

UNIVERSIDADE FEDERAL DE ALAGOAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

FLÁVIO YURI AQUINO DE OLIVEIRA

**UM MODELO COMPUTACIONAL DE RECOMENDAÇÃO DE TRATAMENTO
TERAPÊUTICO PERSONALIZADO PARA PACIENTE COM QUADRO SÉPTICO**

Maceió

2023

FLÁVIO YURI AQUINO DE OLIVEIRA

**UM MODELO COMPUTACIONAL DE RECOMENDAÇÃO DE TRATAMENTO
TERAPÊUTICO PERSONALIZADO PARA PACIENTE COM QUADRO SÉPTICO**

Dissertação apresentada em cumprimento parcial dos requisitos para o grau de Mestre em Ciências da Informática da Universidade Federal de Alagoas, Instituto de Computação.

Orientadores: Prof. Dr. Rafael de Amorim Silva e Prof. Dr. Bruno Almeida Pimentel

Maceió

2023

Catálogo na Fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 - 1767

O48m Oliveira, Flávio Yuri Aquino de.
Um modelo computacional de recomendação de tratamento terapêutico personalizado para paciente com quadro séptico / Flávio Yuri Aquino de Oliveira . – 2023.
75 f. : il.

Orientador: Rafael de Amorim Silva.
Co-orientador: Bruno Almeida Pimentel.
Dissertação (mestrado em informática) - Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2023.

Bibliografia: f. 53-59.
Apêndices: f. 60-75.

1. SEPSE. 2. Aprendizado do computador. 3. Prescrição terapêutica. I. Título.

CDU: 004.85:616.94



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE ALAGOAS
INSTITUTO DE COMPUTAÇÃO
Av. Lourival Melo Mota, S/N, Tabuleiro do Martins, Maceió - AL, 57.072-970
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO (PROPEP)
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Folha de Aprovação

FLÁVIO YURI AQUINO DE OLIVEIRA

UM MODELO COMPUTACIONAL DE RECOMENDAÇÃO DE TRATAMENTO TERAPÊUTICO PERSONALIZADO PARA PACIENTE COM QUADRO SÉPTICO

A COMPUTATIONAL MODEL FOR RECOMMENDING PERSONALIZED
THERAPEUTIC TREATMENT FOR PATIENTS WITH SEPTIC

Dissertação submetida ao corpo docente do
Programa de Pós-Graduação em Informática
da Universidade Federal de Alagoas e
aprovada em 21 de dezembro de 2023.

Banca Examinadora:

Documento assinado digitalmente

 **RAFAEL DE AMORIM SILVA**
Data: 24/12/2023 14:21:34-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. RAFAEL DE AMORIM SILVA
UFAL – Instituto de Computação
Orientador

Documento assinado digitalmente

 **BRUNO ALMEIDA PIMENTEL**
Data: 22/12/2023 18:45:02-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. BRUNO ALMEIDA PIMENTEL
UFAL – Instituto de Computação
Examinador Interno

Documento assinado digitalmente

 **DIEGO DERMEVAL MEDEIROS DA CUNHA MATOS**
Data: 22/12/2023 14:22:59-0300
Verifique em <https://validar.iti.gov.br>

**Prof. Dr. DIEGO DERMEVAL MEDEIROS DA
CUNHA MATOS**
FAMED – Faculdade de Medicina/UFAL
Examinador Interno

Documento assinado digitalmente

 **ALMIR PEREIRA GUIMARAES**
Data: 21/12/2023 17:38:25-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. ALMIR PEREIRA GUIMARAES
UFAL – Instituto de Computação
Examinador Externo

Assinado por: **FILIFE MONTEZ COELHO MADEIRA**
Num. de Identificação: 07660763
Data: 2023.12.21 18:43:55+00'00'

**Prof. Dr. FILIFE MONTEZ COELHO
MADEIRA**

Instituto Politécnico de Santarém Escola Superior de
Gestão e Tecnologia de Santarém/
Santarém-PORTUGAL
Examinador Externo

AGRADECIMENTOS

Primeiramente, agradeço a Deus, meu Aba. Por sua infinita graça e bondade sobre a minha vida, hoje e todos os dias de minha vida, e o sustento no caminho trilhado até este momento. Sem Ele, nada seria possível.

Agradeço profundamente ao meu orientador, Professor Dr. Rafael de Amorim Silva, pela parceria, companheirismo, direcionamento, conselhos, dedicação, sobretudo, pela paciência em vários momentos nesse percurso. Quero que fique registrado a imensa admiração que tenho pelo senhor.

Ao Professor Dr. Bruno Almeida Pimentel pela ajuda e direcionamento sempre que precisei e não foram poucas as vezes. Por nortear em vários momentos de dúvidas pelo qual caminho seguir e disponibilizar seu vasto conhecimento comigo.

Ao Professor Dr. Diego Dermeval M. da C. Matos, membro da banca, que compartilhou ricas contribuições e observações para enriquecer o trabalho. Muito obrigado!

Agradeço imensamente ao Mestre Phydell Pimentel Carvalho, que, mesmo não o conhecendo pessoalmente, me ajudou bastante com trabalho, principalmente pela cessão de dados para que pudesse ser aplicado nessa pesquisa. Sempre muito gentil e solícito. Isso tornou possível esse trabalho e gostaria que pudéssemos contribuir ainda mais com a academia e com a sociedade.

Não posso deixar de agradecer a minha esposa, Sarah, por todo apoio recebido ao longo desta caminhada. Pela parceria, paciência, compreensão nos momentos de ausência, e principalmente, por todo cuidado recebido nesse período, principalmente nessa reta final. Isso foi fundamental e fez toda a diferença.

Aos meus pais, por serem minha base, meu referencial e por me incentivarem a ser melhor. Pelo amor, carinho e todo ensinamento. Sou parte da história de vocês, mas vocês fazem parte de toda a minha história. Não a nada que eu diga ou faça que possa expressar o quão grato sou a vocês.

Eu jamais poderia deixar de agradecer a minha querida irmã, por tanto incentivo, cuidado, ajuda e carinho. Agradeço por estar ao meu lado sempre e incondicionalmente! Obrigado por ser uma amiga incrível!

“Porque dEle e por Ele, e para Ele, são todas as coisas; glória, pois, a Ele eternamente. Amém.” Paulo, Romanos 11:36

RESUMO

A Informática na Saúde provê recursos para a geração, manutenção, e armazenamento dos dados vitais dos pacientes através de prontuários construídos a partir de sistemas. Este estudo aborda a lacuna identificada na literatura quanto ao suporte técnico aos médicos, especificamente para o fornecimento de sugestões medicamentosas para pacientes com quadro séptico de maneira automática e assistida pelo *Machine Learning*. Dentro da área da Saúde, a SEPSE se apresenta como um quadro clínico em que o tempo de resposta para esta condição é vital para o paciente. Se não tratada de maneira eficiente e eficaz, o paciente pode evoluir a óbito em menos de 24h. Neste contexto, o presente trabalho tem como ideia principal proporcionar um modelo de *Machine Learning* para prever prescrição terapêutica de maneira automática para auxiliar o médico em sua tomada de decisão e personalizada para o paciente. Foram encontrados resultados satisfatórios nos modelos testados para a sugestão, tornando-se um caminho promissor para a mitigação de óbitos decorrentes de SEPSE.

Palavras Chave: SEPSE; *Machine Learning*; Prescrição Terapêutica.

ABSTRACT

Health Informatics provides resources for the generation, maintenance, and storage of necessary patient data through medical records built from systems. This study addresses the identified gap in the literature regarding technical support for physicians, specifically for the automatic and Machine Learning-assisted provision of medication suggestions for patients with septic conditions. Within the Health area, SEPSIS presents itself as a clinical condition in which the response time for this condition is vital for the patient. If not treated efficiently and effectively, the patient may die in less than 24 hours. In this context, the main idea of this work is to provide a Machine Learning model to automatically predict therapeutics to assist the doctor in their decision-making and personalized for the patient. Satisfactory results were found in the tested models for the suggestion, becoming a promising path for mitigating deaths resulting from SEPSIS.

Keywords: SEPSIS; Machine Learning; Therapeutic Prescription.

LISTA DE FIGURAS

Figura 1 - Modelo de assistência ao paciente	36
Figura 2 - Desenvolvimento de sistemas por finalidade	38
Figura 3 - Esquema de Machine Learning para sugestão terapêutica.....	46

LISTA DE TABELA

Tabela 1 - Pontuação de avaliação de falha de órgão sequencial rápida (qSOFA)	20
Tabela 2 - Escala pontuação SOFA	20
Tabela 3 - Principais comorbidades apresentadas.....	23
Tabela 4 - Relação motivo de internação hospitalar	23
Tabela 5 - Resultados com Data Augmentation.....	48

LISTA DE ABREVIATURAS E SIGLAS

AVC	Acidente Vascular Cerebral
AVE	Acidente Vascular Encefálico
CVC	Cateter Venoso Central
DM	Diabetes Mellitus
DPOC	Doença Pulmonar Obstrutiva Crônica
DVA	Droga Vasoativa
HAS	Hipertensão Arterial Sistêmica
IA	Inteligência Artificial
ICC	Insuficiência Cardíaca Congestiva
IH	Infecções Hospitalares
IOT	Intubação Orotraqueal
IRAS	Infecções Relacionadas à Assistência em Saúde
IRC	Insuficiência Renal Crônica
ITU	Infecção do Trato Urinário
IV	Intravenosa
MEWS	Sistema de Alerta Precoce Modificado
ML	Machine Learning
qSOFA	Avaliação sequencial de falência de órgãos rápida
RAM	Reação Adversa a Medicamentos
RNC	Rebaixamento do Nível de consciência
SIRS	Síndrome de Resposta Inflamatória Sistêmica
SOFA	avaliação sequencial de falência de órgãos
SVD	Sonda Vesical de Demora
TQT	Traqueostomia
VM	Ventilação Mecânica

SUMÁRIO

1 INTRODUÇÃO	12
1.1 Motivação	13
1.2 Contextualização	13
1.3 Justificativa	14
1.4 Problemática	14
1.5 Objetivos	15
1.5.1 Objetivo Geral	15
1.5.2 Objetivos Específicos	15
1.6 Proposta	16
1.7 Estrutura do trabalho	16
2 FUNDAMENTAÇÃO TEÓRICA	18
2.1 Sepsis	18
2.2 Fármacos	21
2.2.1 Quadro Farmacológico	21
2.2.2 Fatores De Risco	22
2.3 Inteligência Artificial	24
2.3.1 Inteligência Artificial na Medicina	24
2.4 Machine Learning	27
2.4.1 Aprendizado Supervisionado	29
2.4.1.1 KNN (classificado como supervisionado)	29
2.4.1.2 Regressão Logística	30
2.4.1.3 Árvore de Decisão / Decision Tree	30
2.4.1.4 Naive Bayes	31
2.4.1.5 Random Forest	31
2.4.1.6 Gradient Boosting	32
2.4.1.7 Neural Network	32
3 PROPOSTA	34
3.1 Descrição de Relato de Problema	34
3.1.1 Prontuário	34
3.2 Trabalhos relacionados	37
3.3 Relevância	39
4 EXPERIMENTO	40
4.1 Metodologia	40
4.2 Dados	41
4.3 Ferramentas	42
4.4 Cenário	42
4.5 Parâmetros	43
4.6 Roteiro	44
5 RESULTADOS E ANÁLISES	46
5.1 Descrição	46
5.2 Discussão	49
6 CONSIDERAÇÕES FINAIS	50
6.1 Ameaças à validação	50
6.2 Trabalhos futuros	51
REFERÊNCIAS	53
APÊNDICES	60

1 INTRODUÇÃO

A partir da segunda metade do século XX, mais especificamente nas duas primeiras décadas de nosso século, a Informática tem se feito presente em todos os campos do conhecimento, mudando a maneira como as pessoas lidam com a informação. A informática na saúde é vista como uma área emergente em sistemas de informação, tendo alta demanda de desenvolvimento em novas soluções, seja software ou hardware (Carvalho, 2017). Dentro da saúde é um campo de rápido incremento científico que fornece recursos para geração de dados, armazenamento, recuperação e uso da informação, controle, manutenção e arquivamento dos dados vitais dos pacientes para a resolução de problemas e tomada de decisão.

O suporte de registro em papel está evoluindo para o eletrônico, possibilitando inúmeras vantagens (Carvalho, 2017). Esse método de Prescrição Eletrônica existente em sistemas de Prontuário Eletrônico do Paciente (PEP) admite ser melhorado com uso de técnicas de Inteligência Artificial.

Podemos considerar que a ação de recuperação de informação tem uma relação com a Prescrição Médica, que se trata de uma afinidade terapêutica importante entre paciente e médico (Brunton; Lazo; Parker, 2007). Essa prescrição demonstra a capacidade diagnóstica e terapêutica do médico, por meio dos dados dispostos necessários, para dar andamento ao tratamento do paciente.

Considerando uma possível gravidade clínica do paciente, algumas prescrições podem ser bem complexas, gerando, conseqüentemente, uma construção que precisa ser bem minuciosa e precisa, criando ocasião para chances de erros em interações medicamentosas, dosagens e, ainda, devendo-se considerar patogenia específica do paciente. Interações medicamentosas são, frequentemente, a causa de geração de eventos adversos, tendo como resultado o aumento do número de internações hospitalares e o tempo de internação, assim como a geração de custos extras (Morimoto *et al.*, 2004).

Definir certas prescrições pode ser considerado algo muito complexo, sendo possível a afirmação de que uma prescrição médica não pode ter como base apenas regras simples e genéricas, nem considerar unicamente informações padrão a diferentes tipos fisiológicos de pacientes (Carvalho, 2017).

No campo da medicina, é uma realidade a concepção de grandes bancos de dados com informações sobre sintomas, tratamentos, diagnósticos, resultados de exames e curso das doenças para cada paciente. O bom uso desses dados pode fornecer ciência nova, como a relação entre algumas doenças e certos perfis socioculturais, profissionais, local de moradia e hábitos

peçoais. Informações que são respeitadas nos aspectos referentes às doenças. Essas semelhanças podem ser usadas para melhor compreensão das doenças e seus tratamentos (Carvalho, 2005).

Na área da saúde, o uso dos bancos de dados permite a recuperação de informação inserida nesse ambiente, visto que a finalidade desses bancos é proporcionar o acesso à informação neles armazenados. Assim, um algoritmo, para apoiar a prescrição, necessita considerar o tratamento dos dados de entrada de maneira dinâmica, moldado de acordo com cada novo caso que possa surgir e considerando o conhecimento médico para casos resolvidos antes. Mas para que seja possível usar a base de conhecimento médico é necessário modelar, desenvolver e avaliar um método que seja possível extrair os dados imprescindíveis para a constituição de uma proposta de tratamento eficiente e eficaz.

Portanto, este trabalho realizará o desenvolvimento de um algoritmo que será capaz de auxiliar um médico em sua tomada de decisão ao prescrever tratamento terapêutico medicamentoso para remissão de paciente em quadro séptico.

1.1 Motivação

A sepse constitui uma condição clínica que anualmente resulta no óbito de milhões de pessoas ao redor do mundo (WHO, [s. d.]). O tempo para diagnóstico e administração adequada de medicamentos é um fator crucial para evitar a evolução fatal do quadro do paciente (Van der Poll *et al.*, 2017). Nenhuma pessoa está isenta de vivenciar um quadro séptico durante o transcorrer de sua existência. Diante dessa possibilidade e consciente do impacto significativo que tal situação pode acarretar a vida das pessoas, decidimos realizar uma investigação a fim de contribuir para o corpo acadêmico e a sociedade como um todo.

1.2 Contextualização

Embora existam alguns estudos, consoante Begum *et al.* (2011), que abordem o uso da tecnologia para auxiliar na área da saúde, ao realizar um levantamento na literatura, identificamos uma lacuna específica no suporte aos médicos por meio de uma lista de sugestões medicamentosas para pacientes com quadro séptico. Encontramos trabalhos relacionados à predição da condição de SEPSE (Goh *et al.*, 2021), contudo, até o momento, não havia um recurso eficaz que oferecesse assistência substancial ao médico na tomada de decisões em um

intervalo de tempo reduzido, considerando as diversas particularidades apresentadas pelo paciente.

1.3 Justificativa

Ao integrar os conhecimentos de duas disciplinas distintas, tecnologia e saúde, vislumbramos a possibilidade de impactar positivamente milhões de vidas, tanto em escala nacional quanto global. Ao abordarmos uma condição de saúde capaz de causar óbito em curto período, justificamos esta pesquisa na perspectiva de que, ao reduzirmos em 1% o número de mortes, estamos potencialmente salvando milhares de vidas. Além disso, contribuímos para a geração de novo conhecimento acadêmico destinado aos profissionais futuros.

1.4 Problemática

Anualmente, conforme a World Health Organization (WHO, [s. d.]), cerca de 49 milhões de pacientes são diagnosticados com SEPSE ao redor do mundo. Com um número aproximado de 11 milhões de mortes por ano, o quadro séptico é maior causa de mortes em hospitais por todo mundo, chegando a quase 25%. Nos Estados Unidos da América, por exemplo, essa porcentagem chega a quase 50% das mortes em hospitais (Rudd *et al.*, 2020).

Pesquisas conduzidas no Brasil durante o período de 2001 a 2003 revelaram índices de mortalidade oscilando entre 34,4% e 34,7% para pacientes diagnosticados com sepse grave e de 52,2% a 65,3% para aqueles que enfrentavam quadros de choque séptico (Lobo *et al.*, 2019). Em uma pesquisa conduzida pelo Instituto Latino Americano da Sepse (ILAS) para avaliar a prevalência e a letalidade associadas à sepse grave e choque séptico em 2015, foi observada uma taxa global de letalidade de 55% (Machado *et al.*, 2017).

