 68, p. 1–3. Available from Internet: <<https://www.aaai.org/Papers/Workshops/2000/WS-00-05/WS00-05-001.pdf>>. Accessed 19 Sep. 2021.

SANTOS, R. P. D. et al. Automated healthcare-associated infection surveillance using an artificial intelligence algorithm. **Infect Prev Pract**, v. 3, n. 3, p. 100167, set. 2021. Available from Internet: <<https://pubmed.ncbi.nlm.nih.gov/32126097/>>. Accessed 1 Sep. 2021.

SINGER, A. J. et al. The association between length of emergency department boarding and mortality. **Acad. Emerg. Med.**, v. 18, n. 12, p. 1324–1329, dez. 2011.

SWE, T. T. Analysis of tree based supervised learning algorithms on medical data. **Math. Sci. Res. J.**, International Journal of Scientific and Research Publications (IJSRP), v. 9, n. 4, p. 8817, abr. 2019.

TING, K. M.; WITTEN, I. H. Issues in stacked generalization. **J. Artif. Intell. Res.**, v. 10, p. 271–289, maio 1999.

UTGOFF, P. E. Incremental induction of decision trees. **Mach. Learn.**, v. 4, n. 2, p. 161–186, nov. 1989. Available from Internet: <<https://link.springer.com/article/10.1023%2FA%3A1022699900025>>. Accessed 06 Oct. 2021.

VAART, T. van der; VASTAG, G.; WIJNGAARD, J. Facets of operational performance in an emergency room (ER). **Int. J. Prod. Econ.**, v. 133, n. 1, p. 201–211, set. 2011.

WANG, S. et al. Optimizing area under the ROC curve using semi-supervised learning. **Pattern Recognit.**, v. 48, n. 1, p. 276–287, jan. 2015. Available from Internet: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4226543/>>. Accessed 24 Oct. 2021.

WOLPERT, D. H. Stacked generalization. **Neural Netw.**, v. 5, n. 2, p. 241–259, jan. 1992.

YOU DEN, W. J. Index for rating diagnostic tests. **Cancer**, v. 3, n. 1, p. 32–35, jan. 1950. Available from Internet: <[https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/1097-0142\(1950\)3:1%3C32::AID-CNCR2820030106%3E3.0.CO;2-3](https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/1097-0142(1950)3:1%3C32::AID-CNCR2820030106%3E3.0.CO;2-3)>. Accessed 19 Sep. 2021.

APÊNDICE A — GLOSSÁRIO

A.1 Atributos de acordo com os modelos adotados

Os modelos adotados no presente estudo constituem-se de 421 atributos numéricos e categóricos, os quais são nominados a seguir:

- Numéricas: 375 atributos

'NEOMICINA_count_sum', 'ERITROGRAMA - ERITROCITOS_avg', 'FR_count', 'TEMP_avg', 'GASOMETRIA VENOSA - TCO2_min', 'FR_avg', 'hipertensao', 'MAGNESIO - MAGNESIO_max', 'CEFUROXIMA_count_sum', 'FC_max', 'GASOMETRIA ARTERIAL - SO2_max', 'ENFERMEIRO_A__first', 'PIRIMETAMINA_count_sum', 'GASOMETRIA VENOSA - HCO3_avg', 'GASOMETRIA ARTERIAL - PH_count', 'DOR_first', 'ERITROGRAMA - MCHC_avg', 'ALT_max', 'VIA_SE_count_sum', 'ANFOTERICINA_B_count_sum', 'TEMP_first', 'PAS_count', 'ERITROGRAMA - MCHC_min', 'FC_last', 'ERITROGRAMA - HEMOGLOBINA_min', 'GENTAMICINA_count_sum', 'LIDER DE ENFERMAGEM_last', 'SODIO - SODIO_count', 'qtd_hemocultura_positiva', 'DOR_count', 'PENICILINA_BENZATINA_count_sum', 'GASOMETRIA ARTERIAL - PCO2_max', 'ACICLOVIR_count_sum', 'cirurgias_tempo_total', 'CEFTAZIDIMA_count_sum', 'NITROFURANTOINA_count_sum', 'GASOMETRIA VENOSA - HCO3_max', 'GASOMETRIA VENOSA - TCO2_count', 'GASOMETRIA ARTERIAL - FIO2_min', 'GASOMETRIA ARTERIAL - TCO2_max', 'cefaleia', 'EQU - DENSIDADE_avg', 'TEMPO DE PROTROMBINA - RNI_count', 'carcinoma', 'SULFAMETOXASOL_TRIMETOPRIM_count_sum', 'ERITROGRAMA - MCV_count', 'GASOMETRIA VENOSA - %SO2C_count', 'MEROPENEM_count_sum', 'BILIRRUBINAS TOTAL E FRACOES - BILIRRUBINA DIRETA_avg', 'MICAfungina_count_sum', 'UREIA - UREIA_avg', 'EQU - ERITROCITOS_min', 'EQU - DENSIDADE_count', 'VIA_VV_count_sum', 'HGT_first', 'GASOMETRIA VENOSA - SO2_max', 'TEMPO DE PROTROMBINA - RNI_min', 'GLICOSE - GLICOSE_avg', 'GASOMETRIA ARTERIAL - TCO2_count', 'ALT_first', 'CLORANFENICOL_count_sum', 'GASOMETRIA VENOSA - HCO3_min', 'GASOMETRIA ARTERIAL - SO2_count', 'GASOMETRIA ARTERIAL - FIO2_avg', 'ERITROGRAMA - MCV_min', 'FC_first', 'cirurgia_procedimentos_total', 'OXIMETRIA_last', 'GASOMETRIA ARTERIAL - PH_max', 'GASOMETRIA ARTERIAL - FIO2_count', 'POTASSIO - POTASSIO_min', 'BILIRRUBINAS TOTAL E FRACOES - BILIRRUBINA INDIRETA_min',

