UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO E SISTEMAS
CURSO DE ENGENHARIA DE PRODUÇÃO CIVIL

Alissa Emanueli Cabrera Fumagali

Análise do poder preditivo de modelos com diferentes estratégias de compartilhamento de dados de resistência antimicrobiana

Florianópolis
2021

Alissa Emanueli Cabrera Fumagali

Análise do poder preditivo de modelos com diferentes estratégias de compartilhamento de dados de resistência antimicrobiana

Trabalho Conclusão do Curso de Graduação em Engenharia de Produção Civil do Centro de Tecnologia da Universidade Federal de Santa Catarina como requisito para a obtenção do título em Engenharia Civil, habilitação em Engenharia de Produção Civil
Orientador: Prof. Ricardo Faria Giglio, Dr.

Florianópolis
2021

Ficha de identificação da obra elaborada pelo autor,
através do Programa Geração Automática da Biblioteca Universitária da UFSC.

Fumagali, Alissa
Análise do poder preditivo de modelos com diferentes estratégias de compartilhamento de dados de resistência antimicrobiana / Alissa Fumagali; orientador, Ricardo Faria Giglio, 2021.
99 p.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal de Santa Catarina, Centro Tecnológico, Graduação em Engenharia de Produção Civil, Florianópolis, 2021.

Inclui referências.

1. Engenharia de Produção Civil. 2. Vigilância da Resistência Antimicrobiana. 3. Compartilhamento de Dados. 4. Aprendizado de Máquina Automatizado. I. Faria Giglio, Ricardo. II. Universidade Federal de Santa Catarina. Graduação em Engenharia de Produção Civil. III. Título.

Alissa Emanueli Cabrera Fumagali

Análise do poder preditivo de modelos com diferentes estratégias de compartilhamento de dados de resistência antimicrobiana

Este Trabalho Conclusão de Curso foi julgado adequado para obtenção do Título em Engenharia Civil, habilitação em Engenharia de Produção Civil e aprovado em sua forma final pelo Curso de Engenharia de Produção Civil.

Local, 15 de setembro de 2021.

Documento assinado digitalmente
Monica Maria Mendes Luna
Data: 27/09/2021 10:12:12-0300
CPF: 314.104.993-91
Verifique as assinaturas em https://v.ufsc.br

Profa. Mônica Maria Mendes Luna, Dra.
Coordenadora do Curso

Banca Examinadora:

Documento assinado digitalmente
Ricardo Faria Giglio
Data: 27/09/2021 08:46:03-0300
CPF: 304.078.978-36
Verifique as assinaturas em https://v.ufsc.br

Prof. Ricardo Faria Giglio, Dr.
Orientador(a)
Universidade Federal de Santa Catarina

Documento assinado digitalmente
Carlos Ernani Fries
Data: 24/09/2021 15:40:02-0300
CPF: 304.616.699-00
Verifique as assinaturas em https://v.ufsc.br

Prof. Carlos Ernani Fries, Dr.
Avaliador
Universidade Federal de Santa Catarina

Documento assinado digitalmente
Sergio Fernando Mayerle
Data: 24/09/2021 15:42:37-0300
CPF: 344.463.119-72
Verifique as assinaturas em https://v.ufsc.br

Prof. Sérgio Fernando Mayerle, Dr.
Avaliador
Universidade Federal de Santa Catarina

Este trabalho é dedicado a todos da minha família que me apoiaram e tornaram essa conquista possível.

AGRADECIMENTOS

Agradeço incialmente meu orientador, Prof. Dr. Ricardo Giglio, por todos os ensinamentos, direcionamentos, confiança e auxílio que foram imprescindíveis para a construção deste trabalho e para o meu desenvolvimento profissional. Agradeço também a toda a equipe do Grupo 37.78 pelo tempo em que trabalhamos juntos, e especificamente ao Márcio Gregory, Renato Maciel, Robson Zagre, Dominique Ruther, Arthur Reys e Leonardo Rammé por tornarem o desenvolvimento deste trabalho possível.

Gostaria também de agradecer à UFSC pelas oportunidades e vivências que me proporcionou. Agradeço a todos os meus professores, por todo o conhecimento repassado, e aos servidores do meu curso e da universidade. Agradeço, também, a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela oportunidade de realizar intercâmbio acadêmico. Por fim, gostaria de agradecer a algumas entidades – A7, EJEP, GLean e UCC – pelas amizades, contatos profissionais, lembranças e todo o crescimento pessoal e profissional que me ofertaram.

Agradeço aos meus pais, Aurélio e Telma Fumagali, e à minha irmã Giovanna Fumagali por todo o amor, suporte e carinho ao longo da jornada. Sem o apoio recebido, nada disso teria sido possível. Aos demais familiares, pelos conselhos, ensinamentos e momentos compartilhados.

Gostaria de agradecer também a todos os colegas de universidade que me inspiram e me ensinam muito mais do que eu jamais imaginei aprender durante a faculdade. Em especial eu gostaria de agradecer ao Dan e ao Ricardo por terem sido a minha primeira fortaleza em Florianópolis. Agradeço à Gabriela Gonzales e ao David Eller pelo exemplo na busca de novos desafios e por terem me influenciado a me aventurar no campo de Data Science. Agradeço à Maria Fernanda Vieira, à Patricia Vanzella, à Caroline Piesanti, à Carina Souza, à Karina Mezzari e à Laiza Parizotto pela força, pelo apoio, pela sabedoria e por nossa amizade.

RESUMO

O mercado de serviços de saúde vive uma crise de aumento constante dos custos do setor. Segundo Porter e Teisberg (2006), essa crise é resultado dos modelos de remuneração e dos incentivos por eles gerados. Neste modelo os contratantes de serviços de saúde pagam por procedimento realizado, quando na verdade deveriam remunerar de acordo com o valor entregue ao paciente por meio da transformação digital, melhorando da qualidade de serviços. Ao mesmo tempo em que o setor da saúde se transforma, alguns desafios antigos prevalecem, como o desenvolvimento de resistência por parte de bactérias, que ainda causam 25% das mortes do mundo segundo a Organização Mundial da Saúde (OMS). Nesse contexto a OMS lançou ao programa *Global Antimicrobial Resistance Surveillance System* (Glass) em 2015, utilizando da transformação digital para estruturar um sistema de vigilância do desenvolvimento da resistência das bactérias em um contexto mundial. No Brasil, o Ministério da Saúde é responsável por agregar dados à base Glass. Apesar da relevância do tema há uma dificuldade de integrar novos hospitais à base. Dessa forma, o presente trabalho tem o objetivo de analisar se o compartilhamento de informações epidemiológicas entre hospitais pode ser benéfico para os modelos de previsão locais, e assim, identificar ou não o valor da base do Glass para os hospitais do Brasil. Para tanto, foram treinados modelos para seis hospitais brasileiros a partir de uma ferramenta de aprendizado de máquina automatizado, chamada H_2O, alguns modelos utilizavam apenas de dados locais, enquanto outros foram construídos com os dados compartilhados através da base Glass. Com isso, percebe-se que os modelos desenvolvidos a partir dos dados compartilhados possuem um poder de previsão melhor quando testados localmente nos hospitais. No Brasil, a troca de informações entre instituições de saúde é limitada quando comparada a outros países, e o presente trabalho adiciona evidência a favor do compartilhamento de informações para auxiliar na qualidade entregue aos pacientes e na redução de custos.

Palavras-chave: Vigilância da Resistência Antimicrobiana. Compartilhamento de Dados. Aprendizado de Máquina Automatizado.

ABSTRACT

The healthcare market is experiencing a costs crisis. According to Porter and Teisberg (2006), this crisis is a result of the remuneration models and the incentives it generates. In the current model, health service contractors pay per procedure, when in fact they should pay according to the health value delivered to the patient. Health institutions seek to deliver more value to patients, using digital transformation as their ally to drive quality improvement. At the same time as the healthcare sector changes, some old challenges prevail, such as the bacteria resistance development, which still cause 25% of deaths in the world, according to the World Health Organization (WHO). In this context, WHO launched the Global Antimicrobial Resistance Surveillance System (Glass) program in 2015, using digital transformation to structure a surveillance system for the development of bacterial resistance in a worldwide context. In Brazil, the Health Ministry is responsible for adding data to the Glass database. And, despite the relevance of the topic, there is a difficulty in integrating new hospitals into the base. Thus, the present study aims to analyze whether sharing epidemiological data between hospitals can be beneficial for forecasting models, and thus, identify or not the value of the Glass base for hospitals in Brazil. For this purpose, prediction models were trained for six Brazilian hospitals using an automated machine learning tool, called H_2O, some models used only local data, while others were built with data shared through the Glass base. Thereby, the models developed from shared data had better predictive power when tested locally in hospitals. In Brazil, the exchange of information between health institutions is limited when compared to other countries, this is just one of the examples in which sharing data can help improve quality delivered to patients and reduce costs.

Keywords: Antimicrobial Resistance Surveillance. Data Sharing. Automated Machine Learning.

LISTA DE FIGURAS

LISTA DE QUADROS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

PIB Produto Interno Bruto

IBGE Instituto Brasileiro de Geografia e Estatística

OMS Organização Mundial da Saúde

OCDE Organização para Cooperação e desenvolvimento Econômico

VBHC Modelos de Prestação de Serviço de Saúde Baseado em Valor

Covid-19 Doença do coronavírus

Sars-CoV-2 Vírus que causa a doença do coronavírus

URM Uso Racional de Medicamentos

CCIHs Comissões de Controle de Infecções Hospitalares

Glass *Global Antimicrobial Resistence Surveillance System*

BR-Glass Sistema de Vigilância da Resistência Antimicrobiana Brasileira

TSA Teste de Sensibilidade de Antimicrobianos

ML *Machine Learning*

AutoML *Automated Machine Learning*

TPOT *Tree-Based Pipeline Optimization Tool*

GBM *Gradient Boosting Machines*

GLM *Gradient Linear Machines*

DRF *Distributed Random Forest*

AUC *Area Under the Curve*

ROC *Receiver Operating Characteristic Curve*

TP *True Positive*

TN *True Negative*

FN *False Negative*

FP *False Positive*

TPR *True Positive Rate*

FPR *False Positive Rate*

MCC *Matthews Correlation Coefficient*

EDA *Exploratory Data Analysis*

SUMÁRIO

1 INTRODUÇÃO

No primeiro capítulo deste trabalho é feita uma contextualização, para evidenciar a relevância do tema na situação em que o setor da saúde se encontra. Em adição ao contexto, o tema é justificado de acordo com o potencial de impacto que pode exercer sobre a vida do ser humano. Em sequência, os objetivos gerais e específicos são consolidados de forma sucinta assim como, são descritas a estrutura do trabalho, suas limitações e delimitações para um melhor entendimento do seu desenvolvimento.

1.1 CONTEXTUALIZAÇÃO

No Brasil, de 2010 à 2017, os gastos relacionados aos bens e serviços de saúde aumentaram sua participação no Produto Interno Bruto (PIB) em 1.2 pontos percentuais, totalizando R$ 608.3 bilhões de reais em 2017, segundo a Conta-Satélite de Saúde do Instituto Brasileiro de Geografia e Estatística (IBGE). Esta tendência é a realidade em diferentes países e, à medida que os custos do setor da saúde continuam a subir, os desafios relacionados à prestação de serviços de saúde levantam preocupações e recebem mais atenção (PORTER, 2008).

As explicações para esse aumento de custos não são triviais. Kaplan e Porter (2011) sugerem que o envelhecimento da população e o desenvolvimento de novos tratamentos contribuem para o aumento dos custos, mas defendem que a causa principal reside no modelo de remuneração do setor e os incentivos que ele gera. Atualmente os contratantes de serviços de saúde, como seguradoras, governos e pessoas físicas, reembolsam os procedimentos realizados, em vez dos resultados alcançados. O modelo de remuneração vigente estimula a realização de procedimentos desnecessários e as ineficiências administrativas que, juntos, representam cerca de 30% dos recursos gastos em saúde segundo a Organização Mundial da Saúde (OMS) e a Organização para a Cooperação e Desenvolvimento Econômico (OCDE).

Nesse cenário, o *Value Based Health Care* (VBHC), ou, em português, Modelo de Prestação de Serviços de Saúde Baseados em Valor, propõe um novo modelo de negócio que foca no valor entregue ao paciente (PORTER; TEISBERG, 2006). Em seu livro *Redefining Health Care*, Porter e Teisberg (2006) definem valor como resultado obtido em relação ao custo despendido. Na saúde o numerador indica os resultados clínicos que mais importam

para os pacientes, como recuperação funcional e qualidade de vida, enquanto o denominador se relaciona ao gasto total do ciclo de tratamento.

Para entregar um maior valor aos pacientes, os atores da cadeia do VBHC devem melhorar os resultados entregues e reduzir os custos. O interessante é que a busca da melhora da saúde entregue ao paciente tem um efeito significativo na redução de custos. Diagnósticos corretos e antecipados, tratamentos apropriados e menos invasivos impactam na qualidade de vida do paciente e na redução de custos da cadeia de fornecimento de serviços de saúde. Assim, entende-se que a melhora da qualidade dos serviços de saúde é a melhor maneira de reduzir os custos do sistema (PORTER, 2008).

Estudos realizados antes da pandemia causada pelo coronavírus, evento mais disruptivo que o sistema de serviços de saúde mundial já enfrentou, já identificavam a necessidade de se repensar o modelo de negócios e de operações vigentes na saúde através da perspectiva do paciente. Com o surgimento da pandemia essa necessidade urgente tornou-se algo indispensável. É notável que a Covid-19 estimulou a quebra das barreiras que dificultavam o compartilhamento de informações e conhecimentos entre organizações da saúde. Dessa forma, pode-se afirmar que a transformação digital tem sido e continuará sendo uma parte fundamental da transição para o VBHC (DAVIES, 2020).

1.2 JUSTIFICATIVA

As doenças causadas por bactérias, também conhecidas como infecções, segundo a OMS, representavam 25% das mortes em todo o mundo e 45% das mortes em países menos desenvolvidos antes da pandemia da Covid-19. Apesar da aplicação de vacinas e da prescrição de antimicrobianos, as mortes causadas por germes estão longe de serem erradicadas. Isso acontece, principalmente, devido à resistência microbiana, que tende a aumentar mediante ao uso indiscriminado de antimicrobianos (WANNMACHER, 2004).

A resistência microbiana é o fenômeno de sobrevivência e proliferação da bactéria *in vitro* em uma solução de fármaco. A concentração de fármaco é similar a concentração que se encontra no sangue de um paciente que está em tratamento com o medicamento. Sabe-se que a resistência bacteriana ocorre com uma frequência razoável e que o uso inadequado dos antibióticos e o aumento da resistência bacteriana estão fortemente relacionados. Infecções causadas por bactérias resistentes fazem com que o custo do tratamento aumente devido ao

uso de medicamentos mais caros e muitas vezes mais tóxicos, além de aumentar o tempo de internação do paciente e a taxa de mortalidade vinculada ao tipo de infecção, o que reduz o valor entregue ao paciente segundo o conceito de VBHC (GURGEL; CARVALHO, 2008).

A evolução da resistência das bactérias foi pauta da Assembleia Mundial da Saúde organizada pela OMS em 2015, onde foram definidos cinco objetivos principais que visam "garantir, pelo maior tempo possível, a continuidade do tratamento bem-sucedido e da prevenção de doenças infecciosas com medicamentos eficazes e seguros":

a) Melhorar a conscientização e a compreensão da resistência antimicrobiana;
b) Fortalecer o conhecimento por meio de vigilância e pesquisa;
c) Reduzir a incidência de infecção;
d) Otimizar o uso de agentes antimicrobianos;
e) Garantir o investimento sustentável no combate à resistência antimicrobiana.

Nessa mesma assembleia foi lançado o *Global Antimicrobial Resistance Surveillance System* (Glass), em português, Sistema de Vigilância de Resistência Antimicrobiana Global com o objetivo de coletar, padronizar, comparar e analisar dados sobre a resistência bacteriana de diferentes países. O projeto da base brasileira do Glass, conhecida como BR-Glass é de responsabilidade do Ministério da Saúde, porém possui uma baixa adesão por parte dos hospitais nacionais. Sabe-se também que as infecções hospitalares causadas por bactérias resistentes têm um maior impacto em países onde o agravamento da complexidade dos quadros clínicos dos pacientes coexiste com recursos limitados da saúde pública. Nesses casos, é recomendada a instauração de programas de vigilância nacionais e locais (TOLEDO et al., 2012). No Brasil, é responsabilidade das Comissões de Controle de Infecções Hospitalares (CCIHs) coordenar ações relacionadas à epidemiologia hospitalar e promover ações vinculadas aos cinco objetivos traçados pela OMS nas instituições de saúde (SILVA, 2008). Dessa forma, este trabalho propõe identificar se o compartilhamento de informações epidemiológicas entre hospitais pode ser benéfico para os modelos de previsão locais, e assim, identificar ou não o valor da base do BR-Glass para os hospitais do Brasil.

1.3 OBJETIVOS

As subseções que seguem têm o objetivo de esclarecer os objetivos desta monografia. Inicialmente foi estabelecido o objetivo geral que define, de forma ampla, as

questões abordadas no trabalho, em seguida, serão ordenados os objetivos específicos que definem a abordagem específica que será adotada ao longo do texto.

1.3.1 Objetivo Geral

O objetivo geral deste trabalho consiste em analisar como compartilhamento de informações de saúde impacta o poder preditivo de modelos de aprendizagem de máquina nesse setor. Para isto, será comparado o poder preditivo de diferentes modelos de aprendizagem de máquina treinados com dados de epidemiologia hospitalar nacionais, agregados pelo Ministério da Saúde na base do BR-Glass, e dados de epidemiologia hospitalar individuais, fornecidos por diferentes hospitais ao Ministério da Saúde.

1.3.2 Objetivos Específicos

Com o intuito de alcançar o objetivo geral, foram determinados os seguintes objetivos específicos:

a) Determinar métricas de avaliação dos modelos;
b) Aplicar modelos de aprendizagem de máquina automatizada, treinados com dados da epidemiologia hospitalar individual e com dados de epidemiologia nacional, para a predição do resultado do antibiograma;
c) Comparar poder preditivo dos modelos;
d) Avaliar se os compartilhamentos de dados de epidemiologia podem melhorar o poder preditivo de modelos que preveem o resultado de antibiogramas.

1.4 ESTRUTURAÇÃO DO TRABALHO

Esse trabalho é composto por 5 capítulos. No primeiro capítulo é descrito o cenário econômico que evidencia o problema abordado neste trabalho, além de justificar a importância do tema dentro do setor da saúde, e apresentar as delimitações, as limitações, a estrutura e o objetivo do trabalho.

No segundo capítulo é apresentada a fundamentação teórica do trabalho. Nele são abordados o conceito de Modelo de Prestação de Serviços de Saúde Baseados em Valor

(VBHC), aspectos sobre a vigilância da resistência antimicrobiana, aprendizado de máquina, aprendizado de máquina automatizado e métricas para avaliação de modelos de classificação binária.

No terceiro capítulo é explicado, de forma detalhada, a metodologia adotada e seu desenvolvimento. Nele são abordados mais detalhes sobre o cenário de estudo e o procedimento metodológico que é composto pela declaração e refinamento da questão de pesquisa, pela exploração dos dados e pelo desenvolvimento de modelos estatísticos preditivos.

No quarto capítulo os modelos desenvolvidos serão testados e seus resultados serão discutidos a partir da análise das métricas para avaliação de modelos de classificação binária expostas no capítulo dois. E, por fim, no quinto e último capítulo são apresentadas as conclusões e recomendações resultantes deste trabalho.

1.5 DELIMITAÇÕES DO TRABALHO

O presente trabalho tem como objetivo comparar o poder preditivo de modelos de aprendizagem de máquina treinados com e sem o compartilhamento de dados entre instituições da saúde. Para tanto serão utilizados dados da base *Global Antimicrobial Resistance Surveillance System* (Glass), ou, em português, sistema de vigilância da resistência antimicrobiana global, que é uma inciativa da OMS em conjunto com o Ministério da Saúde. Especificamente, serão utilizados apenas os dados das instituições de saúde brasileiras registrados no sistema, que também pode ser chamado de BR-Glass. Nessa base estão disponíveis informações sobre o quadro clínico do paciente e sobre o Teste de Sensibilidade a Antimicrobianos (TSA), também conhecidos como antibiograma. Esse exame laboratorial tem como objetivo determinar o nível de sensibilidade das bactérias em relação aos antibióticos.

Dessa forma, o presente trabalho não contém informações sobre o histórico do paciente ou sobre suas propensões genéticas. A base utilizada é formada por dados coletado em 2018 e 2019, assim, não será analisada a evolução do perfil de resistência das bactérias no país. Além disso o trabalho não tem o objetivo de discutir ou desenvolver o melhor modelo de previsão de resultados de antibiogramas, pois o objetivo deste trabalho é entender o impacto que o compartilhamento de dados epidemiológicos tem sobre o poder de predição de modelos genéricos.

1.6 LIMITAÇÕES DO TRABALHO

As análises realizadas durante o desenvolvimento do trabalho são limitadas por algumas características da base do BR-Glass e como suas informações podem ser utilizadas:

a) A base é composta por seis hospitais localizados na mesma região do país;
b) Teste para realizar inferências estatísticas não serão feitos pela demanda computacional necessária.

2 FUNDAMENTAÇÃO TEÓRICA

O presente capítulo apresentará um suporte teórico à monografia com o objetivo de facilitar o entendimento do leitor. Assim, inicialmente é abordado o tema de *Value Based Health Care* (VBHC), para maior entendimento das tendências e demandas no setor da saúde. Na sequência, a resistência antimicrobiana e seus sistemas de vigilância são explicados, já que os modelos de predição desenvolvidos no capítulo três são construídos a partir de dados de epidemiologia. Com isso, a fundamentação teórica de questões relacionadas à saúde é concluída e inicia-se a parte da fundamentação voltada ao aprendizado de máquina, como suas aplicações na saúde, o aprendizado de máquina automatizado e métricas de avaliação de modelos de classificação binária, que é o tipo de modelo que será desenvolvido no capítulo três.

2.1 O CONCEITO DE *VALUE BASED HEALTH CARE* (VBHC)

Para compreender o conceito de *Value Based Health Care* (VBHC), ou, em português, Modelo de Prestação de Serviços de Saúde Baseados em Valor, primeiramente precisa-se definir o conceito de saúde. Segundo a Organização Mundial da Saúde (OMS) promover saúde é promover "um estado de completo bem-estar físico, mental e social e não apenas a ausência de doença ou enfermidade". A partir da definição desse primeiro conceito torna-se mais fácil compreender o conceito de valor em saúde.

O termo VBHC, foi usado pela primeira vez no livro *Evidence-Based Healthcare and Public Health* por Muir Gray (2009), onde ele definiu o *Value Based Healthcare* como "a situação em que aqueles que pagam por cuidados de saúde exigirão que os resultados das intervenções obtenham maiores benefícios do que qualquer outra alternativa que consuma os mesmos recursos". Desde então, a definição de valor em saúde foi abordada em diferentes discussões e publicações científicas e até hoje não há apenas uma definição que pode ser adotada. Dessa forma, serão exploradas duas perspectivas complementares sobre o conceito de valor para a saúde.

A primeira perspectiva de valor em saúde que será abordada define que valor é a alocação dos recursos de forma a tratar as pessoas que irão mais se beneficiar, reduzir a desigualdade em relação ao acesso a serviços de saúde de qualidade e promover a eficiência

técnica do sistema, entregar mais saúde utilizando menos recursos (MAESENEER, 2015). Essa ideia de valor pode ser desmembrada em quatro aspectos segundo Maeseneer (2015):

a) Valor pessoal: determinado pela relação entre o resultado do procedimento e a perspectiva do paciente em relação às suas metas, valores e expectativas, por exemplo, a cirurgia do joelho pode fornecer maior flexibilidade na articulação, mas, a menos que tenha resolvido o problema que mais incomodava o paciente, tem pouco ou nenhum valor para ele;
b) Valor alocado: relacionado com a distribuição uniforme dos recursos para diferentes subgrupos da população, como por exemplo, fornecer tratamentos para pessoas com câncer e com diabetes, ou para pessoas idosas e para pessoas jovens;
c) Valor de utilização: determinado pela obtenção dos melhores resultados com os recursos disponíveis e por quão bem são utilizados os recursos alocados para investimento em um determinado subgrupo da população. Este aspecto visa identificar e minimizar a desigualdade, por exemplo, no encaminhamento e tratamento de pessoas dos subgrupos mais carentes dessa população;
d) Valor social: relacionado com a contribuição para a coesão social gerada pela intervenção em saúde, estimulando solidariedade, equidade e reconhecimento da diversidade.

A segunda perspectiva, mais conhecida em comparação à primeira, se desenvolveu em Harvard, especificamente na Harvard Business School e na Harvard Medical School com os professores Porter, Teisberg, Kaplan, Bohmer e Christensen. Em seu livro *Redefining Health Care*, Porter e Teisberg (2006) citam que a definição e a mensuração do valor é a primeira etapa da busca da melhoria e é essencial para o desempenho de qualquer organização. O conceito de valor para eles pode ser traduzido pelo resultado obtido em relação ao custo despendido, ou seja, valor é definido por eles como a melhora da saúde entregue ao paciente dividido pela unidade monetária investida.

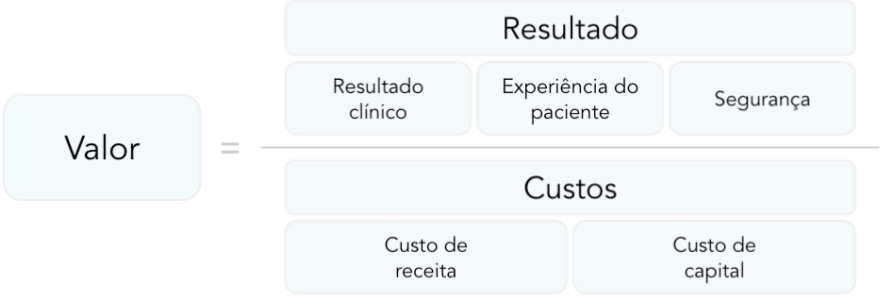
Figura 1 - Representação do conceito de valor em saúde segundo Porter e Teisberg

Fonte: Adaptado Porter e Teisberg (2006)

Em 2015, em um curso ministrado pelo Professor Robert Kaplan sobre custeio de saúde em instituições VBHC foi apresentada uma nova equação, que destrincha os termos da equação em:

a) Resultado ou desfecho clínico que considera saúde da população, taxa de sobrevivência, extensão da recuperação funcional, entre outros;
b) Experiência do paciente que considera conforto, atendimento da equipe, tempo de espera, facilidade de acesso, entre outros;
c) Segurança do paciente que considera erro de diagnóstico, complicações pós operatórias, infecções, entre outros;
d) Custos operacionais ou de receita que considera tempo de tratamento, salários, manutenção do sistema, instalações, entre outros;
e) Custos de capital ou investimentos que considera investimentos em infraestrutura, investimentos em equipamentos, investimento em tecnologia, entre outros.