A distribuição das unidades de terapia intensiva (UTIs) no Brasil foi obtida através da análise do Censo de UTIs da AMIB, que registrava 1.813 UTIs (com um total de 20.731 leitos para pacientes adultos) em todas as regiões do país. A mortalidade na Região Sudeste foi de 51,2%, demonstrando ser menor em comparação com as demais regiões (Centro-Oeste com 70%, Nordeste com 58,3%, Sul com 57,8% e Norte com 57,4%). Além disso, a mortalidade em hospitais vinculados ao sistema público de saúde não diferiu significativamente daquela em hospitais privados (Machado *et al.*, 2017).

Em uma outra pesquisa mais recente (Carvalho *et al.*, 2021), porém de menor proporção, em um Hospital Universitário no interior de Sergipe, HUL (Hospital Universitário de Lagarto),

foram acompanhados 48 pacientes entre 2019 e 2020. Do total de pacientes avaliados, 45,8% (22) evoluíram a óbito.

Como em diversas outras área do conhecimento, na saúde há um grande investimento em sistemas computacionais. Ainda que hajam diferenças entre o investimento nas diferentes regiões do País, o intuito é melhorar o atendimento do paciente e mitigar a taxa de mortalidade (Scheffer *et al.*, 2015).

O problema de pesquisa encontrado é como auxiliar os médicos em sugestão terapêutica, levando em consideração a individualidade de cada paciente no menor tempo possível, visto que o tempo para tratamento é fator de extrema relevância para pacientes em quadro séptico.

Após observar estes cenários, questionamo-nos e nos desafiamos a criar uma ferramenta que pudesse auxiliar os médicos a prescrever tratamentos terapêuticos personalizados e eficientes para paciente de quadro séptico.

1.5 Objetivos

1.5.1 Objetivo Geral

O presente trabalho tem como objetivo a recomendação terapêutica para pacientes com quadro Séptico, levando em consideração sua individualidade no que se refere a patogenias pretéritas que ele possa ter, aplicando o processo de *Machine Learning*. Para isso, coletamos dados de um Hospital Universitário situado no interior de Sergipe, visando uma contribuição segura para a academia, para os profissionais de saúde e para a sociedade de modo geral, a partir de dados legítimos. Conjecturando o melhor modelo para recomendação terapêutica para pacientes com quadro séptico, temos como objetivo técnico usar e comparar modelos de *Machine Learning*.

1.5.2 Objetivos Específicos

- 1) Coletar dados clínicos de pacientes;
- 2) Identificar modelos de *Machine Learning*;
- 3) Elaborar algoritmo baseado em *Machine Learning* para sugestão medicamentosa;
- 4) Verificar aplicação da *Machine Learning* a partir dos dados coletados;
- 5) Analisar as etapas que o dado percorre até seu resultado.

1.6 Proposta

Propomos um modelo de *Machine Learning* que atue como um consultor médico virtual que, ao analisar os dados de saúde do paciente, possa sugerir antibióticos para tratar quadro séptico. Queremos que, por conta de sua capacidade de considerar variáveis específicas do paciente, o modelo possa se tornar uma ferramenta promissora no auxílio à decisão médica, buscando imitar o processo cognitivo humano ao identificar padrões complexos e correlações entre as características do paciente e tratamentos mais eficazes. Isso inclui nuances clínicas, histórico de tratamentos e a possibilidade de resistência bacteriana, personalizando as sugestões. No entanto, sua eficácia depende de atualizações constantes com informações médicas relevantes e validações rigorosas para garantir sua confiabilidade no apoio aos médicos na tomada de decisões clínicas.

1.7 Estrutura do trabalho

A estrutura do presente trabalho desenvolvido objetiva proporcionar uma leitura que facilite a compreensão e o entendimento do tema abordado. Apresentaremos o capítulo 2 – Fundamentação Teórica –, que conceituará assuntos necessários ao entendimento do objeto, proporcionando informações acerca da temática em questão. Dentro da “Fundamentação Teórica” o leitor poderá entender a conceituação de SEPSE e como ela ocorre, os perigos inerentes e sua alta letalidade. Falaremos também sobre “Fármacos”, como funcionam as interações medicamentosas, efeitos e aplicação para cada paciente.

Posteriormente, no capítulo 3, abordaremos o “Relato de Problema”, que demonstrará a natureza do problema que encontramos, o estado atual na literatura e a necessidade de ser investigado e resolvido. Veremos, no capítulo 4, a nossa “Proposta”, que conduzirá o leitor a entender como a Inteligência Artificial tem atuado na área da medicina. Adicionalmente, falaremos sobre *Machine Learning*, abordando conceitos, aplicações, funcionamento; e então entraremos em mais detalhes de algumas técnicas de *Machine Learning* que foram aplicadas no presente trabalho, para entendermos como ela pode auxiliar os médicos a prescreverem terapia medicamentosa para pacientes com SEPSE.

No capítulo 5 elucidaremos o “Experimento”, discorrendo sobre os dados usados nesta pesquisa, as ferramentas usadas nas aplicações das técnicas de *Machine Learning* para a criação de sugestão terapêutica. Também serão abordados o Cenário pensado, Metodologia aplicada e

os Parâmetros para execução. De igual modo, veremos o roteiro de construção do Experimento e o passo a passo para que ele gere a lista de sugestão terapêutica para o médico.

Após essa leitura será posto para o leitor os “Resultados e Análises”, que descreverá como alcançamos os resultados presentes nesta pesquisa, da mesma maneira, explanaremos sobre os achados da pesquisa, verificando quais foram mais relevantes e se mostraram mais promissores dentro do escopo da pesquisa. Por fim, abordaremos as “Considerações finais” referentes à pesquisa, expondo nuances e caminhos percorridos até aqui. Registraremos ameaças que podem validar este trabalho, bem como trabalhos futuros e perspectivas para novas contribuições para a academia e a sociedade.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção será abordada a teoria sobre a qual está fundamentada este trabalho e conceitos que nortearão o entendimento a respeito do tema. A seção será distribuída em objetos que tratam de Saúde e Fármacos para propor uma melhor compreensão e entendimento sobre o assunto.

Nossa intenção aqui não é trazer discussões aprofundadas sobre o conceito de saúde. Interessamo-nos pelos diálogos que a Saúde e os Fármacos possuem e suas interações frente ao trabalho de remissão de pacientes.

2.1 Sepses

A resistência de microrganismos (especialmente bactérias) a diversos agentes antimicrobianos é objeto que exibe forte vínculo com as Infecções Hospitalares (IH). A obtenção de resistência acontece de forma mais aguda no interior dos serviços de saúde do que na comunidade. Esse fato se deve à ampla centralização de indivíduos suscetíveis, à facilitação da difusão pelos procedimentos de trabalho e à decorrência ecológica do uso de antimicrobianos (Lipsitch; Samore, 2002).

As IH (Infecções Hospitalares), também designadas como “Infecções Relacionadas à Assistência em Saúde” (IRAS), concebem um grave risco à saúde de populações em todo o planeta. Entretanto, em pouquíssimas ocasiões elas são abordadas com a devida prioridade em países em desenvolvimento, onde se reúne a maior parte dos episódios e, por subsecutiva, da mortalidade e dos efeitos creditados a IH/IRAS (Pittet *et al.*, 2008).

Em uma revisão sistemática achada em Allegranzi *et al.* (2011), foram identificadas taxas de IH/IRAS duas a vinte vezes maiores nos países de “média e baixa renda”, se comparados àqueles situados nos Estados Unidos e Europa Ocidental. A mesma revisão quantificou taxas de IH/IRAS em unidades neonatais brasileiras, identificando casos nove vezes superior ao de serviços norte-americanos de propriedade semelhante. Os autores concluem sua apreciação corroborando que falta, aos países em desenvolvimento, uma política sólida de desenvolvimento de recursos humanos, retaguarda laboratorial e expertise em análise de dados epidemiológicos.

A principal causa de morte em hospitais, consoante Burdick *et al.* (2020), está ligada diretamente a SEPSE. Que é uma condição potencialmente fatal que surge quando o organismo tenta dar ao corpo uma resposta para uma infecção, mas essa resposta acaba danificando os

próprios tecidos e órgãos do paciente (Chicco; Jurman, 2020). É uma resposta fora de controle contra a infecção, que pode levar à falência de órgãos ou até mesmo à morte. Quanto mais cedo a previsão de que um paciente pode desenvolver um quadro de SEPSE, maior as chances de vida para ele (Van der Poll *et al.*, 2017). Porém, não é tão fácil prever essa possibilidade pois muitos sintomas e sinais apresentados pela SEPSE assemelham-se a outras condições menos críticas.

A natureza desigual de possíveis insultos infecciosos e a variedade de respostas do paciente, muitas vezes, faz com que a identificação da SEPSE seja prolixa, tornando seu reconhecimento tardio, não sendo uma condição uniforme (Chicco, Jurman, 2020; Leligdowicz, Matthay, 2019). Nos Estados Unidos, SEPSE é a principal causa de mortes em hospitais, respondendo por quase metade de todas as mortes hospitalares. Além disso, é uma das condições mais caras para diagnosticar e tratar, estima-se que seu custo em cuidados gire em torno de US\$ 24 bilhões por ano apenas nos Estados Unidos (Rudd *et al.*, 2020). Há estimativas, de acordo com a Organização Mundial da Saúde (OMS), que sugerem que mais de 49 milhões de pacientes são diagnosticados com SEPSES ao redor do mundo e 11 milhões de mortes, anualmente (WHO, [s. d.]). No Brasil a situação de óbitos é ainda maior, segundo uma pesquisa feita em 2015 (Machado *et al.*, 2017), a taxa de mortalidade no país é de 55% dos pacientes diagnosticados com SEPSE. Um estudo menor, realizado em um Hospital Universitário do interior do estado de Sergipe (Carvalho *et al.*, 2021), revela uma taxa de mortalidade de 45,8% entre os pacientes internados com SEPSE. A condição de SEPSE é altamente sensível ao tempo, por este motivo é crucial a previsão precoce, para que haja prevenção da mortalidade do paciente hospitalizado. O diagnóstico e o tratamento de maneira precoce podem atenuar esse quadro. Baseado nas diretrizes médicas internacionais, é recomendado que a ressuscitação precoce de fluidos comece nas primeiras 3 horas para poder estabilizar a hipoperfusão do tecido¹ induzido pela SEPSE. É recomendado que a administração intravenosa de antimicrobianos comece o mais cedo possível, dentro de 1 hora de SEPSE (Rhodes *et al.*, 2017).

É importante notar que a mortalidade em decorrência de SEPSE aumenta significativamente a cada hora de atraso na administração de antimicrobianos (Ferrer *et al.*, 2014). A identificação precoce de SEPSE pode ser algo desafiador, já que o manejo da SEPSE é baseado em abordagem de gestão padronizada e as restrições operacionais na prestação de cuidados podem levar a taxas de mortalidade altas. Um exemplo de como o tempo pode ser crucial é demonstrado pela demora de comunicação entre médicos, enfermeiros e

¹ Contribui para o desenvolvimento de disfunção celular, a qual pode progredir para disfunção de múltiplos órgãos e óbito.

farmacêuticos, que eleva o tempo de resposta, agravando a manutenção da SEPSE. Dessa maneira, a previsão precoce pode dar aos médicos um tempo adicional e decisivo para planejar e executar planos de tratamento para a SEPSE (Goh *et al.*, 2021).

Normalmente a prática clínica utiliza sistemas padronizados de pontuação, como a Síndrome de Resposta Inflamatória Sistêmica (SIRS), que apesar do uso prévio e com base em alguns critérios foi constatado como não específico (Shankar-Hari *et al.*, 2016), o rastreamento ainda continua sendo usado em muitos ambientes clínicos. Seus critérios são a taquicardia (configurada por batimentos acima de 90 por minuto), a febre ou hipotermia (temperaturas acima de 38°C ou abaixo de 36°C) e as células de defesa (presença no sangue menor que 1,200/mm³, maior que 4,000/mm³), avaliação sequencial de falência de órgãos (SOFA) baseada na coleta de sangue e análise detalhada laboratorial (Vincent *et al.*, 1996). O quickSOFA (qSOFA) é fundamentada no SOFA, mas são colhidos mais rapidamente, tornando possível a triagem do paciente e redirecionando para outros exames, caso necessário, e o Sistema de Alerta Precoce Modificado (MEWS) (Subbe *et al.*, 2006) para prever SEPSE ou mortalidade por infecção.

Os médicos são obrigados a normalizar o lactato (processo da transformação da glicose em energia para as células quando não há quantidade suficiente de oxigênio), que é uma importante substância, considerada sinalizadora para o Sistema Nervoso Central, biomarcador de alterações nervosas e de hipoperfusão tecidual², com níveis elevados de lactato para a ressuscitação inicial. Quando um médico suspeita de SEPSE, deve solicitar um ou mais exames de cultura microbiológica de rotina e exames de lactato antes de iniciar a terapia antimicrobiana para atenuar o atraso no início do antimicrobianos.

Tabela 1 - Pontuação de avaliação de falha de órgão sequencial rápida (qSOFA)

Sintomas avaliados	Pontuação
Taxa de respiração ≥ 22 /min	1
Mudança no estado mental	1
Pressão sistólica no sangue ≤ 100 mmHg	1

Fonte: Adaptado (Marik; Taeb, 2017)

Tabela 2 - Escala pontuação SOFA

Sistema de órgão	0	1	2	3	4
Respiratório, PO ₂ /FiO ₂ , mmHg (kPa)	≥ 400 (53.3)	< 400 (53.3)	< 300	< 200	< 100
			-40	(26.7) com suporte respiratório	(13.3) com suporte respiratório

² O lactato alto pode ocorrer por conta da SEPSE e choque séptico, que devido a produção de toxinas pelos microrganismos há uma diminuição da quantidade de oxigênio que chega aos tecidos.

Coagulação, plaquetas, × 10 ³ / mm ³	≥150	<150	<100	<50	<20
Fígado, bilirrubina, mg / dL	<1.2	1.2–1.9	2.0–5.9	6.0–11.9	>12.0
Cardiovascular	MAP	MAP	Dopamina <5 ou dobutamina (qualquer dose)	Dopamina 5,1-15 ou epinefrina ≤0,1 ou norepinefrina ≤0,1	Dopamina > 15 ou epinefrina > 0,1 ou norepinefrina
Sistema nervoso central, Escala de Glasgow	5	13-14	10/dez	06/set	<6
Renal, creatinina, mg/ dL. Produção de urina, mL/d	1.2	1.2-1.9	2.0-3.4	3.5-4.9 <500	>5.0 <200

Fonte: Adaptado (Marik; Taeb, 2017)

A prática médica e estudos anteriores mostram que a detecção precoce da SEPSE é desafiadora (Goh *et al.*, 2021) e os pacientes com SEPSE podem evoluir negativamente de maneira muito rápida, sendo assim, cada minuto conta no diagnóstico de SEPSE e na administração de antimicrobianos.

2.2 Fármacos

Buscando proporcionar melhor entendimento do que é proposto neste trabalho, é importante compreender conceitos sobre fármacos, que ajudarão a compreender detalhes de elementos/cenários que devem ser considerados em um paciente para ajudar o profissional de saúde com o direcionamento de melhor sugestão medicamentosa, objetivando maior nível de efetividade para a cura do paciente.

2.2.1 Quadro Farmacológico

A terapia medicamentosa é um fator que possui várias dimensões, que atinge diretamente os pacientes e que pode gerar problemas consideráveis como: efeitos adversos recorrentes e falha no tratamento, por exemplo (Oliveira, 2020). A realidade comum é o uso concomitante de dois ou mais medicamentos. Essa influência mútua medicamentosa pode acontecer entre medicamento-alimento, medicamento-medicamento e medicamento-droga, permitindo que haja efeitos adversos ou deficiência no tratamento. Quando aplicados de forma concomitante a um paciente, eles podem admitir uma maneira independente ou desempenhar interação entre eles, com a redução ou acréscimo de implicação terapêutica ou tóxica. Muito

embora exista a possibilidade dos danos, os medicamentos também podem ser eficazes e favoráveis, tudo depende de fatores como: as circunstâncias de seu uso, o quadro do paciente e o medicamento em si (Oliveira, 2020).

As reações adversas a medicamentos (RAM) são um dos pontos que coopera com a não eficiência de tratamentos. Segundo a Anvisa (2020), a definição de RAM é: “qualquer resposta prejudicial ou indesejável, não intencional a um medicamento, que ocorre nas doses usualmente empregadas”.

Para que possam atuar de maneira eficaz, os profissionais de saúde necessitam conhecer o quadro clínico atual e histórico de seus pacientes. Sendo assim, a detecção de conduta não favorável está diretamente ligada ao impacto de desfechos terapêuticos (Cruz; Macedo, 2016).

2.2.2 Fatores De Risco

O investimento em educação continuada, identificação precoce, infraestrutura, detecção e início rápido do tratamento é fundamentalmente importante devido aos problemas de saúde pública a nível global que decorrem da SEPSE (Carvalho *et al.*, 2021). Há, no Brasil, uma dificuldade a mais, a depender dos recursos disponíveis no serviço público para qualificar e tratar, da disponibilidade de antibióticos de largo espectro e de profissionais de saúde com insuficiente ou nenhuma formação formal na lida de pacientes com SEPSE.

Assim sendo, caso haja um quadro de disfunção orgânica sem outra elucidação aceitável e com foco infeccioso cabível, o diagnóstico de SEPSE é efetivado e o pacote de tratamento iniciado prontamente após a identificação (Cruz; Macedo, 2016). Devido à incerteza do diagnóstico em atmosfera hospitalar, realizada de maneira empírica, a terapia para SEPSE bacteriana leva ao uso abusivo de antibióticos e ao risco de resistência bacteriana. Para reduzir a resistência bacteriana, uma das táticas mais eficazes é suspender antimicrobianos em casos de infecções não confirmadas ou pouco prováveis, garantindo a segurança do paciente (Carvalho *et al.*, 2021).