'NISTATINA_count_sum', 'ALT_avg', 'PIPERACICLINA_TAZOBACTAM_count_sum',
 'TEMP_stddev', 'MAGNESIO - MAGNESIO_count', 'OXIMETRIA_max', 'GASOMETRIA VENOSA - PH_avg', 'AMOXICILINA_CLAVULANATO_count_sum', 'BILIRRUBINAS TOTAL E FRACOES - BILIRRUBINA TOTAL_avg', 'OXACILINA_count_sum',
 'CLINDAMICINA_count_sum', 'FC_avg', 'TEMPO DE PROTROMBINA - RNI_avg',
 'GASOMETRIA VENOSA - PCO2_count', 'DIUR_count', 'KTPP - TEMPO DE TROMBOPLASTINA_avg', 'PESO_max', 'estertor_respiratorio',
 'qtd_urocultura_sem_crescimento', 'HGT_max', 'FR_min', 'PESO_first', 'DIUR_last',
 'BILIRRUBINAS TOTAL E FRACOES - BILIRRUBINA TOTAL_min',
 'POLIMIXINA_B_count_sum', 'EQU - VOLUME_count', 'GASOMETRIA ARTERIAL - PCO2_avg', 'DOR_avg', 'POTASSIO - POTASSIO_count', 'GASOMETRIA VENOSA - TCO2_avg', 'PAD_last', 'SODIO - SODIO_avg', 'SINAIS_VITAIS_count_total_sum',
 'AMPICILINA_count_sum', 'Idade Hoje', 'diabetes', 'CLARITROMICINA_count_sum',
 'DOXICICLINA_count_sum', 'FC_stddev', 'TEMPO DE PROTROMBINA - TEMPO_min',
 'PAS_last', 'ENFERMEIRO_A__last', 'PAS_first', 'UREIA - UREIA_max', 'ERITROGRAMA - ERITROCITOS_count', 'PROTEINA C REATIVA - PROTEINA C REATIVA_count', 'BILIRRUBINAS TOTAL E FRACOES - BILIRRUBINA INDIRETA_avg',
 'RIFAMPICINA_count_sum', 'ERITROGRAMA - HEMOGLOBINA_max', 'ERITROGRAMA - HEMOGLOBINA_avg', 'PAD_first', 'EQU - DENSIDADE_max',
 'VIA_OF_count_sum', 'PESO_last', 'GASOMETRIA VENOSA - PH_count', 'GASOMETRIA ARTERIAL - HCO3_count', 'ITRACONAZOL_count_sum',
 'VANCOMICINA_count_sum', 'GASOMETRIA VENOSA - PCO2_min', 'ERITROGRAMA - HEMATOCRITO_count', 'TEMP_count', 'GASOMETRIA VENOSA - HCO3_count', 'BILIRRUBINAS TOTAL E FRACOES - BILIRRUBINA TOTAL_count',
 'GASOMETRIA ARTERIAL - PH_avg', 'disenteria', 'PLAQUETAS CONTAGEM - PLAQUETAS_count', 'TEMPO DE PROTROMBINA - ATIVIDADE_max', 'UREIA - UREIA_count', 'TEMP_last', 'TEMPO DE PROTROMBINA - ATIVIDADE_count',
 'OXIMETRIA_first', 'PAD_avg', 'PAS_stddev', 'ERITROMICINA_count_sum', 'GASOMETRIA VENOSA - SO2_min', 'GLICOSE - GLICOSE_max', 'DIUR_first', 'ERITROGRAMA - MCV_avg', 'GLICOSE - GLICOSE_count', 'DOR_stddev', 'GASOMETRIA VENOSA - SO2_count', 'GASOMETRIA VENOSA - PO2_max', 'FR_max',
 'TEMPO DE PROTROMBINA - ATIVIDADE_min', 'ERITROGRAMA - HEMATOCRITO_avg', 'PAS_avg', 'GLICOSE - GLICOSE_min', 'HGT_stddev', 'GASOMETRIA ARTERIAL - HCO3_avg', 'VORICONAZOL_count_sum', 'ALT_last', 'TIGE-