Figura 2 – Representação do conceito de valor em saúde segundo Kaplan.



Fonte: Adaptado Kaplan (2015)

Apesar de suas diferenças, as duas equações consideram que tanto os resultados quanto os custos devem ser mensurados ao nível do paciente e devem considerar todo o ciclo de cuidados para a condição médica específica do paciente, que pode conter diversas

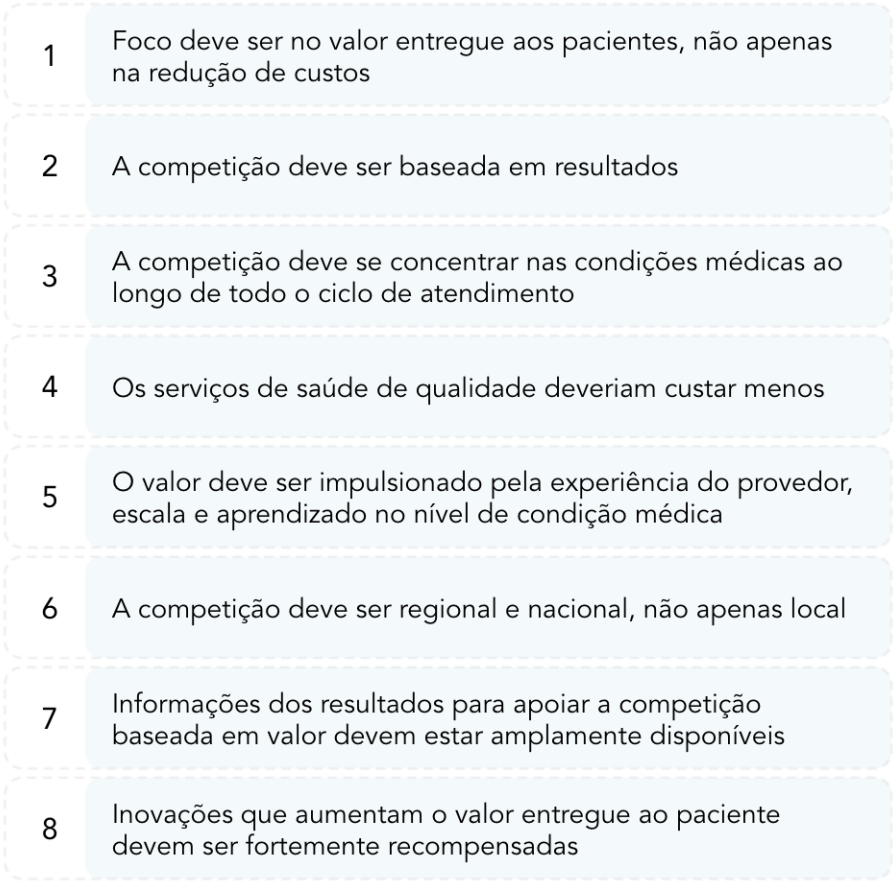
especialidades e internações. A partir do entendimento de diferentes perspectivas sobre o valor em saúde, pode-se perceber que o Modelo de Prestação de Serviços de Saúde Baseados em Valor (VBHC), de maneira geral, propõe um novo modelo que foca no valor entregue ao paciente, independente da perspectiva que pode ser individual ou coletiva (PORTER; TEISBERG, 2006).

O VBHC recebe mais atenção a cada ano, pois ele propõe uma abordagem diferente às dificuldades que o mercado da saúde enfrenta, tais como o aumento dos custos relacionados a serviços de saúde que tendem a continuar crescendo, a variação da qualidade e dos custos em relação a diferentes regiões ou provedores, a realização de procedimentos desnecessários e evitáveis, a lenta disseminação das melhores práticas de atendimento e a resistência à inovação que o setor apresenta.

Normalmente, em um mercado em livre competição há o incentivo para melhorias em relação à qualidade, atendimento, tecnologia e custos devido à competitividade. Os preços são ajustados em relação à qualidade ofertada, o valor entregue aumenta e o mercado cresce atendendo às necessidades dos consumidores. Dessa forma, alguns competidores prosperam enquanto rivais, mais fracos, precisam se reestruturar ou encerram suas atividades. Essa dinâmica de mercado é a realidade da maioria dos setores, mas não da saúde. O que motiva essa diferença não é a ausência de competição, mas a competição em níveis errados e sobre as coisas erradas. Os provedores de serviços de saúde se encontram em um padrão de competição disfuncional, competem para transferir custos de uns para os outros e, assim, aumentar seu poder de negociação. Esse tipo de competição não gera valor aos consumidores, que nesse setor são os pacientes, e são a causa raiz das dificuldades enfrentadas pelo setor que foram citadas anteriormente (PORTER;TEISBERG, 2006).

No livro *Redefining Health Care*, Porter e Teisberg (2006) definem oito princípios fundamentais que viabilizam uma competição baseada em valor no setor da saúde. Estes princípios estão apresentados na Figura 3.

Figura 3 – Princípios fundamentais para uma competição baseada em valor.



Fonte: Adaptado Porter e Teisberg, 2006

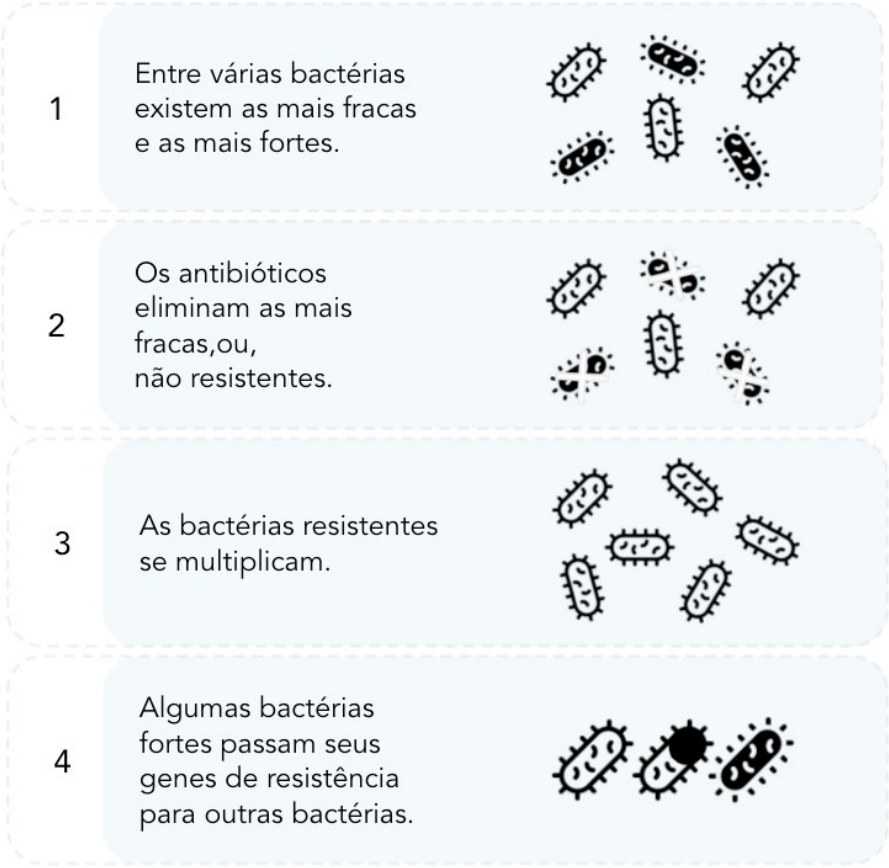
A Covid-19 acelerou o desenvolvimento de alguns destes princípios, principalmente o sétimo, pois incentivou o compartilhamento de informações entre organizações da saúde, e o oitavo, visto que durante a pandemia a velocidade de desenvolvimento de recursos digitais no setor foi notável, mesmo que a transformação digital na saúde seja difícil por ser um dos setores mais regulamentados do mundo e por ser um setor adverso ao risco. Acredita-se que a transformação digital tem sido e continuará sendo uma parte fundamental da transição do setor para o VBHC. Construir sistemas e serviços com segurança, respeitando a privacidade e confidencialidade dos pacientes e favorecendo a transparência e troca de conhecimento entre as organizações fará com que a confiança dos pacientes e das equipes na transformação digital se mantenha. O que tem potencial para promover uma mudança real no comportamento dos cidadãos, dos funcionários da saúde e no próprio sistema alterando o foco para o valor entregue ao paciente (DAVIES, 2020).

2.2 VIGILÂNCIA DA RESISTÊNCIA ANTIMICROBIANA

O desenvolvimento e utilização dos antimicrobianos é considerada a intervenção de saúde pública mais importante do século XX pois contribuíram significativamente para a redução da mortalidade infecciosa. Os antimicrobianos são substâncias naturais ou sintéticas que impedem o crescimento dos micro-organismos podendo provocar a sua morte (SILVA, 2008; MULLER et al., 2015). Os antimicrobianos são eficazes contra bactérias, mofo, fungos e vírus enquanto os antibacterianos são eficazes apenas contra bactérias. Os antibacterianos também são conhecidos como antibióticos. Atualmente, os medicamentos são a principal intervenção terapêutica adotada para a melhora do quadro clínico de um paciente, e, segundo OMS, os antibióticos representam 12% de todas as prescrições ambulatoriais do mundo (GURGEL; CARVALHO, 2008; OLIVEIRA; MUNARETTO, 2010). O Brasil é o quarto país no ranking mundial de consumo de medicamentos, e, segundo o Sindicato das Indústrias Farmacêuticas, 40% do consumo de medicamentos no território nacional é representado pelos antibióticos (MULLER et al., 2015).

As doenças infecciosas sempre foram um problema de saúde pública mundial, porque desde as primeiras utilizações de terapias com antibacterianos houve relatos de resistência bacteriana (GURGEL; CARVALHO, 2008). O termo “resistência bacteriana” refere-se a cepas de microrganismos que têm a capacidade de se multiplicar na presença de antibacterianos, com concentração similar à que se encontra nos seres humanos durante o tratamento (WANNMACHER, 2004). Isso acontece porque os antibióticos atuam sobre as bactérias dizimando totalmente as mais fracas e selecionando despropositadamente as mais fortes, levando à multiplicação de bactérias resistentes, também conhecidas como superbactérias (MULLER et al., 2015).

Figura 4 – Exemplificação do desenvolvimento da resistência bacteriana.



Fonte: Elaborado pela Autora

Segundo os dados da OMS, ainda hoje, as infecções causam 25% das mortes em todo o mundo, e 45% nos países menos desenvolvidos (WANNMACHER, 2004). Dessa forma, a resistência bacteriana é considerada uma preocupação mundial e é objeto das publicações atuais sobre antibacterianos, que são os únicos medicamentos que influenciam no caso clínico de pacientes infectados e no ecossistema em que está inserido (WANNMACHER, 2004). Segundo a OMS, mais de 50% das prescrições de antibióticos são inapropriadas e 2/3 dos antibióticos são usados sem prescrição médica (WANNMACHER, 2004), dados que suportam a opinião de Gurgel e Carvalho (2008) sobre os quatro principais fatores que acarretam o desenvolvimento de superbactérias:

a) Prescrição arbitrária de antimicrobianos;
b) Uso abusivo ou inadequado de antimicrobianos;

c) Fácil transmissão de patógenos resistentes entre continentes devido à globalização;
d) Falta de um sistema global de vigilância epidemiológica da resistência bacteriana aos antibacterianos que gere informação para a tomada de decisões e elaboração de políticas terapêuticas e reguladoras.

A resistência bacteriana tem um enorme impacto sobre a economia pois, do ponto de vista do prescritor, há o custo da ineficácia da terapia convencional com a eventual perda de pacientes. Em relação ao paciente, existe o próprio custo da doença, que pode ser a morte, além do custo de medicamentos alternativos, usualmente mais caros. Para o sistema público de saúde há um aumento dos gastos, o que desequilibra o uso de seus recursos, normalmente, escassos. A indústria farmacêutica é a única que se beneficia através do desenvolvimento de novos patógenos através do lucro de novos medicamentos capazes de suprimir as bactérias resistentes (WANNMACHER, 2004).

De forma sucinta, as principais consequências da resistência bacteriana são o aumento dos custos e do tempo de tratamento, pela utilização de medicamentos que são, usualmente, mais caros e mais tóxicos, que influenciam diretamente o tempo de hospitalização e de isolamento do paciente. Além disso, as superbactérias estão diretamente relacionadas com o aumento da frequência e da gravidade das infecções hospitalares e com o aumento da taxa de mortalidade associada a este tipo de infecção (GURGEL; CARVALHO, 2008). Em relação ao aumento de custos, o desenvolvimento de drogas capazes de controlar germes resistentes resultou no incremento dos custos assistenciais, de forma que os antibióticos representam aproximadamente 1/3 dos gastos de medicamentos prescritos em hospitais (SILVA, 2008).

Silva (2008) cita que ainda existe controvérsia sobre a relação entre o uso de antibióticos e o desenvolvimento de resistências em hospitais, mas apesar disso, já se foi comprovado:

a) Prescrição arbitrária de antimicrobianos impacta o padrão de resistência das bactérias;
b) A mudança no uso de antimicrobianos promovem alterações nos padrões de resistência;
c) A resistência bacteriana é mais comum nos germes hospitalares, onde se concentra o maior uso de antibióticos;

d) As áreas hospitalares com maior densidade de uso de antibióticos possuem maior incidência de resistência;
e) A duração da antibioticoterapia eleva os riscos de colonização por bactérias multirresistentes.

Para conter o desenvolvimento de bactérias resistentes a Política Nacional de Medicamentos define o Uso Racional de Medicamentos (URM) como o processo que compreende a prescrição correta, que resulta na indicação ótima relacionada à dosagem, via de administração, duração do tratamento e ao alcance do sucesso clínico com a menor toxicidade para o paciente (SILVA, 2008; OLIVEIRA; MUNARETTO, 2010). Além disso, defende-se que as ações de vigilância epidemiológica devem ser intensificadas, de modo que a indústria farmacêutica, os prescritores de medicamentos e a saúde pública tenham acesso a informações sobre o padrão de resistência de patógenos e sobre a deficiência dos antimicrobianos (GURGEL; CARVALHO, 2008).

No Brasil, o Conselho Federal de Medicina no dia 20 de agosto de 1999 definiu que é responsabilidade das Comissões de Controle de Infecções Hospitalares (CCIHs) definir as estratégias de controle do uso de antibióticos em instituições de saúde, assim sendo, as CCIHs são responsáveis pela implementação de programas de uso racional de antibióticos em hospitais. Estas comissões assumem as principais atividades executivas de planejamento, apoiam a diferentes setores e fomentam a criação de comitês específicos quando necessário (SILVA, 2008).

Como citado anteriormente, o aumento de resistência das bactérias é uma questão mundial e estava na pauta da Assembleia Mundial da Saúde realizada em maio de 2015 pela OMS. Nessa ocasião foi definido um plano de ação global para abordar esse tema com o objetivo de "garantir, pelo maior tempo possível, a continuidade do tratamento bem-sucedido e da prevenção de doenças infecciosas com medicamentos eficazes e seguros, com garantia de qualidade, usados de forma responsável e acessíveis a todos os que deles precisam". Para atingir essa meta foram definidos cinco objetivos estratégicos:

a) Melhorar a conscientização e a compreensão da resistência antimicrobiana;
b) Fortalecer o conhecimento por meio de vigilância e pesquisa;
c) Reduzir a incidência de infecção;
d) Otimizar o uso de agentes antimicrobianos;
e) Garantir o investimento sustentável no combate à resistência antimicrobiana.

Entende-se que um sistema de vigilância da resistência antimicrobiana é fundamental para o atingimento dos cinco objetivos estratégicos, pois fornece as informações necessárias para apoiar ações locais, como em hospitais, nacionais e globais. Sistemas de monitoramento regionais já obtiveram sucesso na coleta de dados ao longo de vários anos, como exemplo pode-se citar o *Central Asian and Eastern European Surveillance of Antimicrobial Resistance* (CAESAR), o *European Antimicrobial Resistance Surveillance Network* (EARS-Net) e o *Latin AmericanAntimicrobial Resistance Surveillance Network* (ReLAVRA). Apesar destes esforços ainda existem lacunas significativas na vigilância de diferentes patógenos devido à falta de metodologias padrão, a ausência de compartilhamento de dados e de cooperação entre os níveis local, nacional e global. Dessa forma foi lançado pela OMS em outubro de 2015 o *Global Antimicrobial Resistance Surveillance System* (Glass), em português, Sistema de Vigilância de Resistência Antimicrobiana Global, com o objetivo de coletar, padronizar, comparar e analisar dados sobre a resistência bacteriana de diferentes países, e assim compartilhar conhecimentos gerados a fim de auxiliar a tomada de decisão no âmbito local, nacional e regional fornecendo informações sobre o desenvolvimento de resistências de bactérias e de deficiências de medicamentos.

O Brasil, por meio do Ministério da Saúde, integrou o sistema Glass ao Plano Nacional de Resistência Antimicrobiana em 2018, agregando, inicialmente, dados epidemiológicos de hospitais do sul do país à base Glass. As informações fornecidas pelo ministério são os resultados do Teste de Sensibilidade a Antimicrobianos (TSA), ou, antibiograma, que indicam a capacidade de uma bactéria resistir aos efeitos dos antibióticos. A adoção dos padrões de vigilância propostos pelo Glass visa a melhora da segurança do paciente, a promoção a otimização do diagnóstico e do uso racional de agentes antimicrobianos.

2.2.1 Teste de Sensibilidade a Antimicrobianos (TSA)

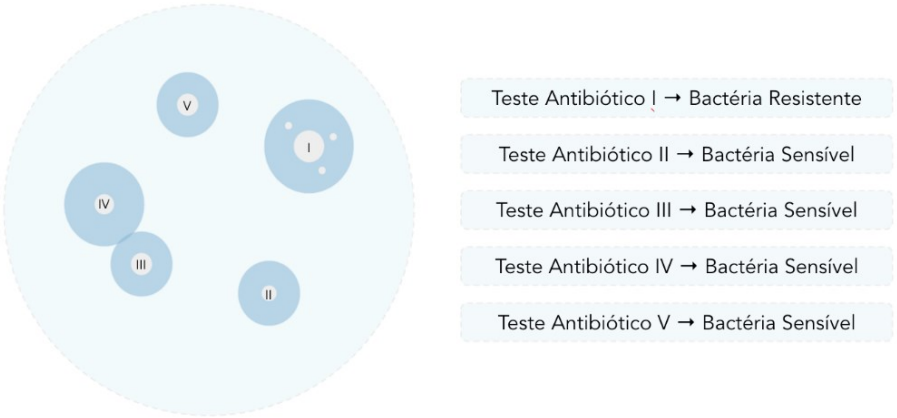
Uma das principais causas do desenvolvimento da resistência de bactérias é a prescrição arbitrária de antibióticos. A prescrição é o processo de escolha e indicação de um tratamento adequado para o paciente e abrange a definição do medicamento, da dosagem e da via de administração. A prescrição tem mais chances de ser feita de forma assertiva quando há

a definição de um diagnóstico preciso e embasado na avaliação fisiológica do paciente (GURGEL; CARVALHO, 2008; OLIVEIRA; MUNARETTO, 2010).

O Teste de Sensibilidade a Antimicrobianos (TSA), também conhecido como antibiograma, é um exame laboratorial que auxilia os médicos a realizar uma prescrição assertiva de fármacos, pois determina o perfil de sensibilidade e resistência de uma bactéria a diferentes antibióticos. Para a realização desse exame é necessário a coleta do material biológico, como, por exemplo, sangue, urina, saliva, catarro, fezes ou célula contaminadas por bactérias. Após a coleta, é feita a cultura, onde a amostra é mantida em um meio favorável para o desenvolvimento e multiplicação das bactérias. Em seguida, o microrganismo é isolado e submetido a testes que têm como objetivo identificar a bactéria que está causando a infecção. Nessa etapa é realizado o antibiograma, onde o material isolado é exposto a diferentes antibióticos para que seja determinada a sua sensibilidade ou resistência a cada um deles. O resultado do antibiograma demora de 3 a 5 dias.

Maier e Abegg (2007) expõem que 66,2% dos médicos entrevistados em seu estudo consideram os resultados do antibiograma valiosos para a escolha do antibiótico, porém, sabe-se que apenas em 12,5% das situações, os antibiogramas são solicitados antes da prescrição. Essa diferença se dá pela carência de recursos de diagnóstico laboratorial, pelas condições socioeconômicas dos pacientes e pela precipitação dos profissionais no momento de realizar a prescrição de antibióticos (OLIVEIRA; MUNARETTO, 2010; MAIER; ABEGG, 2007).

Figura 5 – Exemplificação do teste de sensibilidade a antimicrobianos (TSA).



Fonte: Elaborado pela Autora

2.3 APRENDIZADO DE MÁQUINA (ML)

Arthur Samuel utilizou o termo Aprendizado de Máquina pela primeira vez em 1959 e o definiu como um “campo de estudo que dá aos computadores a capacidade de aprender sem serem explicitamente programados” (ADVANI, 2020). O aprendizado de máquina, em inglês, *Machine Learning* (ML) é uma subárea da ciência da computação que estuda o desenvolvimento de algoritmos computacionais projetados para imitar a inteligência humana que aprende a partir das informações fornecidas por seu entorno (BURKOV, 2019; NAQA; MURPHY, 2015). Pode-se dizer que o ML usa de dados para aprender e responder perguntas. A parte de “uso de dados” é conhecida como treinamento do algoritmo e a parte de “responder perguntas” representa as inferências e previsões que são resultados do algoritmo, o que conecta essas duas partes é o modelo. Os modelos são treinados para fazer previsões cada vez melhores, aprendendo através do conjunto de dados fornecido e quando implantado, prevendo sobre dados não vistos anteriormente (YUFENG, 2017).

Figura 6 – Exemplificação do aprendizado de máquina.

Fonte: Elaborado pela Autora

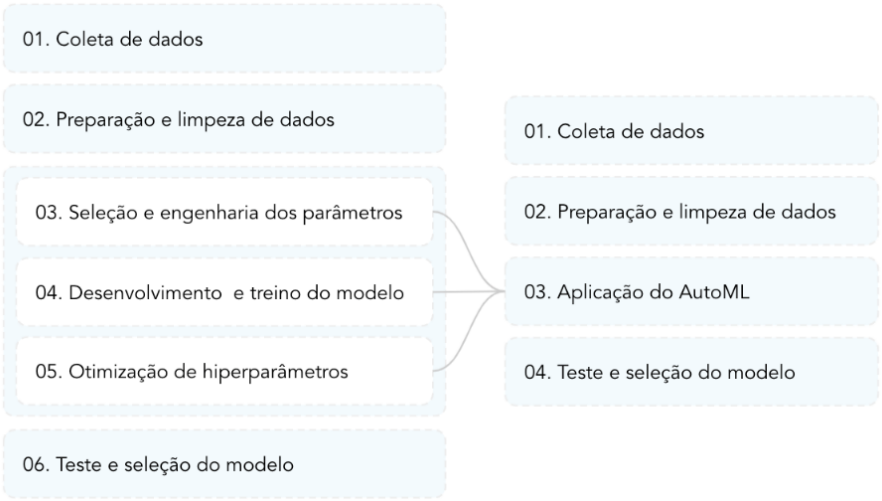
Em relação ao uso de ML no setor da saúde, o aprendizado de máquina é uma das principais técnicas que têm auxiliado na interpretação e transformação de grandes conjuntos de dados eletrônicos em conhecimento acionável. Em geral, o uso de modelos de aprendizado de máquina tem potencial para melhorar a segurança do paciente, melhorar a qualidade do atendimento e reduzir os custos de saúde, ou seja, pode auxiliar na transição e fortalecer o VBHC. Apesar do aprendizado de máquina ter potencial de gerar resultados no setor da saúde, essa técnica não é amplamente utilizada por hospitais. Uma das razões para isso é a diferença entre os conhecimentos necessários para o dia a dia de um profissional da saúde e os conhecimentos necessários para desenvolver e implementar um algoritmo de predição. Nesse contexto, o Aprendizado de Máquina Automatizado, em inglês, *Automated Machine Learning*

(AutoML) pode ser útil, pois facilita a aplicação de técnicas de aprendizagem de máquina e reduz a demanda por profissionais especializados (WARING; LINDVALL; UMETON, 2020).

2.3.1 Aprendizado de Máquina Automatizado (AutoML)

O Aprendizado de Máquina Automatizado (AutoML) foi desenvolvido para atender a demanda por ferramentas de ML mais acessíveis e que permitissem extrair valor e informação dos dados de forma mais ágil e com menos esforços. As ferramentas de AutoML simplificam o processo de treinamento e ajuste do modelo, pois fornecem uma interface simples para processar um grande volume de dados e treinar muitos modelos, o que é útil tanto para um profissional experiente em aprendizado de máquina, quanto para um iniciante. Em resumo, as ferramentas de AutoML automatizam a maioria das etapas do desenvolvimento e implementação de um modelo de aprendizagem de máquina, oferecendo uma única função para substituir um processo que normalmente é composto por várias linhas de código (BALAJI; ALLEN, 2018; LEDELL, 2020).