O número de pacientes imunossuprimidos, idosos, portadores de doenças crônicas facilitam o desenvolvimento de infecções graves, caracterizando, assim, como fatores de risco (Instituto Latino-Americano de Sepse, 2015).

De acordo com uma pesquisa de Carvalho *et al.* (2021), todos os diagnósticos foram feitos de maneira empírica e as comorbidades mais prevalentes são a hipertensão arterial, a Diabetes Mellitus, a cardiopatia e a disfunção renal, conforme tabela a seguir:

Tabela 3 - Principais comorbidades apresentadas

	PACIENTES	PORCENTAGEM
Hipertensão arterial	31	23,86%
Diabetes Mellitus	28	21,54%
Insuficiência Cardíaca/ Cardiopatia	19	14,61%
Disfunção Renal	14	10,77%
Problemas pulmonares (DPOC, ex-tabagista, asma)	7	5,38%
Doenças psiquiátricas (Alzheimer, Síndrome demencial, Transtorno de humor)	7	5,38%
AVE	5	3,84%
Neoplasia	3	2,31%
Disfunção hepática	3	2,31%
Ausência de comorbidade	1	0,77%
Outros: Síndrome de imobilidade, hipotireoidismo, dislipidemia, úlcera gástrica, convulsão, febre reumática, ex-etilista, deficiência visual, hemiplegia, anemia, gastrite, encefalopatia	12	9,23%

Fonte: Carvalho *et al.* (2021)

Como complicações mais prevalentes, justificando motivo inicial de internação hospitalar, foram destacados os problemas de cunho respiratórios, como pneumonia e insuficiência respiratória, como visto na tabela abaixo:

Tabela 4 - Relação motivo de internação hospitalar

	FREQUÊNCIA	PORCENTAGEM
Problemas Respiratórios	14	29,2%
Sepse/septicemias	6	12,5%
RNC – Rebaixamento do Nível de Consciência	6	12,5%
Cetoacidose diabética	3	6,3%
AVE – Acidente Vascular Encefálico	3	6,3%
Desconforto abdominal, Abdome agudo obstrutivo, diarreia	3	6,3%
Insuficiência renal ou Infecção do trato urinário	3	6,3%
Lesão cutânea, abscesso cutâneo, celulite, furúnculo	3	6,3%
Insuficiência ou doença hepática	2	4,2%
Desconforto torácico	2	4,2%
Transtorno mental e comportamental	1	2,1%
Artrite piogênica, osteomielite	1	2,1%
Hemoptise	1	2,1%

Fonte: Carvalho *et al.* (2021)

Em pacientes de risco, um elevado apontador de suposição deve ser mantido. Em pacientes suspeitos, deve-se considerar histórico detalhado, exames físicos e laboratoriais realizados constantemente. Todos os dados são importantes para situar a apresentação e a origem da SEPSE. Demonstrando, assim, que outros fatores como patogenias preexistentes devem ser consideradas ao cuidar de um paciente com quadro clínico de SEPSE (Armstrong; Betzold; May, 2017).

2.3 Inteligência Artificial

Nesta seção abordaremos conceitos de IA que foram empregados nessa pesquisa e que nortearão o leitor no que diz respeito à Inteligência Artificial e sua aplicação, sobretudo, para área médica.

2.3.1 Inteligência Artificial na Medicina

A Inteligência Artificial, consoante Russell e Norvig (2013), é um ramo da ciência da computação, cujo alvo é fazer com que os computadores “pensem” ou se comportem de forma inteligente, similar à humana. Ela tem se mostrado uma ferramenta³ que pode ser aplicada nas mais diversas áreas que afetam diretamente a vida das pessoas.

A literatura organiza as abordagens da IA em duas extensões: processo de *pensamento/raciocínio* e *comportamento*. Sobre o processo de *pensamento/raciocínio* é definido como o esforço de tentar fazer com que os computadores pensem como seres humanos, algo como o aprendizado ao longo do tempo e, em decorrência desse aprendizado, sejam capazes de tomar decisões e de resolver problemas, podendo perceber o ambiente a sua volta, raciocinando e interagindo com ele. Por outro lado, o *comportamento* está intrinsecamente ligado à maneira como se espera que os computadores tenham um melhor desempenho em tarefas que hoje são desenvolvidas por humanos, tendo uma performance inteligente e superior aos realizados por estes.

Para o conhecimento, percepção, aprendizagem, solução de problemas, planejamento e raciocínio, a Inteligência Artificial precisa de algoritmos. Pode-se dizer que um dispositivo dotado de Inteligência Artificial tem a competência de deliberar pela melhor entre opções

³ Dr. Dario Gil, vice-presidente de IA da *IBM RESEARCH* defende que a IA é uma ferramenta a ser utilizada por humanos.

preestabelecidas, sendo estas opções preestabelecidas enriquecidas invariavelmente pelo próprio sistema. Uma propriedade da Inteligência Artificial é o “aprendizado” do sistema pelo aumento do banco de informações.

Um dos exploradores do estudo de Inteligência Artificial na área médica, Bleich (1969), expôs a possibilidade de influência do equilíbrio hidroeletrólítico de pacientes por meio de atuações de um sistema computadorizado. É a metodologia de se empregar sistemas computacionais mirando amparar e agilizar a administração dos serviços de saúde, cuidados clínicos, investigação médica, tratamento e treinamento (Hobbs, 2001). Esse cenário exige o bom emprego das tecnologias de computação, como a mineração de dados e de comunicações, para otimizar o processamento de informações em saúde em todas as fases, como a coleta, armazenamento, recuperação efetiva (no seu devido tempo e lugar) e exame, e auxílio à decisão para os administradores, médicos, pesquisadores e educadores na medicina. O uso da Inteligência Artificial na prática médica tem se tornado crescente. Está cada vez mais comum o uso de dispositivos em rotinas assistenciais. Estes dispositivos unidos ao paciente auxiliam no monitoramento de apontadores metabólicos como níveis de pressão arterial, taxas de glicemia e ritmo cardíaco.

Dispositivos para monitoramento cardíaco acoplados com desfibrilador subcutâneo podem provocar atuações automáticas, como, por exemplo, oferecer uma descarga elétrica, para que aconteça resposta à normalidade dos batimentos cardíacos (Bleich, 1969). No monitoramento de coeficientes glicêmicos sanguíneo podem provocar atuações automáticas de aplicação de insulina em circunstâncias que possam afetar a vida do paciente. A Inteligência Artificial tem sido empregada em sistemas de prontuários eletrônicos hospitalares e está tornando a dispensação de medicamentos mais segura. Estes softwares, ao realizarem intersecções da prescrição médica com dados do paciente, impedem interações ou doses inapropriadas, fazendo com que a prescrição medicamentosa seja segura.

O Watson Health, sistema computacional desenvolvido pela IBM ([s. d.]), está mudando positivamente o tratamento do câncer. Isto graças à admirável quantidade de dados médicos acrescidos dos mais importantes núcleos de pesquisa e tratamento de câncer no mundo (IBM, [s. d.]). Por meio do processamento destes dados, é possível alcançar a estimativa de risco e desenvolvimento de pacientes com câncer. O Watson Health sugere condutas associadas às probabilidades de sucesso baseado na análise do amplo número de casos armazenados com diagnósticos, tratamentos efetivados e resultados adquiridos.

O sistema de oncologia da IBM vem sendo usado em algumas instituições de saúde dos EUA. Mas não é unicamente no tratamento oncológico que o Watson Health está sendo

empregado. Na tentativa de identificar medicamentos eficazes por meio de experimentos clínicos em alterações genômicas conferidos com a literatura médica, foi desenvolvida a plataforma *Watson Genomics* para auxiliar os estudos. Existe, nos Estados Unidos, o projeto NCI-MATCH (National Cancer Institute, 2022a), programa de pesquisa em câncer realizado pelo Instituto Nacional de Câncer, independentemente do tipo de câncer, conhecido como MATCH, que se refere a um experimento clínico sobre o câncer. No presente e mais ainda no futuro, a medicina certamente necessitará ponderar a eficácia medicamentosa em condições específicas.

A Medicina de Precisão (*Precision Medicine Initiative*) está em andamento nos Estados Unidos, por meio de um programa que prevê colocar uma base de dados genéticos de um milhão de pessoas para aferir a eficácia de drogas (National Cancer Institute, 2022b). O conceito de Medicina de Precisão tem como base de estratégias a prevenção e tratamento que consideram a individualidade variável/personalizada de acordo com a anamnese e histórico do paciente. Esta apreciação na medicina não é nova, mas a perspectiva de se dedicar largamente esse conceito foi, de maneira drástica, aperfeiçoada pelo recente desenvolvimento de bancos de dados biológicos de grande escala (o sequenciamento do genoma humano). Para este tipo de trabalho, as plataformas computacionais são a base para analisar grandes conjuntos de dados.

O uso de tecnologia adotada pela área da saúde inclui próteses, máquinas, equipamentos para diagnóstico e intervenção, prontuários, comunicação, entre outros, inclusive a produção artificial de células humanas. E estes avanços tornaram-se expressivos com a introdução da informática e dispositivos cada vez mais sofisticados, que trazem como benefícios a rapidez no diagnóstico e tratamento de doenças (Bleich, 1969).

A fim de aperfeiçoar a qualidade de tomadas de decisões clínicas por parte de seus profissionais, a área de saúde mostra-se um terreno altamente propício para que a IA seja aplicada. E isto é possível devido a algumas situações que são favoráveis a isso (Bleich, 1969), como a fatura de dados crescentes, diversas possibilidades de aproveitamentos práticos e o rápido desenvolvimento de métodos de análise de *Big Data* (que ajudam a desvendar informações importantes e latentes). Estima-se que uso da IA permite a diminuição de custos e a ampliação da eficiência na gestão do processo clínico (Long *et al.*, 2017), além de evitar desperdício de tempo, reduzir o tempo de espera e aumentar a velocidade da prestação de serviços da saúde (Dyas *et al.*, 2015).

Em uma pesquisa realizada em 2015 (Miotto *et al.*, 2016) feita com um programa de IA, *Deep Patient*, com dados de aproximadamente 700.000 pacientes, quando aplicado aos novos dados resultaram predição de doenças que superaram em muito as expectativas (Knight, 2017).

Sem qualquer instrução especial, descobriram-se padrões ocultos nos dados e pareceu advertir a disposição dos sujeitos a desenvolver um vasto tipo de doenças, por exemplo, o cancro de fígado. Surpreendentemente (para os programadores), foi capaz de antecipar doenças psiquiátricas, como a esquizofrenia, uma vez que esse distúrbio é de difícil diagnóstico. A investigação foi descontinuada, tendo em vista que os investigadores não entenderam o padrão de acerto do programa, portanto, não conseguiriam elucidar quando eventualmente errasse (Knight, 2017).

A IA pode ser aplicada a vários tipos de dados de saúde (Jiang *et al.*, 2017):

Não estruturados: incluem o *deep learning*, constam no processamento de linguagem natural de dados desestruturados.

Estruturados: incluem a técnica do *machine learning*, como, por exemplo, os sistemas de redes neurais.

Na prática clínica, a IA tem achado espaço para produzir implicações bastante concretas no apoio à efetivação de juízos de prognose, sugestão de diagnósticos e prescrição de tratamentos no âmbito de díspares especialidades médicas (Topol, 2019).

2.4 Machine Learning

O *Machine learning* é um algoritmo computacional (Ertel, 2017) de análise de dados para a constituição de modelos analíticos, representando um dos principais pilares da Inteligência Artificial. Esta programação se alicerça no conceito de que o sistema computacional pode tomar decisões, aprender com dados e identificar padrões com o mínimo de intervenção humana. A “máquina” pode ser ator principal no seu aprendizado e capaz de se adaptar sem interferência humana quando exposta a novos dados.

Os algoritmos de *Machine Learning* operam ao buscar aprender com extensos conjuntos de dados, independentemente de sua estrutura, visando identificar padrões que possam ser aplicados tanto em classificações quanto em regressões. Nesse modelo, a estatística é bastante empregada, sendo uma divisão da matemática que emprega teorias probabilísticas para a esclarecimento de eventos, de estudos e de experimentos (Carrilho Junior, 2007). O seu alvo é obter, organizar e analisar dados, determinar suas correlações e extrair delas suas implicações para exposição e explicação do passado, assim como prever e organizar o futuro, por exemplo, seu uso na categorização de textos de acordo com o contexto.

Considere que o *Machine Learning* é uma abordagem da aprendizagem em que uma máquina, por meio de um método indutivo, constrói um classificador, aprendendo a partir de

um conjunto de exemplos pré-classificados (Feldman; Sanger, 2007). Isso constitui em aprender as características das categorias de uma classe de dados de treinamento pré-classificado na nomenclatura de aprendizagem e o seu método de aprendizagem é uma instância de aprendizagem supervisionada. A maior parte dos trabalhos atuais sobre classificação está concentrada na abordagem da aprendizagem de máquina, que necessita de apenas um conjunto de instâncias de treinamento, que precisa ser feito de maneira manual e não é muito difícil de produzir (Feldman; Sanger, 2007). Há várias abordagens do classificador de aprendizagem, alguns deles são criados especificamente para categorização, enquanto outros são variantes de algoritmos mais gerais.

Os objetivos e metas a serem alcançados são definidos na etapa da identificação do problema. Para isso, os dados necessitam ser pensados de maneira muito sólida pois seu resultado será usado nas etapas seguintes e o sucesso depende dos conjuntos de dados selecionados e das metodologias aqui aplicadas.

Tendo como finalidade preparar essas informações selecionadas para a fase de mineração de dados, aplicando o pré-processamento nos dados (Feldman; Sanger, 2007), é necessário o uso de vários métodos que possibilitam extração e integração, transformação, limpeza, seleção e redução de volume destes dados, mas de maneira que o sentido e o valor dos dados não sejam alterados.

A extração de conhecimento, que ocorre na fase de mineração de dados, é feita a partir de grandes volumes de dados, valendo-se das tecnologias derivadas dos “bancos de dados, inteligência artificial, aprendizado de máquina, redes neurais, estatística, reconhecimento de padrões, sistemas baseados em conhecimento, recuperação de informação, computação de alto desempenho e visualização de dados” (Morais; Ambrósio, 2007). Para esta finalidade, conforme Moraes e Ambrósio (2007), são utilizados algoritmos apropriados, que poderão ser aplicados reiteradas vezes para o melhor entendimento e objetivando responder os questionamentos levantados na identificação do problema.

A fase de pós-processamento baseia-se no exame de qualidade da grande quantidade de modelos descobertos na etapa de mineração de dados. Considerando este cenário, o conhecimento extraído necessita ser analisado e fundamentado em medidas de qualidade e desempenho, elegendo o que é importante, relevante ou interessante ao usuário. A partir de então, será analisado todo o procedimento que gerou o volume de informações. Caso seja detectado que não está aceitável/suficiente, é preciso que seja repetido todo o processo, para que novos dados sejam trabalhados de novas maneiras, na procura dos objetivos propostos no início (Morais; Ambrósio, 2007).

Existem algumas técnicas usadas em *Machine Learning* para trabalhar os dados. Abordaremos de maneira superficial o Aprendizado Não Supervisionado, pois, por causa da natureza do problema, a técnica de Aprendizado Supervisionado se mostra mais propícia a resolução. Com relação a isso, Paixão *et al.* (2022) aborda a diferença primordial entre os modelos de aprendizagem com supervisão e sem supervisão reside no método de treinamento. Nos casos de aprendizado não supervisionado, o modelo de aprendizado de máquina extrai as características dos dados e constrói uma representação sem ter conhecimento prévio sobre os rótulos associados a cada conjunto de dados.

Em outras palavras, ele identifica heurísticamente os padrões nas informações de classe. E essa ausência de supervisão para o algoritmo pode ser benéfica, permitindo que ele analise padrões previamente não considerados. Diferentemente da abordagem supervisionada, que versaremos a seguir.

2.4.1 Aprendizado Supervisionado

No aprendizado supervisionado, segundo Paixão *et al.* (2022), o modelo de aprendizado de máquina possui conhecimento dos rótulos atribuídos aos dados, o que significa que as amostras estão devidamente classificadas. O processo de treinamento é baseado na comparação entre os resultados obtidos pelo modelo e os rótulos pré-classificados. Esse procedimento é repetido até alcançar um nível mínimo de erro. São exemplos de classificação supervisionado: KNN, Regressão Logística, Árvore de Decisão, Naive Bayes, Random Forest, Gradient Boosting e Neural Network.

2.4.1.1 KNN (classificado como supervisionado)

Apesar de termos apresentado anteriormente o KNN como não supervisionado, o algoritmo de *K-Nearest Neighbors* (KNN) representa uma abordagem de aprendizado supervisionado. A metodologia de classificação se concentra na identificação de grupos de indivíduos que compartilham características semelhantes, seguida pelo agrupamento subsequente (*clustering*) (Rosa, 2003). Em linhas gerais, o conceito central envolve determinar o número (K) de exemplos rotulados mais próximos ao exemplo não classificado. Com base nos rótulos desses exemplos mais próximos, a decisão sobre a classe do exemplo não rotulado é tomada (Ferrero, 2009).

A escolha do valor de K (o número de vizinhos mais próximos considerados pelo algoritmo) é deixada para o usuário, sendo aconselhável optar por um número ímpar para evitar empates na classificação. Se duas classes, A e B, apresentam vários pontos em seus domínios, quando um ponto desconhecido x é considerado, a classificação desse ponto é determinada pela quantidade de pontos cujas distâncias são as menores possíveis em relação às classes A e B (Rosa, 2003).