CICLINA_count_sum', 'TGP_ALT_ - TGP_max', 'GASOMETRIA ARTERIAL - HCO3_max', 'bronquite', 'LIDER DE ENFERMAGEM_count', 'ERITROGRAMA - MCHC_count', 'PAD_max', 'ERITROGRAMA - MCV_max', 'FR_first', 'dispneia', 'qtd_bacteroscopico_positiva', 'EQU - PH_max', 'qtd_bacteriologico_sem_crescimento', 'PAD_count', 'EQU - ERITROCITOS_count', 'EQU - PH_min', 'ALT_stddev', 'TGO_AST_ - TGO_avg', 'qtd_hemocultura_sem_crescimento', 'AMOXICILINA_count_sum', 'FC_count', 'DIUR_min', 'mialgia', 'CEFEPIME_count_sum', 'qtd_urocultura_positiva', 'PROTEINA C REATIVA - PROTEINA C REATIVA_min', 'TGP_ALT_ - TGP_min', 'LIDER DE ENFERMAGEM_first', 'VIA_TO_count_sum', 'BILIRRUBINAS TOTAL E FRACOES - BILI DIRETA_max', 'ENFERMEIRO_A_count', 'PROTEINA C REATIVA - PROTEINA C REATIVA_max', 'PESO_count', 'GASOMETRIA ARTERIAL - PH_min', 'GASOMETRIA VENOSA - PO2_count', 'FR_stddev', 'POTASSIO - POTASSIO_avg', 'OXIMETRIA_min', 'BILIRRUBINAS TOTAL E FRACOES - BILIRRUBINA TOTAL_max', 'OXIMETRIA_stddev', 'laudo_exame_imagem_count', 'CREATININA - CREATININA_min', 'FLUCONAZOL_count_sum', 'GASOMETRIA ARTERIAL - SO2_avg', 'CREATININA - CREATININA_max', 'GASOMETRIA VENOSA - %SO2C_min', 'PAS_min', 'GASOMETRIA VENOSA - PH_max', 'HGT_count', 'EQU - VOLUME_min', 'HGT_avg', 'cirrose', 'VIA_VE_count_sum', 'CEFAZOLINA_count_sum', 'NORFLOXACINA_count_sum', 'PAS_max', 'FC_min', 'PAD_stddev', 'KTTP - TEMPO DE TROMBOPLASTINA_min', 'qtd_bacteroscopico_negativa', 'DIUR_stddev', 'GASOMETRIA VENOSA - PCO2_max', 'GASOMETRIA ARTERIAL - TCO2_avg', 'AMPICILINA_SULBACTAM_count_sum', 'VIA_GT_count_sum', 'ERITROGRAMA - ERITROCITOS_max', 'HGT_last', 'ERITROGRAMA - HEMATOCRITO_min', 'ERITROGRAMA - MCHC_max', 'GASOMETRIA VENOSA - TCO2_max', 'BILIRRUBINAS TOTAL E FRACOES - BILIRRUBINA INDIRETA_max', 'GASOMETRIA ARTERIAL - PCO2_min', 'BILIRRUBINAS TOTAL E FRACOES - BILI DIRETA_min', 'PESO_min', 'TGO_AST_ - TGO_count', 'ERITROGRAMA - HEMATOCRITO_max', 'VIA_EV_count_sum', 'TEMPO DE PROTROMBINA - RNI_max', 'obesidade', 'BILIRRUBINAS TOTAL E FRACOES - BILIRRUBINA DIRETA_min', 'GASOMETRIA ARTERIAL - TCO2_min', 'DOR_last', 'EQU - PH_avg', 'DIUR_avg', 'EQU - VOLUME_avg', 'qtd_bacteriologico_positiva', 'IVERMECTINA_count_sum', 'BILIRRUBINAS TOTAL E FRACOES - BILIRRUBINA DIRETA_max', 'PESO_stddev', 'CREATININA - CREATININA_avg', 'PLAQUETAS CONTAGEM - PLAQUETAS_avg', 'TGO_AST_ - TGO_max', 'OXIMETRIA_avg', 'CIPROFLOXACINA_count_sum',