Figura 7 – Otimização do desenvolvimento de um modelo de ML



Fonte: Adaptado Waring, Lindvall, Umenton (2020) e Advani (2020)

Para compreender o funcionamento das ferramentas de AutoML é indispensável o entendimento do pipeline do aprendizado de máquina. A primeira etapa para o desenvolvimento de um modelo de previsão é a coleta de dados. Essa etapa é muito

importante, pois a qualidade e a quantidade dos dados coletados irão impactar diretamente no quão bom será o modelo. A segunda etapa é composta pela preparação e limpeza dos dados. Isso pode envolver a transformação e normalização dos dados, além da preparação para desenvolvimento do modelo (WARING; LINDVALL; UMETON, 2020; ADVANI, 2020). Para o desenvolvimento de um modelo a partir de uma ferramenta de AutoML, a base deve ser dividida em base de treino, base de validação e base de teste. A terceira etapa do pipeline é a seleção e engenharia dos recursos, que em inglês é conhecido como *Feature Selection* e *Feature Engineering*. Uma *feature*, em português, recursos ou atributos, são uma representação de uma propriedade ou característica de um objeto. Em uma base de dados tabular, por exemplo, as linhas representam as observações e as colunas representariam os atributos. A seleção das *features* é o processo de escolha dos atributos que serão utilizados na construção de modelo, enquanto a engenharia dos atributos é o processo de usar o conhecimento existente na base de dados para criar novos atributos com o objetivo de melhorar o desempenho do modelo (WARING; LINDVALL; UMETON, 2020; ADVANI, 2020).

Figura 8 – Exemplificação do conceito de atributos.

Fonte: Elaborado pela Autora

A partir dessa etapa os processos acontecem de uma maneira iterativa, no qual o modelo é construído, treinado, validado, otimizado, testado e, enfim, selecionado. A etapa de

desenvolvimento e treino do modelo engloba a definição de quais equações matemáticas serão utilizadas para a realização da previsão, e o treino do modelo, que é a adequação dos parâmetros da equação, de acordo com a base de teste. Em sequência é feito a otimização dos hiperparâmetros, responsáveis por ajustes externos e complementares à equação dos modelos, utilizando a base de validação. Com isso, os modelos estão definidos e eles podem ser testados com a base de treino, que não foi utilizada para a determinação dos modelos, ou seja, uma base que contém dados imparciais. Essa etapa tem como propósito identificar se o modelo tem um bom poder preditivo quando aplicado a dados genéricos (WARING; LINDVALL; UMETON, 2020; ADVANI, 2020).

Atualmente, existem várias soluções de AutoML, como Auto_ml, Auto-sklearn, Tree *Based Pipeline Optimization Tool* (TPOT) e H_2O. Essas soluções utilizam técnicas de modelagem difundidas por uma biblioteca de aprendizado de máquina de código aberta, a scikit-learn, que é amplamente utilizada no meio. Apesar disso, os métodos utilizados para a automação e os métodos de avaliação dos modelos diferem imensamente (BALAJI; ALLEN, 2018). Para o desenvolvimento deste trabalho será utilizada a biblioteca, de código aberto, H_2O.

2.3.1.1 H_2O AutoML

O H_2O AutoML é "um algoritmo de aprendizado supervisionado altamente escalável, totalmente automatizado, que automatiza o processo de treinamento de uma grande seleção de modelos em uma única função" (LEDELL, 2020). A eficiência do algoritmo H_2O AutoML está vinculada ao treinamento eficiente de algoritmos de ML. Ele tem a capacidade de gerar muitos modelos em um curto espaço de tempo, sendo que alguns modelos podem ser gerados em apenas milissegundos. Apesar do algoritmo ser automatizado, ele permite uma série de configurações que são expostas como parâmetros das funções, tais como o tempo de parada, o grau de validação *k-fold* e quais tipos de modelos serão considerados. A solução foi lançada em junho de 2017 na versão H_2O v3.12.0.1 e está disponível para uso em Python, R, Java, Scala e na interface web GUI (LEDELL, 2020; BALAJI; ALLEN, 2018).

As otimizações fornecidas pelo H_2O AutoML são as mesmas que foram indicadas na Figura 7. A engenharia dos atributos realizada pela solução inclui *one-hot encoding* para modelos *XGBoost*, separação de dados categóricos em grupos, para modelos baseados em

árvores de decisão, como *Gradient Boosting Machines*(GBM) e *Random Forests* e imputação e normalização quando necessário. Após o pré-processamento dos atributos são gerados os modelos básicos, o H_2O AutoML é capaz de gerar diferentes modelos, como modelos de aprendizado profundo com a técnica de *Deep Neural Network*, ou modelos como o *XGBoost Gradient Boosting Machines* (GBM), *o H20 Gradient Boosting Machines* (GBM), o *Random Forests* (Distributed Random Forest (DRF)), e o *Gradient Linear Machines* (GLM). Em sequência são gerados dois modelos do tipo *Stacked Ensemble* com funções da própria biblioteca H_2O. Esses modelos são uma junção dos modelos base a partir de um algoritmo de meta-aprendizagem para combinar da melhor forma as diferentes previsões (LEDELL, 2020). No Apêndice D as diferentes técnicas de desenvolvimentos de modelos exploradas pela ferramenta H_2O AutoML são explicadas de forma resumida para uma maior compreensão do leitor. Em relação a otimização de hiper parâmetros dos modelos, o algoritmo suporta dois métodos: *cartesian grid search e random grid search* (BALAJI; ALLEN, 2018).

O resultado da execução do H_2O AutoML é uma tabela que lista todos os modelos desenvolvidos e compara os seus desempenhos segundo algumas métricas de avaliação que auxiliam na determinação do melhor modelo. Essa tabela de resultados é chamada *leaderboard* e os modelos listados nela podem ser facilmente exportados para uso em um ambiente de produção, o que facilita o processo de teste dos modelos (LEDELL, 2020).

2.4 AVALIAÇÃO DE MODELOS DE CLASSIFICAÇÃO BINÁRIA

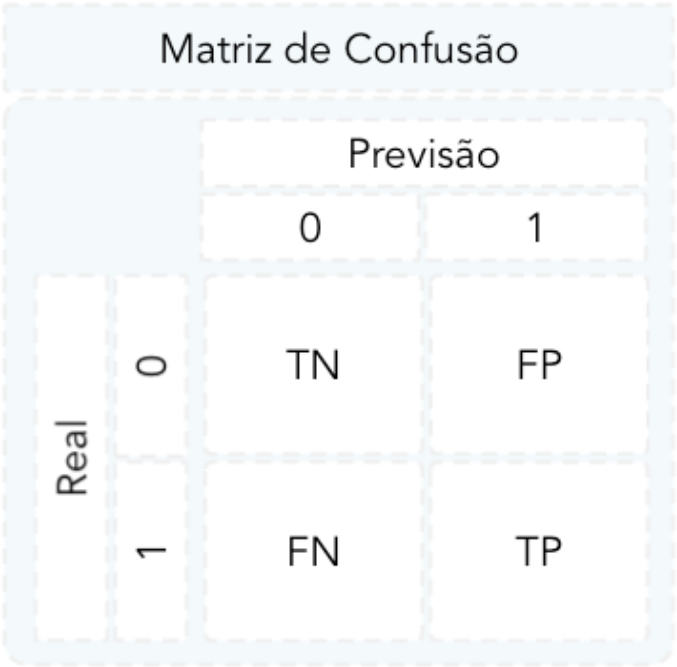
Um modelo de aprendizagem de máquina de classificação binária tem como objetivo prever uma classe, dentre duas opções. No caso estudado neste trabalho, por exemplo, uma bactéria pode ser categorizada como resistente à um antibiótico, o que quer dizer que ela sobrevive mesmo quando exposta a um determinado antibiótico, ou ela pode ser sensível à ele, ou seja, a bactéria não sobrevive quando exposta a um determinado antibiótico. A avaliação de um modelo de classificação binária é feita a partir da comparação entre a classe prevista pelo modelo e as classes verdadeiras obtidas através base de teste do modelo. Apesar de diferentes métricas serem utilizadas para essa avaliação, todas elas têm o objetivo de mensurar quão distante a predição feita pelo modelo está dos dados verídicos. Nessa sessão são abordadas diferentes métricas que são analisadas ao decorrer do trabalho como instrumento de comparação do poder preditivo dos modelos (KUNUMI, 2020).

2.4.1 Matriz de Confusão

A matriz de confusão é uma tabela que resume os resultados do modelo de classificação, essa tabela indica quantos exemplos existem em cada um dos seguintes grupos:

a) Falso positivo (FP) são casos em que o modelo previu a classe 1, que nesse trabalho é a classe das bactérias sensíveis à um antibiótico específico, mas a classe real era 0, ou seja, a bactéria na verdade resiste a esse antibiótico;
b) Falso negativo (FN) são casos em que o modelo previu a classe 0, que nesse trabalho é a classe das bactérias resistentes à um antibiótico específico, mas a classe real era 1, ou seja, a bactéria na verdade é sensível a esse antibiótico;
c) Verdadeiro positivo (TP) são casos em que o modelo previu corretamente a classe 1, ou seja, o modelo identificou corretamente que a bactéria é sensível a um determinado antibiótico;
d) Verdadeiro negativo (TN) são casos em que o modelo previu corretamente a classe 0, ou seja, o modelo identificou corretamente que a bactéria é resistente a um determinado antibiótico.

Figura 9 – Exemplificação da matriz de confusão.



Fonte: Elaborado pela Autora

A matriz de confusão permite uma visualização clara de quantos exemplos foram classificados em cada uma das classes, o que auxilia na interpretação do modelo e na identificação de tendências do modelo.

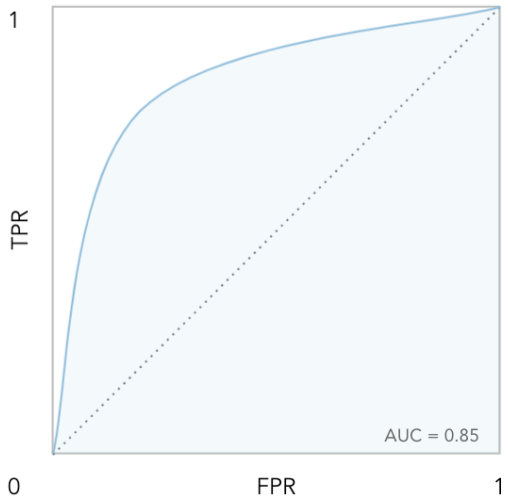
2.4.2 Área sob a curva ROC (AUC)

A curva *Receiver Operating Characteristic curve* (ROC), em português, curva das características operacionais do receptor, é uma métrica para análise de performance dos modelos de classificação que combina a taxa de previsões verdadeiras positivas, em inglês, *True Positive Rate* (TPR), e a taxa de previsões falsas positivas, em inglês, *False Positive Rate* (FPR) para resumir o desempenho da performance do modelo (BURKOV, 2019). A taxa de previsão verdadeira positiva (TPR) e a taxa de previsão falsa positiva (FPR) podem ser definidas como:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{1}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP+TN}} \tag{2}$$

Figura 10 – Curva ROC



Fonte: Adaptado Burkov (2019)

A área hachurada sob a curva ROC, como indicado na Figura 10, em inglês é chamada de *Area Under the Curve* (AUC). Um modelo de classificação perfeito tem o AUC

igual a 1, pois sua taxa de previsões verdadeiras positivas é igual à 100% e a taxa de previsões falsas positivas é igual à 0%, ou seja, o modelo prevê corretamente todas as classes. Um modelo com um AUC maior que 0.5 realiza uma previsão melhor que um modelo aleatório médio e se um modelo possui um AUC menor que 0.5, isso é um indicador de que tem algo errado com o modelo e ele não é confiável (BURKOV, 2019).

2.4.3 *Matthews Correlation Coefficient* (MCC)

O *Matthews Correlation Coefficient* (MCC) é uma métrica utilizada no aprendizado de máquina para avaliar modelos de classificação, ele leva em consideração a quantidade de verdadeiros positivos (TP), falsos positivos (FP), verdadeiros negativos (TN), falsos negativos (FN), além disso é uma métrica independente da distribuição entre as classes da base de treino, ou seja, pode ser usado mesmo que uma classe seja maior que as outras.

$$\mathrm{MCC} = \frac{\mathrm{TP}*\mathrm{TN}-\mathrm{FP}*\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})*(\mathrm{TP}+\mathrm{FN})*(\mathrm{TN}+\mathrm{FP})*(\mathrm{TN}+\mathrm{FN})}} \quad (3)$$

O MCC é um indicador de correlação que varia entre -1 e +1 e algumas propriedades interessantes podem ser facilmente interpretadas a partir da função (3). Quando o modelo prevê todas as classes corretamente, ou seja, FP e FN são iguais a 0, o valor de MCC é igual a 1, o que significa uma previsão perfeita. Em uma situação em que o modelo prevê todas as classes incorretamente, ou seja, TP e TN são iguais a 0, o valor de MCC é igual a -1, o que representa uma previsão inversa. Quando o MCC é igual a 0 entende-se que o modelo apresenta um resultado semelhante a uma previsão aleatória (PEDREGOSA et al., 2011).

2.4.4 Sensibilidade e Especificidade

Na área da saúde, a sensibilidade de um teste diagnóstico representa o percentual de resultados positivos, ou seja, de TP, dentre todas as pessoas que apresentem essa condição clínica, ou seja, a soma entre TP e *False Negative* (FN). Na área de estudos do aprendizado de máquina essa métrica é conhecida como revocação, em inglês, *recall*.

$$\text{Sensibilidade} = \text{Revocação} = \frac{\text{TP}}{\text{TP+FN}} \quad (4)$$

Outra métrica utilizada para analisar a capacidade de diagnóstico de um exame é a especificidade, que é definida por:

$$\text{Especificidade} = \frac{\text{TN}}{\text{TN+FP}} \quad (5)$$

Um exemplo de aplicação dessas métricas no setor da saúde pode ser dado através de um teste de diagnóstico. A sensibilidade é a porcentagem de resultados positivos dentre todas as pessoas infectadas pela doença testada, enquanto a especificidade é a probabilidade de o resultado do teste ser negativo para as pessoas que não estão infectadas. Um teste diagnóstico ideal possui a sensibilidade e especificidade igual a 100%.

2.4.5 F1

A métrica F1, também conhecida como *F-score* ou *F-measure* pode ser interpretada como uma média ponderada, também conhecida como média harmônica, entre a previsão e a revocação de um modelo (PEDREGOSA et al., 2011). A precisão, em inglês, *precision*, é definida pela razão entre a quantidade de TP e o total de observações previstas como positivas, ou seja, a soma entre TP e FP. Uma das características da média ponderada é que se qualquer um dos termos, a precisão ou a revocação, se aproximem de 0 o valor de F1 também se aproximará de 0, ou seja, quando um modelo possui um F1 bom, ele possui uma precisão alta e uma revocação alta.

$$\text{Precisão} = \frac{\text{TP}}{\text{TP+FP}} \quad (6)$$

$$\text{F1} = 2 * \frac{\text{Precisão} * \text{Revocação}}{\text{Precisão+Revocação}} \quad (7)$$

Assim, o F1 é considerado um bom resumo da qualidade do modelo, mas não traz informações acuradas como a precisão, a revocação, e a matriz de confusão.

3 METODOLOGIA

Nesse capítulo são expostas mais informações sobre o cenário de estudos e são apresentados os aspectos metodológicos e os procedimentos que foram executados no decorrer do estudo para alcançar os objetivos propostos.

3.1 CENÁRIO DE ESTUDO

A presente pesquisa tem como base os dados do BR-Glass, que foram disponibilizados pelo Ministério da Saúde ao Grupo 3778. Os dados em questão foram coletados por seis hospitais da região sul do Brasil e contém informações sobre exames laboratoriais chamados de antibiogramas realizados no ano de 2018 e 2019 e sobre o quadro clínico dos pacientes. O projeto de desenvolvimento do BR-Glass começou em 2015 e está sendo tocado pelo Ministério da Saúde. Apesar disso, o Ministério vem tendo dificuldades em integrar novos hospitais na base. Há duas razões principais para isso: a primeira seria a dificuldade de integração dos dados, pois apesar dos dados necessários existirem nos sistemas dos hospitais, eles não se encontram juntos e organizadas da forma que o Ministério solicita. A segunda razão é a falta de interesse dos hospitais, por não verem valor na existência do programa BR-Glass. Com este trabalho busca-se identificar se o compartilhamento de informações epidemiológicas entre hospitais pode ser benéfico para os modelos de previsão locais, e assim, identificar ou não o valor da base do BR-Glass para os hospitais do Brasil.

3.2 ROTEIRO METODOLÓGICO

O embasamento metodologia do presente estudo se baseou no livro *The Art of Data Science* escrito por Peng e Matsui (2015). Segundo eles existem cinco atividades principais vinculadas ao processo de análise de dados:

a) Declarar e refinar a questão da pesquisa;
b) Explorar dados;
c) Construir modelos estatísticos;
d) Interpretar resultados;
e) Comunicar resultados.

A interpretação dos resultados da análise é explorada no capítulo 4 deste trabalho e a comunicação dos resultados, está sendo executada por meio deste trabalho como um todo. Dessa forma essas duas etapas não são exploradas nesse capítulo.

3.3 DESENVOLVIMENTO

3.3.1 Declaração e Refinamento da Questão de Pesquisa

De acordo com Leek e Peng (2015) existem apenas seis tipos de perguntas a serem declaradas, perguntas descritivas, exploratórias, inferenciais, preditivas, casuais e mecanicistas. Neste trabalho está sendo abordada uma questão exploratória. Esse tipo de questão é respondido através da análise de dados que permite a identificação de padrões, tendências e relações entre as variáveis. Elas também são conhecidas como análises geradoras de hipóteses, por que em vez de testar uma hipótese ela tem o objetivo de identificar padrões que apoiam a proposição de uma hipótese (PENG; MATSUI, 2015).

A etapa de refinamento deve considerar as principais cinco características que definem uma boa questão de pesquisa:

a) A questão deve ser de interesse e um segmento;
b) A questão ainda não pode ter sido respondida;
c) A questão deve originar-se de uma hipótese plausível;
d) Ser uma questão passível de resposta;
e) A questão deve ser específica.

Dessa forma, a proposta do presente trabalho de comparar o poder preditivo de modelos de aprendizagem de máquina treinados com dados epidemiológicos nacionais compartilhados e treinados com, apenas, os dados epidemiológicos das instituições locais. E, assim, identificar as vantagens ou desvantagens do compartilhamento de dados epidemiológicos entre hospitais.

3.3.2 Exploração de Dados

A exploração de dados, tem como objetivo determinar se tem algum problema com os dados da base que será utilizada nas análises e se a questão da pesquisa pode ser

respondida com os dados disponíveis nessa base. Além disso, é nessa etapa que se identifica quais serão os procedimentos aplicados ao banco de dados para obtenção dos resultados desejados (PENG; MATSUI, 2015).

A base do BR-Glass é composta por dados clínicos agregados aos resultados de antibiogramas. Cada linha da base representa parte do resultado do antibiograma de um paciente, sendo uma combinação do identificador do paciente, da bactéria que o infecta e do antibiótico que está sendo testado no exame. Antes de executar a análise exploratória dos dados foi realizado um benchmark com Pillonetto et al. (2020), pois o estudo publicado por eles utilizou a mesma base de dados que está sendo utilizada neste estudo. Além disso, médicos do Grupo 3778 foram contactados para melhor entendimento dos dados de epidemiologia. A partir dessas informações foi feita uma limpeza da base na qual foram retirados os testes que não são praticados em um cenário real hospitalar e que atrapalhariam o poder de predição do modelo, além do que, alguns atributos foram simplificados.

3.3.2.1 Tratamento da Base BR-Glass

A primeira alteração realizada na base foi em relação à idade dos pacientes, pois algumas idades estavam em dias, outras em meses e outras em anos. A primeira conversão foi unificar a unidade de medidas em anos, para em seguida criar categorias de idades que interferem na determinação de uma prescrição médica. Dessa forma, as categorias de idade utilizadas são:

a) Recém-nascido: com até 28 dias de vida;
b) Lactante: entre 28 dias e 2 anos de vida;
c) Criança: entre 2 e 12 anos de vida;
d) Adolescente: entre 12 e 18 anos de vida;
e) Adulto: entre 18 e 60 anos de vida;
f) Idoso: com mais de 60 anos de vida.

Em seguida, os dados relacionados com os antibiogramas foram tratados. Cada linha da base representa um teste específico que determina o perfil de sensibilidade de uma bactéria em relação a um antibiótico, o perfil pode ser categorizado como “sensível”, “intermediário” e “resistente”. Todos os resultados “intermediários” foram alterados para resistentes, porque, de acordo com a aplicação clínica os resultados intermediários não indicam uma boa combinação para prescrição.

A terceira alteração foi em relação aos nomes antibióticos. Alguns dos registros possuíam espaços ao final do registro, o que foi excluído, além de que alguns nomes de antibióticos foram corrigidos e outros foram excluídos como é mostrado no Quadro 1 do Apêndice A. Os nomes dos microrganismos também foram tratados: foi retirado o espaço ao fim dos registros e alterada a nomenclatura de alguns, de acordo com o padrão estabelecido para facilitar a análise de dados entre as diferentes instituições. As alterações na nomenclatura dos microrganismos estão dispostas no Quadro 2 do Apêndice A.

Em sequência, com base no conhecimento dos especialistas em epidemiologia, foram excluídos da base os registros que não estão de acordo com a prática clínica, tais como:

a) Antimicrobianos que não são utilizados em quadros com um grupo específico de bactérias;
b) Antimicrobianos que não são utilizados em quadros com bactérias da família *Enterobacteriaceae*;
c) Antimicrobianos que não são utilizados em quadros com um gênero específico de bactérias;
d) Antimicrobianos que não são utilizados como materiais de coleta específicos;
e) Antimicrobianos que não são compatíveis com algumas bactérias específicas.

Todas essas alterações da base foram efetuadas com o objetivo de aproximar os registros à realidade da prática clínica e podem ser observados no Apêndice A nos Quadros 3, 4, 5 e 6, respectivamente. Após a limpeza da base foi executada a análise exploratória dos dados defendido por Peng e Matsui (2015).

3.3.2.2 Análise Exploratória dos Dados

A *Exploratory Data Analysis* (EDA), em português, análise exploratória dos dados é um processo amplamente difundido na área de estudo do aprendizado de máquina. O objetivo do EDA é criar familiaridade com os dados e reduzir a carga de trabalho durante o processo de análise (COX, 2017). Dessa forma, nessa subseção são abordadas características da base BR-Glass após tratamento.

Cada linha da base representa uma combinação única do paciente, da bactéria que o infecta e o antibiótico que está sendo testado no antibiograma. A base contém dados clínicos

do paciente e seus resultados do antibiograma, esses dados estão distribuídos nas colunas em forma de atributos que são utilizados para a construção do modelo:

a) id_paciente: número único de identificação do paciente no sistema de origem;
b) estabelecimento: instituição de saúde de onde se originam os dados;
c) idade_categ: idade do paciente em categorias;
d) sexo: gênero do paciente que pode ser definido em masculino, feminino e ignorado;
e) atendimento: tipo de atendimento, que é classificado em ambulatorial ou hospitalar;
f) area_isolada: agrupamento do material biológico testado;
g) material_isolado: material biológico testado;
h) microrganismo: microrganismo analisado;
i) grupo: coloração gram, definida em gram positivo ou gram negativo;
j) familia: família do microrganismo analisado;
k) antibiotico: nome do antibiótico testado.

Tabela 1 – Características das bases individuais dos hospitais da base BR-Glass

Hospital	**Número de registros**	**Número de pacientes**	**Número de microrganismos testados**	**Número de antibióticos testados**	**Taxa de sensibilidade**
h1	45 126	5 778	66	35	76.38%
h2	8 641	839	34	31	69.94%
h3	109 868	16 240	60	33	71.39%
h4	55 602	5 435	58	41	61.45%
h5	71 193	2 698	50	31	70.67%
h6	996	199	1	5	69.45%
Total	**227 425**	**31 189**	**104**	**45**	**69.93%**

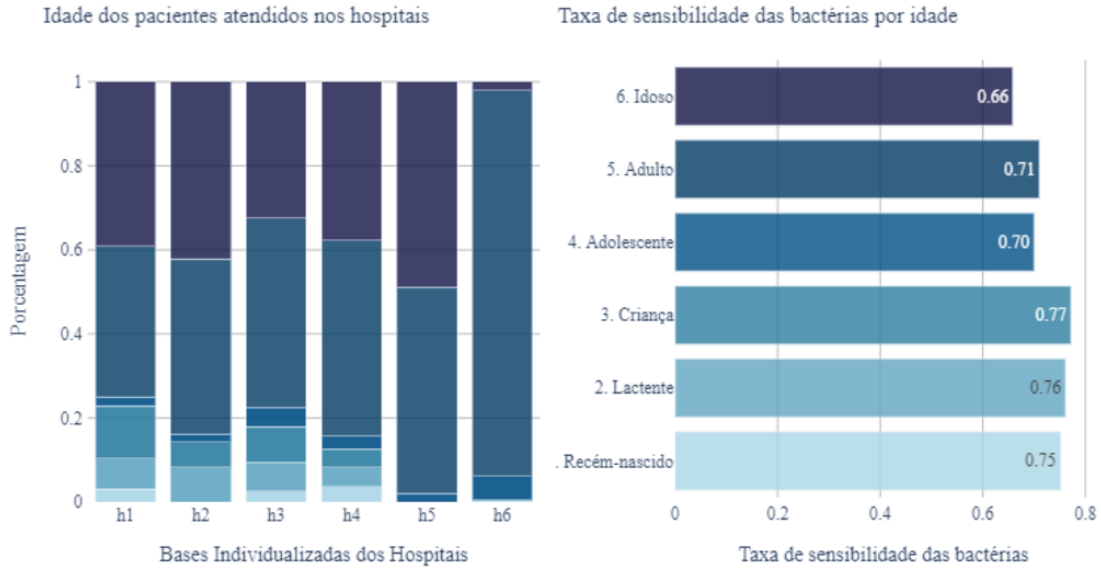
Fonte: Elaborado pela Autora

A base de dados possui 227.425 registros, que contém dados de 31.189 pacientes de 6 hospitais diferentes. Nesses dados existe o registro de 13 tipos de áreas isoladas, como por exemplo trato gastrointestinal, sistema nervoso central e trato respiratório, e 57 possibilidades de material isolado, tais como lavado gástrico, linfonodo e fragmento de pulmão. Os registros dispostos na base testam 45 antibióticos diferentes, sendo que nenhum dos hospitais testam todos esses antibióticos, por exemplo o hospital ‘h1’ testa 35 dos 45 antibióticos dispostos da base BR-Glass. Em relação aos microrganismos a base registra 104 microrganismos pertencentes a 21 famílias, e, novamente, nenhuma das instituições estudadas testa os 104

microrganismos, por exemplo, o hospital ‘h6’ testa apenas um desses microrganismos. Além disso, sabe-se que das combinações de testes feitos entre bactéria e antibiótico, em 69.93% dos resultados, as bactérias mostraram-se sensíveis aos antibióticos. Essa taxa será referida como “Taxa de Sensibilidade” ao decorrer do trabalho. O valor da taxa de sensibilidade da base completa é obtido a partir da média ponderada entre a taxa de sensibilidade e o número de registros de cada hospital da base. Apesar dos hospitais possuírem um perfil semelhante, percebe-se com os dados expostos na Tabela 1 que a quantidade de registros e a quantidade de pacientes disponibilizados por cada um deles difere. No que diz respeito aos antibióticos há uma certa semelhança em relação aos antibióticos testados. Apenas o hospital apelidado de ‘h6’ realiza um teste específico que envolve apenas um microrganismo e 5 antibióticos. Em relação ao perfil de sensibilidade dos microrganismos, há uma diferença nos perfis identificados em cada um dos hospitais e na taxa de sensibilidade de cada um deles, entre a menor e maior taxa existe uma variação de 15 pontos percentuais.