2.4.1.2 Regressão Logística

A Regressão Logística, de acordo com Belfiore (2015), é um método estatístico que visa criar um modelo a partir de um conjunto de dados, permitindo a previsão dos valores de uma variável categórica, frequentemente binária. Esse modelo é gerado a partir de variáveis independentes contínuas e/ou binárias.

Com base nesse modelo, é viável calcular ou antever a probabilidade de um evento acontecer, considerando uma observação aleatória. Esse método permite:

- Modelar a probabilidade de um evento conforme os valores das variáveis independentes, sejam elas categóricas ou contínuas;
- Prever o impacto do conjunto de variáveis na variável categórica dependente;
- Estimar a probabilidade de um evento ocorrer em comparação com a probabilidade de não ocorrer para uma observação escolhida aleatoriamente;
- Classificar observações ao estimar a probabilidade de uma observação pertencer a uma categoria específica.

2.4.1.3 Árvore de Decisão / Decision Tree

As Árvores de Decisão são reconhecidas como uma das maneiras mais simples de desenvolver algoritmos de aprendizado (Samuel, 1959; Mitchell, 1997). Os modelos de regressão e de classificação baseados em árvores de decisão podem ser utilizados como uma opção viável ou como complemento a outros métodos estatísticos, como regressão, agrupamento e classificação de dados univariados e multivariados. A simplicidade na construção e interpretação desses modelos figura como um dos atrativos distintivos desse algoritmo (Samuel, 1959; Mitchell, 1997).

Os métodos das árvores de decisão empregam regras para dividir os dados, por exemplo, baseiam-se em regras binárias para a partição dos dados (Breiman, 2001). Em modelos

fundamentados em uma estrutura de árvore, o primeiro nó é denominado raiz e, por convenção, encontra-se no topo da árvore. A partir da raiz, são gerados ramos sucessivos que se conectam a outros nós, denominados nós internos e essas conexões persistem até alcançar um nó terminal, também conhecido como folha.

2.4.1.4 Naive Bayes

Conforme mencionado por Panda e Patra (2008), o algoritmo de Naive Bayes é uma extensão do Teorema de Bayes operando sob a premissa da independência entre atributos. A facilidade de implementação e a ausência da necessidade de esquemas complexos de estimação de parâmetros iterativos tornam o modelo prontamente aplicável a conjuntos extensos de dados. Panda e Patra (2008) destacam a robustez e a facilidade de interpretação do modelo, destacando que geralmente obtém bons resultados, embora não seja necessariamente o classificador ideal em todas as aplicações.

Em Amudha e Rauf (2011), encontramos uma definição mais detalhada, descrevendo o classificador Naive Bayes como um dos métodos de aprendizagem amplamente utilizado para prever as probabilidades de pertencimento de instâncias a uma classe específica. Este método presume a independência condicional de classe e se fundamenta no Teorema de Bayes.

2.4.1.5 Random Forest

O algoritmo *Random Forest* se configura como uma técnica de classificação em conjunto. Consoante Amudha e Rauf (2011), esse método envolve a construção de múltiplas árvores de decisão binárias, sendo que a predição da classe para cada nova instância é determinada pela maioria das predições de classe provenientes do conjunto de árvores. Notavelmente, neste método, os atributos escolhidos em cada nó interno e os registros selecionados para o crescimento de cada árvore diferem do procedimento padrão de conjunto.

De acordo com Kulkarni e Lowe (2016), a geração de cada árvore na *Random Forest* segue o seguinte processo: a partir de um conjunto de treinamento, um subconjunto aleatório é compilado e utilizado para a construção da árvore. Além disso, destaca-se que a taxa de erro da floresta pode ser avaliada por meio de uma métrica rápida, utilizando dados *Out-Of-Bag* (OOB). No entanto, é possível também inserir um conjunto de casos de teste na floresta para obter uma taxa de erro adicional.

2.4.1.6 Gradient Boosting

O método *Gradient Boosting* é caracterizado pela generalização para problemas tanto de classificação quanto a problemas de regressão, combinando as previsões de um conjunto de classificadores fracos, cuja taxa de erro é ligeiramente inferior a de uma classificação aleatória (Hastie; Tibshirani; Friedman, 2008). Seu objetivo principal, dada uma função perda e um algoritmo base, é identificar um modelo aditivo que minimize a função perda (Kuhn; Johnson, 2013).

Na forma mais simples do método *Gradient Boosting*, ele é tipicamente inicializado por um modelo associado ao melhor palpite para a resposta de interesse como a média. Os resíduos são então calculados e um algoritmo base é aplicado a esses resíduos, que representam a resposta de interesse. Este algoritmo base ajusta um modelo que minimiza uma determinada função perda. Esse modelo é adicionado ao anterior para atualizar o valor predito e esse processo continua até que um número pré-especificado de iterações seja alcançado (Kuhn; Johnson, 2013).

Embora o método *Gradient Boosting* possa ser aplicado a vários algoritmos, a árvore de decisão é frequentemente escolhida como algoritmo base devido à sua capacidade de ser usada como um classificador fraco, construindo uma árvore com poucas divisões (profundidade reduzida) (Kuhn; Johnson, 2013). Nesse contexto, um *Gradient Boosting* básico apresenta dois hiperparâmetros: a profundidade da árvore, relacionada ao número de divisões em cada árvore, e o número de iterações, “m”, diretamente relacionado ao número de árvores no modelo *boosting* final, ambos podendo ser otimizados por validação cruzada k-fold (James *et al.*, 2014; Kuhn, Johnson, 2013).

2.4.1.7 Neural Network

As redes neurais constituem um método de aprendizado de máquina amplamente aplicado em diversas áreas, como reconhecimento de voz, reconhecimento de padrões, diagnósticos médicos e detecção de fraudes (Abiodun *et al.*, 2018).

Essas redes fundamentam-se na utilização de muitos elementos de processamento interconectados, denominados neurônios. O termo “neural” é adotado devido à semelhança funcional entre uma rede neural e o cérebro humano, ambos capazes de aprender a realizar tarefas por meio de exemplos (Abiodun *et al.*, 2018).

A estrutura de uma rede neural consiste na organização dos neurônios em camadas interconectadas. Geralmente, há uma camada de entrada que contém a informação a ser processada pela rede, seguida por camadas ocultas responsáveis por extrair informações relevantes. Por fim, existe uma camada de saída que representa o resultado da rede como um todo. Normalmente, um neurônio em uma camada está conectado a todos os neurônios da camada anterior.

Com o intuito de evitar viés na avaliação de uma rede neural, os dados são divididos em três categorias: dados de treinamento, empregados no aprendizado dos parâmetros da rede; dados de validação, utilizados para comparar diferentes configurações durante o treinamento (como variações no número de camadas ocultas); e dados de teste, empregados para a avaliação final da rede escolhida com base nos dados de validação (Ripley, 1996).

3 PROPOSTA

Ao interagir com os dados de saúde do paciente, o modelo de *Machine Learning* age como um consultor médico virtual, desta maneira, sugerindo ao médico uma listagem de antibióticos para tratamento de quadro séptico de seu paciente. A capacidade de considerar variáveis específicas do paciente torna esse modelo em uma ferramenta promissora no suporte à decisão médica, embora ainda precise ser continuamente aprimorado e validado para garantir a confiabilidade e eficácia clínica.

O modelo busca imitar o processo cognitivo, identificando padrões complexos e as correlações entre as características do paciente e os antibióticos, buscando tratamento mais eficiente para infecções sépticas. Essas considerações devem refletir características clínicas sutis do paciente, respostas anteriores de outros tratamentos e até mesmo a probabilidade de resistência bacteriana, dessa maneira, personalizando e adaptando a sugestão ao quadro específico de cada paciente.

Entretanto, a eficácia de modelo depende de constantes atualizações com informações clínicas relevantes. A integração contínua de novos conhecimentos médicos e validação rigorosa das sugestões são fundamentais para assegurar que o modelo seja uma ferramenta confiável no apoio à tomada de decisões para auxílio aos médicos.

3.1 Descrição de Relato de Problema

Nesta seção abordaremos como funciona o uso de dados de prontuário visando à otimização para o melhor tratamento de um paciente com quadro de SEPSE.

3.1.1 Prontuário

Um importante ponto desta documentação é o prontuário do paciente, apontamento definido pelo Conselho Federal de Medicina (CFM), no Artigo 1º da Resolução de nº 1.638/2002, como sendo

[...] o documento único constituído de um conjunto de informações, sinais e imagens registradas, geradas a partir de fatos, acontecimentos e situações sobre a saúde do paciente e a assistência a ele prestada, de caráter legal, sigiloso e científico, que possibilita a comunicação entre membros da equipe multiprofissional e a continuidade da assistência prestada ao indivíduo (Brasil, 2002).

Trata-se de um documento multidisciplinar (Pinto, 2014) que acompanha o histórico do paciente e é bastante abrangente devido aos vários tipos de profissionais (como médicos, fisioterapeutas, enfermeiros) que minutam elementos e acessam seus teores ali registrados.

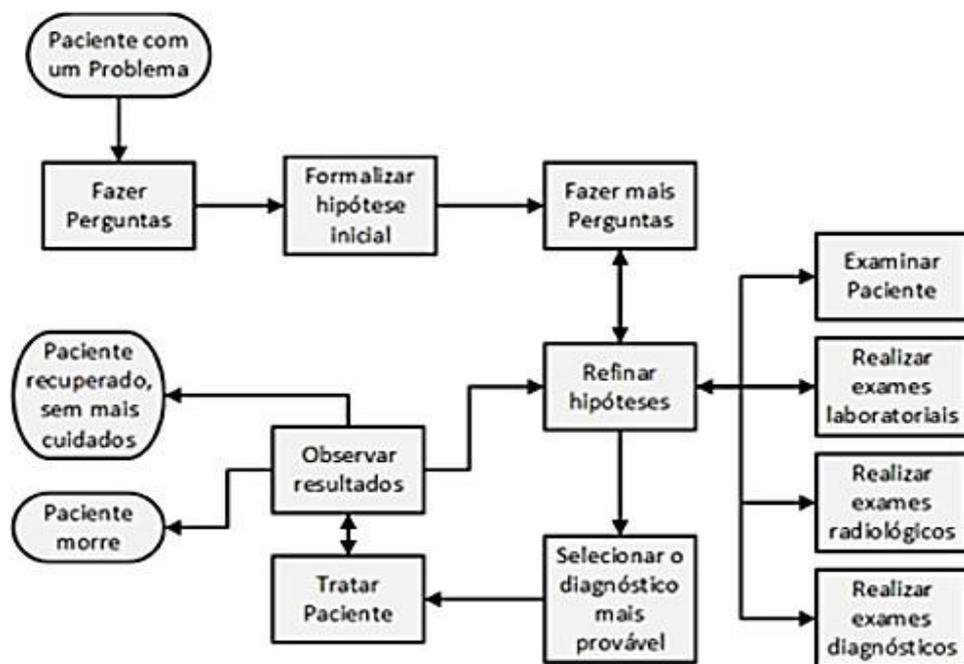
Dentro do prontuário do paciente é importante que tenha toda e qualquer informação alusiva à saúde e ao histórico de um paciente. Ele deve ser singular, universal e disponível para os mais díspares alvos (profissionais de saúde que estejam cuidando do paciente, pesquisadores, por exemplo), tendo em vista que seu armazenamento deve ser fornecido de modo a garantir a proteção e o sigilo de seu conteúdo. Nos casos de pesquisas médicas, ou áreas correlatas, é salutar a facilitação de acesso, fornecendo subsídios confiáveis a respeito do pesquisado. Assim sendo, podemos inferir que em um prontuário “todas as informações relativas a uma pessoa doente e às ações de cuidados e tratamentos a ela dispensados, a fim de que seja possível gerenciar o curso da patologia identificando os sintomas, causas e os remédios para solucioná-los” (Pinto, 2014).

A quantidade de grupos e profissionais que normalmente acolhem um paciente, especialmente se internado, é muito alta (Pinto, 2014). Do mesmo modo como todas as influências mútuas ocorrem na procura de resolver a dificuldade desse paciente, o prontuário é percebido como o ponto principal de conversação entre os membros envolvidos e a configuração de registro que posteriormente, acatando as legislações vigentes, será guardado e resguardado do acesso impróprio.

Na saúde, consoante Tavares *et al.* (2012), os cuidados acerca de um paciente compreendem diferentes áreas de formação, atendendo tanto o campo de humanas como serviço social e psicologia, assim como os cursos de saúde: medicina, fisioterapia, enfermagem, nutrição, terapia ocupacional, entre outros. Logo, admitindo o contato de diferentes formações de profissionais com o paciente, todo esse intercâmbio deve estar, de modo pleno, documentado em seu prontuário, significando que toda e qualquer interação tem a sua importância, porém todas fazem parte de um sistema informacional maior. O prontuário do paciente é considerado multidisciplinar (Pinto, 2014), ainda que a decisão sobre o diagnóstico do paciente decorra do médico. Ao redor do médico há um grupo cujas tarefas são dispostas de forma a acolher todas as necessidades de cuidados do paciente e a satisfazer as disposições desse médico para prover o melhor tratamento. Estes profissionais, em suas mais diversas interações, produzem dados e informações que, mesmo particularmente em cada especialidade, podem cooperar com o diagnóstico e com a constituição do histórico, mas, nesse aspecto, o paciente é visto como algo maior, se comparado a cada área isolada. Há uma soma de todas elas.

Caso a pessoa não possua um prontuário, tudo se inicia com a abertura do prontuário, logo, a partir de subsídios preexistentes, o profissional de saúde pode ter conhecimento do histórico do paciente ou de suas queixas recentes e, em seguida, dar início ao processo de apontamento das informações. Nele, todos os profissionais precisam explicar os procedimentos e exames conforme a legislação, podendo ser por meio eletrônico, de preferência, ou papel, de acordo com cada instituição de saúde. Para representar o modelo de assistência ao paciente e os métodos envolvidos durante um único atendimento veremos o esquema a seguir (Figura 1):

Figura 1 - Modelo de assistência ao paciente



Fonte: Adaptado (Carvalho, 2017)

O prontuário tem em seu desígnio o escopo de prover muitas benfeitorias que podem ser alcançadas, caso um profissional de saúde tenha acesso ao histórico de saúde do paciente. Conforme exposto por Possari (2007), muitos podem se beneficiar deste ato, destacando aqui exemplos destes benefícios:

Para o paciente – Os dados coletados e guardados possibilitariam atendimento, descoberta do diagnóstico e tratamento mais rápido, eficiente e econômico. Em caso de transferência, permite uma melhor comunicação entre as equipes das clínicas envolvidas, possibilitando a sequência do tratamento. Os registros em seu prontuário podem simplificar seu atendimento, dispensando maiores questionamentos ou até mesmo alguns exames, o que poderia reduzir o custo de seu atendimento e melhorar o aproveitamento do tempo desse

atendimento. Em caso de algum problema, pode ser considerado como prova a respeito do atendimento para todos os envolvidos.

Para o hospital – Melhores prontuários permitem menor tempo de internação e melhor utilização das vagas, aumentando o faturamento para o hospital. Evita a repetição de exames e serviços médicos. Serve como prova de todo o atendimento em caso de processos e acusações por motivos de negligência, imprudência ou imperícia. Serve para entender os resultados alcançados e o padrão de atendimento que foi prestado.

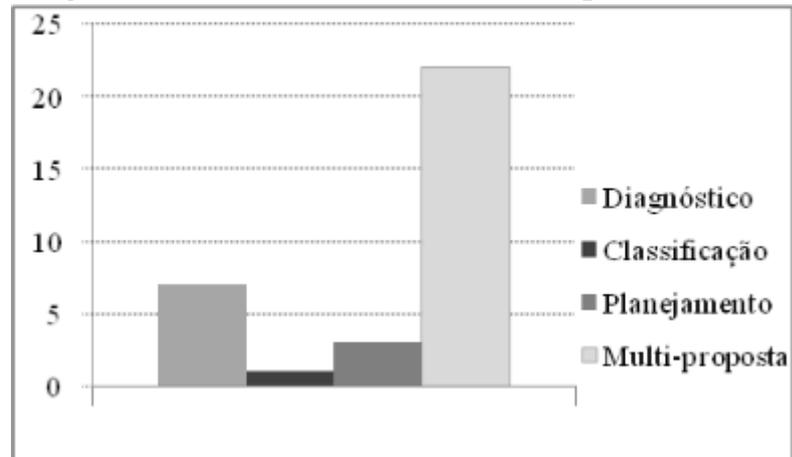
Para a equipe de saúde – Permite a integração e intercomunicação entre os profissionais que prestam atendimento, com a finalidade da continuidade da prestação dos serviços. Contribui para a análise da evolução do tratamento, verificando quais condutas obtêm melhores resultados e se o problema, origem do atendimento, foi sanado.

Para o ensino e pesquisa – Possibilita o conhecimento de casos antecedentes de uma enfermidade e suas informações, facilita o estudo do diagnóstico. É campo para pesquisa, fornecendo dados como incidências, prevalências, morbidade e mortalidade, além de permitir a comparação de condutas terapêuticas e verificar se foram efetivas.

De acordo com o que foi demonstrado até aqui, é possível verificar a necessidade de construir diálogos entre a Informática e a Saúde na perspectiva do acesso à informação do paciente para prover, da melhor maneira possível, sua saúde. Estes diálogos podem cooperar para as tomadas de decisões por parte dos médicos, dos enfermeiros, dos fisioterapeutas, dentre outros profissionais da saúde, para políticas públicas de saúde e outros. Vale ressaltar que em tempos de explosão e exploração de informações dessa área, elas precisam ser acessadas a quem de direito, com qualidade e segurança.