'EQU - ERITROCITOS_max', 'GASOMETRIA VENOSA - PCO2_avg', 'PROTEINA C REATIVA - PROTEINA C REATIVA_avg', 'GASOMETRIA VENOSA - %SO2C_avg', 'TEMPO DE PROTROMBINA - ATIVIDADE_avg', 'ALBENDAZOL_count_sum', 'SODIO - SODIO_max', 'SULFADIAZINA_count_sum', 'GASOMETRIA VENOSA - PH_min', 'FEBRE_COMPUTED_max', 'GASOMETRIA VENOSA - PO2_avg', 'CEFALEXINA_count_sum', 'DOR_min', 'TGP_ALT_ - TGP_count', 'CREATININA - CREATININA_count', 'EQU - VOLUME_max', 'TETRACICLINA_count_sum', 'VIA_VA_count_sum', 'AZITROMICINA_count_sum', 'PAD_min', 'KTP - TEMPO DE TROMBOPLASTINA_max', 'GASOMETRIA ARTERIAL - HCO3_min', 'POTASSIO - POTASSIO_max', 'UREIA - UREIA_min', 'GASOMETRIA VENOSA - PO2_min', 'GASOMETRIA VENOSA - %SO2C_max', 'TGP_ALT_ - TGP_avg', 'GASOMETRIA ARTERIAL - PCO2_count', 'BILIRRUBINAS TOTAL E FRACOES - BILIRRUBINA DIRETA_count', 'PESO_avg', 'ALT_count', 'MAGNESIO - MAGNESIO_avg', 'MAGNESIO - MAGNESIO_min', 'tosse', 'KTP - TEMPO DE TROMBOPLASTINA_count', 'GASOMETRIA ARTERIAL - FIO2_max', 'ERITROGRAMA - HEMOGLOBINA_count', 'GASOMETRIA VENOSA - SO2_avg', 'FR_last', 'BILIRRUBINAS TOTAL E FRACOES - BILIRRUBINA INDIRETA_count', 'febre', 'ALT_min', 'PLAQUETAS CONTAGEM - PALQUETAS_min', 'EQU - DENSIDADE_min', 'TGO_AST_ - TGO_min', 'CEFOXITINA_count_sum', 'BILIRRUBINAS TOTAL E FRACOES - BILI DIRETA_count', 'OXIMETRIA_count', 'CEFTRIAXONA_count_sum', 'EQU - PH_count', 'ERITROGRAMA - ERITROCITOS_min', 'DIUR_max', 'LINEZOLIDA_count_sum', 'VIA_IM_count_sum', 'GASOMETRIA ARTERIAL - SO2_min', 'SODIO - SODIO_min', 'VIA_IP_count_sum', 'ERTAPENEM_count_sum', 'VIA_VO_count_sum', 'TEMP_min', 'HGT_min' e 'TEMP_max'.

- Categóricas: 46 atributos

'bacteriologico_escarro', 'bacteriologico_abcesso', 'bacteriologico_cateter', 'bacteriologico_liquido', 'bacteriologico_tecido', 'bacteriologico_secrecao', 'bacteriologico_fragmento', 'bacteriologico_colecao', 'bacteriologico_ferida', 'bacteriologico_outros_materiais', 'agg_hemocultura', 'agg_urocultura', 'HIV 1 E 2 - HIV 1 E 2', 'HIV 1 E 2 - HIV', 'TGO_AST_ - METODOLOGIA', 'EQU - COR', 'BILIRRUBINAS TOTAL E FRACOES - METODOLOGIA', 'POTASSIO - METODOLOGIA', 'SODIO - METODOLOGIA', 'EQU - CETONA', 'EQU - GLICOSE', 'EQU - FILAMENTOS DE MUCO', 'KTP - METODOLOGIA', 'MAGNESIO - METODOLOGIA', 'EQU - CILINDROS',

'UREIA - METODOLOGIA', 'EQU - CELULAS EPITELIAIS', 'GASOMETRIA ARTERIAL - VENTILACAO', 'EQU - PIGMENTOS BILIARES', 'TGP_ALT_ - METODOLOGIA', 'EQU - HEMOGLOBINA', 'EQU - LEVEDURAS', 'EQU - CRISTAIS', 'EQU - ASPECTO', 'LACTATO ARTERIAL _ACIDO LACTICO_ - METODOLOGIA', 'EQU - UROBILINOGENIO', 'EQU - PROTEINAS', 'CREATININA - METODOLOGIA', 'GLICOSE - METODOLOGIA', 'EQU - BACTERIAS', 'EQU - NITRITO', 'TEMPO DE PROTROMBINA - METODOLOGIA', 'GASOMETRIA VENOSA - VENTILACAO', 'GLICOSE - MATERIAL', 'CORONAVIRUS 2019 DETECCAO POR PCR_COVID19_ - RESULTADO' e 'SEXO'.