Quando o perfil dos pacientes atendidos em cada um dos hospitais é analisado observa-se que a maioria dos pacientes testados em todas as instituições são de adultos ou idosos, sendo que os hospitais ‘h5’ e ‘h6’ não testam recém-nascidos, lactantes e crianças, como é mostrado no gráfico à esquerda da Figura 11. Quando calculada a taxa de sensibilidade para cada uma das categorias de idade, observa-se uma tendência de redução de sensibilidade, ou seja, aumento de testes que identificam um padrão de resistência nas bactérias em pessoas mais velhas, como é mostrado na no gráfico à direita da Figura 11.

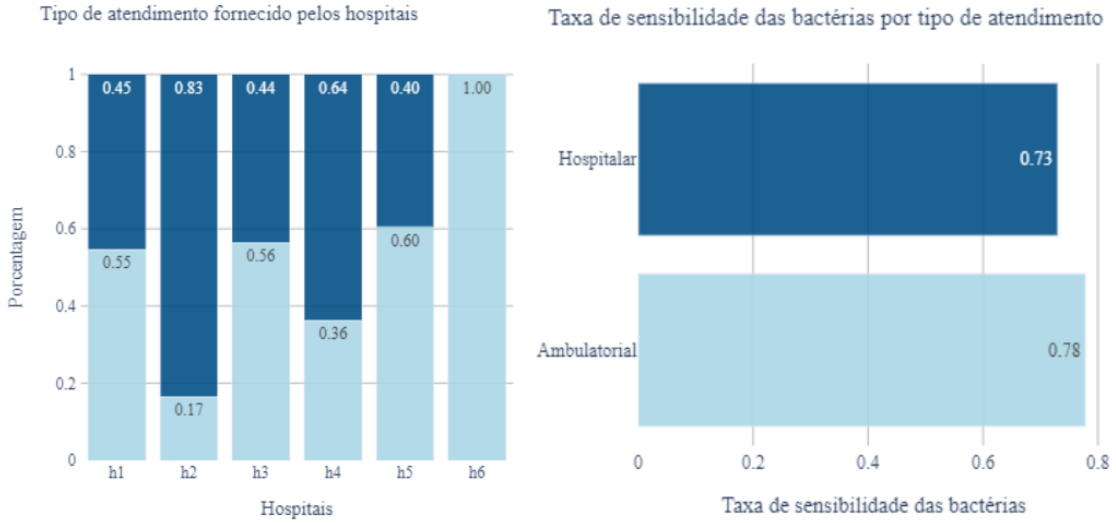
Figura 11 - Análise da idade dos pacientes de cada uma das instituições e da taxa de sensibilidade para cada uma das faixas etárias utilizadas



Fonte: Elaborado pela Autora

Em relação ao tipo de atendimento, duas categorias são apresentadas nos dados: a hospitalar que é vinculada ao atendimento que necessita de internação e a ambulatorial, quando o atendimento é pontual, como em uma consulta. A partir das análises indicadas na Figura 12 foram identificadas características de atendimento semelhante entre os hospitais ‘h1’, ‘h3’, ‘h5’ e ‘h6’. Neles a solicitação de antibiogramas é feita majoritariamente em atendimentos do tipo ambulatorial. Já nos hospitais ‘h2’ e ‘h4’, os antibiogramas são solicitados majoritariamente nos atendimentos do tipo hospitalar. Em adição à essa informação, nos atendimentos que necessitam de internação, as bactérias testadas possuem uma menor taxa de sensibilidade, ou seja, apresentam resistência a antibióticos em um maior número de testes, em comparação com os testes oriundos de atendimentos ambulatoriais, o que confirma a afirmação de Silva (2008) de que a resistência bacteriana é mais comum nos germes hospitalares, onde há um maior uso de antibióticos.

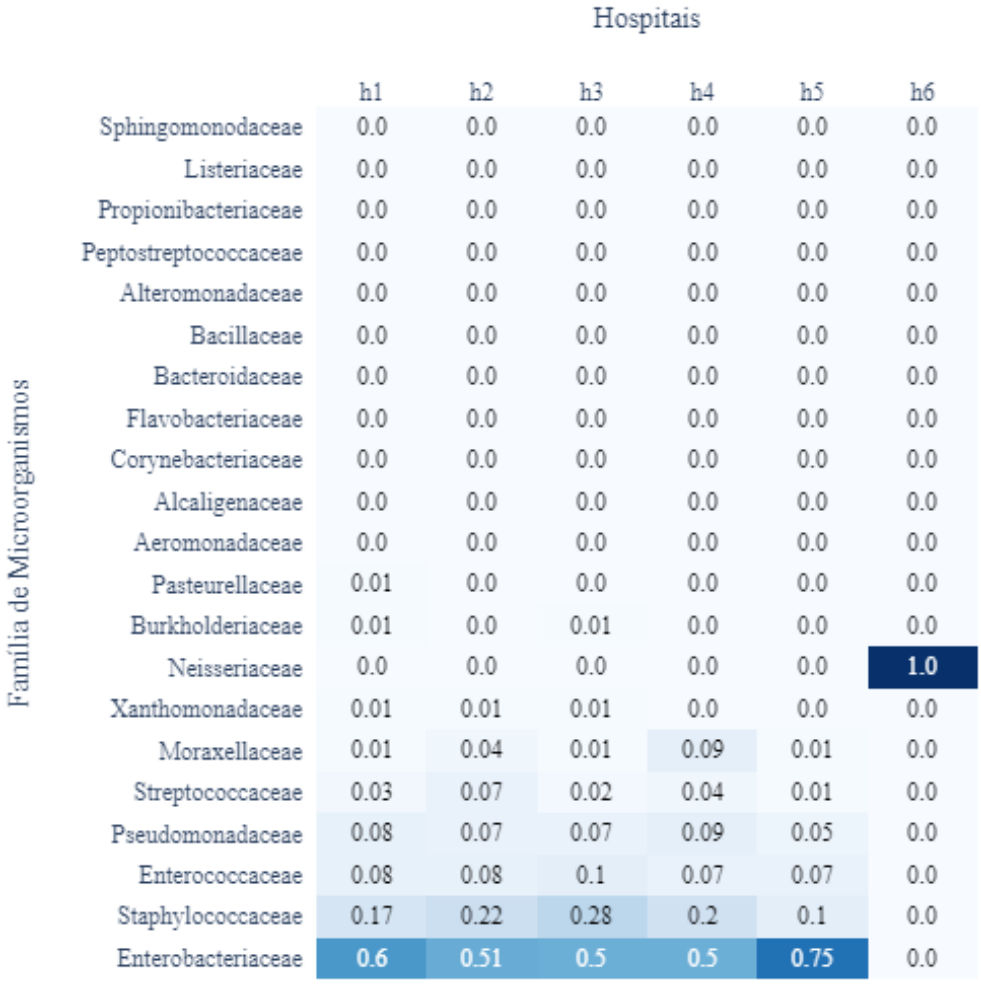
Figura 12 – Análise do tipo de atendimento recebido pelos pacientes analisados em cada uma das instituições e da taxa de sensibilidade para cada um dos tipos de atendimento ofertados



Fonte: Elaborado pela Autora

Em relação aos antibiogramas realizados pelas instituições presentes na base, o hospital ‘h6’ é o único que realiza exames que envolvem apenas um microrganismo específico que pertence à família Neisseriaceae, enquanto as demais instituições de saúde realizam uma maior variedade de testes. Apesar de testarem diferentes microrganismos que pertencem a diferentes famílias, percebe-se na Figura 13, que algumas famílias específicas são testadas de forma mais frequentes, ou seja, são famílias que ocasionam as infecções mais comuns. Nos hospitais ‘h1’, ‘h2’, ‘h3’, ‘h4’ e ‘h5’ a maior parte dos testes realizados envolviam bactérias da família *Enterobacteriaceae*.

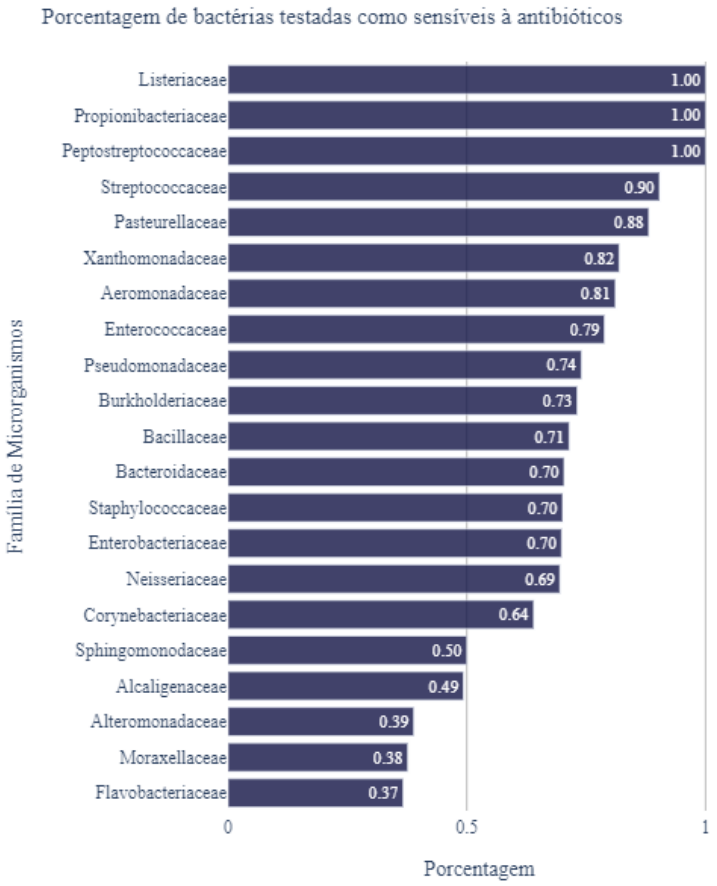
Figura 13 – Famílias de microrganismos mais testados nos hospitais.



Nota: Valores decimais são a porcentagem de testes realizados com bactérias pertencentes às famílias em cada uma das instituições.
Fonte: Elaborado pela Autora

Quando calculada a taxa de sensibilidade de cada uma das famílias de microrganismos, foi identificado que a família a que pertence o microrganismo impacta na probabilidade de uma bactéria ser resiste ou não aos antibióticos, como mostra a Figura 14. Como exemplo pode-se citar a família Listeriaceae que se mostrou sensível a todos antibióticos testados, ou seja, não sobreviveu em presença de nenhum dos antibióticos utilizados durante a execução do Teste de Sensibilidade a Antimicrobianos (TSA).

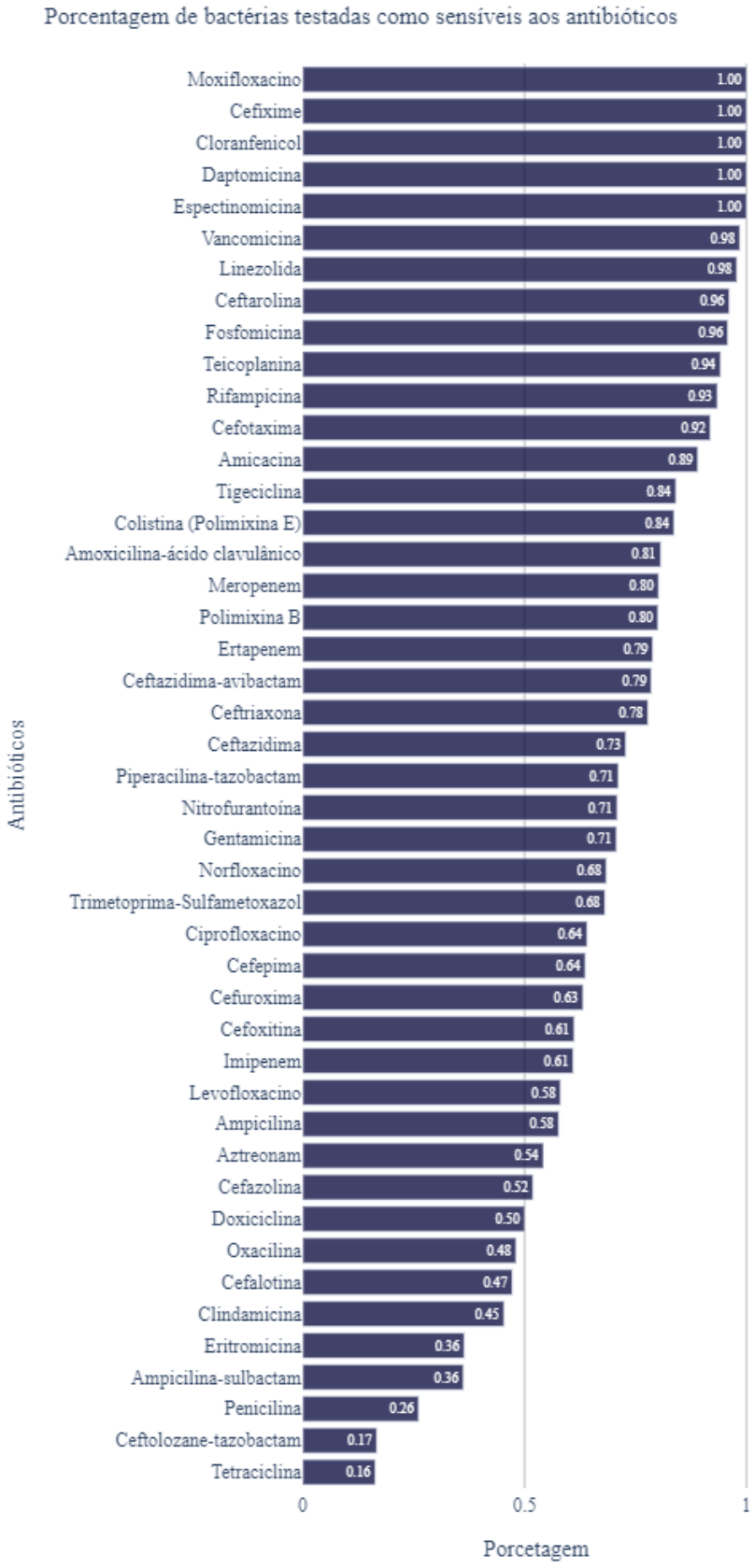
Figura 14 – Taxa de sensibilidade média das famílias dos microrganismos



Fonte: Elaborado pela Autora

A mesma análise e inferência pode ser feita com os antibióticos, pois existem alguns antibióticos que são considerados “mais fortes” e outros que são considerados “mais fracos”. Os mais fortes são capazes de suprimir o desenvolvimento de qualquer bactéria, mas não devem ser utilizados em todos os casos, porque muitas das vezes um antibiótico “mais fraco” é suficiente. Os antibióticos mais potentes, normalmente, possuem uma maior toxidade, além de que o uso contínuo deles pode resultar na sobrevivência e multiplicação das bactérias ainda mais resistentes.

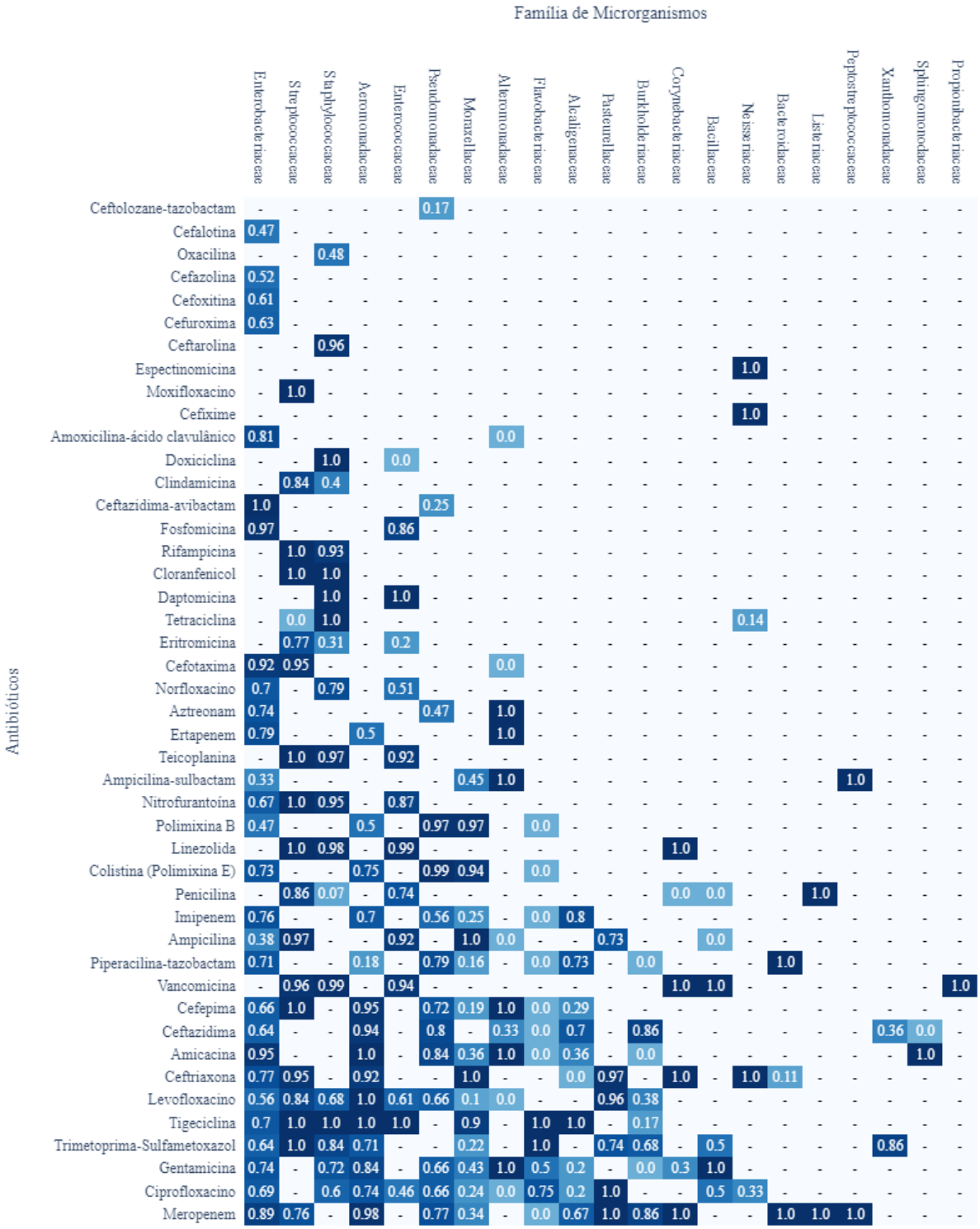
Figura 15 – Taxa de sensibilidade média dos microrganismos na presença dos antibióticos



Fonte: Elaborado pela Autora

Quando se realiza uma análise cruzada entre família de microrganismos e antibióticos é possível perceber quais combinações que não fazem sentido em uma realidade clínica (Figura 16 está marcada como "-", cor mais clara do gráfico). Também se percebe aquelas combinações em que o antibiótico se mostrou eficaz no combate a certa família de microrganismos em 100% dos testes, categorizados como 1 na Figura 16, e outras combinações, categorizadas como 0, indicam que o antibiótico não obteve sucesso no combate de uma certa família de bactérias em todos os testes executados.

Figura 16 – Taxa de sensibilidade média apresentada pelas famílias de microrganismos em presença de antibióticos



Fonte: Elaborado pela Autora

Com o EDA ficou claro que os resultados do Teste de Sensibilidade a Antimicrobianos (TSA) realizado a partir de uma coleta de um paciente são representados em

várias linhas da base de dados, pois em um TSA, o mesmo microrganismo identificado na cultura é testado com um conjunto de antibióticos.

Dessa maneira, como para a construção dos modelos estatísticos é necessário separar a base em treino, validação e teste. Para realizar a divisão da base em três foi respeitada a premissa que os dados de um mesmo paciente permaneceriam em uma mesma base. Assim, a separação de 70%, 15%, 15% para as bases de treino, validação e teste, respectivamente, foi feita com base no atributo chamado de id_paciente e não na quantidade de registros da base. Em outras palavras, os códigos de id_paciente registrados na base de treino não estarão presentes na base de validação, nem na base de teste, e o contrário também é verdade. Além disso, para possibilitar a realização dos estudos, sete bases de dados diferentes foram criadas. Sendo uma delas contendo todos os dados do BR-Glass, que chamar-se-á de "Base Compartilhada", e outras seis bases, cada uma delas contendo os dados de cada uma das instituições cadastradas no BR-Glass de forma separada, que serão chamadas de "Bases Individualizadas". Como exemplo, a base de teste da base compartilha é a junção de todas as bases de teste derivadas das bases individualizadas.

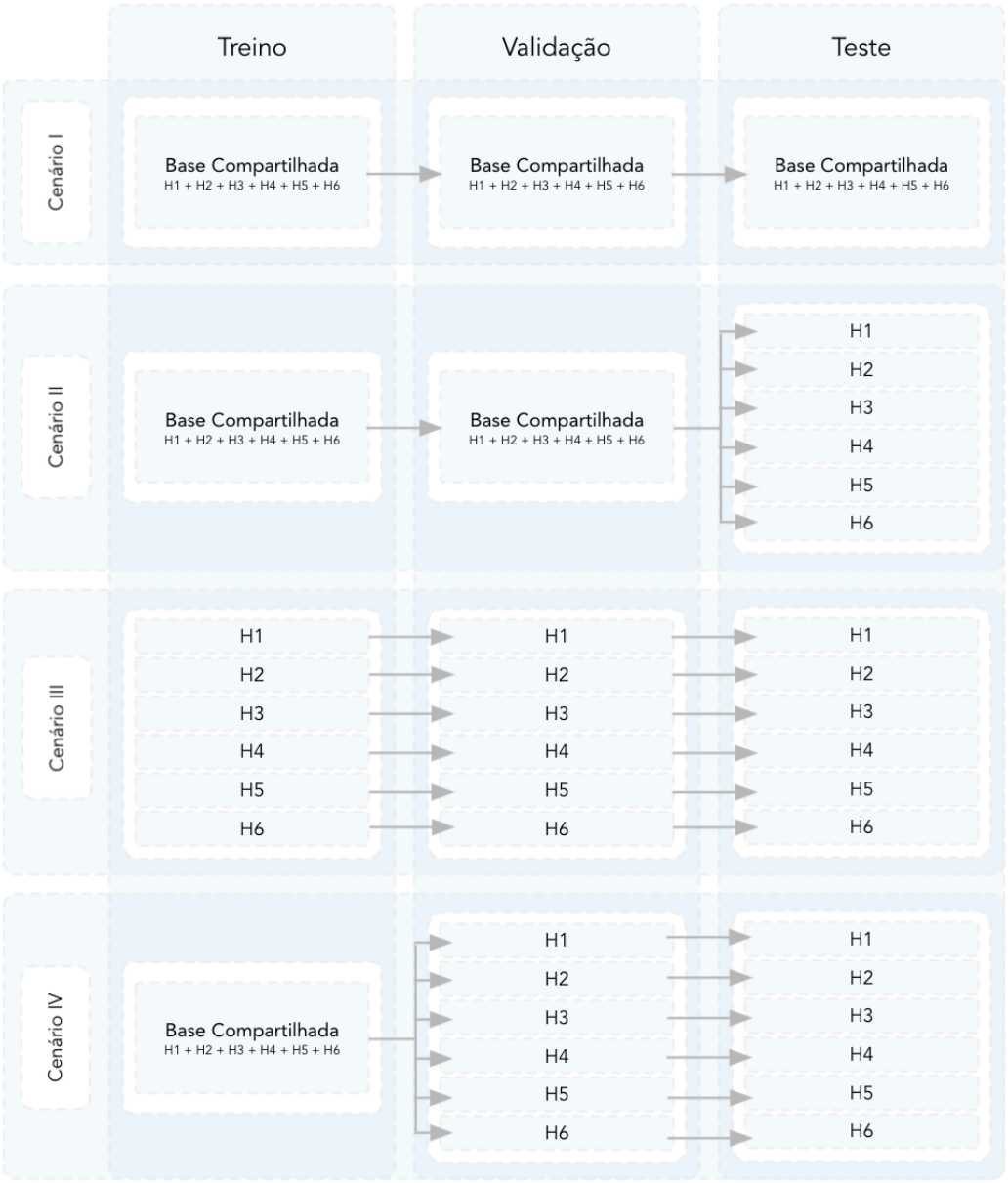
3.3.3 Construção dos modelos estatísticos

O processo de desenvolvimento de um modelo estatístico envolve a determinação de uma estrutura específica aos dados e a criação de um resumo dos dados, nessa sub sessão é explicado a estrutura utilizada para o desenvolvimento dos modelos (PENG; MATSUI, 2015). Para tanto é importante relembrar que o objetivo deste trabalho é comparar o poder preditivo de modelos de aprendizagem de máquina treinados com dados epidemiológicos nacionais, disponíveis na base do BR-Glass, e treinados apenas com os dados epidemiológicos das instituições locais, e para isso são utilizados os dados de diferentes hospitais que também estão disponíveis na base do BR-Glass.

A comparação de poder preditivo é feita em quatro cenários diferentes e em todos os cenários os modelos são desenvolvidos com a ferramenta H_2O AutoML. No primeiro cenário a Base Compartilhada, que contém informações de seis hospitais diferente, é dividida em base de treino, base de validação e base de teste. Nesse cenário são desenvolvidos modelos treinados e validados com a base de treino e a base de validação derivadas da Base

Compartilhada, respectivamente. Após desenvolvimento dos modelos eles são testados na base de teste derivada da Base Compartilhada.