3.2 Trabalhos relacionados

Com a evolução tecnológica crescente nas mais diversas áreas do cotidiano humano, a área da saúde vem se mostrando um campo bem receptivo e com largo espaço para ser pesquisado, principalmente no que concerne a sugestão de prescrição terapêutica medicamentosa. Begum *et al.* (2011) exploram uma pesquisa abrangente sobre os mais recentes avanços na aplicação de técnicas de Inteligência Artificial (IA) na área da saúde. Dos 34 sistemas analisados por eles, apenas 12 incluem planejamentos terapêuticos como um de seus objetivos. Notavelmente, entre esses 12, apenas 3 sistemas são totalmente dedicados a essa finalidade específica. Isso evidencia uma atual tendência no desenvolvimento de sistemas com propostas multifacetadas, conforme ilustrado no gráfico apresentado na Figura 2:

Figura 2 - Desenvolvimento de sistemas por finalidade

Fonte: Begum *et al.* (2011)

A abordagem adotada neste trabalho parte do pressuposto de que o diagnóstico do paciente já tenha sido estabelecido, SEPSE, concentrando-se, assim, no desenvolvimento do planejamento terapêutico para o tratamento da condição atual do paciente. Neste sentido, buscou-se na literatura soluções que se enquadrassem nesse contexto e um em particular merece ser destacado.

Foi encontrado um trabalho intitulado “Apoio à Decisão com Base em Casos para Pacientes com Diabetes Tipo 1 em Terapia com Bomba de Insulina”, de Marling, Shubrook e Schwartz (2008), que destaca um sistema inovador à época, responsável pela gestão diária da dose de insulina necessária para indivíduos que enfrentam Diabetes Tipo 1 e utilizam a bomba de infusão de insulina. Diariamente, o paciente registra seu nível de glicose atual por meio da Web. Com base nesse dado e na compreensão dos hábitos do paciente, o sistema consulta sua base de conhecimento em busca de situações semelhantes e, então, oferece ao médico opções de doses terapêuticas empregadas em contextos comparáveis, visando uma possível recomendação ao paciente (Marling; Shubrook; Schwartz, 2008). Este sistema propõe ajustes na dosagem de insulina com base no nível de glicose e em eventos ou circunstâncias comunicadas pelo paciente.

Mais recentemente, encontramos um trabalho (Datorre, 2017) que fala sobre a implementação conjunta de programas de computador para prescrição, juntamente com um sistema cooperativo que visa desenvolver um sistema colaborativo de prescrição de medicamentos. O sistema do trabalho em questão possibilita a automação de processos relacionados à gestão de prescrições médicas.

Há ainda um outro trabalho (Teles, 2022) em que é proposto o desenvolvimento de um protótipo para um sistema de apoio à decisão clínica no campo da oncologia, denominado SiReMA. Trata-se de um sistema de recomendação de medicamentos anticancerígenos.

Porém, ao realizar a pesquisa bibliográfica, não foi identificada nenhuma solução que se propusesse a contribuir integralmente com o que estamos apresentando. Encontram-se apenas sistemas que compartilham alguns aspectos de funcionamento, aplicação de técnicas ou objetivos semelhantes aos delineados neste trabalho. Uma característica distintiva deste estudo é a incorporação de dados reais provenientes da rotina hospitalar.

3.3 Relevância

Conforme demonstrado ao longo deste capítulo, conseguimos observar que Inteligência Artificial pode ser uma grande aliada para dar suporte a profissionais de saúde, sobretudo o médico, no que concerne ao auxílio da criação de uma sugestão medicamentosa personalizada para um paciente. Vimos também que há diversas técnicas de *Machine Learning* que podem ser adotadas na elaboração de um algoritmo que possa executar esse apoio, mitigando problemas de interações medicamentosas e potencializando as chances de remissão para o paciente.

Não obstante, potencialmente, o sucesso desse algoritmo pode levar a três resultados importantes: a redução de custos por parte dos hospitais, considerando os valores investidos na recuperação de um paciente com SEPSE; a contribuição com novo conhecimento gerado; e por último, mas não menos importante, a quantidade de vidas que podem ser salvas diretamente.

4 EXPERIMENTO

Nesta seção está descrita a metodologia de experimento do trabalho desenvolvido.

4.1 Metodologia

A abordagem adotada por este trabalho caracteriza-se como pesquisa-ação demonstrando caráter duplo: pesquisa e ação. O seu objetivo é caracterizado pela pesquisa como ferramenta para ampliação do conhecimento científico do pesquisador (Thiollent, 1986). Já a ação é uma proposta de ferramenta feita pelo pesquisador com base no conhecimento adquirido em sua pesquisa, objetivando uma melhoria para a comunidade acadêmica e a sociedade em geral onde a pesquisa é feita.

Neste tipo de abordagem, o pesquisador explora processos complexos. O enfoque do pesquisador está em compreender o problema e em suas ações empreendidas deve gerar resolução dentro de um ambiente real específico, contribuindo para a geração de conhecimento científico associado às teorias que fundamentam a pesquisa. O posicionamento do pesquisador não é de um observador, mas de um ator envolvido no processo que busca soluções, propõe ações e interfere ativamente na pesquisa.

O resultado esperado na pesquisa-ação está na geração de novo conhecimento. Esse é um papel fundamental na promoção de mudança eficaz e compreensão científica. É um tipo de pesquisa (Thiollent, 1986) que estreita associação das práticas com a resolução de um problema coletivo.

A primeira etapa desta pesquisa (Gerhardt; Silveira, 2009) foi fundamentada em fontes bibliográficas buscando alguns contextos diferentes, mas que trabalham em conjunto, como, por exemplo, destacar conhecimento sobre SEPSE, seus conceitos, riscos, como reconhecer e seus tratamentos. Entender o funcionamento de fármacos administrados em pessoas com quadro séptico. Por último, mas não menos importante, o avanço das pesquisas com Inteligência Artificial, principalmente na área da medicina, e como funcionam as técnicas de *Machine Learning* para que pudéssemos aplicar de maneira mais coerente e correta em nossa pesquisa.

Desse modo, o levantamento de informações foi realizado por meio de um levantamento bibliográfico (Gerhardt; Silveira, 2009), considerando artigos científicos, sites, livros e revistas publicados entre 2019 e 2023. Como critério de inclusão para a seleção dos artigos foram coletados os artigos cujos temas abordam: SEPSE, algoritmo, *machine learning*, prescrição medicamentosa, predição, fármacos e inteligência artificial em diversas possibilidades de uso e

aplicação, publicados em português ou inglês no período supracitado em sites como Google Acadêmico, ScienceDirect, Sielo, IEEE, ACM, Scopus, PubMed, Nature, entre outros. O crivo observado para exclusão foi a literatura em língua estrangeira, que não a inglesa, ou o português de artigos que não abordavam o tema ou abordavam apenas parcialmente. O conteúdo selecionado foi classificado nas seguintes matérias: autor, material literário e relevância com o tema da pesquisa. Dos estudos foram extraídos os padrões: SEPSE. Os resultados serão apresentados através da construção de ferramenta sobre o tema, destacando a sugestão de prescrição medicamentosa de acordo com o levantamento feito.

4.2 Dados

Para o trabalho desta pesquisa, inicialmente foram usados dados de 48 pacientes com quadro séptico internados no Hospital Universitário de Lagarto, em Sergipe – Brasil. Houve aprovação pelo conselho de revisão institucional — Comitê de Ética CAAE nº 16847919.8.0000.5546 – Número do parecer: 3.496.684 — da Fundação Universidade Federal de Sergipe (UFS).

Os dados precisaram passar por um processo de tratamento antes de serem submetidos a experimentação. Os dados vieram categorizados por classificação, como “comorbidades”, “motivo de internação”, se houve “sedação”, “antibióticos” usados no tratamento, entre outros. Porém, cada classificação dessa poderia ter mais de um item.

Inicialmente fizemos um filtro para poder constatar que cada item seria único. No *dataset* original, os itens de cada classe estavam todos descritos em uma única célula, por exemplo, em comorbidades era descrito “DM, HAS, Cardiopata”, mas outro paciente poderia ter “DPOC, HAS, DM, ICC em comorbidade”. O que fizemos para tratar este cenário foi separar cada item em uma coluna, desse modo “DM”, “HAS”, “DPOC” e demais itens passaram a ser características separadas.

Um fato importante foi a padronização dos dados, pois alguns tinham caractere de “espaço” antes e/ou depois do item ou em excesso na escrita. Tinham itens que significavam a mesma coisa e por este motivo foi ajustado (uns estavam como siglas e outros escritos por extenso, por exemplo) para que o resultado obtido fosse o mais otimizado possível.

Depois de todos os ajustes, limpeza e tratamento o *dataset* passou a ter 48 amostras (os pacientes) e 136 características (colunas). Dentre os dados apresentados, alguns não tinham aderência para o que se propõe nesta pesquisa, como por exemplo, “número do paciente”, “data de nascimento”, “data da coleta”, “leito”, “data admissão hospitalar”, “data admissão UTI”,

“data óbito”, “data alta” e “data de finalização de coleta”. Além dessas características, outras 03 (três) também não foram usadas para aplicação do *Machine Learning*, pois a técnica usada para pré-processamento foi a *one-hot-encoding*, são elas: “peso”, “idade” e “tempo de permanência hospitalar”.

Após o tratamento aplicado, restaram 123 características como colunas e para cada paciente, essas colunas de características de valor único passaram a ser categorizadas como binária. Essa é uma técnica de pré-processamento conhecida como “*one-hot-encoding*”, que terá como valores possíveis apenas “0” e “1”, indicando se aquela característica está presente ou não em determinada amostra (paciente).

4.3 Ferramentas

Foi usado o *Google Colab* como ambiente de desenvolvimento e o Python como linguagem de programação para a aplicação dos modelos de *Machine Learning*. Para que os códigos fossem executados e os dados manipulados corretamente, foi necessária a importação de algumas bibliotecas, sendo elas: *numpy*, *pandas*, *scipy* e *sklearn*. Essa última, para que algoritmos de *Machine Learning* fossem implementados de maneira mais prática. O *dataset* originalmente tem seus dados em uma planilha de Excel, produto Microsoft. O fato de o ambiente de desenvolvimento ser o *Google Colab* acaba por outorgar a responsabilidade da plataforma pelo processamento de dados e demais processamentos computacionais necessários para o desenvolvimento.

Ainda que o processamento seja realizado por parte da plataforma online, é importante registrar que a elaboração e os testes serão realizados por meio de um notebook Dell Inspiron 15 5566, com processador 2.7 GHz Dual-Core Intel Core i5, memória RAM de 16 GB 2666 MHz DDR4, Intel HD Graphics 620 128 MB e armazenamento SSD -Kingston A400 480GB.

4.4 Cenário

O cenário aqui exposto tem a finalidade de predição de quais antibióticos teriam maiores chances, se usados, de fazer com que o paciente não evolua à óbito.

Para que o modelo de *Machine Learning* pudesse ter um desempenho satisfatório, os dados obtidos foram originados em um ambiente hospitalar real, mais especificamente o setor de UTI para aprendizagem e teste.

4.5 Parâmetros

Para esta pesquisa diversas “features” estão presentes na aplicação do modelo de *Machine Learning*, principalmente por se tratar de uma base de dados de casos reais. Então, aqui será apresentado parâmetros que foram usados para formulação do modelo com base nas suposições iniciais, buscando alcançar resultados que atendam às demandas do problema de pesquisa.

Por conta da opção pela técnica de *one-hot-encoding*, todas as características/features serão encaradas como binárias: 0 ou 1. Ou o paciente tem aquela característica ou não tem. Como dito anteriormente, diversas características estão presentes no *dataset* usado e aqui vamos listá-las por classificação:

COMORBIDADES: alzheimer, anemia, anemia grave, asma, AVC, AVE, cardiopata, convulsão, deficiência visual, demência, disfunção hepática, disfunção renal, dislipidemia, DM, doença psiquiátrica, DPOC, encefalopatia pós anoxia, etilista, etilista crônico, ex-tabagista, febre reumática, gastrite, HAS, hemiplegia a direita, hipotireoidismo, ICC, insuficiência cardíaca, neoplasia, retinopatia diabética, síndrome de imobilidade, síndrome demencial, transtorno de humor, úlcera gástrica, usuário de drogas;

SUORTE ADMISSIONAL: CVC, diálise, IOT, SVD, TQT, VM, sedação, DVA;

DVA: noradrenalina, dobutamina, amiodarona, vasopressina;

SEDAÇÃO: dexmedetomidina, etomidato, fentanil, midazolam;

MOTIVO INTERNAÇÃO: acidose metabólica, abdômen agudo, abdômen agudo obstrutivo, abscesso cutâneo, artrite piogénica, AVC prévio, AVE prévio, celulite em outros locais, cetoacidose diabética, descompensação respiratória, desconforto torácico, diarreia, dispneia, doença hepática tóxica, DPOC, edema agudo de pulmão, febre, furúnculo, hemoptise, hepatite não classificada, insuficiência hepática crônica, insuficiência renal aguda, insuficiência respiratória, IRC, ITU, lesão com edema, náuseas, osteomielite de fêmur distal direita, pneumonia bacteriana, pneumonia bronco aspirativa, RNC, SEPSE, sonolência, taquidispneia, transtornos comportamentais devido ao uso de álcool, transtornos mentais devido ao uso de álcool, vômito;

FOCO DA SEPSE: abdominal, articular, bacteremia, cateter venoso central, partes moles, pele, pulmonar, respiratório, sangue, urinário;

ANTIBIOTICO USADO EM TRATAMENTO: Amicacina 1g IV , Amicacina 500mg IV , Ampicilina + Sulbactam 3g IV, Azitromicina 500mg IV , Cefepime 1g IV , Cefepime 2g IV , Ceftriaxona 1g, Ceftriaxona 2g IV , Ciprofloxacino 400mg IV , Clindamicina

600mg IV, Fluconazol 200mg IV , Fluconazol 400mg IV , Meropenem 1g IV, Meropenem 2g IV, Meropenem 500mg IV, Metronidazol 500mg IV , Piperacilina + Tazobactam 2,25mg IV , Piperacilina + Tazobactam 4,5g IV, Polimixina B 1.000.000 UI IV, Polimixina B 500.000UI IV , Polimixina B 750.000UI IV , Teicoplanina 200mg IV , Teicoplanina 400mg IV, Vancomicina 1g IV.

Além dessas características acima citadas há também a de **sexo**, em que 0 indica sexo feminino e 1 sexo masculino e a de **óbito**, com 1 indicando o óbito e 0 a alta do paciente após tratamento.

4.6 Roteiro

Após todo o tratamento feito, salvamos a planilha com os dados no ambiente do Google Drive, desse modo não seria necessário fazer upload a cada vez que fosse executar o processo, mas apenas usar a autenticação do próprio google e configurar o caminho de leitura do arquivo salvo.

Abaixo vamos descrever passo a passo de como o modelo foi criado:

- 1 – Todas as bibliotecas necessárias para a manipulação dos dados foram importadas no início do projeto;
- 2 – Usamos a autenticação do *Google Colab* para efetuar login e conseguir usar o arquivo com os dados no Google Drive passando o caminho e o nome do arquivo;
- 3 – Visualizamos se todas as colunas/características estavam presentes na hora da importação;
- 4 – Criamos um *dataframe* com todas as colunas selecionadas;
- 5 – Verificamos se havia alguma coluna com dado faltante;
- 6 – Separamos as características entre “*feature*” e “*target*”;
- 7 – Nosso “*target*” são as colunas que possuem o termo “Antibiótico”. As demais são as “*features*”;
- 8 – Após a separação, usamos as bibliotecas “*sklearn*” e selecionamos alguns modelos de *Machine Learning* para saber quais resultados seriam gerados;
- 9 – Aplicamos o “*cross-validation*” para poder avaliar o desempenho dos modelos de *Machine Learning*;
- 10 – Depois da validação, uma das técnicas de *Machine Learning* foi escolhida baseada nos resultados demonstrados;
- 11 – Um paciente é selecionado aleatoriamente dentre os que estavam no *dataframe*

para ser um “novo” paciente;

12 – Caso o “novo” paciente tenha a *feature* “óbito” como valor Verdadeiro (1) alteramos ele para Falso (0), pois não queremos que o paciente evolua para “óbito”;

13 – Com isso é gerada uma lista com o(s) antibiótico(s) que é (são) sugerido(s) ao médico, indicando que sua administração possui maior probabilidade de manter o paciente vivo.

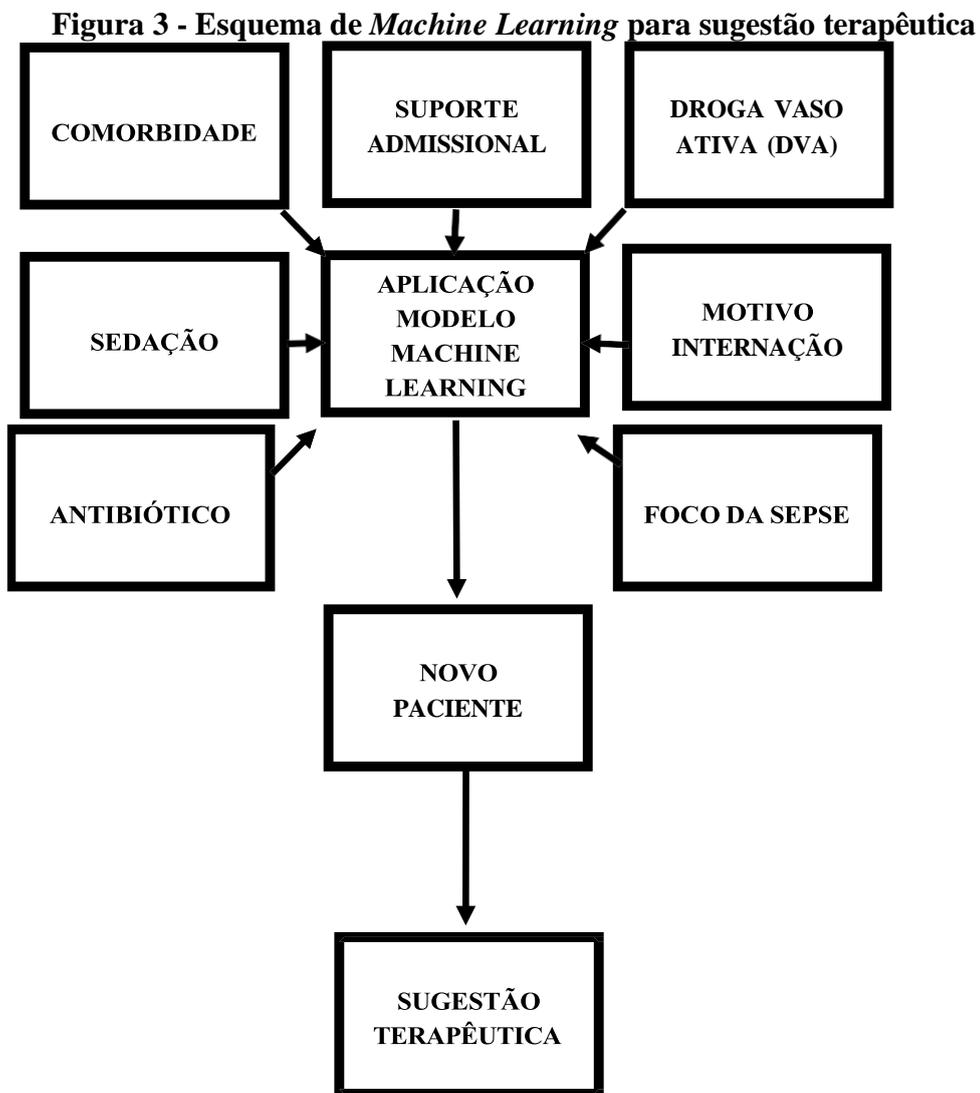
5 RESULTADOS E ANÁLISES

Nesta seção nos debruçaremos sobre os resultados obtidos de acordo com o cenário que foi proposto após implementação de técnica de *Machine Learning*.

5.1 Descrição

Inicialmente, registramos que trabalhamos com 48 amostras (quantidade de pacientes) e que ao todo havia 123 colunas entre “*features*” e “*target*”, conforme falado na seção anterior.

Para chegar aos resultados obtidos, contamos com o esquema seguinte (Figura 3), que ilustra como funciona o experimento do presente trabalho:



Fonte: elaborado pelo autor (2023)

Cada classe está representada por um dos quadrados do esquema acima, que são submetidos a alguns modelos de *Machine Learning* diferentes. Desta maneira, é possível verificar qual modelo tem nível de acurácia e desvio padrão mais satisfatório antes que a lista de sugestão medicamentosa seja gerada. Após verificação, um “novo” paciente aleatório é selecionado. O paciente pertence ao *dataset* usado para treinamento. Caso esse paciente possua “1” como valor em sua *feature* “óbito” esse valor é alterado para “0”, pois a indicação de medicamento(s) tem como objetivo a evolução do paciente para alta hospitalar. Só então, a lista de sugestão medicamentosa é gerada para auxiliar o médico.

Inicialmente temos 48 amostras disponíveis e quantidade de características, 123, que são as *features* disponíveis. Para aumentar a representatividade dos dados, foi usada a técnica de *Data Augmentation*, uma abordagem amplamente utilizada no campo de *Machine Learning*, especialmente quando a quantidade de dados reais disponíveis para treinamento de um modelo é limitada, dado que o modelo não é capaz de adquirir conhecimentos relevantes o bastante para efetuar a classificação do problema (Van Dyk; Meng, 2001).

Essa técnica consiste em criar novos exemplos de treinamento a partir dos dados existentes, aplicando transformações e perturbações controladas nos exemplos de treinamento originais. Dessa maneira, é possível incrementar o volume de dados, promovendo, conseqüentemente, o balanceamento da variável alvo em um conjunto de dados inicialmente desproporcional (Van Dyk; Meng, 2001).

O *Data Augmentation* (DA) pode ser empregada tanto em imagens quanto em dados tabulares, sendo este último o enfoque deste trabalho. Tornando possível sintetizar amostras (novas ou duplicadas) com base nas distribuições e recursos de classe.

No contexto desta pesquisa, que é usada a codificação *one-hot-encoding*, o aumento de dados pode ser realizado alterando os exemplos de treinamento de maneiras que preservem suas etiquetas ou rótulos originais, a fim de diversificar e aumentar a quantidade de dados disponíveis para o treinamento de modelos de aprendizado de máquina.

Após a aplicação do *Data Augmentation*, os dados foram submetidos ao esquema acima demonstrado. Dessa vez o número de amostras é de 144, com a mesma quantidade de *features*/características e o resultado obtido segue a seguir⁴:

⁴ Há uma captura de tela dos resultados gerados pelo código em “Apêndice” desse trabalho.

Tabela 5 - Resultados com *Data Augmentation*

	ACURÁCIA MÉDIA	DESVIO PADRÃO
DECISION TREE	0,95	0,08
EXTRATREE	0,94	0,09
EXTRATREES	0,94	0,09
RANDOMFOREST	0,94	0,09
KNEIGHBORS	0,46	0,14
REGRESSÃO LOGÍSTICA	nan	nan
ÁRVORE DE DECISÃO	0,95	0,08
KNN	0,46	0,14
NAIVE BAYES	0,62	0,07
RANDOM FOREST	0,94	0,09
GRADIENT BOOSTING	nan	nan
NEURALNETWORK	0,91	0,11

Fonte: elaborado pelo autor (2023)

Faz-se necessário elucidar que, no método estatístico “*cross-validation*”⁵ usado, algumas abordagens diferentes foram utilizadas e por este motivo “*Random Forest*” e “Árvore de Decisão/*Decision Tree*” apareceram duas vezes, por exemplo. Na parte superior da tabela foi usada a abordagem *support multiclass-multioutput*, indicando que o modelo tem a capacidade de lidar, ao mesmo tempo, com múltiplas classes e saídas. Suportando assim, a complexa tarefa de prever múltiplas previsões para uma única amostra.

Na parte inferior da tabela, em que as técnicas de ML estão repetidas, os resultados da abordagem referente ao *multioutput classifier*, que é uma abordagem específica projetada para trabalhar explicitamente com problemas em que diversas variáveis destino sejam esperadas.

Ao estabelecer um comparativo entre eles, é possível observar que, após a aplicação da técnica de *Data Augmentation*, houve um aumento significativo de acurácia entre os modelos apresentados. Pode-se observar também que o desvio padrão apresentado pelos modelos diminuíram, o que sugere uma maior consistência nos resultados obtidos, de modo que são mais estáveis e previsíveis se comparados com os primeiros resultados.

⁵ Estimador que consiste em os exemplos serem aleatoriamente divididos em k partições mutuamente exclusivas (folds) de tamanho aproximadamente igual a n/k exemplos.

5.2 Discussão

É possível perceber que o modelo *Decision Tree*, com abordagem *support multiclass-multioutput*, teve resultado mais satisfatório⁶ se comparado a Acurácia e Desvio Padrão apresentados frente aos outros modelos. Enquanto *ExtraTree*, *ExtraTrees*, *RandomForest* (em ambas as abordagens) e *Árvore de Decisão* apresentam valores iguais para Acurácia e Desvio padrão, sendo, ainda assim, valores satisfatórios. Isso demonstra que o desempenho entre os modelos baseados em *Árvores* teve um desempenho consistente ao lidar com os dados, o que gerou resultados robustos para esta pesquisa.

Por outro lado, os modelos de *Kneighbors*, *KNN* e *Naive Bayes* tiveram desempenhos bem abaixo e insatisfatórios quando comparados aos modelos baseados em *Árvores*. Este é um indicativo que não seria uma boa opção escolhê-los, ao menos com esta configuração de dados, por não ser tão adequado ou robusto para fazer sugestão terapêutica para o médico.

Diferentemente dos modelos apresentados até este momento, a *Regressão Logística* e *Gradient Boosting* apresentaram valores “nan” para acurácia média e para o desvio padrão. Isso pode indicar que algo está causando problemas no cálculo que envolve essas métricas. O que sugere que esses modelos não devem ser adotados, pois não conseguiriam gerar uma lista de sugestão terapêutica para auxiliar o médico dado o conjunto de dados em que ele trabalhou.

Já o desempenho apresentado pela *Neural Network* pode ser considerado razoável, mas não tão alto quanto àqueles que foram baseados em *Árvores*. Ele não foi insuficiente quando comparado ao *Naive Bayes*, inclusive apresentando um resultado superior. Mas os seus valores ficam aquém quando confrontados com os modelos baseados em *Árvores*, o que não o tornaria primeira opção de uso, visto o conjunto de dados que foi utilizado.

Diante do exposto, foram usados como métricas para comparação entre os modelos, o percentual de acurácia e o desvio padrão presente em cada uma das abordagens de *Machine Learning* abordadas nesse estudo. Então, a combinação dessas duas variáveis foram ponderadas e sopesadas para perceber qual o modelo teve melhor desempenho diante do cenário e das características disponíveis nesse estudo.

Desta maneira, o trabalho aponta um caminho que pode ser seguido de maneira que os médicos possam ser auxiliados em seu dia a dia no que tange a importância de uma eficiente e eficaz prescrição medicamentosa terapêutica para remissão de quadro séptico de seus pacientes.

⁶ Há um gráfico em “Anexos” demonstrando a comparação entre os modelos aqui usados.

6 CONSIDERAÇÕES FINAIS

Devido à natureza e à sensibilidade da pesquisa, a ausência de dados robustos e o tempo disponível são variáveis não laboram a nosso favor. A falta de dados em grande escala, concretos e reais, na literatura dificulta nos concentrarmos mais especificamente em como avançar no auxílio aos médicos em prol da mitigação de óbitos decorrente de SEPSE dentro de hospitais de maneira mais eficiente e eficaz.

Apesar das limitações na literatura, conseguimos fazer a coleta dos dados clínicos de pacientes com quadro séptico em um Hospital Universitário cedido por um outro pesquisador. Ainda assim, a base real que conseguimos obter necessitava passar por diversos tratamentos para que pudesse ser manejada de maneira correta e trazer resultados que pudessem ser considerados satisfatórios devido à sua finalidade de aprendizagem e predição para tratamento séptico.

Quanto à proposta criada no início da pesquisa, foi percebido o alcance em sua totalidade da predição terapêutica medicamentosa para paciente acometidos por quadro séptico. O objetivo de identificar os modelos de *Machine Learning* para predição e elaboração de algoritmo para sugestão medicamentosa foi alcançado. Foi verificado, também, a aplicação desses modelos a partir dos dados coletados, verificando a sua acurácia e analisando os resultados em cada etapa percorrida por este trabalho, visando a aplicação de método para auxílio médico, buscando a remissão de seus pacientes.

6.1 Ameaças à validação

Conforme citado no início desta seção, devido às limitações temporais e de quantidade de dados coletados no espaço dessa pesquisa, não houve a possibilidade de implantar o algoritmo junto a uma equipe médica no cenário ideal para sua aplicação, a UTI. Embora os dados partam inicialmente de casos reais, não houve a possibilidade de implementar as sugestões terapêuticas em um ambiente real. Por exemplo, na chegada de um novo paciente, imputar as novas características desse paciente no modelo e aguardar a sugestão terapêutica para auxiliar o médico. Após crivo médico, caso a aprovação feita por ela seja concedida, fazer a aplicação de sugestão medicamentosa e avaliar a evolução do paciente, percebendo, assim, se o modelo é eficiente e eficaz no auxílio à alta hospitalar desse paciente. O intuito com essa validação é poder mitigar a chance de óbito ao máximo de pessoas que forem possíveis e estejam acometidas por quadro séptico.

É importante ressaltar que, em tese, essa aplicação não deve gerar riscos ao paciente. Isso parte do pressuposto que o uso do tratamento, em sua totalidade ou de maneira parcial, está condicionado a aprovação por parte do profissional de saúde, que deve ter a capacidade de discernimento se o seu uso está de acordo com a finalidade e que não causará prejuízos ao paciente.

Um outro fato importante a ser mencionado e que pode ameaçar a validação deste trabalho é que, ao aplicar o *Data Augmentation*, é necessário ter cuidado para não introduzir padrões artificiais nos dados que possam confundir o modelo durante o treinamento. As transformações aplicadas devem ser escolhidas com base no conhecimento do domínio e na compreensão das características dos dados, de forma a manter a relevância e a representatividade dos exemplos gerados.

Além disso, é fundamental observar que, embora o *Data Augmentation* seja uma técnica valiosa para aumentar a variedade e a quantidade de dados disponíveis para treinamento, ele não substitui a necessidade de dados de alta qualidade e representativos. É um método complementar que pode melhorar o desempenho dos modelos, especialmente em situações em que os conjuntos de dados são limitados, como foi o caso desta pesquisa.

O *Data Augmentation* é uma técnica útil para criar exemplos de treinamentos adicionais a partir de dados reais existentes – visto que a quantidade de dados disponíveis era hipossuficiente para treinamento dos métodos de *Machine Learning* –, dessa maneira foram aplicadas transformações controladas nos exemplos de treinamento, mas que precisa de muito cuidado e atenção.

6.2 Trabalhos futuros

Por conta das limitações impostas, sobretudo considerando a quantidade pequena de dados e tempo não muito amplo de pesquisa, desejamos, em trabalhos futuros, conseguir uma base de dados real maior e mais robusta. Se possível, nas mais diversas partes do país. Para que de maneira heterogênea possamos fazer com que a aplicação de *Machine Learning* compreenda como os mais diversos profissionais abordam os cuidados na tentativa de remissão de seus pacientes. Para isso, buscaremos parcerias com hospitais públicos, particulares e universitários para que seja oferecido à academia e à sociedade uma contribuição significativa, segura e que impacte positivamente a vida de diversas pessoas a nível mundial.

Outro fator preponderante, caso não haja disponibilização de dados na literatura, é a importância de possuir mais tempo para fazer não apenas o levantamento, mas, além do

algoritmo, criar um sistema que possa ser usado por médicos nas unidades de UTI e poder verificar quanto os dados que foram levantados inicialmente estão ajudando outros médicos (na sugestão do tratamento) e pacientes (na remissão de quadro de saúde). E, dessa maneira, retroalimentando o sistema que poderá aprender mais e ajudar, cada vez mais, as partes envolvidas, o médico, no auxílio prescritivo, e o paciente, com a remissão de seu quadro de saúde.

REFERÊNCIAS

ABIODUN, O. I. *et al.* State-of-the-art in artificial neural network applications: a survey.

Heliyon, [S. l.], v. 4, n. 11, pe. 00938, nov. 2018. DOI:

<https://doi.org/10.1016/heliyon.2018.e00938>. Disponível em:

<https://www.sciencedirect.com/science/article/pii/S2405844018332067>. Acesso em: 24 jul. 2023.

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA (Brasil). **Bulas de medicamentos:**

confira as orientações para alteração. Brasília, DF: ANVISA, 11 dez. 2020. Disponível em:

<https://www.gov.br/anvisa/pt-br/assuntos/noticias-anvisa/2020/bulas-de-medicamentos-confira-as-orientacoes-para-alteracao>. Acesso em: 15 jul. 2023.

ALLEGIANZI, B. *et al.* Burden of endemic health-care-associated infection in developing countries: systematic review and meta-analysis. **The Lancet**. [S. l.], v. 377, n. 9761, p. 228-

241, jan. 2011. DOI: [https://doi.org/10.1016/S0140-6736\(10\)61458-4](https://doi.org/10.1016/S0140-6736(10)61458-4). Disponível em:

<https://www.sciencedirect.com/science/article/abs/pii/S0140673610614584>. Acesso em: 12 jun. 2023.

AMUDHA, P.; RAUF, H. A. Performance analysis of data mining approaches in intrusion

detection. *In: IEEE INTERNATIONAL CONFERENCE ON PROCESS AUTOMATION, CONTROL AND COMPUTING, 2011, Índia. Anais [...]. Índia: IEEE, 2011, p. 1-6.*

Disponível em: <https://ieeexplore.ieee.org/abstract/document/5978878>. Acesso em: 12 jun. 2023.

ARMSTRONG, B. A.; BETZOLD, R. D.; MAY, A. K. Sepsis and Septic Shock Strategies.

Surgical Clinics of North America, [S. l.], v. 97, n.6, p. 1339-1379, 2017. DOI:

<https://doi.org/10.1016/j.suc.2017.07.003>.

BEGUM, S. *et al.* Case-Based Reasoning Systems in the Health Sciences: A Survey Recent

Trends and Developments. **IEEE Trans. on Systems, Man and Cybernetics-Part C:**

Applications e Reviews, [S. l.], v. 41, n. 4, p. 421-434, ago. 2011. Disponível em:

[https://www.researchgate.net/publication/224178122_Case-](https://www.researchgate.net/publication/224178122_Case-Based_Reasoning_Systems_in_the_Health_Sciences_A_Survey_of_Recent_Trends_and_Developments)

[Based_Reasoning_Systems_in_the_Health_Sciences_A_Survey_of_Recent_Trends_and_Developments](https://www.researchgate.net/publication/224178122_Case-Based_Reasoning_Systems_in_the_Health_Sciences_A_Survey_of_Recent_Trends_and_Developments). Acesso em: 20 jun. 2023.

BELFIORE, P. **Estatística:** aplicada a administração, contabilidade e economia com Excel e

SPSS. 1. ed. Rio de Janeiro: GEN LTC, 2015

BLEICH, H. Computer Evaluation of Acid-Base Disorders. **The Journal of Clinical**

Investigation, [S. l.], v. 48, n. 9, p. 1689-1696, 1969. DOI:<https://doi.org/10.1172/JCI106134>.