Figura 17 – Lógica utilizada para a construção e teste dos modelos



Fonte: Elaborado pela Autora

No segundo cenário os modelos desenvolvidos no cenário I, modelos treinados e validados com os dados de seis hospitais diferentes, são testados, individualmente, nas bases de teste das seis instituições derivadas das Bases Individualizadas. Ou seja, o cenário II testa o poder de previsão de um modelo que se beneficia do compartilhamento de dados em uma situação local.

No terceiro cenário são desenvolvidos modelos para cada um dos hospitais separadamente. Nesse cenário as Bases Individualizada são separadas em treino, validação e teste. Com isso, são desenvolvidas modelos com a base de treino e validação derivadas das Bases Individualizadas, e então os modelos resultantes são aplicados nas bases de teste, também derivadas das Bases Individualizadas. Dessa forma, esse cenário simula uma situação de não compartilhamento de informações entre as instituições.

No quarto e último cenário são construídos modelos treinados com a base teste derivada da Base Compartilhada, mas validados com seis bases de validação diferentes derivadas das Bases Individualizadas. Ou seja, no quarto cenário o treino dos modelos é feito com os dados da Base Compartilhada, enquanto o ajuste dos hiper parâmetros é feito com os dados das Bases Individualizadas. Como resultado temos modelos específicos para cada um dos hospitais que são aplicados nas bases testes derivados das Bases Individualizadas. Esse cenário simula uma outra forma de utilização dos dados compartilhados para desenvolvimento de modelos preditivos, em comparação com o cenário II.

3.3.3.1 Critério de Parada do AutoML

Após definida a estrutura de comparação e análise que seria usada para o desenvolvimento dos modelos estatísticos foi considerada à utilização e aplicação dos parâmetros obrigatórios da ferramenta de AutoML utilizada. O único parâmetro obrigatório da função de desenvolvimentos de modelos da ferramenta H2O AutoML é o argumento que define o seu critério de parada. Existem duas opções para determinar o fim a execução da função, pode-se determinar a quantidade de modelos a serem desenvolvidos, e quando eles estiverem prontos o algoritmo para, ou pode-se determinar o tempo em que a função irá rodar. Neste estudo foi utilizado o critério de parada nomeado como “max_runtime_secs”, é o argumento especifica o tempo em que a função do AutoML será executada, antes de começar a treinar os modelos finais do *Stacked Ensemble* (LEDELL, 2020).

Para definir o “max_runtime_secs” que é utilizado durante o experimento foram desenvolvidos testes no Cenário I com diferentes tempos máximos de processamento: 10 minutos, 1 hora, 2 horas e 5 horas. O intuito desses testes é compreender qual a interferência do parâmetro no poder de predição dos modelos, dessa forma pode-se comparar a quantidade de modelos gerados de acordo com o tempo máximo definido:

a) Em 10 minutos foram gerados 20 modelos base e 2 do tipo *Stacked Ensemble*;
b) Em 1 hora foram gerados 31 modelos base e 2 do tipo *Stacked Ensemble*;
c) Em 2 horas foram gerados 80 modelos base e 2 do tipo *Stacked Ensemble*;
d) Em 5 horas foram gerados 125 modelos base e 2 do tipo *Stacked Ensemble*.

O resultado da execução do H2O AutoML é uma tabela que lista todos os modelos desenvolvidos que é chamada de *leaderboard*. O *leaderboard* de modelos de classificação binária, como no caso exposto nesse trabalho, é ordenado segundo o indicador ROC AUC de forma decrescente. O *leaderboard* também traz outras informações e métricas dos modelos desenvolvidos, é importante salientar que as métricas trazidas no *leaderboard* são calculadas a partir da aplicação dos modelos nas bases de validação.

Os 10 melhores modelos gerados em cada uma das configurações de tempo de execução citados acima, segundo a classificação do *leaderboard*, foram aplicados a base treino utilizada no cenário I, que é a base treino derivada da base compartilhada. Com isso, foi possível mensurar o poder de previsão dos modelos quando aplicados a novos dados de acordo com as métricas abordadas no capítulo 2. Em sequência, foi selecionado o melhor modelo, com o maior valor de ROC AUC, para cada um dos tempos de execução testados.

Dessa forma, foi possível comparar a diferença entre o poder preditivo do melhor modelo para cada um dos valores de "max_runtime_secs" utilizados, é interessante pontuar que todos os 4 melhores modelos selecionados pertencem ao tipo *Gradient Boosting Machines* (GBM).

Tabela 2 – Comparação do poder preditivo dos modelos segundo o tempo máximo de execução do AutoML

Run Time	**AUC**	**F1**	**MCC**	**Precisão**	**Revocação**
10 min	0.704	0.860	0.468	0.805	0.922
1 hr	0.711	0.863	0.483	0.809	0.925
2 hrs	0.710	0.861	0.477	0.808	0.922
5 hrs	0.709	0.861	0.474	0.808	0.921

Fonte: Elaborado pela Autora

Com as informações expostas na Tabela 2 é possível afirmar que o aumento do tempo máximo de execução do AutoML interfere diretamente na quantidade de modelos que são gerados, mas que, para o caso específico analisado neste trabalho, não parece afetar no desempenho dos melhores modelos, pois não há uma variação significativa nas métricas

apresentadas. Baseado nisso ao decorrer do experimento foram utilizados dois "max_runtime_secs" diferentes, 10 minutos e duas horas, pois se enquadravam dentro do orçamento existente para o experimento.

3.3.3.2 Modelos Estatísticos do Cenário I

No cenário I são gerados modelos de predição que são treinados e validados com bases que derivam da base compartilhada, para desenvolvimento desses modelos o H_2O AutoML foi executado em 10 minutos e em duas horas. Com isso obteve-se 104 modelos para o cenário I, os 10 melhores de cada execução do AutoML segundo *o leaderboard* estão dispostos no Apêndice B.

3.3.3.3 Modelos Estatísticos do Cenário II

No cenário II são utilizados os modelos gerados no cenário I, o que muda nesse cenário em comparação com o cenário anterior é a aplicação dos modelos que será abordada na próxima sucessão.

3.3.3.4 Modelos Estatísticos do Cenário III

No cenário III são gerados modelos de predição que são treinados e validados com bases que derivam da base individualizadas, ou seja, são gerados modelos para cada instituição de saúde individualmente. Para desenvolvimento desses modelos o H_2O AutoML foi executado em 10 minutos e em duas horas. Com isso obteve-se:

a) Para o hospital 'h1' foram gerados 159 modelos;
b) Para o hospital 'h2' foram gerados 187 modelos;
c) Para o hospital 'h3' foram gerados 156 modelos;
d) Para o hospital 'h4' foram gerados 118 modelos;
e) Para o hospital 'h5' foram gerados 158 modelos;
f) Para o hospital 'h6' foram gerados 420 modelos.

Totalizando 1.198 modelos gerados para o cenário III. Após análise dos modelos gerados foram selecionados os 10 melhores para cada uma das instituições segundo o *leaderboard*, com maior valor de ROC AUC, que estão dispostos no Apêndice B.

3.3.3.5 Modelos Estatísticos do Cenário IV

No cenário IV são gerados modelos de predição que são treinados com parte da base compartilhada e validados com bases que derivam das bases individualizadas, ou seja, são gerados modelos para cada instituição de saúde individualmente que se beneficiam do compartilhamento de dados. Para desenvolvimento desses modelos o H_2O AutoML foi executado em 10 minutos e em duas horas. Com isso obteve-se:

a) Para o hospital ‘h1’ foram gerados 108 modelos;
b) Para o hospital ‘h2’ foram gerados 98 modelos;
c) Para o hospital ‘h3’ foram gerados 92 modelos;
d) Para o hospital ‘h4’ foram gerados 91 modelos;
e) Para o hospital ‘h5’ foram gerados 93 modelos;
f) Para o hospital ‘h6’ foram gerados 111 modelos.

Totalizando 593 modelos gerados para o cenário IV. Os 10 melhores, para cada uma das instituições, segundo o *leaderboard*, estão dispostos no Apêndice B.

4 RESULTADOS E DISCUSSÕES

Para analisar o poder preditivo dos modelos é imprescindível que sua aplicação seja feita em uma base de dados que não foi utilizada em sua construção. Por isso, desde o começo do desenvolvimento dos modelos foram separadas bases para teste. Devido ao grande número de modelos gerados, foi priorizado a aplicação dos 10 melhores modelos de cada execução do AutoML ordenados no *leaderboard*, de acordo com o maior valor de AUC, para cada um dos cenários e instituições. Em outras palavras, foram aplicados todos os modelos listados no Apêndice B. Dessa forma, primeiramente, foram aplicados os 10 melhores modelos gerados com o tempo de execução do AutoML configurado em 10 minutos e em sequência foram aplicados os 10 melhores modelos gerados com o tempo de execução do AutoML configurado em 2 horas. A partir da aplicação dos modelos nas bases testes foram calculadas métricas que possibilitam avaliar o poder preditivo de cada um dos modelos, essas métricas foram calculadas segundo as funções definidas na biblioteca sckit-learn. O melhor modelo para cada um dos cenários foi determinado através do maior valor de AUC quando se aplica o modelo à base teste. Apesar de a métrica AUC ser utilizada para determinar qual é o modelo com o melhor poder preditivo, outras métricas serão calculadas para fornecer uma melhor análise do poder preditivo dos modelos.

4.1 PODER PREDITIVO DOS MODELOS ESTATÍSTICOS

4.1.1 Cenário I

O cenário I simula uma situação em que os modelos são treinados, validados e testados na base de dados compartilhada, que contém informações de seis instituições de saúde diferentes. Nas Tabelas 3 e 4 estão registradas as métricas calculadas a partir dessas aplicações. Pode-se observar que o tempo de execução não afetou o desempenho dos modelos, pois os valores das métricas de desempenho se assemelham. A coluna nomeada “Index” das tabelas de mensuração do poder preditivo indica a ordem dos modelos apresentadas no *leaderboard* resultante do AutoML, isso quer dizer que o modelo com index 1 é o melhor modelo segundo o H_2O AutoML. Porém, pode-se notar que o melhor modelo segundo o *leaderboard* não foi o modelo que obteve um melhor desempenho quando aplicado

à base teste. Nota-se que a performance dos 20 modelos aplicados ao cenário I é semelhante, pois não há grande variação entre os valores dos indicadores utilizados.

A partir da aplicação dos modelos na base teste e da análise das métricas foi determinado que o melhor modelo gerado no cenário I foi gerado com o tempo de execução de 2 horas, é do tipo *Gradient Boosting Machines* (GBM), o qual pode ser identificado a partir do index 4 na Tabela 4.

Tabela 3 – Mensuração do poder preditivo dos modelos gerados para o cenário I com o tempo de execução do AutoML configurado em 10 minutos

Cenário I – 10 minutos						
Index	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
7	GBM	0.704	0.468	0.860	0.922	0.486
2	Stacked Ensemble	0.703	0.473	0.862	0.931	0.475
1	Stacked Ensemble	0.703	0.473	0.862	0.931	0.475
5	XGBoost	0.700	0.463	0.860	0.926	0.473
8	GBM	0.699	0.460	0.859	0.925	0.473
3	GBM	0.697	0.465	0.861	0.932	0.463
4	DRF	0.696	0.462	0.860	0.932	0.459
9	GBM	0.684	0.447	0.859	0.939	0.428
6	GBM	0.683	0.447	0.859	0.940	0.427
10	XGBoost	0.679	0.438	0.857	0.939	0.419

Fonte: Elaborado pela Autora

Tabela 4 – Mensuração do poder preditivo dos modelos gerados para o cenário I com o tempo de execução do AutoML configurado em 2 horas

Cenário I – 2 horas						
Index	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
4	GBM	0.710	0.477	0.861	0.922	0.497
3	GBM	0.709	0.478	0.862	0.923	0.495
6	GBM	0.708	0.477	0.862	0.925	0.490
1	Stacked Ensemble	0.707	0.476	0.862	0.925	0.489
10	GBM	0.707	0.472	0.860	0.921	0.403
2	Stacked Ensemble	0.704	0.473	0.862	0.928	0.480
7	GBM	0.704	0.472	0.861	0.928	0.480
5	XGBoost	0.701	0.468	0.861	0.929	0.473
8	DRF	0.600	0.469	0.862	0.934	0.464
9	XGBoost	0.698	0.463	0.860	0.929	0.468

Fonte: Elaborado pela Autora

4.1.2 Cenário II

Os mesmos modelos que foram aplicados no cenário I foram replicados no cenário II, porém no cenário II, os modelos são aplicados nas bases testes das bases individualizadas, ou seja, nesse cenário busca-se analisar qual é o poder preditivo de um modelo desenvolvido a partir da base compartilhada, que se beneficia da troca de informações entre as instituições, quando aplicado a uma situação local. Como explicado anteriormente, foram aplicados 20 modelos, os 10 melhores modelos gerados com o tempo de execução igual a 10 minutos e os 10 melhores modelos gerados com o tempo de execução do AutoML configurados em 2 horas, segundo a priorização resultante no *leaderboard*. Os 20 modelos foram testados nos 6 hospitais da base, gerando 120 predições diferentes. As métricas resultantes de todas essas predições podem ser observadas no Apêndice C e na Tabela 5 estão dispostos os melhores modelos encontrados para cada um dos hospitais e as métricas para avaliá-los em termos de poder de predição.

Nota-se que o modelo do tipo *Gradient Boosting Machines* (GBM), desenvolvido com o tempo máximo de execução configurado em 2 horas e que possui o index igual a 3, possui o melhor desempenho quando aplicado às bases dos hospitais ‘h1’, ‘h2’ e ‘h4’.

Tabela 5 – Modelos com melhor poder preditivo do cenário II para cada um dos hospitais e suas métricas de avaliação.

Melhores Modelos do Cenário II								
	RunTime	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
h1	2 hrs	3	GBM	0.729	0.560	0.918	0.969	0.489
h2	2 hrs	3	GBM	0.842	0.714	0.920	0.946	0.738
h3	2 hrs	3	GBM	0.698	0.479	0.870	0.946	0.450
h4	2 hrs	3	GBM	0.805	0.635	0.861	0.914	0.695
h5	2 hrs	2	Stacked Ensemble	0.732	0.569	0.890	0.971	0.492
h6	2 hrs	1	Stacked Ensemble	0.937	0.827	0.933	0.874	1.000

Fonte: Elaborado pela Autora

4.1.3 Cenário III

A mesma lógica aplicada aos cenários I e II foi replicada no cenário III. O cenário III simula uma situação em que não há compartilhamento de informação entre as instituições. Dessa forma os modelos são treinados, validados e testados em âmbito local, ou seja, são

treinados, validados e testados apenas com os dados do hospital em questão. Seguindo a metodologia, no cenário III foram aplicados 120 modelos de predição diferentes, 20 para cada um dos 6 hospitais da base. As métricas calculadas a partir das predições resultantes de cada um dos modelos podem ser observadas no Apêndice C e na Tabela 6 estão dispostos os melhores modelos encontrados um para cada um dos hospitais, ressaltando que nesse cenário os modelos foram desenvolvidos especificamente para apenas um hospital, utilizando apenas dados locais.

Tabela 6 – Modelos com melhor poder preditivo do cenário III para cada um dos hospitais e suas métricas de avaliação.

Melhores Modelos do Cenário III								
	RunTime	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
h1	10 min	6	GBM	0.723	0.520	0.909	0.950	0.496
h2	2 hrs	9	GBM	0.790	0.585	0.879	0.884	0.697
h3	10 min	1	Stacked Ensemble	0.669	0.414	0.857	0.935	0.401
h4	10 min	7	GBM	0.766	0.573	0.842	0.921	0.611
h5	2 hrs	10	DRF	0.742	0.516	0.869	0.903	0.580
h6	2 hrs	1	XGBoost	0.937	0.827	0.933	0.874	1.000

Fonte: Elaborado pela Autora

4.1.4 Cenário IV

O cenário IV é muito similar ao cenário II. Nele é simulada uma situação em que as instituições de saúde utilizam dados externos, de outras instituições, para treinar um modelo preditivo de epidemiologia. Porém nesse cenário a validação do modelo, que faz parte da função da ferramenta H_2O AutoML, é feita com as bases individuais. Em resumo, os modelos são treinados com a base compartilhada, mas são validados e testados com as bases individuais. Dessa forma, para o cenário IV foram aplicados 120 modelos de predição diferentes, 20 para cada um dos 6 hospitais da base. As métricas calculadas a partir das predições resultantes de cada um dos modelos podem ser observadas no Apêndice C e na Tabela 7 estão dispostos os melhores modelos encontrados um para cada um dos hospitais, ressaltando que nesse cenário os modelos foram desenvolvidos especificamente para apenas um hospital, utilizando de dados internos e externos.

Tabela 7 – Modelos com melhor poder preditivo do cenário IV para cada um dos hospitais e suas métricas de avaliação

Melhores Modelos do Cenário IV								
	RunTime	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
h1	2 hrs	8	DRF	0.770	0.583	0.916	0.939	0.601
h2	2 hrs	7	GBM	0.852	0.713	0.817	0.925	0.779
h3	2 hrs	4	DRF	0.717	0.494	0.870	0.928	0.506
h4	2 hrs	3	GBM	0.796	0.624	0.858	0.923	0.668
h5	10 min	3	DRF	0.776	0.576	0.882	0.906	0.646
h6	2 hrs	3	GMB	0.927	0.805	0.921	0.854	1.000

Fonte: Elaborado pela Autora

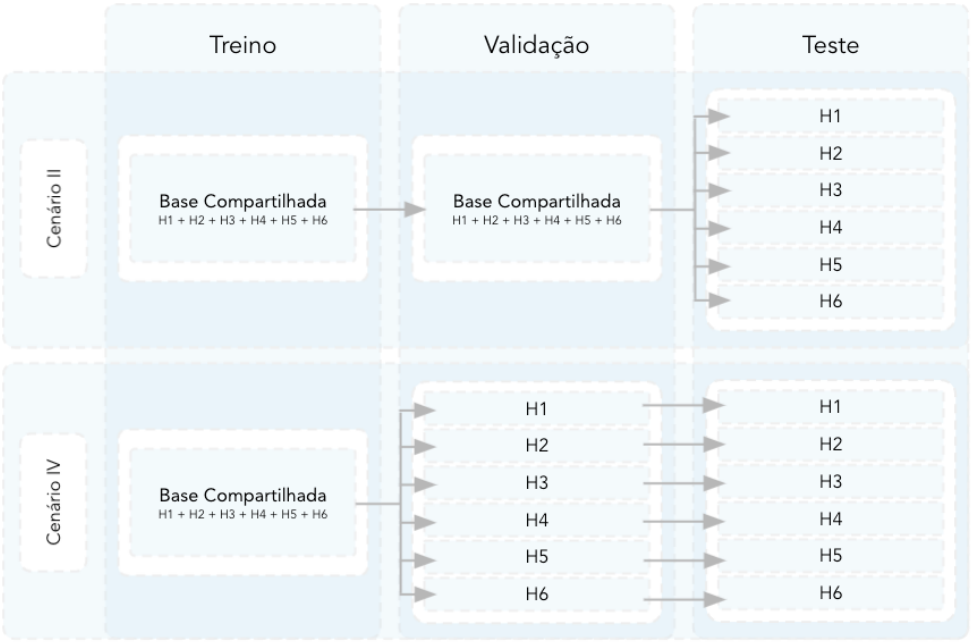
4.2 COMPARAÇÃO DOS CENÁRIOS

A partir do teste dos modelos desenvolvidos a partir da ferramenta de AutoML foi possível definir o melhor modelo para cada um dos cenários e, quando necessário, para cada um dos hospitais. A análise e comparação do poder preditivo dos modelos dos diferentes cenários será feita com base no melhor modelo encontrado a partir da realização dos testes que foram executados. O trabalho foi desenvolvido dessa forma, pois o foco dele não é identificar qual o melhor modelo matemático e estatístico para realizar a previsão dos resultados de um antibiograma, mas sim para analisar qual o impacto do compartilhamento de informações no poder de previsão dos modelos.

4.2.1 Comparação entre cenários com compartilhamento de dados

Os cenários II e IV simulam situações em que as instituições de saúde utilizam de dados externos para a construção de modelos de previsão. Sendo que no cenário II os modelos são treinados e validados com bases derivadas da base compartilhada e no cenário IV os modelos são treinados com bases derivadas da base compartilhada, mas são validados com bases derivadas das bases individualizadas, como mostrado na Figura 18.

Figura 18 – Metodologia de desenvolvimento de modelos estatísticos nos cenários com compartilhamento de informações



Fonte: Elaborado pela Autora

Tabela 8 – Modelos com melhor poder preditivo do cenário IV para cada um dos hospitais e suas métricas de avaliação.

Hospital	Cenário	AUC	MCC	F1	Sensibilidade	Especificidade
h1	II	0.729	0.560	0.918	0.969	0.489
	IV	0.770	0.583	0.916	0.939	0.601
h2	II	0.842	0.714	0.920	0.946	0.738
	IV	0.852	0.713	0.817	0.925	0.779
h3	II	0.698	0.479	0.870	0.946	0.450
	IV	0.717	0.494	0.870	0.928	0.506
h4	II	0.805	0.635	0.861	0.914	0.695
	IV	0.796	0.624	0.858	0.923	0.668
h5	II	0.732	0.569	0.890	0.971	0.492
	IV	0.776	0.576	0.882	0.906	0.646
h6	II	0.937	0.827	0.933	0.874	1.000
	IV	0.927	0.805	0.921	0.854	1.000

Fonte: Elaborado pela Autora

Quando comparados os dois cenários percebe-se que, para todos os hospitais, os melhores modelos do cenário II e do cenário IV possuem um poder preditivo semelhante. No geral pode-se dizer que os modelos gerados no cenário II são mais indicados, pois necessitam de um menor esforço computacional para serem desenvolvidos, já que no cenário II foram testados 20 modelos enquanto no cenário IV foram aplicados 120 modelos para a obtenção

desses resultados. Além disso, no cenário II não seria necessário fazer a validação com dados individuais, só seria utilizada a base compartilhada, ou seja, a base disponível no BR-Glass, o que facilita o acesso aos dados. Ainda assim, será utilizado o modelo com melhor desempenho dentre eles para a comparação com os demais cenários, já que, os dois representam técnicas executáveis para a construção de modelos estatísticos.

4.2.2 Comparação entre situações com e sem o compartilhamento de dados

O objetivo geral deste trabalho consiste em analisar como compartilhamento de informações de saúde impacta o poder preditivo de modelos de aprendizagem de máquina nesse setor. Dessa forma, dois cenários principais de comparação foram estipulados: o cenário III, chamado de 'cenário individualizado' na Tabela 9, nos quais os modelos são desenvolvidos apenas com dados locais e o 'cenário compartilhado', que pode ser um modelo do cenário II ou do cenário IV, que é o exemplo de uma situação em que os modelos são desenvolvidos a partir do compartilhamento de informações. Para que o poder preditivo dos modelos desses cenários fosse interpretado de maneira mais fácil também foi desenvolvido o cenário I, chamado de cenário do 'Ministério da Saúde' na Tabela 9, esse cenário representa uma situação em que o ministério da saúde utiliza dos dados do BR-Glass para treinar, validar e fazer predições para diferentes instituições de saúde, o desempenho do melhor modelo pertencente ao cenário do 'Ministério da Saúde' irá servir como um parâmetro para análise dos demais modelos.

Tabela 9 – Comparação de modelos construídos apenas com dados locais e modelos construídos com dados compartilhados

Hospital	Cenário	AUC	MCC	F1	Sensibilidade	Especificidade
Ministério da Saúde		0.710	0.447	0.861	0.922	0.497
h1	Individualizado	0.723	0.520	0.909	0.950	0.496
	Compartilhado	0.770	0.583	0.916	0.939	0.601
h2	Individualizado	0.790	0.585	0.879	0.884	0.697
	Compartilhado	0.852	0.713	0.817	0.925	0.779
h3	Individualizado	0.669	0.414	0.857	0.935	0.401
	Compartilhado	0.717	0.494	0.870	0.928	0.506
h4	Individualizado	0.766	0.573	0.842	0.921	0.611
	Compartilhado	0.805	0.635	0.861	0.914	0.695
h5	Individualizado	0.742	0.516	0.869	0.903	0.580
	Compartilhado	0.776	0.576	0.882	0.906	0.646

h6	Individualizado	0.937	0.827	0.933	0.874	1.000
	Compartilhado	0.937	0.827	0.933	0.874	1.000

Fonte: Elaborado pela Autora

A partir das métricas expostas na Tabela 9 pode-se dizer que o compartilhamento de informações epidemiológicas entre instituições de saúde é benéfico para a melhora do poder preditivo dos modelos estatísticos. Para todos os hospitais da base, o cenário que utilizava do compartilhamento de dados realizou previsões melhores ou idênticas aos cenários que não utilizavam de dados externos à instituição, mesmo com a diferença entre o tamanho e a variedade das amostras fornecidas por cada uma das instituições. Vale ressaltar que a única instituição em que o modelo do cenário compartilhado se igualou ao modelo do cenário individualizado foi para o hospital 'h6' que possui o menor número de registros da base, representando 995 dos 227.425 registros e que possui um perfil único em comparação com as demais instituições de saúde.

Tabela 10 – Modelos com melhor poder preditivo do cenário IV para cada um dos hospitais e suas métricas de avaliação

Hospital	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
h1	+ 6.50%	+ 12.12%	+ 0.77%	- 1.16%	+ 21.17%
h2	+ 7.85%	+ 21.88%	+ 4.32%	+ 4.64%	+ 11.76%
h3	+ 7.34%	+ 19.32%	+ 1.52%	- 0.75%	+ 25.87%
h4	+ 5.09%	+ 10.82%	+ 2.26%	+ 0.22%	+ 9.33%
h5	+ 4.58%	+ 11.63%	+ 1.50%	+ 0.33%	+ 11.38%
h6	-	-	-	-	-

Fonte: Elaborado pela Autora

Na tabela 10 foi calculado a diferença do desempenho entre os modelos individualizados e compartilhados para cada um dos hospitais e cada um dos indicadores. Percebe-se que as instituições de saúde, mesmo que com tamanhos de bases diferentes, obtiveram um ganho proporcional semelhante de AUC quando aplicados os modelos que utilizavam do compartilhamento de dados, com exceção do hospital 'h6' que tem um perfil diferente dos demais, como foi citado anteriormente. Além disso, pode-se notar que quando há um ganho grande em relação à especificidade há uma perda em relação a sensibilidade do modelo.

Apesar de haver uma diferença entre o poder preditivo do cenário individualizado e do cenário compartilhado é importante salientar que a diferença não é muito representativa, e que seria necessário um maior número de hospitais e de registros, além de que, a aplicação de

algumas técnicas estatísticas para inferir que o compartilhamento de informações na saúde é interessante para a melhoria do poder de predição de modelos estatísticos em todos os casos. Mas, para esse caso, pode-se dizer que o compartilhamento de informações é benéfico e pode ser mais bem explorado pelas organizações.

5 CONCLUSÃO E RECOMENDAÇÕES

Neste capítulo são descritas as considerações finais deste trabalho, em que são revisados os resultados atingidos ao longo do desenvolvimento da pesquisa, além das limitações e sugestões para futuros estudos sobre o assunto em questão.

Este trabalho tem como objetivo analisar como compartilhamento de informações de saúde impacta o poder preditivo de modelos de aprendizagem de máquina, comparando o poder preditivo de diferentes modelos de aprendizagem de máquina treinados com dados de epidemiologia hospitalar nacionais e dados de epidemiologia hospitalar individuais. Para tanto, foi feita uma limpeza na base de dados seguindo orientações de médicos e especialistas da área de epidemiologia. Em seguida, foram estabelecidos diferentes cenários de análise, sendo um deles uma base comparativa (Cenário I), dois deles que se beneficiam de dados compartilhados para de construção de modelos (Cenário II e IV) e um último cenário em que os modelos são desenvolvidos exclusivamente com dados locais (Cenário III). Para o treinamento, validação e teste dos modelos foi utilizada a ferramenta H_20 AutoML, já que o objetivo do presente trabalho não era definir o melhor modelo de predição de resultados dos antibiogramas, mas sim comparar o impacto que o compartilhamento de dados tem sobre o poder de predição dos modelos. Através da utilização dessa ferramenta foram desenvolvidos diferentes modelos, e aquele que apresentou o maior valor de AUC quando aplicado à base teste foi selecionado como o melhor modelo para um determinado hospital para cada um dos cenários pré-definidos.

Dessa forma, cada um dos seis hospitais analisados possuía um modelo de referência para o cenário II, cenário III e cenário IV. Com isso, inicialmente foram comparados os cenários II e IV para definir qual das duas metodologias de desenvolvimento de modelos utilizando do compartilhamento de dados resultava em um melhor poder de predição de acordo com o valor de AUC quando aplicado à base teste. Assim, foi possível comparar o poder preditivo dos modelos desenvolvidos no cenário III, que representa um cenário com

dados individualizados, com o poder preditivo dos modelos desenvolvidos em um cenário com dados compartilhados, representados pelo melhor modelo gerado nos cenários II e IV.

Ao fim da aplicação da metodologia proposta, foi atingido o objetivo principal deste trabalho, que é a comparação o poder preditivo de diferentes modelos de aprendizagem de máquina treinados com dados de epidemiologia hospitalar nacionais e dados de epidemiologia hospitalar individuais. Através dessa comparação, pode-se concluir que modelos que se beneficiam do compartilhamento de dados, na situação estudada, possuem um poder de preditivo melhor do que aqueles que não se beneficiam.

Ainda que o objetivo desta pesquisa foi atingido, observou-se algumas limitações ao longo de seu desenvolvimento. Dessa forma, recomenda-se para trabalhos futuros uma análise que contemplem um maior número de hospitais, pacientes e registros, já que este trabalho possui a limitação de analisar apenas hospitais da região sul do país. Além disso, neste trabalho foi respondida uma questão do tipo exploratória, como definido no capítulo três, também conhecidas como questões geradoras de hipótese, mas é possível fazer um estudo de inferência para entender se a realidade evidenciada nesse trabalho pode ou não ser generalizada, já que nesse trabalho não foi feito devido às limitações práticas e computacionais.

Por meio dos resultados alcançados e dos procedimentos adotados, pode-se considerar que o estudo agregou contribuições ao tema abordado, enriquecendo a literatura existente e fomentando o aprofundamento da pesquisa no tema. Além disso, o estudo sugere formas em que as organizações e o governo possam, em conjunto, entregar um maior valor ao paciente, assunto relevante no setor estudado.

Por fim, é importante salientar a relevância do Trabalho de Conclusão de Curso no processo de formação do estudante, pois possibilita a aplicação dos conhecimentos adquiridos ao longo da graduação e garante a capacidade de análise e resolução de problemas dos futuros profissionais.

REFERÊNCIAS

ADVANI, V. What is Machine Learning? How Machine Learning Works and future of it?. **Medium**. Disponível em: https://www.mygreatlearning.com. Acesso em: 10 jun. 2021.

BALAJI, A.; ALLEN, A. Benchmarking automatic machine learning frameworks. **Cornell University**, v. 1808.06492, 2018. Disponível em: https://arxiv.org/abs/1808.06492. Acesso em: 19 ago. 2021.

BURKOV, A. **The hundred-page machine learning book**. Andriy Burkov, 2019. Disponível em: https://leanpub.com/theMLbook. Acesso em: 19 ago. 2021.
COX, V. Exploratory data analysis. **Translating Statistics to Make Decisions**. Apress, Berkeley, CA, 2017. p. 47-74.

DAVIES, M. Emerging Smarter: Rethink Healthcare on a time of COVID-19. **IBM Corporation**. Disponível em: https://www.ibm.com/uk-en/marketing/emergingsmarter-rethinkinghealthcare/. Acesso em: 01 maio 2021.

DE MAESENEER, J. European expert panel on effective ways of investing in health: opinion on primary care. **Primary health care research & development**, v. 16, n. 2, p. 109-110, 2015.

DE OLIVEIRA, K. R.; MUNARETTO, P. Uso racional de antibióticos: responsabilidade de prescritores, usuários e dispensadores. **Revista Contexto & Saúde**, v. 10, n. 18, p. 43-51, 2010.

EL NAQA, I.; MURPHY, M. J. What is machine learning?. M**achine learning in radiation oncology**. Springer, Cham, 2015. p. 3-11.

GRAY, J. A. M. **Evidence-based healthcare and public health: how to make decisions about health services and public health**. Churchill Livingstone, 2009.

GURGEL, T. C.; CARVALHO, W. S. A assistência farmacêutica e o aumento da resistência bacteriana aos antimicrobianos. **Lat. Am. J. Pharm**, v. 27, n. 1, p. 118-23, 2008.

KAPLAN, R. S.; PORTER, M. E. The Big Idea: How to Solve the Cost Crisis in Health Care. 2011. **Harvard Business Review**, 2018.

KUMUNI. Métricas de Avaliação em Machine Learning: Classificação. **Medium**. Disponível em: https://medium.com/kunumi/métricas-de-avaliação-em-machine-learning-classificação-49340dcdb198. Acesso em: 10 jun. 2021.

LEDELL, E.; POIRIER, S. H2o automl: Scalable automatic machine learning. **Proceedings of the AutoML Workshop at ICML**. 2020.

LEEK, J. T.; PENG, R. D. What is the question?. **Science**, v. 347, n. 6228, p. 1314-1315, 2015.

MAIER, C. R.; ABEGG, M. A. Avaliação da utilização de antibióticos por profissionais de saúde e pela população na cidade de Toledo, Paraná, Brasil. **Arquivos de Ciências da Saúde da UNIPAR**, v. 11, n. 1, 2007.

MULLER, P. de S. G. *et al*. Regulamentação para a venda de antibióticos no brasil e sua aceitação pela população. **Acta Biomédica Brasiliensia**, v. 6, n. 1, p. 91-100, 2015.
PEDREGOSA, F. *et al*. Scikit-learn: Machine learning in Python. **The Journal of machine Learning Research**, v. 12, p. 2825-2830, 2011.

PENG, R. D.; MATSUI, E. The art of data science. **A Guide for Anyone Who Works with Data. Skybrude Consulting, LLC**, 2015.

PILLONETTO, M. et al. SMART-CDSS An artificial intelligence system for antimicrobial prescription support. **Gates Open Res**, v. 4, 2020.

PORTER, M. E. Value-based health care delivery. **Annals of surgery**, v. 248, n. 4, p. 503-509, 2008.

PORTER, Michael E.; TEISBERG, E. O. **Redefining health care: creating value-based competition on results**. Harvard business press, 2006.

SILVA, E. U. A importância do controle da prescrição de antimicrobianos em hospitais para melhoria da qualidade, redução de custos e controle da resistência bacteriana. **Prática Hospitalar**, v. 10, n. 57, p. 101-6, 2008.

TOLEDO, P. V. M. *et al*. Surveillance programme for multidrug-resistant bacteria in healthcare-associated infections: an urban perspective in South Brazil. **Journal of Hospital Infection**, v. 80, n. 4, p. 351-353, 2012.

WANNMACHER, L. Uso indiscriminado de antibióticos e resistência microbiana: uma guerra perdida. **Uso racional de medicamentos: temas selecionados**, v. 1, n. 4, p. 1-6, 2004.
WARING, J.; LINDVALL, C.; UMETON, R. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. **Artificial Intelligence in Medicine**, v. 104, p. 101822, 2020.

YUFENG, G. What is Machine Learning?. **Medium**. Disponível em: https://towardsdatascience.com/what-is-machine-learning-8c6871016736. Acesso em: 10 jun. 2021.

APÊNDICE A – Limpeza da Base BR-Glass

Neste apêndice estão dispostos os materiais de apoio que foram desenvolvidos por profissionais da área da saúde e de epidemiologia para aproximar os dados dispostos na base BR-Glass da prática clínica e, dessa forma, fazer com que os modelos desenvolvidos tenham uma melhor previsibilidade para os testes que realmente são executados e para os antibióticos que realmente são prescritos. Nos Quadros 1 e 2 estão listadas as correções feitas em relação à nomenclatura dos microrganismos e antibióticos, além disso foram excluídos da base os antibióticos que não são utilizados na prática clínica.

Em sequência foram feitas alterações mais específicas na base, onde variáveis combinadas são analisadas e, assim, se determina se a combinação é válida ou não na prática clínica. Nessas análises foram consideradas variáveis tais como antimicrobiano, grupo gram, família, gênero e espécie do microrganismo, e também o material de coleta a ser analisado. Nas planilhas o número 0 vai evidenciar combinações que não estão de acordo com a prática clínica e, por isso, foram retiradas da base para a execução das análises e desenvolvimento dos modelos. Especificamente na Planilha 6, em alguns cruzamentos é identificado o número 3, estes casos devem ser excluídos apenas quando o atendimento for do tipo ambulatorial.

Quadro 1 – Correção da nomenclatura dos antibióticos e exclusão dos antibióticos que não são utilizados na prática clínica

Antibiótico	**Correção**	**Excluir?**
Ácido Fusidico		Excluir
Ácido Nalidíxico		Excluir
Amoxicilina-Clavulanato	Amoxicilina-ácido clavulânico	
Sulbactam-Ampicilina	Ampicilina-sulbactam	
Clinafloxacino		Excluir
Estreptomicina		Excluir
Gatifloxacina		Excluir
Gentamicina (de alta concentração)		Excluir
Tobramicina		Excluir
Sulfazotrim	Trimetoprima-Sulfametoxazol	

Fonte: Elaborado pela Autora

Quadro 2 – Correção e normalização da nomenclatura microrganismos.

Microrganismo	**Correção**
Acinetobacter baumannii	Acinetobacter baumannii Complexo
Aeromonas caviae	Aeromonas hydrophila Complexo
Aeromonas hydrophila	Aeromonas hydrophila Complexo
Aeromonas hydrophila/caviae	Aeromonas hydrophila Complexo
Burkholderia cepacia	Burkholderia cepacia Complexo
Burkholderia spp.	Burkholderia cepacia Complexo
Burkholderia vietnamiensis	Burkholderia cepacia Complexo
Citrobacter diversus	Citrobacter koseri
Citrobacter rodentium	Citrobacter spp.
Enterobacter aerogenes	Klebsiella aerogenes
Enterobacter aglomerans	Pantoea agglomerans
Enterobacter asburiae	Enterobacter cloacae Complexo
Enterobacter cloacae	Enterobacter cloacae Complexo
Escherichia coli (produtora de ESBL)	Escherichia coli
Haemophilus Influenzae não capsulado	Haemophilus influenzae
Klebsiella oxytoca (produtora de ESBL)	Klebsiella oxytoca
Klebsiella ozaenae	Klebsiella pneumoniae
Klebsiella pneumoniae (produtora de ESBL)	Klebsiella pneumoniae
Klebsiella pneumoniae ssp ozaenae	Klebsiella pneumoniae
Klebsiella pneumoniae ssp pneumoniae	Klebsiella pneumoniae
Salmonella Groupo	Salmonella spp.
Salmonella Thompson	Salmonella spp.
Serratia fonticola	Serratia spp.
Serratia odorifera	Serratia spp.
Shigella sonnei	Shigella spp.
Staphylococcus aureus (MRSA)	Staphylococcus aureus
Staphylococcus aureus ssp aureus	Staphylococcus aureus
Staphylococcus auricularis	Staphylococcus coagulase-negativa
Staphylococcus capitis	Staphylococcus coagulase-negativa

Staphylococcus caprae	Staphylococcus coagulase-negativa
Staphylococcus cohnii	Staphylococcus coagulase-negativa
Staphylococcus cohnii ssp cohnii	Staphylococcus coagulase-negativa
Staphylococcus cohnii ssp urealyticus	Staphylococcus coagulase-negativa
Staphylococcus equorum	Staphylococcus coagulase-negativa
Staphylococcus haemolyticus	Staphylococcus coagulase-negativa
Staphylococcus hominis ssp hominis	Staphylococcus coagulase-negativa
Staphylococcus hyicus	Staphylococcus coagulase-negativa
Staphylococcus intermedius	Staphylococcus coagulase-negativa
Staphylococcus lentus	Staphylococcus coagulase-negativa
Staphylococcus lugdunensis	Staphylococcus coagulase-negativa
Staphylococcus pseudintermedius	Staphylococcus coagulase-negativa
Staphylococcus schleiferi	Staphylococcus coagulase-negativa
Staphylococcus sciuri	Staphylococcus coagulase-negativa
Staphylococcus simulans	Staphylococcus coagulase-negativa
Staphylococcus spp.	Staphylococcus coagulase-negativa
Staphylococcus vitulinus	Staphylococcus coagulase-negativa
Staphylococcus warneri	Staphylococcus coagulase-negativa
Staphylococcus xylosus	Staphylococcus coagulase-negativa
Streptococcus anginosus	Streptococcus viridans Grupo
Streptococcus anginosus Grupo	Streptococcus viridans Grupo
Streptococcus bovis Grupo	Streptococcus viridans Grupo
Streptococcus constellatus	Streptococcus viridans Grupo
Streptococcus gallolyticus ssp gallolyticus	Streptococcus viridans Grupo
Streptococcus gordonii	Streptococcus viridans Grupo
Streptococcus intermedius	Streptococcus viridans Grupo
Streptococcus mitis	Streptococcus viridans Grupo
Streptococcus mitis/oralis	Streptococcus viridans Grupo
Streptococcus oralis	Streptococcus viridans Grupo
Streptococcus parasanguinis	Streptococcus viridans Grupo
Streptococcus salivarius	Streptococcus viridans Grupo

Streptococcus sanguinis	Streptococcus viridans Grupo
Streptococcus sanguis	Streptococcus viridans Grupo
Streptococcus spp.	Streptococcus viridans Grupo
Streptococcus suis II	Streptococcus suis
Streptococcus uberis	Streptococcus viridans Grupo
Streptococcus vestibularis	Streptococcus viridans Grupo
Achromobacter spp.	Achromobacter spp.

Fonte: Elaborado pela Autora.

Quadro 3 – Identificação das combinações entre antimicrobianos e grupo e família que não são utilizadas na prática clínica.

Antimicrobianos	**Grupo**		**Família**
	Gram Negativo	**Gram Positivo**	**Enterobacteriaceae**
Amicacina	1	1	1
Amoxicilina-ácido clavulânico	1	1	1
Ampicilina	1	1	1
Ampicilina-sulbactam	1	1	1
Aztreonam	1	0	1
Cefalotina	1	1	1
Cefazolina	1	1	1
Cefepima	1	1	1
Cefotaxima	1	1	1
Cefoxitina	1	1	1
Ceftarolina	1	1	1
Ceftazidima	1	1	1
Ceftriaxona	1	1	1
Cefuroxima	1	1	1
Ciprofloxacino	1	1	1
Clindamicina	0	1	0
Cloranfenicol	0	1	1
Colistina (Polimixina E)	1	1	1
Daptomicina	0	1	0

Doxiciclina	1	1	1
Eritromicina	0	1	0
Ertapenem	1	1	1
Fosfomicina	1	1	1
Gentamicina	1	1	1
Imipenem	1	1	1
Levofloxacino	1	1	1
Linezolida	0	1	0
Meropenem	1	1	1
Nitrofurantoína	1	1	1
Norfloxacino	1	1	1
Oxacilina	0	1	0
Penicilina	0	1	0
Piperacilina-tazobactam	1	1	1
Polimixina B	1	0	1
Rifampicina	0	1	0
Teicoplanina	0	1	0
Tetraciclina	1	1	0
Tigeciclina	1	1	1
Trimetoprima-Sulfametoxazol	1	1	1
Vancomicina	0	1	0
Amoxicilina	1	1	1
Azitromicina	0	1	0
Cefaclor	1	1	1
Cefadroxila	1	1	1
Cefalexina	1	1	1
Claritromicina	0	1	0
Penicilina G Benzatina	0	1	0
Penicilina G Cristalina	0	1	0
Penicilina G Procaína	0	1	0
Penicilina V - fenoximetílico	0	1	0

Fonte: Elaborado pela Autora.

Quadro 4 – Identificação das combinações entre antimicrobianos e gênero que não são utilizadas na prática clínica.

Antimicrobianos	**Gênero**						
	Salmonella	**Pseudomonas**	**Acinetobacter**	**Staphylococcus**	**Enterococcus**	**Burkholderia**	**Stenotrophomonas**
Amicacina	0	1	1	1	0	0	0
Amoxicilina-ácido clavulânico	1	0	0	1	1	0	0
Ampicilina	1	0	0	1	1	0	0
Ampicilina-sulbactam	1	0	1	1	1	0	0
Aztreonam	1	1	0	0	0	0	0
Cefalotina	0	0	0	1	0	0	0
Cefazolina	0	0	0	1	0	0	0
Cefepima	1	1	1	1	0	0	1
Cefotaxima	0	0	0	1	0	0	0
Cefoxitina	1	0	0	1	0	0	0
Ceftarolina	1	1	0	1	0	0	0
Ceftazidima	1	1	0	1	0	1	1
Ceftriaxona	0	0	0	1	0	0	0
Cefuroxima	1	0	0	1	0	0	0
Ciprofloxacino	1	1	1	1	1	0	1
Clindamicina	0	0	0	1	0	0	0
Cloranfenicol	1	0	0	1	0	0	0
Colistina (Polimixina E)	1	1	1	0	0	0	1
Daptomicina	0	0	0	1	1	0	0
Doxiciclina	1	0	1	1	1	0	1
Eritromicina	0	0	0	1	1	0	0
Ertapenem	1	0	0	1	1	0	0

Fosfomicina	1	0	0	1	1	0	0
Gentamicina	0	1	1	1	0	0	0
Imipenem	1	1	1	1	1	0	0
Levofloxacino	1	1	1	1	1	1	0
Linezolida	0	0	0	1	1	0	0
Meropenem	1	1	1	1	0	1	0
Nitrofurantoína	1	0	0	1	1	0	0
Norfloxacino	1	0	0	1	1	0	0
Oxacilina	0	0	0	1	0	0	0
Penicilina	0	0	0	1	1	0	0
Piperacilina-tazobactam	1	1	1	1	1	0	0
Polimixina B	1	1	1	0	0	0	1
Rifampicina	0	0	0	1	0	0	0
Teicoplanina	0	0	0	1	1	0	0
Tetraciclina	0	0	0	1	0	0	0
Tigeciclina	1	0	1	1	1	1	1
Trimetoprima-Sulfametoxazol	1	0	1	1	0	1	1
Vancomicina	0	0	0	1	1	0	0
Amoxicilina	1	0	0	1	1	0	0
Azitromicina	0	0	0	1	1	0	0
Cefaclor	0	0	0	1	0	0	0
Cefadroxila	0	0	0	1	0	0	0
Cefalexina	0	0	0	1	0	0	0
Claritromicina	0	0	0	1	1	0	0
Penicilina G Benzatina	0	0	0	1	1	0	0
Penicilina G Cristalina	0	0	0	1	1	0	0
Penicilina G Procaína	0	0	0	1	1	0	0
Penicilina V - fenoximetílico	0	0	0	1	1	0	0

Fonte: Elaborado pela Autora.

Quadro 5 – Identificação das combinações entre antimicrobianos e espécie que não são utilizadas na prática clínica.

Antimicrobianos	Espécie			
	Streptococcus pyogenes	Streptococcus agalactiae	Streptococcus viridans	Streptococcus pneumoniae
Amicacina	0	0	0	0
Amoxicilina-ácido clavulânico	1	1	1	1
Ampicilina	1	1	1	1
Ampicilina-sulbactam	1	1	1	1
Aztreonam	0	0	0	0
Cefalotina	1	1	1	0
Cefazolina	1	1	1	0
Cefepima	1	1	1	1
Cefotaxima	1	1	1	1
Cefoxitina	0	0	0	0
Ceftarolina	1	1	0	0
Ceftazidima	0	0	0	0
Ceftriaxona	1	1	1	1
Cefuroxima	1	1	1	0
Ciprofloxacino	0	0	0	0
Clindamicina	1	1	1	1
Cloranfenicol	1	1	1	1
Colistina (Polimixina E)	0	0	0	0
Daptomicina	1	1	1	1
Doxiciclina	1	1	1	1
Eritromicina	1	1	1	1
Ertapenem	1	1	1	1
Fosfomicina	0	0	0	0

Gentamicina	0	0	0	0
Imipenem	1	1	0	1
Levofloxacino	1	1	1	1
Linezolida	1	1	1	1
Meropenem	1	1	1	1
Nitrofurantoína	1	1	0	0
Norfloxacino	1	1	0	0
Oxacilina	0	0	0	0
Penicilina	1	1	1	1
Piperacilina-tazobactam	1	1	1	1
Polimixina B	0	0	0	0
Rifampicina	1	1	1	1
Teicoplanina	1	1	1	1
Tetraciclina	0	0	0	1
Tigeciclina	1	1	1	1
Trimetoprima-Sulfametoxazol	0	0	0	0
Vancomicina	1	1	1	1
Amoxicilina	1	1	1	1
Azitromicina	1	1	1	1
Cefaclor	1	1	1	0
Cefadroxila	1	1	1	0
Cefalexina	1	1	1	0
Claritromicina	1	1	1	1
Penicilina G Benzatina	1	1	1	1
Penicilina G Cristalina	1	1	1	1
Penicilina G Procaína	1	1	1	1
Penicilina V - fenoximetílico	1	1	1	1

Fonte: Elaborado pela Autora.

Quadro 6 – Identificação das combinações entre antimicrobianos e materiais coletados para análise que não são utilizadas na prática clínica.

	Materiais de coleta																											
Antimicrobianos	**EENT**	**Secreção ocular**	**Swab de ouvido**	**Ferida cirúrgica**	**Lesão**	**Raspado de Ulceras**	**Secreção**	**Secreção de ferida**	**Secreção de Lesão**	**Swab**	**Swab de Lesão de pele**	**Biopsia**	**Fragmentos de tecido ósseo**	**Peça cirurgica**	**Tecido**	**Ponta de cateter**	**Sangue**	**Liquor**	**Fezes**	**Aspirado brônquico**	**Aspirado Traqueal**	**Escarro**	**Lavado bronco alveolar**	**Lavado bronquico**	**Secreção traqueal**	**Secreção orofaringea**	**Swab de orofaringe**	**Urina**	**Urina/ Pielonefrite**
Amicacina	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Amoxicilina-ácido clavulânico	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1
Ampicilina	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ampicilina-sulbactam	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0
Aztreonam	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	3	3
Cefalotina	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1
Cefazolina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
Cefepima	0	0	0	3	0	3	3	3	3	3	3	3	3	3	3	0	0	0	3	3	3	3	3	3	3	3	3	3
Cefotaxima	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
Cefoxitina	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
Ceftarolina	0	0	0	3	0	3	3	3	3	3	3	3	3	3	3	0	0	0	0	3	3	3	3	3	3	3	0	0
Ceftazidima	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
Ceftriaxona	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cefuroxima	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1

Ciprofloxacino	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1
Clindamicina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0
Cloranfenicol	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Colistina (Polimixina E)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
Daptomicina	0	0	0	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0
Doxiciclina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0
Eritromicina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0
Ertapenem	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	3	3	3	3	3	3	3	3	3	3	3	3
Fosfomicina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Gentamicina	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Imipenem	0	0	0	3	3	3	3	3	3	3	3	3	3	3	3	1	3	0	3	3	3	3	3	3	3	3	3	3	3	3
Levofloxacino	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1
Linezolida	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	0	0
Meropenem	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
Nitrofurantoína	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Norfloxacino	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	
Oxacilina	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Penicilina	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	
Piperacilina-tazobactam	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	0	0	
Polimixina B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Rifampicina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Teicoplanina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1
Tetraciclina	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tigeciclina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0
Trimetoprima-Sulfametoxazol	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1
Vancomicina	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
Amoxicilina	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1
Azitromicina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0
Cefaclor	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Cefadroxila	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Cefalexina	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Claritromicina	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0
Penicilina G Benzatina	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	0	0
Penicilina G Cristalina	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1	1	1	1	1	1	0	0
Penicilina G Procaína	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1	1	1	1	1	1	0	0
Penicilina V - fenoximetílico	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	0	0

Fonte: Elaborado pela Autora.

APÊNDICE B – *Leaderboars* Resultantes do Algoritmo de AutoML

Para o desenvolvimento deste trabalho foi utilizada a ferramenta de aprendizado de máquina automatizada H_2O AutoML. Uma das saídas principais dessa ferramenta é o *leaderboard*, uma tabela que contém todos os modelos gerados durante o período de execução do algoritmo de AutoML. Em conjunto com a identificação dos modelos gerados também são dispostas algumas métricas de avaliação do modelo que permite que seja determinado o melhor modelo entre eles, em modelos de classificação binária como é o caso deste trabalho, a métrica utilizada para definir o melhor modelo é a AUC, dessa forma, as demais métricas não serão expostas.

O cálculo das métricas dispostas no *leaderboard* são calculadas a partir da validação cruzada, ou em inglês, *cross validation* que é feita em 5 partes, seguindo a configuração padrão da ferramenta. Vale ressaltar que essas métricas são calculadas a partir das bases utilizadas para a construção do modelo, ou seja, a base de treino e a base de validação, dessa forma, quando os modelos forem aplicados as bases de treinos a tendência é que os valores dos indicadores reduzam.

Tabela 11 – 20 Melhores modelos segundo *leaderboard* para cada um dos casos analisados.

Cenário	Hospital	Run Time	Index	Algoritmo	AUC
I e II	-	10 min	1	Stacked Ensemble	0.874
I e II	-	10 min	2	Stacked Ensemble	0.874
I e II	-	10 min	3	GBM	0.871
I e II	-	10 min	4	DRF	0.871
I e II	-	10 min	5	XGBoost	0.869
I e II	-	10 min	6	GBM	0.867
I e II	-	10 min	7	GBM	0.867
I e II	-	10 min	8	GBM	0.866
I e II	-	10 min	9	GBM	0.864
I e II	-	10 min	10	XGBoost	0.861
I e II	-	2 hrs	1	Stacked Ensemble	0.876
I e II	-	2 hrs	2	Stacked Ensemble	0.876
I e II	-	2 hrs	3	GBM	0.874
I e II	-	2 hrs	4	GBM	0.874
I e II	-	2 hrs	5	XGBoost	0.874
I e II	-	2 hrs	6	GBM	0.873
I e II	-	2 hrs	7	GBM	0.873
I e II	-	2 hrs	8	DRF	0.873
I e II	-	2 hrs	9	DRF	0.873
I e II	-	2 hrs	10	XGBoost	0.873

I e II	-	2 hrs	10	GBM	0.873
III	h1	10 min	1	Stacked Ensemble	0.878
III	h1	10 min	2	Stacked Ensemble	0.878
III	h1	10 min	3	GBM	0.877
III	h1	10 min	4	GBM	0.876
III	h1	10 min	5	GBM	0.876
III	h1	10 min	6	GBM	0.876
III	h1	10 min	7	GBM	0.876
III	h1	10 min	8	GBM	0.876
III	h1	10 min	9	GBM	0.876
III	h1	10 min	10	GBM	0.876
III	h1	2 hrs	1	Stacked Ensemble	0.880
III	h1	2 hrs	2	Stacked Ensemble	0.880
III	h1	2 hrs	3	GBM	0.878
III	h1	2 hrs	4	XGBoost	0.877
III	h1	2 hrs	5	XGBoost	0.877
III	h1	2 hrs	6	GBM	0.877
III	h1	2 hrs	7	XGBoost	0.877
III	h1	2 hrs	8	GBM	0.877
III	h1	2 hrs	9	XGBoost	0.877
III	h1	2 hrs	10	GBM	0.876
III	h2	10 min	1	Stacked Ensemble	0.907
III	h2	10 min	2	GBM	0.907
III	h2	10 min	3	Stacked Ensemble	0.907
III	h2	10 min	4	GBM	0.905
III	h2	10 min	5	GBM	0.905
III	h2	10 min	6	GBM	0.903
III	h2	10 min	7	GBM	0.902
III	h2	10 min	8	GBM	0.902
III	h2	10 min	9	GBM	0.901
III	h2	10 min	10	GBM	0.901
III	h2	2 hrs	1	Stacked Ensemble	0.908
III	h2	2 hrs	2	Stacked Ensemble	0.906
III	h2	2 hrs	3	GBM	0.905
III	h2	2 hrs	4	GBM	0.905
III	h2	2 hrs	5	GBM	0.904
III	h2	2 hrs	6	GBM	0.903
III	h2	2 hrs	7	GBM	0.902
III	h2	2 hrs	8	GBM	0.901
III	h2	2 hrs	9	GBM	0.901
III	h2	2 hrs	10	XGBoost	0.901
III	h3	10 min	1	Stacked Ensemble	0.851
III	h3	10 min	2	Stacked Ensemble	0.851
III	h3	10 min	3	GBM	0.850
III	h3	10 min	4	GBM	0.849
III	h3	10 min	5	DRF	0.849
III	h3	10 min	6	GBM	0.849
III	h3	10 min	7	GBM	0.849
III	h3	10 min	8	GBM	0.847

III	h3	10 min	9	GBM	0.847
III	h3	10 min	10	GBM	0.844
III	h3	2 hrs	1	Stacked Ensemble	0.852
III	h3	2 hrs	2	Stacked Ensemble	0.851
III	h3	2 hrs	3	GBM	0.851
III	h3	2 hrs	4	GBM	0.850
III	h3	2 hrs	5	GBM	0.850
III	h3	2 hrs	6	DRF	0.850
III	h3	2 hrs	7	GBM	0.849
III	h3	2 hrs	8	GBM	0.849
III	h3	2 hrs	9	GBM	0.849
III	h3	2 hrs	10	XGBoost	0.848
III	h4	10 min	1	Stacked Ensemble	0.895
III	h4	10 min	2	Stacked Ensemble	0.895
III	h4	10 min	3	GBM	0.894
III	h4	10 min	4	GBM	0.894
III	h4	10 min	5	GBM	0.894
III	h4	10 min	6	GBM	0.894
III	h4	10 min	7	GBM	0.892
III	h4	10 min	8	XGBoost	0.892
III	h4	10 min	9	DRF	0.892
III	h4	10 min	10	GBM	0.891
III	h4	2 hrs	1	Stacked Ensemble	0.895
III	h4	2 hrs	2	Stacked Ensemble	0.895
III	h4	2 hrs	3	GBM	0.894
III	h4	2 hrs	4	GBM	0.893
III	h4	2 hrs	5	GBM	0.893
III	h4	2 hrs	6	GBM	0.892
III	h4	2 hrs	7	DRF	0.892
III	h4	2 hrs	8	XGBoost	0.891
III	h4	2 hrs	9	XGBoost	0.891
III	h4	2 hrs	10	XGBoost	0.890
III	h5	10 min	1	Stacked Ensemble	0.884
III	h5	10 min	2	Stacked Ensemble	0.883
III	h5	10 min	3	GBM	0.883
III	h5	10 min	4	GBM	0.880
III	h5	10 min	5	GBM	0.880
III	h5	10 min	6	GBM	0.880
III	h5	10 min	7	GBM	0.879
III	h5	10 min	8	GBM	0.879
III	h5	10 min	9	DRF	0.879
III	h5	10 min	10	GBM	0.879
III	h5	2 hrs	1	Stacked Ensemble	0.883
III	h5	2 hrs	2	Stacked Ensemble	0.883
III	h5	2 hrs	3	GBM	0.883
III	h5	2 hrs	4	GBM	0.882
III	h5	2 hrs	5	GBM	0.880
III	h5	2 hrs	6	GBM	0.880

III	h5	2 hrs	7	GBM	0.879
III	h5	2 hrs	8	GBM	0.879
III	h5	2 hrs	9	GBM	0.879
III	h5	2 hrs	10	DRF	0.878
III	h6	10 min	1	XGBoost	0.949
III	h6	10 min	2	XGBoost	0.949
III	h6	10 min	3	Deep Learning	0.947
III	h6	10 min	4	XGBoost	0.944
III	h6	10 min	5	Deep Learning	0.943
III	h6	10 min	6	Deep Learning	0.943
III	h6	10 min	7	GBM	0.943
III	h6	10 min	8	DRF	0.942
III	h6	10 min	9	Stacked Ensemble	0.942
III	h6	10 min	10	GBM	0.942
III	h6	2 hrs	1	XGBoost	0.949
III	h6	2 hrs	2	XGBoost	0.949
III	h6	2 hrs	3	XGBoost	0.949
III	h6	2 hrs	4	XGBoost	0.949
III	h6	2 hrs	5	XGBoost	0.949
III	h6	2 hrs	6	XGBoost	0.949
III	h6	2 hrs	7	XGBoost	0.949
III	h6	2 hrs	8	XGBoost	0.949
III	h6	2 hrs	9	XGBoost	0.949
III	h6	2 hrs	10	Deep Learning	0.948
IV	h1	10 min	1	Stacked Ensemble	0.874
IV	h1	10 min	2	Stacked Ensemble	0.874
IV	h1	10 min	3	GBM	0.872
IV	h1	10 min	4	DRF	0.871
IV	h1	10 min	5	GBM	0.871
IV	h1	10 min	6	GBM	0.866
IV	h1	10 min	7	GBM	0.866
IV	h1	10 min	8	XGBoost	0.866
IV	h1	10 min	9	GBM	0.865
IV	h1	10 min	10	GBM	0.864
IV	h1	2 hrs	1	Stacked Ensemble	0.876
IV	h1	2 hrs	2	Stacked Ensemble	0.876
IV	h1	2 hrs	3	GBM	0.876
IV	h1	2 hrs	4	GBM	0.874
IV	h1	2 hrs	5	GBM	0.874
IV	h1	2 hrs	6	GBM	0.874
IV	h1	2 hrs	7	GBM	0.874
IV	h1	2 hrs	8	DRF	0.873
IV	h1	2 hrs	9	GBM	0.873
IV	h1	2 hrs	10	GBM	0.873
IV	h2	10 min	1	Stacked Ensemble	0.872
IV	h2	10 min	2	Stacked Ensemble	0.872
IV	h2	10 min	3	DRF	0.871
IV	h2	10 min	4	GBM	0.870
IV	h2	10 min	5	GBM	0.867

IV	h2	10 min	6	GBM	0.866
IV	h2	10 min	7	GBM	0.865
IV	h2	10 min	8	GBM	0.864
IV	h2	10 min	9	XGBoost	0.862
IV	h2	10 min	10	XGBoost	0.862
IV	h2	2 hrs	1	Stacked Ensemble	0.876
IV	h2	2 hrs	2	Stacked Ensemble	0.876
IV	h2	2 hrs	3	GBM	0.875
IV	h2	2 hrs	4	GBM	0.875
IV	h2	2 hrs	5	GBM	0.874
IV	h2	2 hrs	6	DRF	0.873
IV	h2	2 hrs	7	GBM	0.873
IV	h2	2 hrs	8	GBM	0.872
IV	h2	2 hrs	9	GBM	0.872
IV	h2	2 hrs	10	GBM	0.872
IV	h3	10 min	1	Stacked Ensemble	0.874
IV	h3	10 min	2	Stacked Ensemble	0.873
IV	h3	10 min	3	DRF	0.871
IV	h3	10 min	4	GBM	0.870
IV	h3	10 min	5	XGBoost	0.869
IV	h3	10 min	6	GBM	0.867
IV	h3	10 min	7	GBM	0.867
IV	h3	10 min	8	GBM	0.862
IV	h3	10 min	9	GBM	0.861
IV	h3	10 min	10	XGBoost	0.861
IV	h3	2 hrs	1	Stacked Ensemble	0.876
IV	h3	2 hrs	2	Stacked Ensemble	0.876
IV	h3	2 hrs	3	GBM	0.875
IV	h3	2 hrs	4	DRF	0.873
IV	h3	2 hrs	5	XGBoost	0.873
IV	h3	2 hrs	6	GBM	0.873
IV	h3	2 hrs	7	XGBoost	0.873
IV	h3	2 hrs	8	XGBoost	0.873
IV	h3	2 hrs	9	XGBoost	0.872
IV	h3	2 hrs	10	GBM	0.872
IV	h4	10 min	1	Stacked Ensemble	0.873
IV	h4	10 min	2	Stacked Ensemble	0.873
IV	h4	10 min	3	GBM	0.871
IV	h4	10 min	4	DRF	0.871
IV	h4	10 min	5	GBM	0.870
IV	h4	10 min	6	XGBoost	0.867
IV	h4	10 min	7	GBM	0.866
IV	h4	10 min	8	GBM	0.866
IV	h4	10 min	9	GBM	0.864
IV	h4	10 min	10	XGBoost	0.861
IV	h4	2 hrs	1	Stacked Ensemble	0.876
IV	h4	2 hrs	2	Stacked Ensemble	0.875
IV	h4	2 hrs	3	GBM	0.874

IV	h4	2 hrs	4	GBM	0.874
IV	h4	2 hrs	5	GBM	0.873
IV	h4	2 hrs	6	GBM	0.873
IV	h4	2 hrs	7	XGBoost	0.873
IV	h4	2 hrs	8	XGBoost	0.873
IV	h4	2 hrs	9	DRF	0.873
IV	h4	2 hrs	10	XGBoost	0.872
IV	h5	10 min	1	Stacked Ensemble	0.873
IV	h5	10 min	2	Stacked Ensemble	0.873
IV	h5	10 min	3	DRF	0.871
IV	h5	10 min	4	GBM	0.871
IV	h5	10 min	5	GBM	0.870
IV	h5	10 min	6	GBM	0.869
IV	h5	10 min	7	GBM	0.867
IV	h5	10 min	8	GBM	0.866
IV	h5	10 min	9	GBM	0.866
IV	h5	10 min	10	GBM	0.864
IV	h5	2 hrs	1	Stacked Ensemble	0.876
IV	h5	2 hrs	2	Stacked Ensemble	0.876
IV	h5	2 hrs	3	GBM	0.874
IV	h5	2 hrs	4	GBM	0.874
IV	h5	2 hrs	5	GBM	0.874
IV	h5	2 hrs	6	GBM	0.874
IV	h5	2 hrs	7	DRF	0.873
IV	h5	2 hrs	8	XGBoost	0.872
IV	h5	2 hrs	9	GBM	0.872
IV	h5	2 hrs	10	GBM	0.872
IV	h6	10 min	1	Stacked Ensemble	0.874
IV	h6	10 min	2	Stacked Ensemble	0.874
IV	h6	10 min	3	DRF	0.871
IV	h6	10 min	4	GBM	0.871
IV	h6	10 min	5	GBM	0.871
IV	h6	10 min	6	XGBoost	0.868
IV	h6	10 min	7	GBM	0.867
IV	h6	10 min	8	GBM	0.867
IV	h6	10 min	9	GBM	0.866
IV	h6	10 min	10	GBM	0.864
IV	h6	2 hrs	1	Stacked Ensemble	0.876
IV	h6	2 hrs	2	Stacked Ensemble	0.876
IV	h6	2 hrs	3	GBM	0.875
IV	h6	2 hrs	4	GBM	0.874
IV	h6	2 hrs	5	GBM	0.874
IV	h6	2 hrs	6	GBM	0.874
IV	h6	2 hrs	7	DRF	0.873
IV	h6	2 hrs	8	GBM	0.873
IV	h6	2 hrs	9	XGBoost	0.873
IV	h6	2 hrs	10	GBM	0.873

Fonte: Elaborado pela Autora

APÊNDICE C – Teste e Poder de Previsão dos Modelos Gerados

No capítulo 3 deste trabalho foi citado que, apesar do AutoML ter gerado inúmeros modelos para os diferentes cenários propostos, apenas os dez primeiros segundo o *leaderboard*, ou seja, os modelos que apresentavam os 10 melhores valores de AUC foram aplicados a base teste. Porém, o algoritmo do H_2O AutoML foi executado duas vezes, com dois valores diferentes de tempo de processamento, 10 minutos e 2 horas. Dessa forma, foram aplicadas a todas as bases de teste dos cenários 20 modelos, sendo que eles variam de acordo com os hospitais e os cenários. Neste apêndice serão apresentadas métricas para a interpretação do poder preditivo destes 20 modelos testados.

Tabela 12 – Mensuração do poder preditivo dos modelos gerados para o cenário I.

Cenário I							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	7	GBM	0.704	0.468	0.860	0.922	0.486
10 min	2	Stacked Ensemble	0.703	0.473	0.862	0.931	0.475
10 min	1	Stacked Ensemble	0.703	0.473	0.862	0.931	0.475
10 min	5	XGBoost	0.700	0.463	0.860	0.926	0.473
10 min	8	GBM	0.699	0.460	0.859	0.925	0.473
10 min	3	GBM	0.697	0.465	0.861	0.932	0.463
10 min	4	DRF	0.696	0.462	0.860	0.932	0.459
10 min	9	GBM	0.684	0.447	0.859	0.939	0.428
10 min	6	GBM	0.683	0.447	0.859	0.940	0.427
10 min	10	XGBoost	0.679	0.438	0.857	0.939	0.419
2 hrs	4	GBM	0.710	0.477	0.861	0.922	0.497
2 hrs	3	GBM	0.709	0.478	0.862	0.923	0.495
2 hrs	6	GBM	0.708	0.477	0.862	0.925	0.490
2 hrs	1	Stacked Ensemble	0.707	0.476	0.862	0.925	0.489
2 hrs	10	GBM	0.707	0.472	0.860	0.921	0.403
2 hrs	2	Stacked Ensemble	0.704	0.473	0.862	0.928	0.480
2 hrs	7	GBM	0.704	0.472	0.861	0.928	0.480
2 hrs	5	XGBoost	0.701	0.468	0.861	0.929	0.473
2 hrs	8	DRF	0.600	0.469	0.862	0.934	0.464
2 hrs	9	XGBoost	0.698	0.463	0.860	0.929	0.468

Fonte: Elaborado pela Autora

Tabela 13 – Mensuração do poder preditivo dos modelos gerados para o cenário II quando testados no hospital h1.

Cenário II – Hospital h1

Run Time	Index	Algoritmo	AUC	MCC	F1	Sensibilidade	Especificidade
10 min	4	DRF	0.715	0.544	0.916	0.973	0.456
10 min	7	GBM	0.711	0.529	0.913	0.968	0.454
10 min	1	Stacked Ensemble	0.711	0.532	0.914	0.970	0.452
10 min	2	Stacked Ensemble	0.711	0.531	0.914	0.970	0.452
10 min	5	XGBoost	0.707	0.520	0.912	0.967	0.446
10 min	8	GBM	0.707	0.524	0.913	0.970	0.443
10 min	3	GBM	0.706	0.528	0.914	0.972	0.439
10 min	9	GBM	0.696	0.509	0.911	0.971	0.421
10 min	6	GBM	0.696	0.517	0.912	0.975	0.416
10 min	10	XGBoost	0.683	0.496	0.910	0.977	0.388
2 hrs	3	GBM	0.729	0.560	0.918	0.969	0.489
2 hrs	4	GBM	0.723	0.548	0.916	0.968	0.478
2 hrs	2	Stacked Ensemble	0.720	0.549	0.916	0.971	0.469
2 hrs	1	Stacked Ensemble	0.719	0.546	0.916	0.970	0.469
2 hrs	18	DRF	0.719	0.548	0.916	0.971	0.467
2 hrs	7	GBM	0.718	0.545	0.916	0.971	0.465
2 hrs	6	GBM	0.717	0.542	0.915	0.970	0.464
2 hrs	10	GBM	0.716	0.541	0.915	0.970	0.461
2 hrs	5	XGBoost	0.712	0.539	0.915	0.973	0.452
2 hrs	9	XGBoost	0.707	0.531	0.914	0.973	0.442

Fonte: Elaborado pela Autora

Tabela 14 – Mensuração do poder preditivo dos modelos gerados para o cenário II quando testados no hospital h2.

Cenário II – Hospital h2							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	4	DRF	0.818	0.682	0.913	0.950	0.687
10 min	1	Stacked Ensemble	0.811	0.668	0.909	0.948	0.674
10 min	2	Stacked Ensemble	0.811	0.668	0.909	0.948	0.674
10 min	3	GBM	0.803	0.653	0.905	0.944	0.662
10 min	5	XGBoost	0.802	0.645	0.903	0.938	0.667
10 min	8	GBM	0.794	0.625	0.897	0.931	0.656
10 min	6	GBM	0.787	0.628	0.900	0.946	0.628
10 min	9	GBM	0.782	0.624	0.899	0.949	0.616
10 min	7	GBM	0.782	0.609	0.894	0.935	0.628
10 min	10	XGBoost	0.780	0.614	0.896	0.943	0.616
2 hrs	3	GBM	0.842	0.714	0.920	0.946	0.738
2 hrs	6	GBM	0.836	0.703	0.917	0.943	0.728
2 hrs	1	Stacked Ensemble	0.834	0.704	0.918	0.948	0.720
2 hrs	4	GBM	0.831	0.691	0.914	0.939	0.723
2 hrs	2	Stacked Ensemble	0.830	0.702	0.918	0.951	0.710
2 hrs	7	GBM	0.822	0.679	0.911	0.941	0.702
2 hrs	8	DRF	0.821	0.691	0.916	0.955	0.687
2 hrs	10	GBM	0.820	0.674	0.909	0.938	0.702
2 hrs	5	XGBoost	0.812	0.667	0.909	0.944	0.679

2 hrs	9	XGBoost	0.801	0.660	0.908	0.95	0.646

Fonte: Elaborado pela Autora

Tabela 15 – Mensuração do poder preditivo dos modelos gerados para o cenário II quando testados no hospital h3.

Cenário II – Hospital h3							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	7	GBM	0.687	0.449	0.863	0.936	0.438
10 min	1	Stacked Ensemble	0.682	0.454	0.866	0.948	0.416
10 min	2	Stacked Ensemble	0.682	0.454	0.866	0.948	0.416
10 min	8	GBM	0.680	0.441	0.862	0.940	0.421
10 min	4	DRF	0.680	0.457	0.867	0.954	0.407
10 min	5	XGBoost	0.680	0.443	0.863	0.943	0.417
10 min	3	GBM	0.674	0.444	0.865	0.952	0.396
10 min	6	GBM	0.656	0.419	0.861	0.959	0.353
10 min	10	XGBoost	0.656	0.415	0.860	0.956	0.355
10 min	9	GBM	0.655	0.418	0.861	0.959	0.351
2 hrs	3	GBM	0.698	0.479	0.870	0.946	0.450
2 hrs	10	GBM	0.697	0.467	0.867	0.936	0.457
2 hrs	4	GBM	0.606	0.469	0.867	0.939	0.453
2 hrs	1	Stacked Ensemble	0.692	0.469	0.869	0.946	0.437
2 hrs	6	GBM	0.689	0.463	0.867	0.945	0.434
2 hrs	2	Stacked Ensemble	0.688	0.465	0.868	0.949	0.426
2 hrs	9	XGBoost	0.686	0.460	0.867	0.948	0.423
2 hrs	10	GBM	0.685	0.458	0.867	0.947	0.423
2 hrs	5	XGBoost	0.682	0.456	0.867	0.949	0.415
2 hrs	8	DRF	0.681	0.462	0.869	0.956	0.407

Fonte: Elaborado pela Autora

Tabela 16 – Mensuração do poder preditivo dos modelos gerados para o cenário II quando testados no hospital h4.

Cenário II – Hospital h4							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	4	DRF	0.793	0.619	0.857	0.922	0.664
10 min	1	Stacked Ensemble	0.789	0.608	0.852	0.913	0.666
10 min	2	Stacked Ensemble	0.789	0.608	0.852	0.913	0.666
10 min	3	GBM	0.788	0.603	0.850	0.908	0.668
10 min	8	GBM	0.787	0.599	0.848	0.902	0.673
10 min	7	GBM	0.784	0.596	0.847	0.906	0.662
10 min	5	XGBoost	0.780	0.592	0.847	0.912	0.649
10 min	6	GBM	0.776	0.588	0.846	0.917	0.635
10 min	10	XGBoost	0.772	0.583	0.845	0.921	0.624
10 min	9	GBM	0.772	0.579	0.843	0.914	0.630
2 hrs	3	GBM	0.805	0.635	0.861	0.914	0.695

2 hrs	1	Stacked Ensemble	0.801	0.629	0.859	0.913	0.690
2 hrs	4	GBM	0.801	0.627	0.858	0.912	0.690
2 hrs	6	GBM	0.798	0.623	0.857	0.913	0.682
2 hrs	2	Stacked Ensemble	0.797	0.622	0.857	0.916	0.677
2 hrs	7	GBM	0.796	0.619	0.856	0.913	0.679
2 hrs	8	DRF	0.795	0.621	0.857	0.919	0.672
2 hrs	10	GBM	0.794	0.615	0.854	0.911	0.678
2 hrs	5	XGBoost	0.789	0.610	0.853	0.918	0.661
2 hrs	9	XGBoost	0.784	0.603	0.851	0.920	0.648

Fonte: Elaborado pela Autora

Tabela 17 – Mensuração do poder preditivo dos modelos gerados para o cenário II quando testados no hospital h5.

Cenário II – Hospital h5							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	8	GBM	0.719	0.517	0.877	0.94	0.492
10 min	4	DRF	0.715	0.540	0.884	0.970	0.461
10 min	3	GBM	0.714	0.528	0.881	0.963	0.464
10 min	6	GBM	0.709	0.511	0.877	0.957	0.461
10 min	7	GBM	0.708	0.503	0.875	0.952	0.464
10 min	1	Stacked Ensemble	0.706	0.517	0.879	0.965	0.448
10 min	2	Stacked Ensemble	0.706	0.517	0.879	0.965	0.448
10 min	10	XGBoost	0.701	0.506	0.877	0.963	0.439
10 min	5	XGBoost	0.700	0.507	0.877	0.965	0.436
10 min	9	GBM	0.698	0.495	0.874	0.958	0.439
2 hrs	2	Stacked Ensemble	0.732	0.569	0.890	0.971	0.492
2 hrs	6	GBM	0.731	0.551	0.885	0.958	0.505
2 hrs	3	GBM	0.729	0.557	0.887	0.966	0.492
2 hrs	10	GBM	0.727	0.549	0.885	0.962	0.492
2 hrs	8	DRF	0.727	0.561	0.888	0.971	0.483
2 hrs	7	GBM	0.725	0.544	0.884	0.961	0.489
2 hrs	1	Stacked Ensemble	0.722	0.550	0.886	0.970	0.473
2 hrs	4	GBM	0.721	0.541	0.884	0.963	0.480
2 hrs	5	XGBoost	0.718	0.542	0.884	0.969	0.467
2 hrs	9	XGBoost	0.707	0.524	0.880	0.969	0.445

Fonte: Elaborado pela Autora

Tabela 18 – Mensuração do poder preditivo dos modelos gerados para o cenário II quando testados no hospital h6.

Cenário II – Hospital h6							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	1	Stacked Ensemble	0.937	0.827	0.933	0.874	1.000
10 min	2	Stacked Ensemble	0.937	0.827	0.933	0.874	1.000
10 min	3	GBM	0.937	0.827	0.933	0.874	1.000
10 min	4	DRF	0.937	0.827	0.933	0.874	1.000

10 min	6	GBM	0.937	0.827	0.933	0.874	1.000
10 min	7	GBM	0.937	0.827	0.933	0.874	1.000
10 min	8	GBM	0.937	0.827	0.933	0.874	1.000
10 min	9	GBM	0.937	0.827	0.933	0.874	1.000
10 min	5	XGBoost	0.916	0.792	0.923	0.874	0.957
10 min	10	XGBoost	0.757	0.596	0.888	0.961	0.553
2 hrs	1	Stacked Ensemble	0.937	0.827	0.933	0.874	1.000
2 hrs	2	Stacked Ensemble	0.937	0.827	0.933	0.874	1.000
2 hrs	3	GBM	0.937	0.827	0.933	0.874	1.000
2 hrs	4	GBM	0.937	0.827	0.933	0.874	1.000
2 hrs	5	XGBoost	0.937	0.827	0.933	0.874	1.000
2 hrs	6	GBM	0.937	0.827	0.933	0.874	1.000
2 hrs	7	GBM	0.937	0.827	0.933	0.874	1.000
2 hrs	8	DRF	0.937	0.827	0.933	0.874	1.000
2 hrs	9	XGBoost	0.937	0.827	0.933	0.874	1.000
2 hrs	10	GBM	0.937	0.827	0.933	0.874	1.000

Fonte: Elaborado pela Autora

Tabela 19 – Mensuração do poder preditivo dos modelos gerados para o cenário III quando testados no hospital h1.

Cenário III – Hospital h1							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	6	GBM	0.723	0.520	0.909	0.950	0.496
10 min	7	GBM	0.722	0.517	0.908	0.949	0.495
10 min	4	GBM	0.722	0.518	0.909	0.950	0.493
10 min	1	Stacked Ensemble	0.717	0.514	0.909	0.954	0.480
10 min	9	GBM	0.700	0.501	0.909	0.963	0.437
10 min	5	GBM	0.696	0.497	0.908	0.965	0.427
10 min	3	GBM	0.690	0.488	0.907	0.966	0.415
10 min	10	GBM	0.684	0.476	0.906	0.965	0.402
10 min	2	Stacked Ensemble	0.665	0.466	0.906	0.978	0.353
10 min	8	GBM	0.661	0.457	0.905	0.976	0.347
2 hrs	3	GBM	0.710	0.508	0.908	0.957	0.463
2 hrs	2	Stacked Ensemble	0.704	0.502	0.908	0.960	0.449
2 hrs	8	GBM	0.698	0.499	0.908	0.964	0.433
2 hrs	9	XGBoost	0.698	0.497	0.908	0.963	0.432
2 hrs	1	Stacked Ensemble	0.698	0.496	0.908	0.963	0.432
2 hrs	6	GBM	0.692	0.488	0.907	0.963	0.421
2 hrs	5	XGBoost	0.688	0.476	0.905	0.962	0.414
2 hrs	4	XGBoost	0.669	0.458	0.904	0.971	0.367
2 hrs	10	GBM	0.666	0.464	0.905	0.976	0.356
2 hrs	7	XGBoost	0.657	0.453	0.904	0.979	0.336

Fonte: Elaborado pela Autora

Tabela 20 – Mensuração do poder preditivo dos modelos gerados para o cenário III quando testados no hospital h2.

Cenário III – Hospital h2							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	9	GBM	0.784	0.583	0.881	0.896	0.672
10 min	6	GBM	0.781	0.573	0.878	0.890	0.672
10 min	4	GBM	0.780	0.582	0.883	0.904	0.656
10 min	8	GBM	0.778	0.563	0.873	0.881	0.674
10 min	2	GBM	0.778	0.582	0.884	0.909	0.646
10 min	7	GBM	0.777	0.577	0.882	0.905	0.649
10 min	10	GBM	0.775	0.572	0.880	0.902	0.649
10 min	1	StackedEnsemble	0.774	0.575	0.882	0.909	0.639
10 min	3	StackedEnsemble	0.770	0.570	0.882	0.911	0.628
10 min	5	GBM	0.769	0.571	0.882	0.915	0.623
2 hrs	9	GBM	0.790	0.585	0.879	0.884	0.697
2 hrs	5	GBM	0.784	0.582	0.881	0.894	0.674
2 hrs	3	GBM	0.783	0.584	0.882	0.900	0.667
2 hrs	8	GBM	0.783	0.584	0.882	0.900	0.667
2 hrs	2	StackedEnsemble	0.779	0.574	0.879	0.896	0.662
2 hrs	6	GBM	0.773	0.571	0,880	0.905	0.641
2 hrs	4	GBM	0.772	0.583	0.887	0.923	0.621
2 hrs	1	StackedEnsemble	0.771	0.580	0.885	0.921	0.621
2 hrs	7	GBM	0.770	0.559	0.876	0.894	0.646
2 hrs	10	XGBoost	0.769	0.577	0.885	0.922	0.616

Fonte: Elaborado pela Autora

Tabela 21 – Mensuração do poder preditivo dos modelos gerados para o cenário III quando testados no hospital h3.

Cenário III – Hospital h3							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	1	Stacked Ensemble	0.668	0.414	0.857	0.935	0.402
10 min	2	Stacked Ensemble	0.667	0.410	0.856	0.933	0.402
10 min	10	GBM	0.655	0.397	0.855	0.942	0.369
10 min	8	GBM	0.655	0.384	0.851	0.930	0.380
10 min	4	GBM	0.651	0.395	0.856	0.947	0.355
10 min	7	GBM	0.649	0.385	0.853	0.943	0.354
10 min	6	GBM	0.640	0.380	0.854	0.954	0.325
10 min	3	GBM	0.634	0.375	0.854	0.959	0.309
10 min	5	DRF	0.624	0.363	0.853	0.964	0.284
10 min	9	GBM	0.619	0.359	0.853	0.969	0.268
2 hrs	5	GBM	0.665	0.406	0.855	0.934	0.395
2 hrs	1	Stacked Ensemble	0.663	0.408	0.856	0.938	0.389
2 hrs	2	Stacked Ensemble	0.659	0.398	0.854	0.936	0.383
2 hrs	4	GBM	0.658	0.401	0.856	0.941	0.375
2 hrs	6	DRF	0.657	0.397	0.855	0.940	0.373
2 hrs	9	GBM	0.649	0.380	0.851	0.937	0.362
2 hrs	7	GBM	0.644	0.389	0.856	0.954	0.334
2 hrs	3	GBM	0.638	0.382	0.855	0.958	0.318

2 hrs	10	XGBoost	0.629	0.367	0.853	0. 960	0.298
2 hrs	8	GBM	0.623	0.361	0.853	0.964	0.282

Fonte: Elaborado pela Autora

Tabela 22 – Mensuração do poder preditivo dos modelos gerados para o cenário III quando testados no hospital h4.

Cenário III – Hospital h4							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	7	GBM	0.766	0.573	0.842	0.921	0.611
10 min	10	GBM	0.764	0.573	0.842	0.927	0.601
10 min	1	Stacked Ensemble	0.760	0.568	0.841	0.930	0.590
10 min	5	GBM	0.759	0.564	0.839	0.926	0.593
10 min	8	XGBoost	0.757	0.564	0.840	0.932	0.581
10 min	6	GBM	0.755	0.561	0.839	0.931	0.579
10 min	4	GBM	0.755	0.563	0.840	0.935	0.574
10 min	3	GBM	0.754	0.562	0.839	0.937	0.570
10 min	2	Stacked Ensemble	0.745	0.556	0.838	0.949	0.541
10 min	9	DRF	0.739	0.548	0.836	0.952	0.526
2 hrs	3	GBM	0.763	0.570	0.841	0.926	0.600
2 hrs	6	GBM	0.760	0.566	0.840	0.928	0.592
2 hrs	4	GBM	0.757	0.561	0.839	0.928	0.586
2 hrs	2	Stacked Ensemble	0.756	0.565	0.840	0.936	0.575
2 hrs	7	DRF	0.755	0.562	0.839	0.934	0.575
2 hrs	5	GBM	0.754	0.564	0.840	0.938	0.571
2 hrs	8	XGBoost	0.753	0.560	0.839	0.936	0.571
2 hrs	1	Stacked Ensemble	0.752	0.559	0.839	0.937	0.567
2 hrs	10	XGBoost	0.752	0.557	0.838	0.935	0.568
2 hrs	9	XGBoost	0.734	0.536	0.832	0.948	0.520

Fonte: Elaborado pela Autora

Tabela 23 – Mensuração do poder preditivo dos modelos gerados para o cenário III quando testados no hospital h5.

Cenário III – Hospital h5							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	1	Stacked Ensemble	0.736	0.516	0.872	0.916	0.555
10 min	5	GBM	0.735	0.507	0.868	0.906	0.564
10 min	7	GBM	0.730	0.499	0.866	0.906	0.555
10 min	9	DRF	0.730	0.506	0.869	0.915	0.545
10 min	2	Stacked Ensemble	0.730	0.502	0.868	0.911	0.549
10 min	8	GBM	0.729	0.500	0.867	0.910	0.549
10 min	10	GBM	0.728	0.499	0.867	0.911	0.545
10 min	4	GBM	0.727	0.494	0.865	0.906	0.549
10 min	3	GBM	0.724	0.509	0.873	0.931	0.517
10 min	6	GBM	0.711	0.485	0.868	0.929	0.492

2 hrs	10	DRF	0.742	0.516	0.869	0.903	0.580
2 hrs	9	GBM	0.732	0.502	0.867	0.906	0.558
2 hrs	7	GBM	0.729	0.502	0.868	0.912	0.545
2 hrs	8	GBM	0.728	0.499	0.867	0.911	0.545
2 hrs	4	GBM	0.728	0.501	0.868	0.914	0.542
2 hrs	3	Stacked Ensemble	0.723	0.511	0.874	0.935	0.511
2 hrs	1	Stacked Ensemble	0.722	0.494	0.868	0.918	0.527
2 hrs	2	GBM	0.719	0.501	0.871	0.931	0.508
2 hrs	5	GBM	0.717	0.500	0.872	0.933	0.502
2 hrs	6	GBM	0.680	0.480	0.872	0.971	0.389

Fonte: Elaborado pela Autora

Tabela 24 – Mensuração do poder preditivo dos modelos gerados para o cenário III quando testados no hospital h6.

Cenário III – Hospital h6							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	1	XGBoost	0.937	0.827	0.933	0.874	1.000
10 min	2	XGBoost	0.937	0.827	0.933	0.874	1.000
10 min	4	XGBoost	0.937	0.827	0.933	0.874	1.000
10 min	7	GBM	0.937	0.827	0.933	0.874	1.000
10 min	9	Stacked Ensemble	0.937	0.827	0.933	0.874	1.000
10 min	6	Deep Learning	0.917	0.783	0.910	0.835	1.000
10 min	3	Deep Learning	0.908	0.762	0.898	0.816	1.000
10 min	5	Deep Learning	0.908	0.762	0.898	0.816	1.000
10 min	8	DRF	0.908	0.762	0.898	0.816	1.000
10 min	10	GBM	0.908	0.762	0.898	0.816	1.000
2 hrs	1	XGBoost	0.937	0.827	0.933	0.874	1.000
2 hrs	2	XGBoost	0.937	0.827	0.933	0.874	1.000
2 hrs	3	XGBoost	0.937	0.827	0.933	0.874	1.000
2 hrs	4	XGBoost	0.937	0.827	0.933	0.874	1.000
2 hrs	5	XGBoost	0.937	0.827	0.933	0.874	1.000
2 hrs	6	XGBoost	0.937	0.827	0.933	0.874	1.000
2 hrs	7	XGBoost	0.937	0.827	0.933	0.874	1.000
2 hrs	8	XGBoost	0.937	0.827	0.933	0.874	1.000
2 hrs	9	XGBoost	0.937	0.827	0.933	0.874	1.000
2 hrs	10	Deep Learning	0.917	0.783	0.910	0.835	1.000

Fonte: Elaborado pela Autora

Tabela 25 – Mensuração do poder preditivo dos modelos gerados para o cenário IV quando testados no hospital h1.

Cenário IV – Hospital h1							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	2	Stacked Ensemble	0.760	0.567	0.914	0.939	0.581
10 min	1	Stacked Ensemble	0.758	0.569	0.915	0.943	0.573
10 min	4	DRF	0.758	0.571	0.915	0.945	0.571

10 min	3	GBM	0.758	0.569	0.915	0.944	0.572
10 min	5	GBM	0.743	0.554	0.914	0.951	0.535
10 min	7	GBM	0.741	0.544	0.912	0.946	0.535
10 min	10	GBM	0.736	0.535	0.910	0.945	0.528
10 min	9	GBM	0.735	0.538	0.911	0.949	0.521
10 min	8	XGBoost	0.733	0.536	0.911	0.950	0.515
10 min	6	GBM	0.730	0.539	0.912	0.954	0.507
2 hrs	8	DRF	0.770	0.583	0.916	0.939	0.601
2 hrs	5	GBM	0.769	0.579	0.915	0.938	0.599
2 hrs	1	Stacked Ensemble	0.765	0.576	0.915	0.94	0.590
2 hrs	2	Stacked Ensemble	0.764	0.575	0.915	0.942	0.586
2 hrs	3	GBM	0.759	0.574	0.916	0.947	0.571
2 hrs	9	GBM	0.756	0.563	0.913	0.941	0.571
2 hrs	6	GBM	0.756	0.569	0.915	0.946	0.566
2 hrs	7	GBM	0.756	0.570	0.916	0.947	0.565
2 hrs	4	GBM	0.751	0.573	0.917	0.955	0.547
2 hrs	10	GBM	0.747	0.560	0.915	0.951	0.543

Fonte: Elaborado pela Autora

Tabela 26 – Mensuração do poder preditivo dos modelos gerados para o cenário IV quando testados no hospital h2.

Cenário IV – Hospital h2							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	3	DRF	0.841	0.697	0.913	0.926	0.756
10 min	1	Stacked Ensemble	0.826	0.673	0.907	0.925	0.728
10 min	2	Stacked Ensemble	0.826	0.673	0.907	0.925	0.728
10 min	4	GBM	0.819	0.657	0.902	0.919	0.720
10 min	7	GBM	0.806	0.633	0.896	0.916	0.697
10 min	6	GBM	0.800	0.626	0.895	0.919	0.682
10 min	5	GBM	0.799	0.625	0.896	0.921	0.677
10 min	10	XGBoost	0.794	0.616	0.893	0.919	0.669
10 min	8	GBM	0.792	0.610	0.891	0.917	0.667
10 min	9	XGBoost	0.791	0.617	0.895	0.926	0.656
2 hrs	7	GBM	0.852	0.713	0.917	0.925	0.779
2 hrs	1	Stacked Ensemble	0.850	0.712	0.917	0.927	0.774
2 hrs	6	DRF	0.850	0.708	0.915	0.923	0.776
2 hrs	4	GBM	0.849	0.716	0.919	0.935	0.763
2 hrs	2	Stacked Ensemble	0.848	0.708	0.916	0.927	0.768
2 hrs	5	GBM	0.845	0.692	0.909	0.911	0.779
2 hrs	3	GBM	0.844	0.710	0.918	0.937	0.751
2 hrs	10	GBM	0.828	0.653	0.897	0.894	0.761
2 hrs	8	GBM	0.822	0.644	0.895	0.894	0.751
2 hrs	9	GBM	0.821	0.665	0.905	0.925	0.718

Fonte: Elaborado pela Autora

Tabela 27 – Mensuração do poder preditivo dos modelos gerados para o cenário IV quando testados no hospital h3.

Cenário IV – Hospital h3							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	1	Stacked Ensemble	0.714	0.488	0.869	0.927	0.501
10 min	3	DRF	0.705	0.480	0.869	0.935	0.475
10 min	8	GBM	0.700	0.478	0.870	0.941	0.459
10 min	7	GBM	0.682	0.447	0.864	0.942	0.422
10 min	2	Stacked Ensemble	0.681	0.456	0.867	0.951	0.412
10 min	5	XGBoost	0.662	0.430	0.863	0.958	0.366
10 min	6	GBM	0.662	0.420	0.861	0.951	0.372
10 min	10	XGBoost	0.660	0.418	0.860	0.952	0.369
10 min	4	GBM	0.660	0.428	0.863	0.960	0.361
10 min	9	GBM	0.648	0.409	0.860	0.964	0.331
2 hrs	4	DRF	0.717	0.494	0.870	0.928	0.506
2 hrs	2	Stacked Ensemble	0.710	0.484	0.869	0.930	0.491
2 hrs	10	GBM	0.703	0.483	0.870	0.940	0.466
2 hrs	8	XGBoost	0.698	0.470	0.867	0.937	0.460
2 hrs	9	XGBoost	0.698	0.464	0.865	0.932	0.464
2 hrs	5	XGBoost	0.684	0.455	0.866	0.946	0.422
2 hrs	7	XGBoost	0.681	0.451	0.865	0.947	0.416
2 hrs	1	Stacked Ensemble	0.675	0.451	0.867	0.957	0.393
2 hrs	6	GBM	0.675	0.448	0.866	0.955	0.395
2 hrs	3	GBM	0.671	0.446	0.866	0.958	0.384

Fonte: Elaborado pela Autora

Tabela 28 – Mensuração do poder preditivo dos modelos gerados para o cenário IV quando testados no hospital h4.

Cenário IV – Hospital h4							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	4	DRF	0.789	0.612	0.854	0.921	0.657
10 min	5	GBM	0.789	0.603	0.850	0.905	0.673
10 min	7	GBM	0.784	0.595	0.847	0.905	0.662
10 min	9	GBM	0.784	0.59	0.844	0.895	0.673
10 min	1	Stacked Ensemble	0.781	0.602	0.851	0.927	0.636
10 min	3	GBM	0.775	0.601	0.852	0.942	0.607
10 min	10	XGBoost	0.772	0.579	0.843	0.914	0.631
10 min	2	Stacked Ensemble	0.770	0.592	0.849	0.942	0.599
10 min	6	XGBoost	0.765	0.578	0.844	0.933	0.597
10 min	8	GBM	0.764	0.576	0.844	0.933	0.595
2 hrs	3	GBM	0.796	0.624	0.858	0.923	0.668
2 hrs	9	DRF	0.794	0.619	0.856	0.920	0.668
2 hrs	5	GBM	0.792	0.617	0.856	0.922	0.661
2 hrs	4	GBM	0.791	0.611	0.853	0.912	0.670
2 hrs	8	XGBoost	0.784	0.604	0.852	0.921	0.647
2 hrs	10	XGBoost	0.783	0.601	0.850	0.920	0.646

2 hrs	2	Stacked Ensemble	0.783	0.608	0.854	0.932	0.633
2 hrs	7	XGBoost	0.781	0.600	0.850	0.924	0.639
2 hrs	1	Stacked Ensemble	0.777	0.599	0.851	0.935	0.619
2 hrs	6	GBM	0.775	0.593	0.849	0.931	0.618

Fonte: Elaborado pela Autora

Tabela 29 – Mensuração do poder preditivo dos modelos gerados para o cenário IV quando testados no hospital h5.

Cenário IV – Hospital h5							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	3	DRF	0.776	0.576	0.882	0.906	0.646
10 min	2	Stacked Ensemble	0.773	0.559	0.875	0.890	0.655
10 min	1	Stacked Ensemble	0.769	0.552	0.873	0.889	0.649
10 min	4	GBM	0.765	0.548	0.873	0.891	0.639
10 min	8	GBM	0.754	0.528	0.868	0.890	0.618
10 min	9	GBM	0.753	0.527	0.869	0.891	0.614
10 min	7	GBM	0.752	0.531	0.871	0.899	0.605
10 min	6	GBM	0.726	0.518	0.876	0.937	0.514
10 min	5	GBM	0.724	0.541	0.883	0.959	0.489
10 min	10	GBM	0.697	0.501	0.876	0.965	0.429
2 hrs	7	DRF	0.775	0.572	0.881	0.902	0.649
2 hrs	1	Stacked Ensemble	0.773	0.574	0.883	0.910	0.636
2 hrs	9	GBM	0.765	0.554	0.876	0.901	0.630
2 hrs	5	GBM	0.765	0.557	0.878	0.906	0.624
2 hrs	6	GBM	0.740	0.547	0.882	0.941	0.539
2 hrs	2	Stacked Ensemble	0.732	0.534	0.880	0.941	0.524
2 hrs	8	XGBoost	0.721	0.506	0.873	0.933	0.508
2 hrs	3	GBM	0.714	0.547	0.885	0.976	0.451
2 hrs	4	GBM	0.711	0.538	0.884	0.974	0.448
2 hrs	10	GBM	0.707	0.527	0.881	0.971	0.442

Fonte: Elaborado pela Autora

Tabela 30 – Mensuração do poder preditivo dos modelos gerados para o cenário IV quando testados no hospital h6.

Cenário IV – Hospital h6							
Run Time	**Index**	**Algoritmo**	**AUC**	**MCC**	**F1**	**Sensibilidade**	**Especificidade**
10 min	3	GBM	0.927	0.805	0.921	0.854	1.000
10 min	4	GBM	0.927	0.805	0.921	0.854	1.000
10 min	5	GBM	0.927	0.805	0.921	0.854	1.000
10 min	6	GBM	0.927	0.805	0.921	0.854	1.000
10 min	8	GBM	0.927	0.805	0.921	0.854	1.000
10 min	10	GBM	0.927	0.805	0.921	0.854	1.000
10 min	1	Stacked Ensemble	0.908	0.762	0.898	0.816	1.000
10 min	2	Stacked Ensemble	0.908	0.762	0.898	0.816	1.000

10 min	7	DRF	0.908	0.762	0.898	0.816	1.000
10 min	9	XGBoost	0.908	0.762	0.898	0.816	1.000
2 hrs	3	GBM	0.927	0.805	0.921	0.854	1.000
2 hrs	4	GBM	0.927	0.805	0.921	0.854	1.000
2 hrs	5	GBM	0.927	0.805	0.921	0.854	1.000
2 hrs	6	GBM	0.927	0.805	0.921	0.854	1.000
2 hrs	8	GBM	0.927	0.805	0.921	0.854	1.000
2 hrs	10	GBM	0.927	0.805	0.921	0.854	1.000
2 hrs	1	Stacked Ensemble	0.908	0.762	0.898	0.816	1.000
2 hrs	2	Stacked Ensemble	0.908	0.762	0.898	0.816	1.000
2 hrs	7	DRF	0.908	0.762	0.898	0.816	1.000
2 hrs	9	XGBoost	0.908	0.762	0.898	0.816	1.000

Fonte: Elaborado pela Autora

APÊNDICE D – Diferentes Modelos de Previsão Explorados

Gradient Boosting Machines (GBM)

A técnica de aprendizado de máquina denominada *Gradient Boosting* é útil em relação a problemas de regressão e classificação. Ela combina diferentes algoritmos de predição categorizados como fracos de forma a gerar uma predição mais confiável. Normalmente, para a construção dos modelos são utilizadas diversas árvores de decisão e a partir delas é obtido um modelo de classificação ou regressão genérico que otimiza a função de perda. O objetivo do algoritmo é gerar uma sequência de modelos fracos, em que cada um tem como objetivo minimizar o erro do modelo anterior. As melhorias implementadas nos modelos são quantificadas em uma taxa de aprendizagem que determina o impacto que cada uma das árvores terá no modelo final (LEDELL, 2020; BALAJI; ALLEN, 2018).

Extreme Gradient Boosting Machines (XGBoost)

Algoritmos *XGBoost*, como são conhecidos, possuem o mesmo objetivo dos algoritmos GBM: a otimização da função de perda. Além disso, pode-se afirmar que o XGBoost é uma aplicação específica de algoritmos GBM que tende a entregar aproximações mais precisas em um menor tempo. O XGBoost usa da regressão avançada (L1 e L2) para melhorar os recursos de generalização do modelo e utiliza de computação paralela para desenvolver e comparar os diferentes modelos de uma forma mais eficiente. A técnica também calcula derivadas parciais secundárias da função de perda para obter mais informações sobre a direção dos gradientes e, assim, otimizar a função de perda (LEDELL, 2020; BALAJI; ALLEN, 2018).

Distributed Random Forest (DRF)

Random Forest, que em português significa floresta aleatória, é um tipo de algoritmo que cria diversas áreas de decisão de maneira aleatória, variando na seleção da amostra e na seleção das variáveis consideradas. Cada árvore apresentará uma previsão diferente, e assim, em problemas de classificação, a categoria que for prevista o maior número de vezes será a

previsão final do algoritmo. Novas árvores continuam sendo geradas até que não haja melhora do desempenho do poder de predição final (LEDELL, 2020; BALAJI; ALLEN, 2018).

Deep Learning

Algoritmos de *Deep Learning*, também nomeados de *Neural Networks*, são conhecidos em português como Aprendizado de Máquina Profundo ou como Redes Neurais. Diferente dos demais algoritmos apresentados nesse apêndice, os algoritmos do tipo Deep Leaning são categorizados como algoritmos de aprendizagem não supervisionada, ou seja, eles são capazes de identificar padrões a partir dos dados sem possuir um objetivo específico. Esses padrões são identificados a partir das redes neurais convulsionais que são várias camadas de processamento de dados não lineares que resultam em uma inferência (LEDELL, 2020; BALAJI; ALLEN, 2018).

Stacked Ensemble

Modelos da categoria *Ensemble Learning,* que, na tradução literal para o português, significa aprendizado conjunto, agrupam diferentes modelos de previsão mais fracos para obter um modelo composto com melhor poder preditivo. Foram explorados alguns algoritmos dessa categoria neste apêndice, tais como: *Gradient Boosting Machines* (GBM), *Extreme Gradient Boosting Machines (XGBoost)* e *Distributed Random Forest* (DRF). Entre os principais métodos de *Ensemble Learning* estão o *Bagging*, *Boosting* e *Stacking*. O algoritmo nomeado como *Stacked Ensemble* que é citado no desenvolvimento deste estudo deriva do método *Stacking* que agrupa modelos heterogêneos treinados em paralelo, enquanto os demais métodos de *Ensemble Learning* agrupam modelos homogêneos. A ferramenta H_2O AutoML, utilizada para o desenvolvimento deste trabalho, treina dois modelos *Stacked Ensemble* ao final do processo de AutoML: um que inclui todos os modelos gerados e outro composto pelo melhor modelo de cada uma das famílias de algoritmos desenvolvidos (LEDELL, 2020; BALAJI; ALLEN, 2018).