BRASIL. Conselho Federal de Medicina. **Resolução CFM nº 1638/02, de 10 de julho de**

2002. Define prontuário médico e torna obrigatória a criação da Comissão de Revisão de

Prontuários nas instituições de saúde. Brasília: Diário Oficial da União, 2002. Disponível em:

<https://sistemas.cfm.org.br/normas/visualizar/resolucoes/BR/2002/1638>. Acesso em: 04 fev. 2022.

BREIMAN, L. Florestas aleatórias. **Machine Learning**, Países Baixos, v. 45, n. 1, p. 5-32,

2001. Disponível em: <https://link.springer.com/article/10.1023/A:1010933404324>. Acesso

em: 14 nov. 2023.

BRUNTON, L. L.; LAZO, J. S.; PARKER, K. L., **Goodman & Gilman: As Bases Farmacológicas da Terapêutica**. 11. ed. Rio de Janeiro: McGraw-Hill, 2007.

BURDICK, H. *et al.* Validation of a machine learning algorithm for early severe sepsis prediction: a retrospective study predicting severe sepsis up to 48 h in advance using a diverse dataset from 461 US hospitals. **BMC Medical Informatics and Decision Making**, [S. l.], v. 20, n. 276, 2020. Disponível em: <https://bmcmmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01284-x>. Acesso em: 14 nov. 2023.

CARRILHO JUNIOR, J. R. **Desenvolvimento de uma Metodologia para Mineração de Textos**. 2007. Dissertação (Mestrado em Engenharia Elétrica) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2007. Disponível em: https://www.maxwell.vrac.puc-rio.br/11675/11675_1.PDF. Acesso em: 12 jun. 2023.

CARVALHO, L. A. V. **Datamining: A mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração**. Rio de Janeiro: Ciência Moderna, 2005

CARVALHO, P. P. *et al.* Avaliação de pacientes críticos com suspeita de sepse em um hospital universitário. **Research, Society and Development**, [S. l.], v. 10, n. 3, p. e39410313570, 2021. Disponível em: <https://rsdjournal.org/index.php/rsd/article/view/13570>. Acesso em: 20 mar. 2022.

CARVALHO, R. C. **Aplicação de técnicas de mineração de texto na recuperação de informação clínica em prontuário eletrônico do paciente**. 2017. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências - Universidade Estadual Paulista - UNESP, Marília, 2017. 201f.

CHICCO, D.; JURMAN, G. Survival prediction of patients with sepsis from age, sex, and septic episode number alone. **Scientific Reports**, [S. l.], v. 10, n. 17156, 2020. Disponível em: <https://www.nature.com/articles/s41598-020-73558-3>. Acesso em: 17 ago. 2023.

CRUZ, L. L.; MACEDO, C. C. Perfil epidemiológico da sepse em hospital de referência no interior do Ceará. **Revista De Psicologia**, [S. l.], v. 10, n. 29, p. 71-99, 2016. DOI:10.14295/online.v10i1.385. Disponível em: https://www.researchgate.net/publication/301344885_Perfil_Epidemiologico_da_Sepse_em_Hospital_de_Referencia_no_Interior_do_Ceara. Acesso em: 17 ago. 2023.

DATORRE, F. **Sistema Colaborativo para Prescrição de Medicamentos**. 2017. Dissertação (Mestrado em Bioengenharia) – Universidade Brasil, São Paulo, 2017. Disponível em: https://universidadebrasil.edu.br/portal/_biblioteca/uploads/20200316143203.pdf. Acesso em: 20 dez. 2023.

DYAS, S. R. *et al.* Process-improvement cost model for the emergency department. **Journal of Healthcare Management**, [S. l.], v. 60, n. 6, p. 442-458, 2015. DOI: <https://doi.org/10.1097/00115514-201511000-00011>.

ERTEL, W. **Introduction to Artificial Intelligence**. 2. ed. Londres: Springer, 2017.

FELDMAN, R.; SANGER, J. **The Text Mining Handbook: Advanced Approachs in Analyzing Unstructured Data**. New York: Cambridge University Press, 2007.

FERRER, R. *et al.* Empiric antibiotic treatment reduces mortality in severe sepsis and septic shock from the first hour: results from a guideline-based performance improvement program. **Critical Care Medicine**, [S. l.], v. 42, n. 8, p. 1749-1755, ago. 2014. DOI: <https://doi.org/10.1097/CCM.0000000000000330>.

FERRERO, C. A. **Algoritmo KNN para previsão de dados temporais: funções de previsão e critérios de seleção de vizinhos próximos aplicados a variáveis ambientais em limnologia**. 2009. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) — Instituto de Ciências Matemáticas e de Computação - USP, São Carlos, 2009. DOI: <https://doi.org/10.11606/D.55.2009.tde-19052009-135128>.

GERHARDT, T.; SILVEIRA, D. **Métodos de Pesquisa**. Porto Alegre: Editora da UFRGS, 2009.

GOH, K. H. *et al.* Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. **Nature Communications**, [S. l.], v. 12, n. 711, 2021. Disponível em: <https://www.nature.com/articles/s41467-021-20910-4>. Acesso em: 25 jul. 2023.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference and Prediction**. New York: Springer New York, 2008.

HOBBS, G; R. Data mining and healthcare informatics. **American journal of health behavior**, [S. l.], v. 25, n. 3, p. 285-289, 2001. DOI: <https://doi.org/10.5993/ajhb.25.3.16>.

IBM. **Como o Watson, a plataforma de computação cognitiva da IBM, está ajudando a revolucionar o perfil de cinco verticais econômicas**. New York: IBM Watson, [s. d.]. Disponível em: <http://idgnow.com.br/ti-corporativa/2016/06/20/como-owatson-a-plataforma-de-computacao-cognitiva-da-ibm-esta-ajudando-arevolucionar-o-perfil-de-cinco-verticais-economicas>. Acesso em: 04 fev. 2022.

INSTITUTO LATINO-AMERICANO DE SEPSE. **Sepse: um problema de saúde pública**. Brasília: CFM, 2015. 90 p. Disponível em: <https://ilas.org.br/wp-content/uploads/2022/02/livro-sepse-um-problema-de-saude-publica-cfm-ilas.pdf>. Acesso em: 28 out. 2023.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning**. New York: Springer, 2014.

JIANG, F. *et al.* Artificial intelligence in healthcare: past, present, and future. **Stroke and Vascular Neurology**, [S. l.], v. 2, n. 4, p. 230-243, 2017. DOI: <http://dx.doi.org/10.1136/svn-2017-000101>.

KNIGHT, W. The dark secret at the heart of AI. **MIT Technology Review**. [S. l.], 11 abr. 2017. Disponível em: <https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/>. Acesso em: 28 out. 2023.

KUHN, M.; JOHNSON, K. **Applied Predictive Modeling**. New York: Springer, 2013.

- KULKARNI, A. D.; LOWE, B. Random Forest algorithm for land cover classification. **Computer Science Faculty Publications and Presentations**, [S. l.], v. 4, n. 3, p. 58-63, mar. 2016. Disponível em: <http://hdl.handle.net/10950/341>. Acesso em: 22 out. 2023.
- LELIGDOWICZ, A.; MATTHAY, M. A. Heterogeneity in sepsis: new biological evidence with clinical applications. **Critical Care**, [S. l.], v. 23, n. 80, mar. 2019. DOI:<https://doi.org/10.1186/s13054-019-2372-2>. Disponível em: <https://ccforum.biomedcentral.com/articles/10.1186/s13054-019-2372-2>. Acesso em: 22 ago. 2023.
- LIPSITCH, M.; SAMORE, M. H. Antimicrobial use and antimicrobial resistance: a population perspective. **Emerging Infectious Diseases**, [S. l.], v.8, n. 4, p. 347-54, 2002. DOI: <https://doi.org/10.3201/eid0804.010312>.
- LOBO, S. M.; REZENDE, E.; MENDES, C. L.; OLIVEIRA, M. C. de. Mortalidade por sepse no Brasil em um cenário real: projeto UTIs Brasileiras. **Revista Brasileira de Terapia Intensiva**, [S. l.], v. 31, n. 1, p. 1-4, jan. 2019. Disponível em: <https://www.scielo.br/j/rbti/a/XD867yzfcJGNpnMKhQg8wyb/#> . Acesso em: 20 dez. 2023.
- LONG, E. *et al.* An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. **Nature Biomedical Engineering**, [S. l.], v.1, n.2, p. e0024, 2017. Disponível em: https://www.researchgate.net/publication/313111785_An_artificial_intelligence_platform_for_the_multihospital_collaborative_management_of_congenital_cataracts. Acesso em: 14 ago. 2023.
- MACHADO, F. R. *et al.* The epidemiology of sepsis in Brazilian intensive care units (the Sepsis PREvalence Assessment Database, SPREAD): an observational study. **Lancet Infect Disease**, [S. l.], v. 17, n. 11, p.1180-1189, 2017. DOI: [https://doi.org/10.1016/S1473-3099\(17\)30322-5](https://doi.org/10.1016/S1473-3099(17)30322-5). Disponível em: [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(17\)30322-5/fulltext](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(17)30322-5/fulltext). Acesso em: 22 set. 2023.
- MARIK, P. E.; TAEB, A. M. SIRS, qSOFA and new sepsis definition. **Journal of Thoracic Disease**, [S. l.], v. 9, n. 4, p. 943-945, abr. 2017. DOI:<https://doi.org/10.21037/jtd.2017.03.125>. Disponível em: <https://jtd.amegroups.org/article/view/12738/pdf>. Acesso em: 8 jan. 2023.
- MARLING, C.; SHUBROOK, J.; SCHWARTZ, F. Case-based decision support or patients with type 1 diabetes on insulin pump therapy. In: PROC. EUR. CONF. CASE-BASED REASON: (ECCBR), 9., 2008, Berlim. **Anais [...]**. Berlim: Springer-Verlag, 2008, p. 325-339. Disponível em: <http://oucsace.cs.ohio.edu/~smarthlt/pubs/ECCBR-2008.pdf>. Acesso em: 15 ago. 2023.
- MIOTTO, R. *et al.* Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. **Scientific Reports**, [S. l.], v. 6, n.1, p. e26094, 2016. Disponível em: https://www.researchgate.net/publication/303317731_Deep_Patient_An_Unsupervised_Representation_to_Predict_the_Future_of_Patients_from_the_Electronic_Health_Records. Acesso em: 25 set. 2023.

MITCHELL, T. **Aprendizado de máquina**. 1. ed. McGraw-Hil, 1997.

MORAIS, E. A. M.; AMBRÓSIO, A. P. L. **Mineração de textos**. Goiânia: Instituto de Informática da Universidade Federal de Goiás, 2007. 30p. Disponível em: http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_005-07 Acesso em: 04 fev. 2022

MORIMOTO, T. *et al*, Adverse drug events and medication errors: detection and classification methods. **Quality & Safety in Health Care**, [S. l.], v. 13, n. 4, p. 306-314, 2004. Disponível em: https://www.researchgate.net/publication/8420183_Adverse_drug_events_and_medication_errors_Detection_and_classification_methods. Acesso em: 11 set. 2023.

NATIONAL CANCER INSTITUTE (EUA). **Cancer Treatment Research**. Maryland: NCI, 2022b. Disponível em: <https://www.cancer.gov/research/areas/treatment/pmi-oncology> Acesso em: 04 fev. 2022

NATIONAL CANCER INSTITUTE (EUA). **Clinical trials information for patients and caregivers**. Maryland: NCI, 2022a. Disponível em: <https://www.cancer.gov/about-cancer/treatment/clinical-trials/ncisupported/nci-match> Acesso em: 04 fev. 2022

OLIVEIRA, D. F. *et al*. Fatores associados à baixa adesão ao tratamento farmacológico de pacientes atendidos por um Centro Integrado de Saúde. **Brazilian Journal of Natural Sciences**, [S. l.], v. 3, n. 3, p. 430, 2020. DOI: 10.31415/bjns.v3i3.113. Disponível em: <https://www.bjns.com.br/index.php/BJNS/article/view/113>. Acesso em: 3 set. 2023.

PAIXÃO, G. M. de M.; SANTOS, B. C.; ARAUJO, R. M. de; RIBEIRO, M. H.; MORAES, J. L. de; RIBEIRO, A. L. Machine Learning na Medicina: Revisão e Aplicabilidade. **Arquivos Brasileiros de Cardiologia**, [S. l.], v. 118, n. 1, p. 95, 2022. Disponível em: <https://www.scielo.br/j/abc/a/WMgVngCLbYfJrkmC65VFCkp/?format=pdf>. Acesso em: 20 dez. 2023.

PANDA, M.; PATRA, M. R. A comparative study of data mining algorithms for network intrusion detection. In: IEEE INTERNATIONAL CONFERENCE ON EMERGING TRENDS IN ENGINEERING AND TECHNOLOGY, 1., 2008, [S. l.]. **Anais [...]**. [S. l.]: IEEE, 2008, p. 504-507.

PINTO, V. B. A contribuição da terminologia para a gestão da documentação sanitária em organizações de saúde. In: MEDINFOR, 3., 2014, Salvador. **Anais [...]**. Salvador: UFBA, 2014.

PITTET, D. *et al*. Infection control as a major World Health Organization priority for developing countries. **Journal of Hospital Infection**, [S. l.], v. 68, n. 4, p. 285-292, abr. 2008. DOI: <https://doi.org/10.1016/j.jhin.2007.12.013>.

POSSARI, J. F. **Prontuário do paciente e os registros de enfermagem**. 2. ed. São Paulo: Iátria, 2007. 246 p.

RHODES, A. *et al*. Surviving sepsis campaign: International guidelines for management of sepsis and septic shock: 2016. **Intensive Care Medicine**, [S. l.], v. 43, n. 3, p. 304-377, 2017. DOI: <https://doi.org/10.1007/s00134-017-4683-6>. Disponível em: <https://link.springer.com/article/10.1007/s00134-017-4683-6>. Acesso em: 16 set. 2023.

RIPLEY, B.D. **Pattern Recognition and Neural Networks**. Cambridge: Cambridge University Press, 1996. 354p.

ROSA, J. L. A. **Classificação de dados através da otimização do método KNN-Fuzzy em ambiente de computação paralela**. 2003. Tese (Doutorado em Engenharia Civil) — Universidade Federal do Rio de Janeiro - UFRJ, Rio de Janeiro, 2003.

RUDD, K. E. *et al.* Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. **Lancet Infect Disease**, Londres, v. 395, n. 10219, p. 200-211, jan. 2020. DOI:[https://doi.org/10.1016/S0140-6736\(19\)32989-7](https://doi.org/10.1016/S0140-6736(19)32989-7). Disponível em: [https://www.thelancet.com/article/S0140-6736\(19\)32989-7/fulltext](https://www.thelancet.com/article/S0140-6736(19)32989-7/fulltext). Acesso em: 15 jul. 2023.

RUSSELL, S.; NORVIG, P. **Inteligência Artificial**. 3. ed. Rio de Janeiro: Elsevier Editora LTDA, 2013.

SAMUEL, A. L. Algumas pesquisas em aprendizado de máquina usando o jogo de damas. **IBM Journal of Research and Development**, [S. l.], v. 3, n.3, p. 210-229, 1959.

SCHEFFER, M. *et al.* **Demografia Médica no Brasil 2015**. Departamento de Medicina Preventiva, Faculdade de Medicina da USP. Conselho Regional de Medicina do Estado de São Paulo. Conselho Federal de Medicina. São Paulo: 2015. Disponível em: <https://www.usp.br/agen/wp-content/uploads/DemografiaMedica30nov2015.pdf>. Acesso em: 18 out. 2023.

SHANKAR-HARI, M. *et al.* Developing a new definition and assessing new clinical criteria for septic shock: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). **JAMA**, [S. l.], v. 315, n. 8, p. 775-787, 2016. DOI:<https://doi.org/10.1001/jama.2016.0289>. Disponível em: <https://jamanetwork.com/journals/jama/fullarticle/2492876>. Acesso em: 26 set. 2023.

SUBBE, C. *et al.* Validation of physiological scoring systems in the accident and emergency department. **Emergency Medicine Journal**, [S. l.], v. 23, n. 11, p. 841-845, 2006. DOI:<https://doi.org/10.1136/emj.2006.035816>.

TAVARES, S. O. *et al.* Interdisciplinaridade, Multidisciplinaridade Ou Transdisciplinaridade. In: INTERFACES NO FAZER PSICOLÓGICO, 5., 2012, Santa Maria, **Anais[...]**. Santa Maria: UNIFRA, 2012. Disponível em: https://docplayer.com.br/1454502-Interdisciplinaridade-multidisciplinaridade-ou-transdisciplinaridade-1.html#google_vignette. Acesso em: 08 fev. 2022.

TELES, M. C. G. **Desenvolvimento de componentes de execução e validação humana de modelos de Machine Learning no sistema de recomendação de medicamentos anticancerígenos (SiReMA)**. 2022. Dissertação (Mestrado em Análise de Dados e Sistemas de Apoio à Decisão) – Instituto Superior de Contabilidade e Administração, Politécnico de Coimbra, Coimbra, 2022. Disponível em: https://comum.rcaap.pt/bitstream/10400.26/43436/1/M%c3%b3nica_Teles.pdf. Acesso em: 20 dez. 2023.

THIOLLENT, M. **Metodologia da pesquisa-ação**. 2. ed. São Paulo: Editora Autores Associados, 1986.

TOPOL, E. J. High-performance medicine: the convergence of human and artificial intelligence. **Nature Medicine**, [S. l.], v. 25, n. 1, p. 44-56, jan. 2019. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/30617339/>

VAN DER POLL, T. *et al.* The immunopathology of sepsis and potential therapeutic targets. **Nature Reviews Immunology**, [S. l.], v. 17, p. 407-420, 2017. DOI:<https://doi.org/10.1038/nri.2017.36>.

VAN DYK, D. A.; MENG, X.-L. The Art of Data Augmentation. **Journal of Computational and Graphical Statistics**, [S. l.], v.10, n. 1, p. 1–50, 2001. Disponível em: <https://www.jstor.org/stable/1391021>. Acesso em: 20 jun. 2023.

VINCENT, J. L. *et al.* The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. **Intensive Care Medicine**, [S. l.], v. 22, n. 7, p. 707–710, 1996. DOI: <https://doi.org/10.1007/BF01709751>.

WORLD HEALTH ORGANIZATION. **Improving the Prevention, Diagnosis and Clinical Management of Sepsis**. Geneva: WHO, [s. d.]. Disponível em: <https://www.who.int/activities/improving-the-prevention-diagnosis-and-clinical-management-of-sepsis> Acesso em: 04 fev. 2022.

APÊNDICES

APÊNDICE A

CÓDIGO PARA GERAÇÃO DOS MODELOS DE *MACHINE LEARNING*

```
# -*- coding: utf-8 -*-

"""Augmentados Geral Classificado antibióticos - vFinal.ipynb

Automatically generated by Colaboratory.

Original file is located at

https://colab.research.google.com/drive/1R5eUyia4nAqG1O0wHnPN6rYOLM36Jgxa

"""

# IMPORTAÇÃO DAS BIBLIOTECAS NECESSÁRIAS PARA O PROJETO

import scipy

import numpy as np

import pandas as pd

import sklearn

from statistics import mean, stdev

from sklearn.linear_model import LogisticRegression

from pandas.plotting import scatter_matrix

from sklearn import model_selection

from sklearn.metrics import classification_report

from sklearn.metrics import confusion_matrix

from sklearn.metrics import accuracy_score
```


'COMORBIDADE_Disfuncao_renal',
'COMORBIDADE_Dislipidemia',
'COMORBIDADE_DM',
'COMORBIDADE_Doenca_psiquiatica',
'COMORBIDADE_DPOC',
'COMORBIDADE_Encefalopatia_pos_anoxia',
'COMORBIDADE_Etilista',
'COMORBIDADE_Etilista_crônico',
'COMORBIDADE_Ex-tabagista',
'COMORBIDADE_Febre_reumatica',
'COMORBIDADE_gastrite',
'COMORBIDADE_HAS',
'COMORBIDADE_hemiplegia_a_direita',
'COMORBIDADE_Hipotireoidismo',
'COMORBIDADE_ICC',
'COMORBIDADE_insuficiencia_cardiaca',
'COMORBIDADE_Neoplasia',
'COMORBIDADE_Retinopatia_diabética',
'COMORBIDADE_Sindrome_de_imobilidade',
'COMORBIDADE_Sindrome_demencial',
'COMORBIDADE_Transtorno_de_humor',
'COMORBIDADE_ulcera_gastrica',
'COMORBIDADE_usuario_de_drogas',

#SUPORTE ADMISSSIONAL DO PACIENTE

'SUPORTE_ADMISSIONAL_CVC',

'SUPORTE_ADMISSIONAL_Dialise',

'SUPORTE_ADMISSIONAL_IOT',

'SUPORTE_ADMISSIONAL_SVD',

'SUPORTE_ADMISSIONAL_TQT',

'SUPORTE_ADMISSIONAL_VM',

'SUPORTE_ADMISSIONAL_Sedacao',

'SUPORTE_ADMISSIONAL_DVA',

#DROGA VASO ATIVA USADA NO PACIENTE EM ADMISSÃO

'DVA_Noradrenalina',

'DVA_Dobutamina',

'DVA_Amiodarona',

'DVA_Vasopressina',

#SEDAÇÃO USADA NA ADMISSÃO DO PACIENTE

'SEDACAO_Dexmedetomidina',

'SEDACAO_Etomidato',

'SEDACAO_Fentanil',

'SEDACAO_Midazolam',

#MOTIVO DE INTERNAÇÃO ADMISSIONAL

'MOTIVO_INTERNACAO_Acidose_Metabolica',

'MOTIVO_INTERNACAO_Abdomen_agudo',

'MOTIVO_INTERNACAO_Abdomen_agudo_obstrutivo',

'MOTIVO_INTERNACAO_Abscesso_cutaneo',

'MOTIVO_INTERNACAO_Artrite_Piogenica',

'MOTIVO_INTERNACAO_AVC_PREVIO',

'MOTIVO_INTERNACAO_AVE_PREVIO',

'MOTIVO_INTERNACAO_Celulite_em_outros_locais',

'MOTIVO_INTERNACAO_Cetoacidose_diabetica',

'MOTIVO_INTERNACAO_descompensacao_respiratoria',

'MOTIVO_INTERNACAO_Desconforto_toracico',

'MOTIVO_INTERNACAO_diarreia',

'MOTIVO_INTERNACAO_dispnea',

'MOTIVO_INTERNACAO_Doenca_hepatica_toxica',

'MOTIVO_INTERNACAO_DPOC',

'MOTIVO_INTERNACAO_Edema_agudo_de_pulmao',

'MOTIVO_INTERNACAO_Febre',

'MOTIVO_INTERNACAO_furunculo',

'MOTIVO_INTERNACAO_Hemoptise',

'MOTIVO_INTERNACAO_hepatite_nao_classificada',

'MOTIVO_INTERNACAO_insuficiencia_hepatica_cronica',

'MOTIVO_INTERNACAO_Insuficiencia_Renal_Aguda',
'MOTIVO_INTERNACAO_Insuficiencia_respiratoria',
'MOTIVO_INTERNACAO_IRC',
'MOTIVO_INTERNACAO_ITU',
'MOTIVO_INTERNACAO_Lesao_com_edema',
'MOTIVO_INTERNACAO_nauseas',
'MOTIVO_INTERNACAO_Osteomielite_de_femur_distal_dir.',
'MOTIVO_INTERNACAO_Pneumonia_bacteriana',
'MOTIVO_INTERNACAO_Pneumonia_broncoaspirativa',
'MOTIVO_INTERNACAO_RNC',
'MOTIVO_INTERNACAO_Sepse',
'MOTIVO_INTERNACAO_sonolencia',
'MOTIVO_INTERNACAO_Taquidispneia',

'MOTIVO_INTERNACAO_Transornos_comportamentais_devido_ao_uso_de_alcool',

'MOTIVO_INTERNACAO_Transornos_mentais_devido_ao_uso_de_alcool',

'MOTIVO_INTERNACAO_vomito',

#FOCO DA SEPSE EM PACIENTE

'FOCO_Abdominal',

'FOCO_Articular',

'FOCO_bacteremia',

'FOCO_Cateter_Venoso_Central',

'FOCO_partes_moles',

'FOCO_Pele',

'FOCO_Pulmonar',

'FOCO_respiratorio',

'FOCO_Sangue',

'FOCO_Urinario',

#ANTIBIOTICO USADO EM TRATAMENTO:

'ANTIBIOTICO_Amicacina_1g_IV_-_24/24h',

'ANTIBIOTICO_Amicacina_500mg_IV_-_12/12h',

'ANTIBIOTICO_Ampicilina+_Sulbactam_3g_IV',

'ANTIBIOTICO_Azitromicina_500mg_IV_-_24/24h',

'ANTIBIOTICO_Cefepime_1g_IV_-_8/8h',

'ANTIBIOTICO_Cefepime_2g_IV_-_8/8h',

'ANTIBIOTICO_Ceftriaxona_1g',

'ANTIBIOTICO_Ceftriaxona_2g_IV_-_24/24h',

'ANTIBIOTICO_Ciprofloxacino_400mg_IV_-_12/12h',

'ANTIBIOTICO_Clindamicina_600mg_IV',

'ANTIBIOTICO_Fluconazol_200mg_IV_-_24/24h',

'ANTIBIOTICO_Fluconazol_400mg_IV_-_12/12h',

'ANTIBIOTICO_Meropenem_1g_IV',

'ANTIBIOTICO_Meropenem_2g_IV',

'ANTIBIOTICO_Meropenem_500mg_IV',

'ANTIBIOTICO_Metronidazol_500mg_IV_-_8/8h',

'ANTIBIOTICO_Piperacilina+_Tazobactam_2,25mg_IV_-_12/12h',

'ANTIBIOTICO_Piperacilina+_Tazobactam_4,5g_IV',

'ANTIBIOTICO_Polimixina_B_1.000.000_UI_IV',

```

'ANTIBIOTICO_Polimixina_B_500.000UI_IV_-_8/8h',
'ANTIBIOTICO_Polimixina_B_750.000UI_IV_-_12/12h',
'ANTIBIOTICO_Teicoplanina_200mg_IV_-_24/24h',
'ANTIBIOTICO_Teicoplanina_400mg_IV',
'ANTIBIOTICO_Vancomicina_1g_IV',

'Obito'

]

#DF DAS COLUNAS SELECIONADAS

selecaoportientes = df.filter(items=ColunasSelecionadasPortientes)

#QUANTIDADE LINHAS E COLUNAS DO DF

selecaoportientes.shape

# Função para gerar dados sintéticos one-hot-encoded com base na distribuição já existente

def generate_synthetic_data(dataframe, columns, num_samples):

    synthetic_data = np.zeros((num_samples, len(columns)))

    for i in range(num_samples):

        # Amostra aleatória de uma linha existente

        original_sample = dataframe.sample(1, replace=True).to_numpy()

        synthetic_data[i, :] = original_sample

    synthetic_df = pd.DataFrame(synthetic_data, columns=columns)

    return synthetic_df

# Adicionando 2x o valor da amostra original; A amostra será o dataframe original + duas
vezes o seu tamanho

```

```
num_synthetic_samples = 2 * len(selecao_pacientes)

# Gerando dados sintéticos one-hot-encoded

synthetic_data_one_hot = generate_synthetic_data(selecao_pacientes,
ColunasSelecionadasPacientes, num_synthetic_samples)

# Concatenando dados originais e dados sintéticos

selecao_pacientes_aumentado = pd.concat([selecao_pacientes, synthetic_data_one_hot],
ignore_index=True)

#VERIFICANDO SE HÁ VALOR FALTANTE EM ALGUMA COLUNA

selecao_pacientes_aumentado.isnull().sum()

#QUANTIDADE LINHAS E COLUNAS DO NOVO DF

selecao_pacientes_aumentado.shape

#VERIFICANDO OS DADOS

selecao_pacientes_aumentado.describe()

# Dividindo os dados em recursos (X) e variável alvo (y)

col_names_antibiotic = [col for col in selecao_pacientes_aumentado.columns if
'ANTIBIOTICO' in col]

X = selecao_pacientes_aumentado.drop(col_names_antibiotic, axis=1)

y = selecao_pacientes_aumentado[col_names_antibiotic]
```

```
from sklearn.model_selection import cross_val_score

from sklearn.multioutput import MultiOutputClassifier

import pandas as pd

# Multiclass/multioutput https://scikit-learn.org/stable/modules/multiclass.html

# Support multiclass-multioutput

from sklearn.tree import DecisionTreeClassifier, ExtraTreeClassifier

from sklearn.ensemble import ExtraTreesClassifier, RandomForestClassifier

from sklearn.neighbors import KNeighborsClassifier, RadiusNeighborsClassifier

# Multioutput classifier

from sklearn.linear_model import LogisticRegression

from sklearn.tree import DecisionTreeClassifier

from sklearn.neighbors import KNeighborsClassifier

from sklearn.naive_bayes import GaussianNB

from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier

from sklearn.neural_network import MLPClassifier

# Validação cruzada

classifiers = {

    # Support multiclass-multioutput

    "DecisionTree": DecisionTreeClassifier(),

    "ExtraTree": ExtraTreeClassifier(),

    "ExtraTrees": ExtraTreesClassifier(),

    "RandomForest": RandomForestClassifier(),

    "KNeighbors": KNeighborsClassifier(),
```

```
# Multioutput classifier

"Regressão Logística": MultiOutputClassifier(LogisticRegression()),

"Árvore de Decisão": MultiOutputClassifier(DecisionTreeClassifier()),

"KNN": MultiOutputClassifier(KNeighborsClassifier()),

"Naive Bayes": MultiOutputClassifier(GaussianNB()),

"Random Forest": MultiOutputClassifier(RandomForestClassifier()),

"Gradient Boosting": MultiOutputClassifier(GradientBoostingClassifier()),

"NeuralNetwork": MultiOutputClassifier(MLPClassifier())

}

results = {}

for name, clf in classifiers.items():

    print(name)

    scores = cross_val_score(clf, X, y, cv=6, scoring='accuracy')

    results[name] = (scores.mean(), scores.std())

for name, (mean, std) in results.items():

    print(f"{name}:")

    print(f"Acurácia Média: {mean:.2f}")

    print(f"Desvio Padrão: {std:.2f}\n")
```

```
#GRAFICO COMPARANDO O DESEMPENHO DOS MODELOS

import matplotlib.pyplot as plt

classifier_names = list(results.keys())

mean_accuracies = [result[0] for result in results.values()]

std_accuracies = [result[1] for result in results.values()]

x = np.arange(len(classifier_names))

# Criando um gráfico de barras

fig, ax = plt.subplots(figsize=(10, 6)) width = 0.35

rects1 = ax.bar(x - width/2, mean_accuracies, width, label='Acurácia Média')

rects2 = ax.bar(x + width/2, std_accuracies, width, label='Desvio Padrão')

ax.set_xlabel('Classificadores')

ax.set_ylabel('Acurácia')

ax.set_title('Acurácia Média e Desvio Padrão por Classificador')

ax.set_xticks(x)

ax.set_xticklabels(classifier_names, rotation=45, ha="right")

# Legenda para fora da área do gráfico

ax.legend(bbox_to_anchor=(1.05, 1), loc='upper left')

# Adicionando rótulos nas barras

def autolabel(rects):

    for rect in rects:

        height = rect.get_height()

        ax.annotate('{}' .format(round(height, 2)),
```

```

        xy=(rect.get_x() + rect.get_width() / 2, height),

        xytext=(0, 3),

        textcoords="offset points",

        ha='center', va='bottom')

autolabel(rects1)

autolabel(rects2)

fig.tight_layout()

plt.show()

#Dividindo x e y em conjunto de treinamento e teste para todos o modelos para que possa
testar

#individualmente e envelopar como "clf" para preizer medicamento

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=40)

#A PARTIR DESTE PONTO BASTA SELECIONAR O MODELO QUE TEVE MELHOR
DESEMPENHO PARA QUE ELE POSSA PREDIZER A LISTA DE MEDICAMENTO

#DecisionTree

clf = DecisionTreeClassifier()

clf.fit(X_train, y_train)

clf.score(X_test, y_test)

# ExtraTree

clf = ExtraTreeClassifier()

clf.fit(X_train, y_train)

accuracy_extra_tree = clf.score(X_test, y_test)

```

```
print(f"ExtraTree: Accuracy - {accuracy_extra_tree:.4f}")

# RandomForest

clf = RandomForestClassifier()

clf.fit(X_train, y_train)

accuracy_random_forest = clf.score(X_test, y_test)

print(f"RandomForest: Accuracy - {accuracy_random_forest:.4f}")

# Árvore de Decisão Multioutput

clf = MultiOutputClassifier(DecisionTreeClassifier())

clf.fit(X_train, y_train)

accuracy_multioutput_decision_tree = clf.score(X_test, y_test)

print(f"Árvore de Decisão Multioutput: Accuracy -
{accuracy_multioutput_decision_tree:.4f}")

# Gradient Boosting Multioutput

clf = MultiOutputClassifier(GradientBoostingClassifier())

clf.fit(X_train, y_train)

accuracy_multioutput_gradient_boosting = clf.score(X_test, y_test)

print(f"Gradient Boosting Multioutput: Accuracy -
{accuracy_multioutput_gradient_boosting:.4f}")

# Neural Network Multioutput

clf = MultiOutputClassifier(MLPClassifier())

clf.fit(X_train, y_train)

accuracy_multioutput_neural_network = clf.score(X_test, y_test)

print(f"Neural Network Multioutput: Accuracy -
{accuracy_multioutput_neural_network:.4f}")
```

```

#AQUI FINALIZA AS OPÇÕES DE ESCOLHA

#ESCOLHA ALEATÓRIA DE UM NOVO PACIENTE

new_patient = X.sample(1)

new_patient

#CASO O PACIENTE ACIMA TENHA "1" EM ÓBITO SERÁ SUBSTITUIDO POR "0"

new_patient['Obito'] = 0

new_patient

#DEMONSTRA QUAL(IS) ANTIBIOTICO SERA(O) ADMINISTRADO(S) PARA O
PACIENTE SELECIONADO

y.columns[clf.predict(new_patient).ravel().astype(bool)]

```

APÊNDICE B

Resultado da execução do código que compara o desempenho dos classificadores:

DecisionTree: Acurácia Média: 0.95 Desvio Padrão: 0.08	KNeighbors: Acurácia Média: 0.46 Desvio Padrão: 0.14	Naive Bayes: Acurácia Média: 0.62 Desvio Padrão: 0.07
ExtraTree: Acurácia Média: 0.94 Desvio Padrão: 0.09	Regressão Logística: Acurácia Média: nan Desvio Padrão: nan	Random Forest: Acurácia Média: 0.94 Desvio Padrão: 0.09
ExtraTrees: Acurácia Média: 0.94 Desvio Padrão: 0.09	Árvore de Decisão: Acurácia Média: 0.95 Desvio Padrão: 0.08	Gradient Boosting: Acurácia Média: nan Desvio Padrão: nan
RandomForest: Acurácia Média: 0.94 Desvio Padrão: 0.09	KNN: Acurácia Média: 0.46 Desvio Padrão: 0.14	NeuralNetwork: Acurácia Média: 0.91 Desvio Padrão: 0.11

APÊNDICE C

Resultado da execução do código que gera gráfico comparando desempenho dos classificadores:

