

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
ESCOLA POLITÉCNICA
GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO

**COMO A INTELIGÊNCIA ARTIFICIAL PODE AJUDAR NO
MONITORAMENTO PÓS-CIRÚRGICO**

BRUNO HENRIQUE DE OLIVEIRA CARDOSO

GOIÂNIA
2024

BRUNO HENRIQUE DE OLIVEIRA CARDOSO

**COMO A INTELIGÊNCIA ARTIFICIAL PODE AJUDAR NO
MONITORAMENTO PÓS-CIRÚRGICO**

Trabalho de Conclusão de Curso apresentado à banca examinadora, como requisito parcial para a obtenção do título de graduado em Engenharia da Computação.

Orientador: Fabricio Schlag

GOIÂNIA

2024

BRUNO HENRIQUE DE OLIVEIRA CARDOSO

**COMO A INTELIGÊNCIA ARTIFICIAL PODE AJUDAR NO
MONITORAMENTO PÓS-CIRÚRGICO**

Trabalho de Conclusão de Curso apresentado à banca examinadora, como requisito parcial para a obtenção do título de graduado em Engenharia da Computação.

Orientador: Fabricio Schlag

BANCA EXAMINADORA

Prof. Fabricio Schlag

Prof. Nilson Cardoso Amaral

Prof. José Olímpio Ferreira

Goiânia, 04 de dezembro de 2024

AGRADECIMENTOS

Sou grato aos amigos e familiares que me ajudaram e incentivaram no desenvolvimento deste trabalho e decisões tomadas durante esse tempo.

Um agradecimento especial ao professor Fabricio Schlag pela paciência durante esse tempo em orientação.

E por fim um agradecimento a todos que contribuíram com este trabalho e a realização do mesmo.

RESUMO

A integração da Inteligência Artificial (IA) na saúde está revolucionando o panorama médico, trazendo avanços significativos em diagnósticos, tratamentos e gestão de doenças. Tecnologias como *machine learning* e *deep learning* permitem análises de grandes volumes de dados clínicos, imagens médicas e informações genéticas, identificando padrões complexos e relevantes para a medicina personalizada e eficiência nos cuidados de saúde. Diante deste cenário, este estudo tem como objetivo investigar o impacto da IA na saúde, com foco nas suas aplicações práticas, desafios de implementação e benefícios para diagnóstico, tratamento e gestão de doenças. Para tal, utilizou-se metodologia de uma revisão bibliográfica exploratória, visando mapear o estado atual do conhecimento sobre a IA na saúde, identificar lacunas na literatura e entender tendências emergentes. Ante ao exposto, conclui-se que a IA está transformando a prática médica ao oferecer diagnósticos mais precisos e rápidos, personalizar tratamentos e otimizar a eficiência operacional. Exemplos de sucesso incluem o uso de algoritmos para análise de imagens médicas, monitoramento pós-cirúrgico com IA e a implementação de dispositivos médicos inteligentes, como próteses robóticas e sistemas de monitoramento remoto. Esses avanços melhoram a qualidade de vida dos pacientes e reduzem custos operacionais. A contínua evolução da IA, integrada com dados de dispositivos vestíveis e aplicativos móveis, promete ainda mais inovações, reforçando um sistema de saúde centrado no paciente, mais eficiente e acessível.

Palavras-chave: Inteligência Artificial. Diagnóstico Médico. Medicina Personalizada. *Machine Learning*. Inovação em Saúde.

ABSTRACT

The integration of Artificial Intelligence (AI) in healthcare is revolutionizing the medical landscape, bringing innovative advances in diagnosis, treatment and disease management. Technologies such as machine learning and deep learning enable analysis of large volumes of clinical data, medical images and genetic information, identifying complex patterns relevant to personalized medicine and healthcare efficiency. Given this scenario, this study aims to investigate the impact of AI on healthcare, focusing on its practical applications, implementation challenges and benefits for diagnosis, treatment and disease management. To this end, the methodology of an exploratory bibliographic review was used, mapping the current state of knowledge about AI in healthcare, identifying gaps in the literature and understanding emerging trends. In view of the above, it is concluded that AI is evolving medical practice by offering more accurate and faster diagnoses, personalizing treatments and improving operational efficiency. Successful examples include the use of algorithms for medical image analysis, post-surgical monitoring with AI, and the implementation of smart medical devices such as robotic prosthetics and remote monitoring systems. These advances improve patients' quality of life and reduce operating costs. The continued evolution of AI, integrated with data from wearable devices and mobile apps, promises even more innovations, bolstering a more efficient and accessible patient-centric healthcare system.

Keywords: Artificial Intelligence. Medical diagnostic. Personalized Medicine. Machine learning. Innovation in Health.

SUMÁRIO

Lista de Figuras	8
1 INTRODUÇÃO	8
2 A INTELIGÊNCIA ARTIFICIAL NA SAÚDE.....	9
2.1 O PAPEL DA INTELIGÊNCIA ARTIFICIAL NA TRANSFORMAÇÃO DOS CUIDADOS MÉDICOS	11
2.2 A INTELIGÊNCIA ARTIFICIAL NO MONITORAMENTO PÓS CIRÚRGICO.....	12
3 EXEMPLO DE USO DA INTELIGÊNCIA ARTIFICIAL NA SAÚDE.....	15
3.1 Passos Básicos para o Desenvolvimento de um Modelo no Google Colab	15
3.1.1 Configuração do Ambiente.....	15
3.1.2. Carregamento e Pré-processamento dos Dados.....	16
3.1.3. Exploração e Limpeza dos Dados	16
3.1.4. Construção do Modelo Preditivo	16
3.1.5. Otimização e Visualização dos Resultados.....	16
3.1.6. Biblioteca dos dados	17
3.1.7. Bibliotecas utilizadas no modelo	18
3.2. Implementando o modelo.....	19
3.2.1. Análise Exploratória de Dados (EDA).....	19
3.2.2. Dicionário de Dados e Pré-Processamento	19
3.2.3. Processamento e Ajuste dos Dados.....	19
3.2.4. Exploração de Recursos e Alvo	19
3.2.5. Variáveis Numéricas	20
3.2.6. Variáveis Categóricas	22
3.2.7 Mapa de Correlação	23
3.3. Predição	26
3.3.1 Classificadores do Scikit-Learn.....	26
3.3.2 Algoritmos Modernos (Boosted Trees)	30
3.3.3. Explicabilidade do Modelo	30
4 MÉTODO	32
CONCLUSÃO.....	34
CRONOGRAMA.....	35
APÊNDICE	35
REFERÊNCIAS.....	45

Lista de Figuras

Figura 1: Distribuição dos casos

Figura 2: Média das variáveis utilizadas no modelo

Figura 3: Pairplot

Figura 4: Distribuição de recursos categóricos

Figura 6: Talentos cont vs correlação alvo (ponto-bisserial)

Figura 7: Correlação de características categóricas (V de Cramer)

Figura 8: Matriz de confusão

Figura 9: Curvas ROC

Figura 10: Matriz de confusão modelos Iniciais

Figura 11: Regressão logística

Figura 12: LGBM ajustado

Figura 13: Variáveis principais identificadas pelo método de permutação

1 INTRODUÇÃO

A Inteligência Artificial (IA) está se consolidando como uma ferramenta essencial na transformação da saúde, promovendo avanços significativos nos cuidados médicos. Quando integrada em diversas áreas do setor, pode proporcionar soluções inovadoras que aprimoram desde o diagnóstico até o tratamento e a gestão de doenças. Essa tecnologia permite a análise de enormes volumes de dados clínicos e informações genéticas, identificando padrões que auxiliam na medicina personalizada e na melhoria da eficiência dos cuidados de saúde (Kim et al., 2023).

Segundo Kim et al., (2023) além de melhorar a precisão dos diagnósticos, a IA contribui para a personalização dos tratamentos, a otimização de processos administrativos, e o desenvolvimento de dispositivos médicos inteligentes, elevando a qualidade de vida dos pacientes e a eficiência operacional das instituições de saúde. Desta forma, o presente estudo visa elucidar o seguinte questionamento: “Como a inteligência artificial pode otimizar o diagnóstico, o tratamento e a gestão de doenças na área da saúde, e quais são os desafios e oportunidades associados à sua implementação ampla e efetiva?”

Tem-se como objetivo geral: investigar o impacto da inteligência artificial na saúde, com foco nas suas aplicações práticas, desafios de implementação e benefícios potenciais para o diagnóstico, tratamento e gestão de doenças. E como objetivos específicos: analisar as principais aplicações da inteligência artificial no diagnóstico médico e seus efeitos na precisão e rapidez dos diagnósticos; avaliar o papel da inteligência artificial na personalização de tratamentos médicos e na melhoria da eficácia terapêutica; e identificar os desafios técnicos, éticos e operacionais na implementação da inteligência artificial em ambientes clínicos e sugerir estratégias para superá-los.

A relevância desta pesquisa, justifica-se, pois, o crescente uso da inteligência artificial na saúde representa uma revolução com potencial para redefinir práticas médicas e melhorar significativamente os resultados clínicos. A capacidade da IA de analisar e interpretar grandes volumes de dados em tempo real oferece uma oportunidade única para aprimorar o diagnóstico, personalizar tratamentos e aumentar a eficiência operacional em instituições de saúde.

Para alcançar os objetivos propostos, será realizada uma revisão bibliográfica da literatura disponível sobre a aplicação da inteligência artificial na saúde. A metodologia adotada será do tipo exploratória, visando mapear o

estado atual do conhecimento, identificar lacunas na literatura e entender as tendências emergentes.

2 A INTELIGÊNCIA ARTIFICIAL NA SAÚDE

A inteligência artificial na saúde refere-se à aplicação de tecnologias de IA, como machine learning e deep learning, para melhorar o diagnóstico, tratamento e gestão de doenças. Essas ferramentas permitem que sistemas computacionais analisem grandes volumes de dados clínicos, imagens médicas e informações genéticas para identificar padrões, fazer previsões e tomar decisões com precisão e rapidez. Desde a interpretação de exames de imagem até a personalização de tratamentos com base nas características individuais do paciente, a inteligência artificial está revolucionando a prática médica, oferecendo novas perspectivas para a medicina personalizada e aprimorando a eficiência dos cuidados de saúde (Wu et al., 2022).

Além disso, Chen et al., (2022) destaca que, a inteligência artificial na saúde também está impulsionando a inovação em dispositivos médicos, como próteses robóticas e dispositivos de monitoramento remoto. Essas tecnologias não só melhoram a qualidade de vida dos pacientes, proporcionando maior mobilidade e autonomia, mas também permitem uma gestão mais eficaz das condições crônicas, reduzindo internações hospitalares e custos associados. Com a contínua evolução da inteligência artificial e sua integração cada vez maior na prática clínica, espera-se que o campo da saúde experimente avanços significativos, proporcionando melhores resultados para pacientes e profissionais de saúde. Para os autores:

A inserção entre inteligência artificial (IA) e saúde está redefinindo o panorama médico, gerando benefícios palpáveis em diversas frentes. Próteses robóticas, prontuários eletrônicos, diagnósticos precisos e cirurgias assistidas por robôs são apenas algumas das inúmeras aplicações que estão transformando o setor. Essas inovações não só aprimoram o atendimento ao paciente, mas também contribuem para a redução de custos e a melhoria dos resultados clínicos (Chen et al., 2022, p.18).

Um exemplo marcante dessa evolução é o Hospital 9 de Julho, em São Paulo, que se tornou pioneiro ao adotar a inteligência artificial para ampliar a segurança dos pacientes. Por meio de sensores inteligentes desenvolvidos pela Microsoft, o hospital monitora movimentos não usuais em tempo real, prevenindo

quedas e proporcionando um ambiente mais seguro para os pacientes (Hospital 9 de julho, 2018).

Outro destaque é o Tensorflow, uma biblioteca de IA desenvolvida pelo Google para facilitar o desenvolvimento de soluções inteligentes. Utilizado para o diagnóstico de retinopatia diabética, esse sistema agiliza triagens e prioriza atendimentos em unidades de emergência, otimizando a gestão dos recursos médicos. A IBM também está na vanguarda com o Watson, um programa de IA que utiliza deep learning para auxiliar no tratamento do câncer. Com base em dados clínicos e genéticos dos pacientes, o Watson sugere tratamentos personalizados, aumentando as chances de sucesso terapêutico (Portal Telemedicina, 2019).

Além disso, a análise de imagens médicas, como raios X e ressonância magnética, é agilizada e aprimorada por algoritmos de IA, garantindo diagnósticos mais precisos e rápidos. Esses avanços promovem uma medicina mais assertiva e eficiente, beneficiando tanto profissionais quanto pacientes (Portal Telemedicina, 2019).

Com empresas como Google, IBM, Microsoft e Apple investindo cada vez mais no desenvolvimento de soluções de IA para a saúde, o futuro promete ainda mais inovações. A integração e análise de dados coletados não apenas em ambientes clínicos, mas também por meio de dispositivos *wearables* e aplicativos móveis, prometem revolucionar ainda mais a forma como lidamos com a saúde (Schulze, 2024).

Para Schulze (2024) a inserção da inteligência artificial no cenário da saúde representa uma mudança significativa e multifacetada, impactando tanto os processos clínicos quanto os modelos de negócios no setor. A integração dessa tecnologia implica uma redefinição dos padrões operacionais e um realinhamento das práticas profissionais. A capacidade da inteligência artificial de processar e analisar grandes volumes de dados clínicos, combinada com algoritmos avançados de aprendizado de máquina e *deep learning*, permite a geração de insights valiosos e a identificação de padrões complexos que podem ser cruciais para diagnósticos precisos e tratamentos eficazes.

De acordo com Zaparolli (2022) os profissionais de saúde estão sendo desafiados a adotar uma abordagem mais orientada por dados em suas práticas clínicas. Não se trata de substituir a expertise médica, mas sim de complementá-la com uma visão mais informada e baseada em evidências. Os médicos precisam agora interpretar estatísticas e probabilidades, utilizando as

informações fornecidas pela inteligência artificial para tomar decisões mais fundamentadas em relação ao diagnóstico e ao plano de tratamento de seus pacientes. Isso exige um novo conjunto de habilidades e uma mentalidade adaptável por parte dos profissionais de saúde, que devem estar abertos à integração das tecnologias de IA em sua prática diária.

Além disso, a presença da inteligência artificial está reconfigurando as dinâmicas de trabalho dentro das instituições de saúde. Antes improvável, a consulta de toda a literatura médica para obter insights sobre casos específicos agora se torna viável com o auxílio de algoritmos inteligentes. Essa capacidade de acessar e processar informações em larga escala amplia as possibilidades de diagnóstico e tratamento, ao mesmo tempo em que oferece aos profissionais uma fonte valiosa de referência e suporte em suas decisões clínicas (Zaparolli, 2022).

O surgimento das Health Techs, *startups* focadas em soluções tecnológicas para a saúde, é um reflexo claro desse cenário em evolução. Essas empresas estão na vanguarda da inovação, desenvolvendo e implementando tecnologias como impressão 3D, pílulas digitais, Big Data e algoritmos preventivos para melhorar a eficiência e a acessibilidade dos cuidados de saúde. O investimento substancial nessas *startups* demonstra o reconhecimento do potencial transformador da inteligência artificial no setor de saúde e sua capacidade de impulsionar mudanças positivas em larga escala (Schulze, 2024).

2.1 O PAPEL DA INTELIGÊNCIA ARTIFICIAL NA TRANSFORMAÇÃO DOS CUIDADOS MÉDICOS

A tecnologia permeia variados setores da economia, trazendo inovações e soluções que impactam o dia a dia da sociedade, melhorando sua qualidade de vida. Na área da saúde, especificamente, a tecnologia desempenha um papel fundamental ao tornar os processos mais eficientes, reduzindo os custos operacionais de clínicas, hospitais e consultórios (Zaparolli, 2024).

Segundo Zaparolli (2024) a inteligência artificial (IA) na saúde é um exemplo dessa revolução, buscando simplificar e aprimorar os processos cotidianos. Desde a otimização de procedimentos técnicos e logísticos até o desenvolvimento de novas tecnologias aplicadas à IA, o objetivo é transformar os processos e procedimentos em toda a indústria médica, inclusive o histórico de diagnósticos de cada paciente.

A IA tem sido uma ferramenta promissora na melhoria da medicina de emergência, graças à sua capacidade de digitalização e armazenamento de informações. Diversas aplicações estão sendo utilizadas, como softwares e bases de dados mais eficazes, facilitando a descoberta de doenças e o desenvolvimento de planos de tratamento mais seguros e personalizados. O uso de prontuários eletrônicos agiliza o acesso ao histórico do paciente, além de garantir a segurança e organização dos dados (Zaparolli, 2024).

Por meio da nuvem computacional, a IA possibilita o armazenamento e recuperação eficientes de informações, proporcionando maior segurança aos pacientes e liberando os profissionais de saúde para se concentrarem no atendimento direto aos pacientes. A análise de big data permite um melhor rastreamento do histórico familiar de cada indivíduo, contribuindo para a prevenção e detecção precoce de doenças (Awarahman et al., 2022).

Além disso, a IA tem sido aplicada em tarefas administrativas, oferecendo suporte na extração e análise de grandes volumes de dados, o que possibilita um atendimento mais personalizado e eficiente. A otimização do atendimento, através do agendamento eletrônico e automação de tarefas, promove uma experiência mais humanizada para os pacientes (Awarahman et al., 2022).

Na área diagnóstica, a IA demonstra uma precisão cada vez maior, superando até mesmo especialistas em certos casos. Algoritmos avançados são capazes de analisar imagens médicas e identificar padrões que podem passar despercebidos aos olhos humanos, contribuindo para diagnósticos mais rápidos e precisos (Awarahman et al., 2022).

Desta forma, nota-se que, em cirurgias, a IA tem sido utilizada para aumentar a precisão e reduzir os períodos de internação, através de procedimentos menos invasivos e recuperação mais rápida. O uso de robôs cirúrgicos, aliados à IA, permite uma execução mais segura e eficiente de procedimentos complexos.

2.2 A INTELIGÊNCIA ARTIFICIAL NO MONITORAMENTO PÓS CIRÚRGICO

A Inteligência Artificial (IA) vem revolucionando diversos campos da medicina, incluindo o monitoramento pós-cirúrgico, onde desempenha um papel crucial na detecção precoce de complicações e na previsão dos resultados dos pacientes. O uso de algoritmos de aprendizado de máquina para analisar dados pós-operatórios, como sinais vitais e progressos na recuperação, tem

proporcionado uma abordagem mais proativa e eficiente, capacitando as equipes médicas a intervirem prontamente e melhorar significativamente a segurança do paciente (Pereira, 2024).

De acordo com o Portal Telemedicina (2019) a fase pós-cirúrgica é um período crítico em que a vigilância contínua dos pacientes é essencial para a detecção precoce de complicações, como infecções, sangramentos ou falhas na cicatrização. Tradicionalmente, o monitoramento pós-operatório envolvia visitas regulares ao hospital ou consultas presenciais, onde os médicos avaliavam os pacientes com base em relatos e exames clínicos. Contudo, esse método possui limitações, especialmente em detectar complicações que se desenvolvem rapidamente entre as consultas.

Para Zapparoli (2024) com a introdução da IA, esse cenário mudou drasticamente. Algoritmos avançados podem analisar grandes volumes de dados em tempo real, identificando padrões que podem indicar a deterioração do estado do paciente antes que os sintomas se tornem evidentes. A IA utiliza dados coletados de dispositivos de monitoramento, registros eletrônicos de saúde e até de tecnologias vestíveis, integrando informações como sinais vitais, níveis de atividade e padrões de sono. Esses dados são processados continuamente, permitindo uma resposta mais rápida e precisa a qualquer alteração no estado de saúde do paciente.

Um exemplo promissor do uso da IA no monitoramento pós-cirúrgico é o projeto “Implantação de Chatbot Inteligente para Monitoramento da Jornada de Pacientes em Cirurgias Limpas” desenvolvido pelo Hospital das Clínicas da Universidade Federal de Pernambuco (HC-UFPE) em parceria com a *startup* Aicury. Este projeto, parte de uma iniciativa de saúde digital, visa monitorar a evolução da cicatrização em pacientes que receberam alta após cirurgias limpas, que são procedimentos realizados em condições estéreis sem a necessidade de drenagem e que não penetram em tratos digestivo, respiratório ou urinário (UFPE, 2023).

O projeto utiliza um assistente virtual inteligente, ou *chatbot*, alimentado por tecnologia de IA, para acompanhar a jornada do paciente após a cirurgia. Este *chatbot*, capaz de interagir por texto e voz, fornece lembretes de cuidados, atualizações sobre o estado de saúde e alertas em caso de possíveis complicações. Essa comunicação eficiente e instantânea entre paciente e equipe médica não só melhora a experiência do paciente, mas também promove uma recuperação mais segura e eficiente (UFPE, 2023).

A enfermeira Virgínia Menezes, da Comissão de Controle das Infecções Relacionadas à Assistência à Saúde (CCIRAS) do HC, destacou que a integração entre saúde e tecnologia, promovida pelo uso de IA, é fundamental para melhorar o acompanhamento e a eficiência do atendimento ao paciente. Essa abordagem não apenas reduz a incidência de complicações pós-operatórias e a necessidade de novas internações, mas também diminui os custos associados e melhora a satisfação dos pacientes (UFPE, 2023).

A utilização de IA no monitoramento pós-cirúrgico oferece várias vantagens importantes, conforme descritas por Fagundes (2023):

1. **Deteção Precoce de Complicações:** A IA pode identificar sinais sutis de complicações, permitindo intervenções mais rápidas e eficazes, o que é crucial para prevenir o agravamento das condições do paciente.
2. **Monitoramento Contínuo:** Diferentemente das consultas tradicionais, a IA permite um monitoramento contínuo e em tempo real, garantindo que os dados do paciente sejam avaliados constantemente.
3. **Personalização do Cuidado:** Algoritmos de IA podem aprender com os dados específicos de cada paciente, permitindo a personalização dos alertas e recomendações de cuidados de acordo com as necessidades individuais.
4. **Redução de Custos:** A detecção precoce e o monitoramento eficiente reduzem a necessidade de internações adicionais e procedimentos de emergência, diminuindo os custos gerais do sistema de saúde.
5. **Melhora na Experiência do Paciente:** Ferramentas como *chatbots* proporcionam aos pacientes uma comunicação fácil e acessível com a equipe médica, aumentando a confiança e o conforto durante o processo de recuperação.

O futuro da IA no monitoramento pós-operatório é promissor, com o desenvolvimento contínuo de tecnologias que integrarão ainda mais o cuidado médico com soluções digitais. Espera-se que novas inovações permitam um monitoramento ainda mais preciso e personalizado, com o potencial de incluir análise preditiva mais sofisticada, que pode antecipar complicações com base em dados genéticos, históricos médicos e outros fatores de risco (Fagundes, 2023).

Além disso, Pereira et al., (2024) destaca que a integração de IA com dispositivos de Internet das Coisas (IoT) e sensores vestíveis ampliará as capacidades de monitoramento remoto, fornecendo dados em tempo real sobre

diversos parâmetros de saúde dos pacientes. Essa abordagem promete não apenas melhorar a recuperação pós-cirúrgica, mas também otimizar a gestão de mais doenças crônicas e outras condições de saúde, promovendo um sistema de saúde mais eficiente e centrado no paciente.

3 EXEMPLO DE USO DA INTELIGÊNCIA ARTIFICIAL NA SAÚDE

No âmbito deste trabalho foi abordado temas como inteligência artificial (IA), *machine learning*, entre outros contextos de abordagens, pra fins acadêmicos utilizando *google colab* e *datasets* retirados do Kaggle, foi introduzido um sistema de inteligência artificial que a partir de um *dataset* (conjunto de dados de entrada) realize primeiramente uma análise exploratória dos dados e em sequência realize a predição de casos com problemas cardiovasculares , podendo assim prevenir outros casos similares.

O avanço das tecnologias de aprendizado de máquina tem possibilitado aplicações inovadoras em diversas áreas, incluindo a saúde. Modelos preditivos podem ser utilizados para identificar riscos de problemas cardiovasculares, auxiliando profissionais de saúde a tomar decisões mais informadas e a implementar intervenções preventivas. Uma plataforma amplamente usada para prototipagem e desenvolvimento de tais modelos é o Google Colab.

O Google Colab é um ambiente de desenvolvimento integrado (IDE) baseado na nuvem que permite escrever e executar código Python diretamente em seu navegador. Ele oferece recursos como suporte a bibliotecas populares de aprendizado de máquina, GPU gratuita para aceleração de processamento e integração com o Google Drive para armazenamento de dados. Por ser acessível e fácil de usar, o Colab é uma ferramenta ideal tanto para iniciantes quanto para profissionais.

3.1 Passos Básicos para o Desenvolvimento de um Modelo no Google Colab

3.1.1 Configuração do Ambiente

O primeiro passo é garantir que o ambiente esteja configurado corretamente. Isso inclui a instalação de bibliotecas como *pandas*, *numpy*,

scikit-learn e *tensorflow* (caso necessário). No Colab, muitas dessas bibliotecas já estão pré-instaladas.

3.1.2. Carregamento e Pré-processamento dos Dados

O modelo preditivo requer dados históricos para treinar e testar sua eficácia. No caso de problemas cardiovasculares, *datasets* como o Heart Disease UCI (disponível em plataformas como o Kaggle) podem ser utilizados. O Colab permite carregar os dados diretamente do Google Drive ou de outras fontes na web.

3.1.3. Exploração e Limpeza dos Dados

Nesta etapa, é necessário verificar a qualidade dos dados, identificar valores ausentes e realizar ajustes como normalização, codificação de variáveis categóricas e tratamento de *outliers*.

3.1.4. Construção do Modelo Preditivo

O modelo pode ser construído usando algoritmos como regressão logística, árvores de decisão ou redes neurais, dependendo da complexidade do problema. A biblioteca *scikit-learn* é uma excelente escolha para começar.

3.1.5. Otimização e Visualização dos Resultados

Após a construção inicial, é possível otimizar os hiperparâmetros do modelo e utilizar gráficos para visualizar a performance, como curvas ROC ou matrizes de confusão.

O Google Colab é uma plataforma poderosa para a prototipagem de modelos de aprendizado de máquina na área da saúde. Ele permite que pesquisadores e profissionais explorem dados, treinem modelos e avaliem resultados de maneira rápida e eficiente. Ao desenvolver um modelo para

predição de problemas cardiovasculares, o uso de dados confiáveis e técnicas robustas é essencial para garantir precisão e aplicabilidade no mundo real.

Utilizando o UCI Heart Disease para prever riscos cardiovasculares foi desenvolvido um *notebook* no google Colab. Ele inclui etapas como análise exploratória detalhada para identificar padrões, pré-processamento de dados (tratamento de valores nulos e codificação de variáveis categóricas) e a implementação de modelos como regressão logística, árvores de decisão e outros algoritmos. O desempenho é avaliado por métricas como matriz de confusão, acurácia e curvas ROC, com o objetivo de otimizar previsões e extrair *insights* clínicos significativos.

As doenças cardíacas, ou cardiovasculares, abrangem uma série de condições que afetam o coração e o sistema circulatório. Elas são uma das principais causas de incapacidade no mundo, já que o coração, essencial para o corpo, também influencia o funcionamento de outros órgãos. Entre os tipos mais comuns estão problemas nas artérias coronárias, disfunções nas válvulas cardíacas e o aumento anormal do coração. Essas condições podem levar a insuficiência cardíaca, ataques cardíacos e outros problemas sérios de saúde.

3.1.6. Biblioteca dos dados

O conjunto de dados contém informações para prever doenças cardiovasculares com as seguintes variáveis:

- **Idade (age):** idade em anos.
- **Sexo (sex):** 1 = masculino, 0 = feminino.
- **Dor no peito (cp):** 0 a 3, de angina típica a dor assintomática.
- **Pressão arterial (trestbps):** em mmHg.
- **Colesterol (chol):** em mg/dl.
- **Glicemia jejum (fbs):** 1 = alta (> 120 mg/dl), 0 = normal.
- **ECG em repouso (restecg):** 0 a 2, variando de normal a anormalidades graves.
- **Frequência cardíaca máxima (thalach):** atingida.
- **Angina induzida (exang):** 1 = sim, 0 = não.

- **Depressão ST (oldpeak):** diferença entre exercício e repouso.
- **Inclinação do ST (slope):** 0 a 2, inclinação ascendente a descendente.
- **Vasos (ca):** número de vasos detectados (0-3).
- **Thalassemia (thal):** 1 a 3, defeitos cardíacos fixos, reversíveis ou normais.
- **Alvo (target):** 0 = sem doença, 1 = doença presente.

3.1.7. Bibliotecas utilizadas no modelo

Para o desenvolvimento deste trabalho foi-se empregado as seguintes bibliotecas:

- **Pandas** para manipulação e análise de dados.
 - **Vantagens:** Excelente para manipulação e análise de dados tabulares, suporte para leitura/escrita de diversos formatos (CSV, Excel, SQL), funções integradas para limpeza e transformação de dados.
 - **Desvantagens:** Não é otimizada para grandes conjuntos de dados (problemas de desempenho e uso de memória).
- **NumPy** para operações numéricas.
 - **Vantagens:** Alta performance em operações numéricas devido ao uso de arrays otimizados, integração com outras bibliotecas.
 - **Desvantagens:** Requer aprendizado adicional para manipular arrays multidimensionais.
- **Seaborn e Matplotlib** para visualizações.
 - **Vantagens:** Produzem gráficos de alta qualidade, personalizáveis (Matplotlib); abstração para gráficos mais fáceis de criar (Seaborn).
 - **Desvantagens:** Matplotlib tem uma curva de aprendizado íngreme; Seaborn depende do Matplotlib, limitado em gráficos complexos.
- **Scikit-learn** para modelagem, divisão de dados, e avaliação do desempenho dos modelos.
 - **Vantagens:** Ampla gama de algoritmos de aprendizado de máquina, documentação rica, fácil integração com outras bibliotecas.

- **Desvantagens:** Não suporta GPU, tornando-se menos eficiente em modelos muito grandes ou complexos

3.2. Implementando o modelo

3.2.1. Análise Exploratória de Dados (EDA)

3.2.2. Dicionário de Dados e Pré-Processamento

- O conjunto de dados contém 303 observações e 14 colunas (13 características + variável alvo). Após limpeza, restaram 296 amostras.
- Tipos de dados: seis variáveis numéricas e sete categóricas.

3.2.3. Processamento e Ajuste dos Dados

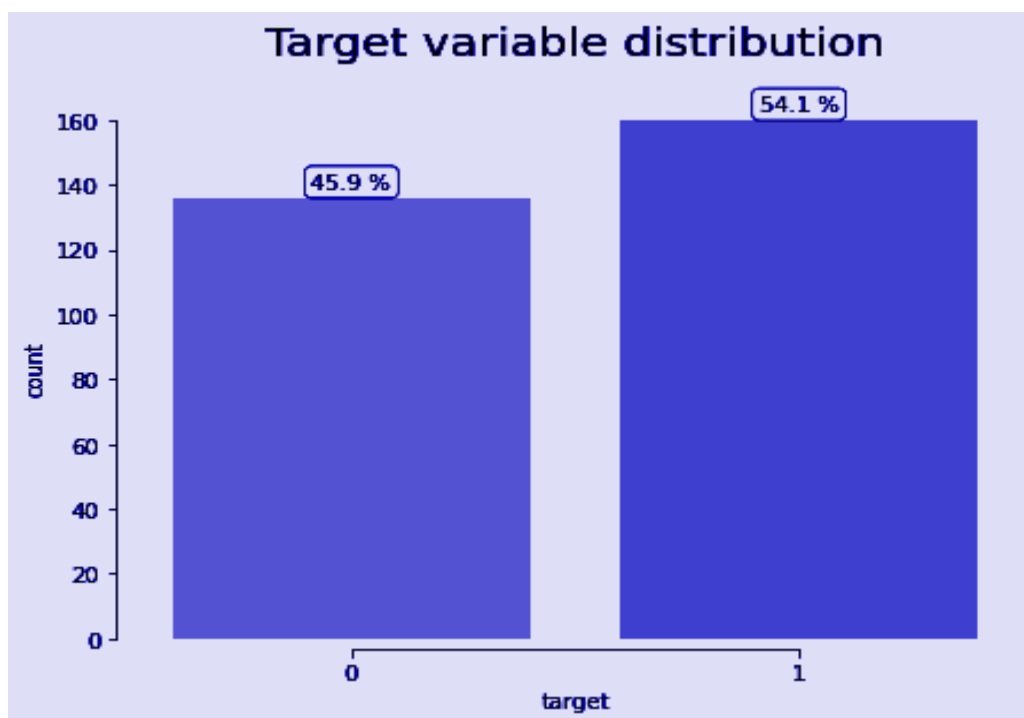
- Linhas defeituosas foram removidas.
- Nomes das colunas foram renomeados para facilitar a compreensão (usando informações do repositório UCL).

3.2.4. Exploração de Recursos e Alvo

1. Distribuição do Alvo:

- a. Equilíbrio razoável: 54% dos casos apresentam doença, 46% não. Conforme figura abaixo

Figura 1: Distribuição dos casos



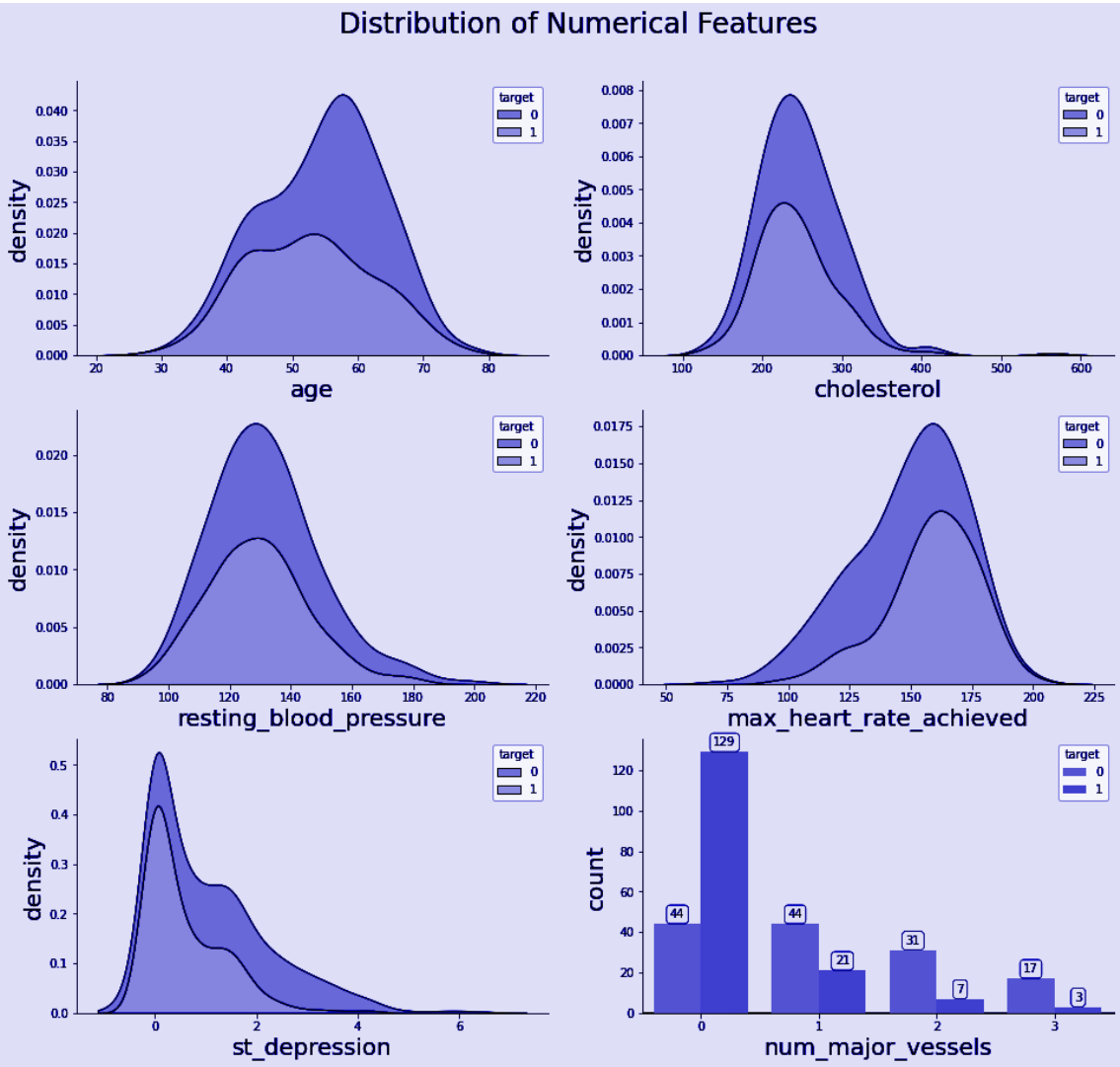
Fonte: O autor

3.2.5. Variáveis Numéricas

Média das variáveis utilizadas no modelo conforme instruções e imagem abaixo

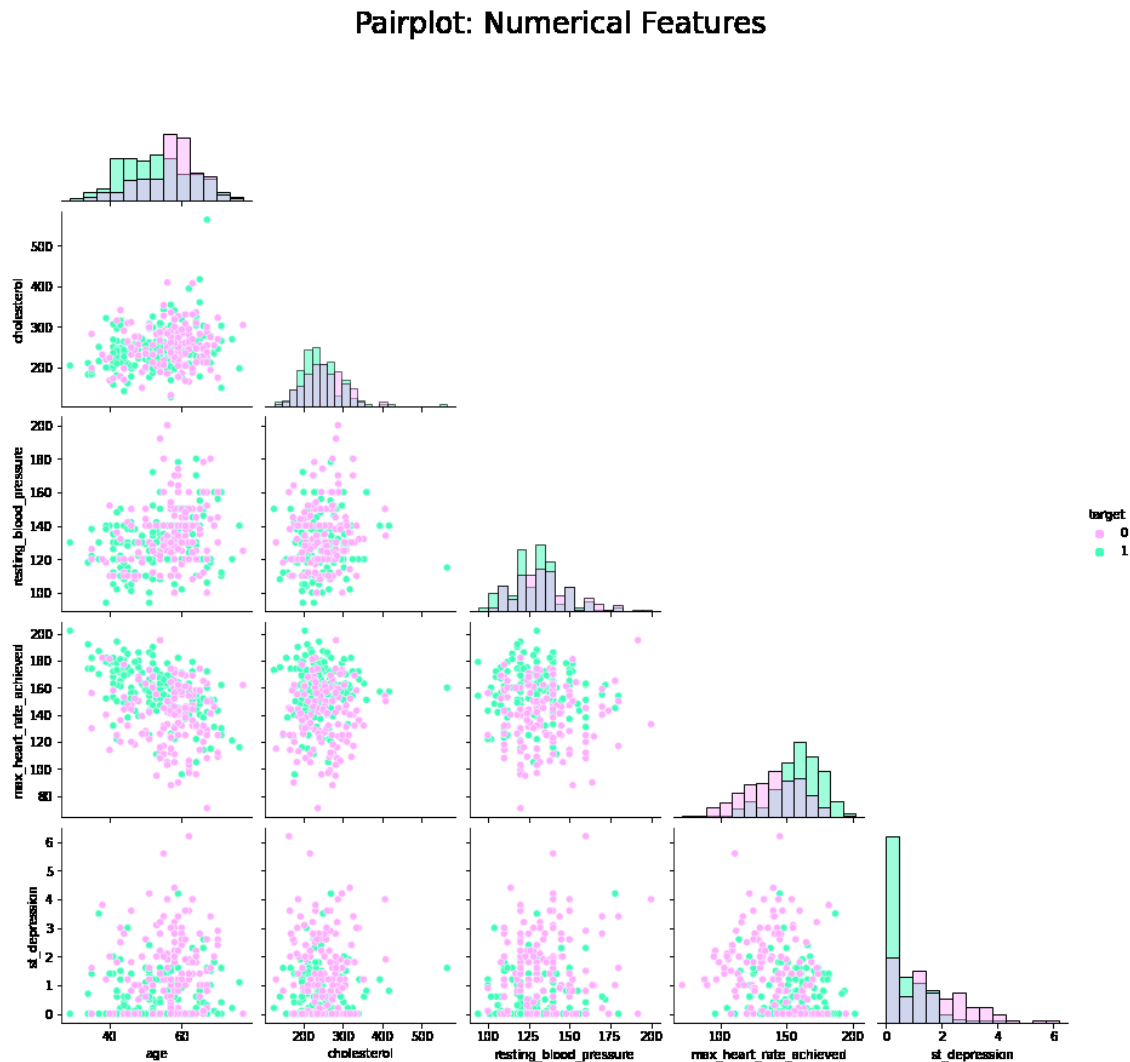
- Idade:** Média de 54,5 anos (mínimo 29, máximo 77).
- Colesterol:** Média de 247,15 mg/dl (saudável: <200 mg/dl)
- Pressão arterial:** Média de 131 mmHg (mínimo 94, máximo 200)
- Frequência cardíaca máxima:** Média de 149,5 bpm.
- Depressão ST:** Média de 1,06 (máximo 6,2).
- Número de vasos maiores:** Média de 0,68 (máximo 3).

Figura 2: Média das variáveis utilizadas no modelo



Fonte: O autor

Figura 3: Pairplot



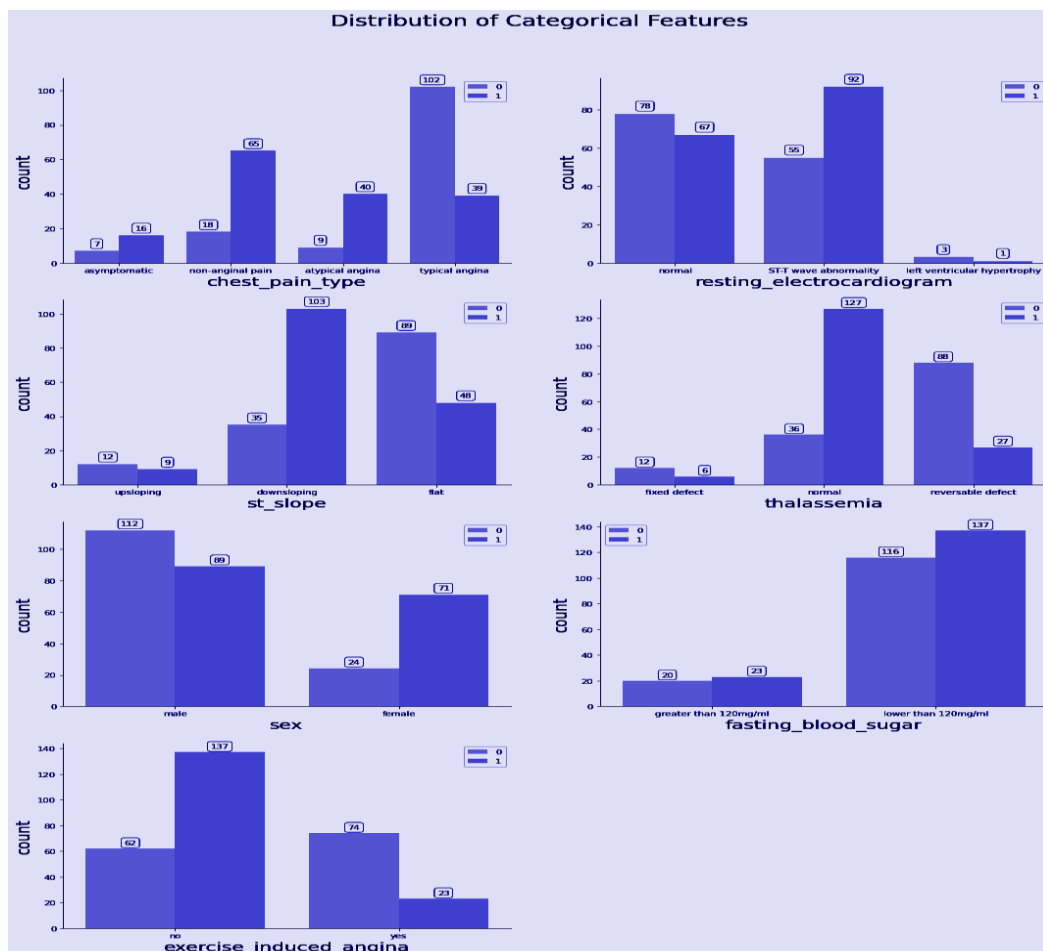
Fonte: <https://www.kaggle.com/code/desalegngeb/heart-disease-predictions>

3.2.6. Variáveis Categóricas

- **Dor no peito:** Angina atípica ou dor não anginosa correlacionam-se com maior probabilidade de doença.
- **Eletrocardiograma em repouso:** Anormalidades ST-T estão associadas à doença.
- **Declive ST:** Declive descendente tem forte relação com doenças cardíacas.
- **Talassemia:** Defeitos normais estão mais associados à doença.

- **Glicose em jejum:** Níveis baixos (<120 mg/ml) aumentam a probabilidade de doença.
- **Angina induzida por exercício:** Ausência correlaciona-se com maior probabilidade de doença.
- **Sexo:** Mulheres têm maior prevalência de doenças cardíacas na amostra.

Figura 4: Distribuição de recursos categóricos



Fonte: O autor

3.2.7 Mapa de Correlação

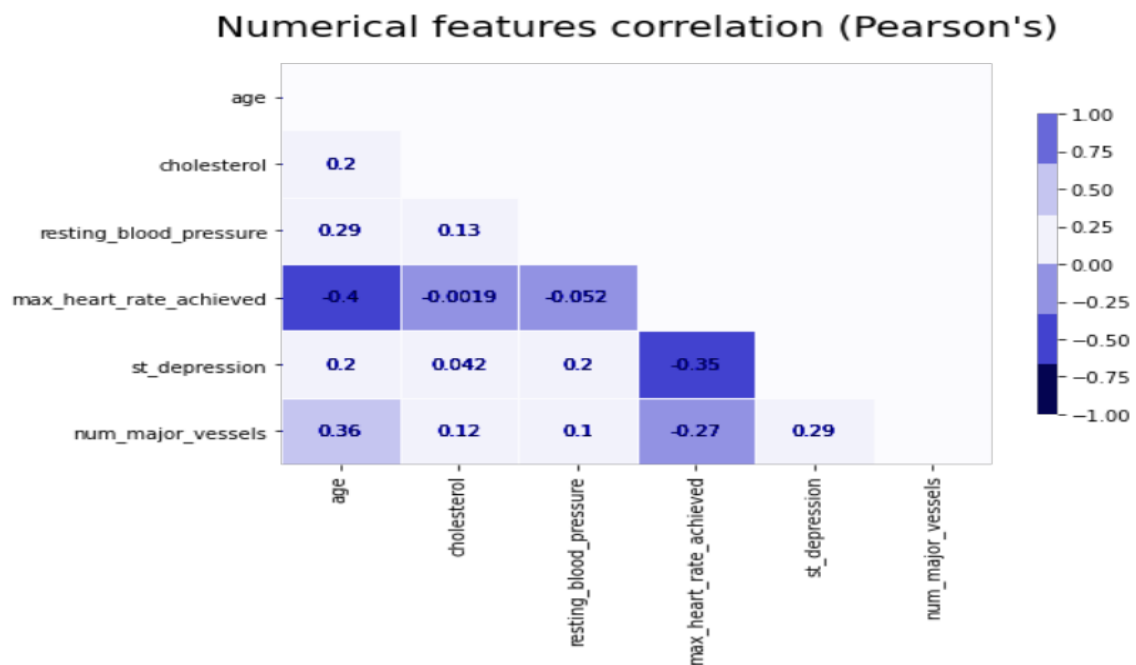
- A correlação entre variáveis é geralmente fraca.
- Principais correlações com a variável alvo:

- **Numéricas:** Número de vasos maiores (-0,47), frequência cardíaca máxima (0,43), depressão ST (-0,43).
- **Categóricas:** Tipo de dor no peito, número de vasos maiores, talassemia e angina induzida por exercício.

- **Resumo do EDA:**

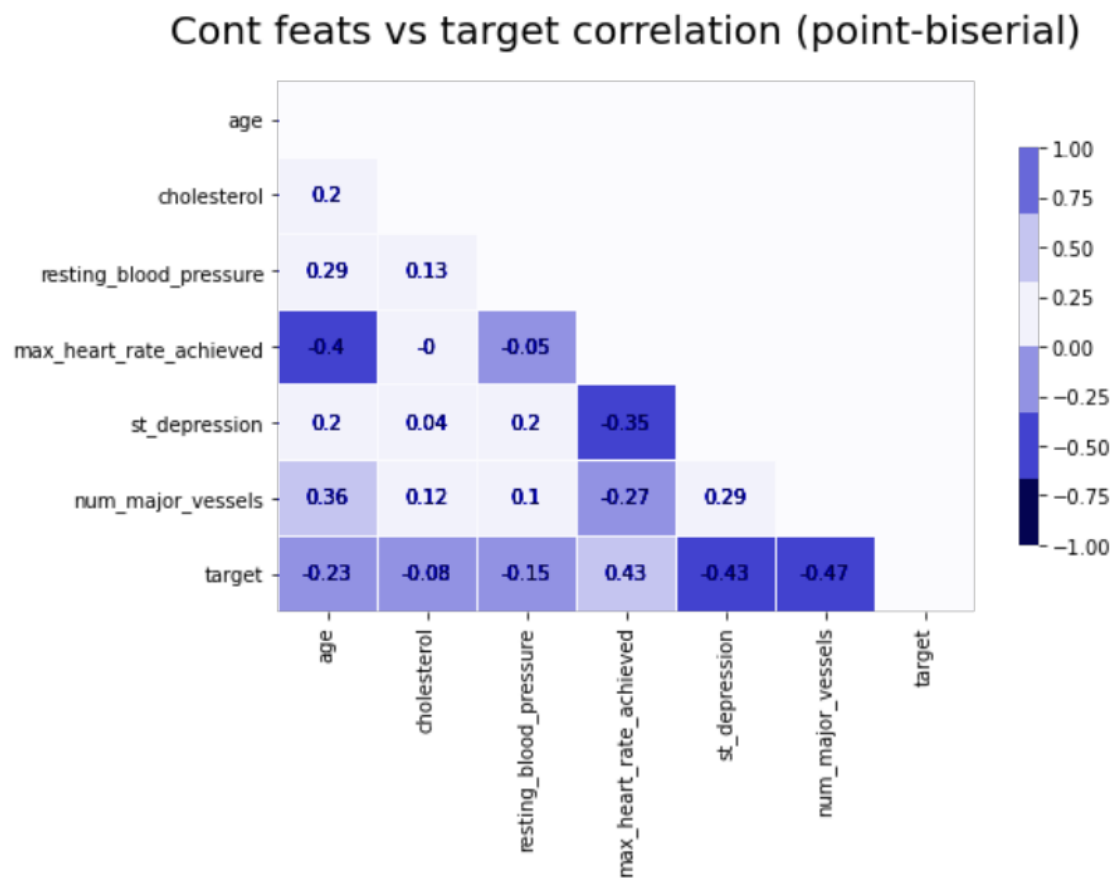
- Dados bem balanceados e sem valores ausentes.
- Principais preditores: tipo de dor no peito, número de vasos maiores, talassemia, angina induzida, frequência cardíaca máxima e depressão ST.

Figura 5: Correlação de características numéricas (Pearson)



Fonte: O autor

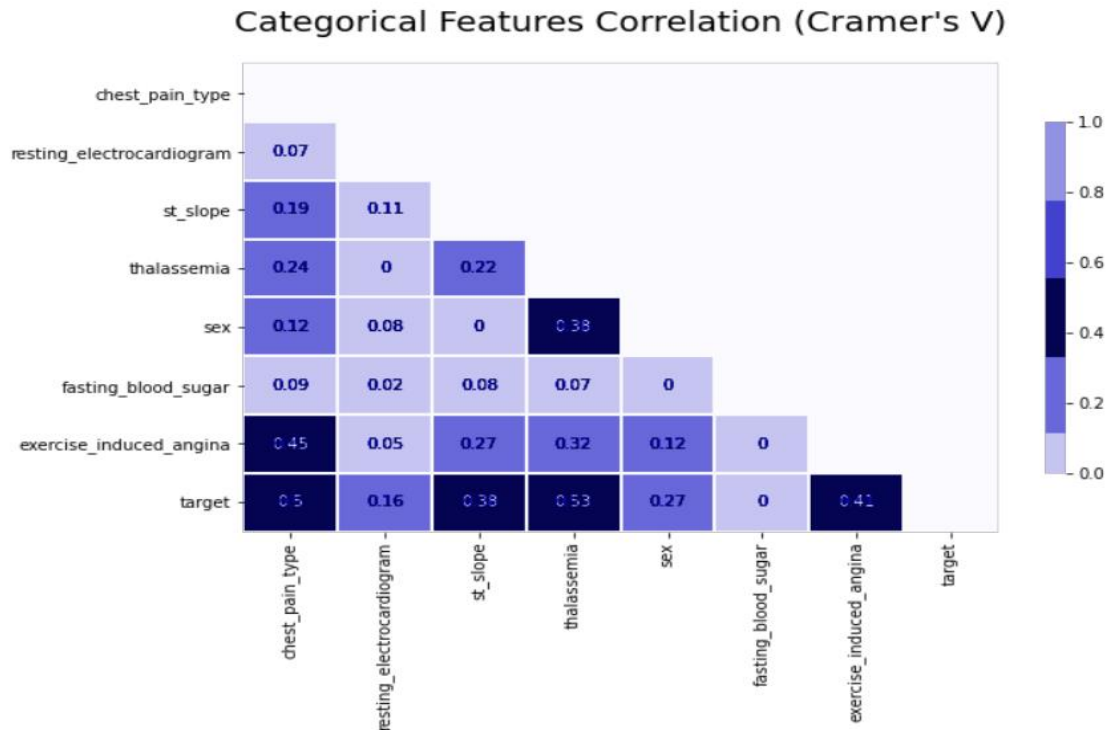
Figura 6: Talentos cont vs correlação alvo (ponto-bisserial)



Fonte: O autor

O V de Cramer é uma medida estatística utilizada para avaliar a força da associação entre duas variáveis categóricas. É derivado do qui-quadrado e ajustado pelo tamanho da amostra e pelo menor número de categorias entre as variáveis analisadas, garantindo uma padronização em um intervalo de 0 a 1. Valores próximos de 0 indicam uma associação fraca ou inexistente, enquanto valores próximos de 1 indicam uma associação forte. É especialmente útil em tabelas de contingência, quando se deseja comparar a dependência entre variáveis categóricas de diferentes dimensões.

Figura 7: Correlação de características categóricas (V de Cramer)



Fonte: O autor

3.3. Predição

3.3.1 Classificadores do Scikit-Learn

- Modelos testados: Regressão Logística, KNN, SVM, Decision Tree, Random Forest, AdaBoost, Gradient Boosting, Naive Bayes, entre outros.
- A matriz de confusão é uma ferramenta usada para avaliar o desempenho de modelos de classificação, apresentando os resultados de forma organizada. Ela é estruturada como uma tabela que compara as previsões feitas pelo modelo com os valores reais, dividindo os resultados em quatro categorias: verdadeiros positivos (quando o modelo acerta ao prever um caso positivo), verdadeiros negativos (quando acerta ao prever um caso negativo), falsos positivos (quando prevê positivo, mas está errado) e falsos negativos (quando prevê negativo, mas deveria ser positivo). Essa matriz é muito útil porque permite analisar os erros e acertos do modelo em detalhes, ajudando a identificar onde ele pode ser melhorado e qual é o impacto desses erros no contexto do problema real.

Figura 8: Matriz de confusão

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

Fonte: O autor

- **Métricas de Avaliação**

- **Acurácia:** Proporção de predições corretas.
- **Recall (sensibilidade):** Identificação de casos positivos.
- **Precisão:** Percentual de positivos preditos corretamente.
- **F1-Score:** Média harmônica entre precisão e recall.

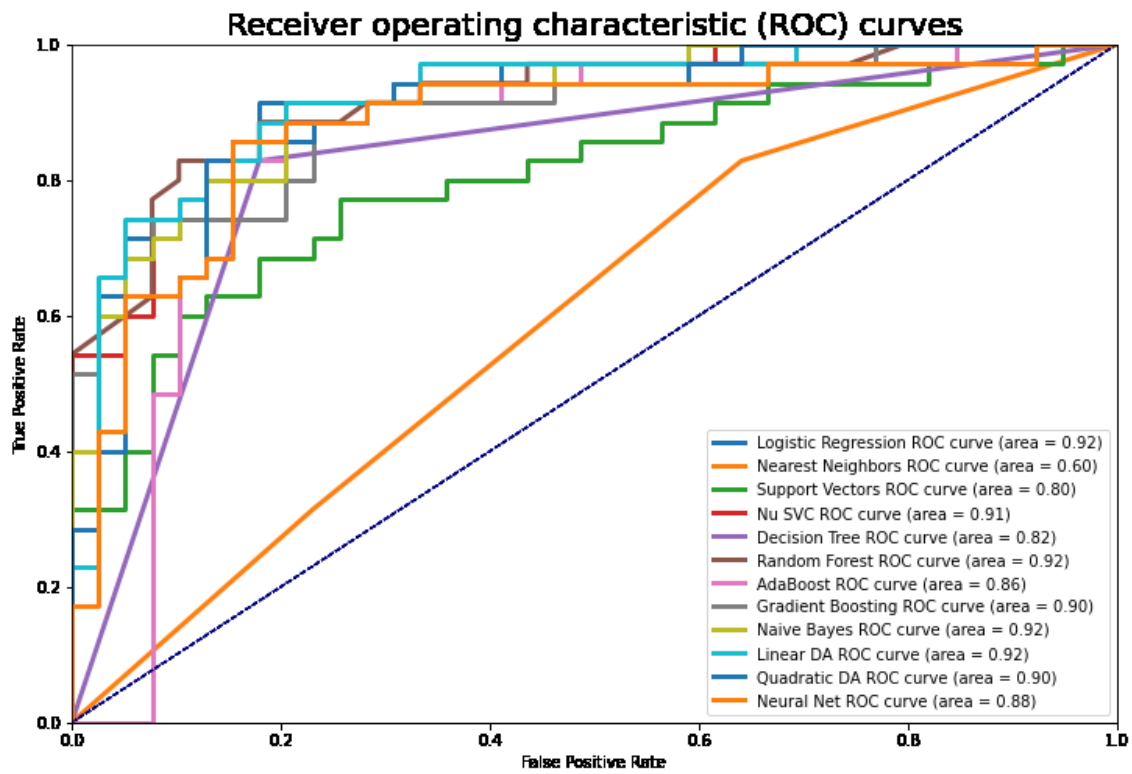
Melhor Modelo Inicial

- A regressão logística apresentou o melhor equilíbrio entre precisão e recall, com o maior F1-Score.

As curvas ROC (*Receiver Operating Characteristic*) são ferramentas usadas na avaliação de modelos de classificação, especialmente em aprendizado de máquina e estatística. Elas mostram a relação entre a taxa de verdadeiros positivos (sensibilidade) e a taxa de falsos positivos (1 - especificidade) para diferentes limiares de decisão. Em termos simples, uma curva ROC ajuda a entender quão bem um modelo consegue distinguir entre classes, como "positivo" e "negativo". Quanto mais próximo a curva estiver do canto superior esquerdo do gráfico, melhor será o desempenho do modelo. O desempenho geral é frequentemente resumido pela área sob a curva (AUC),

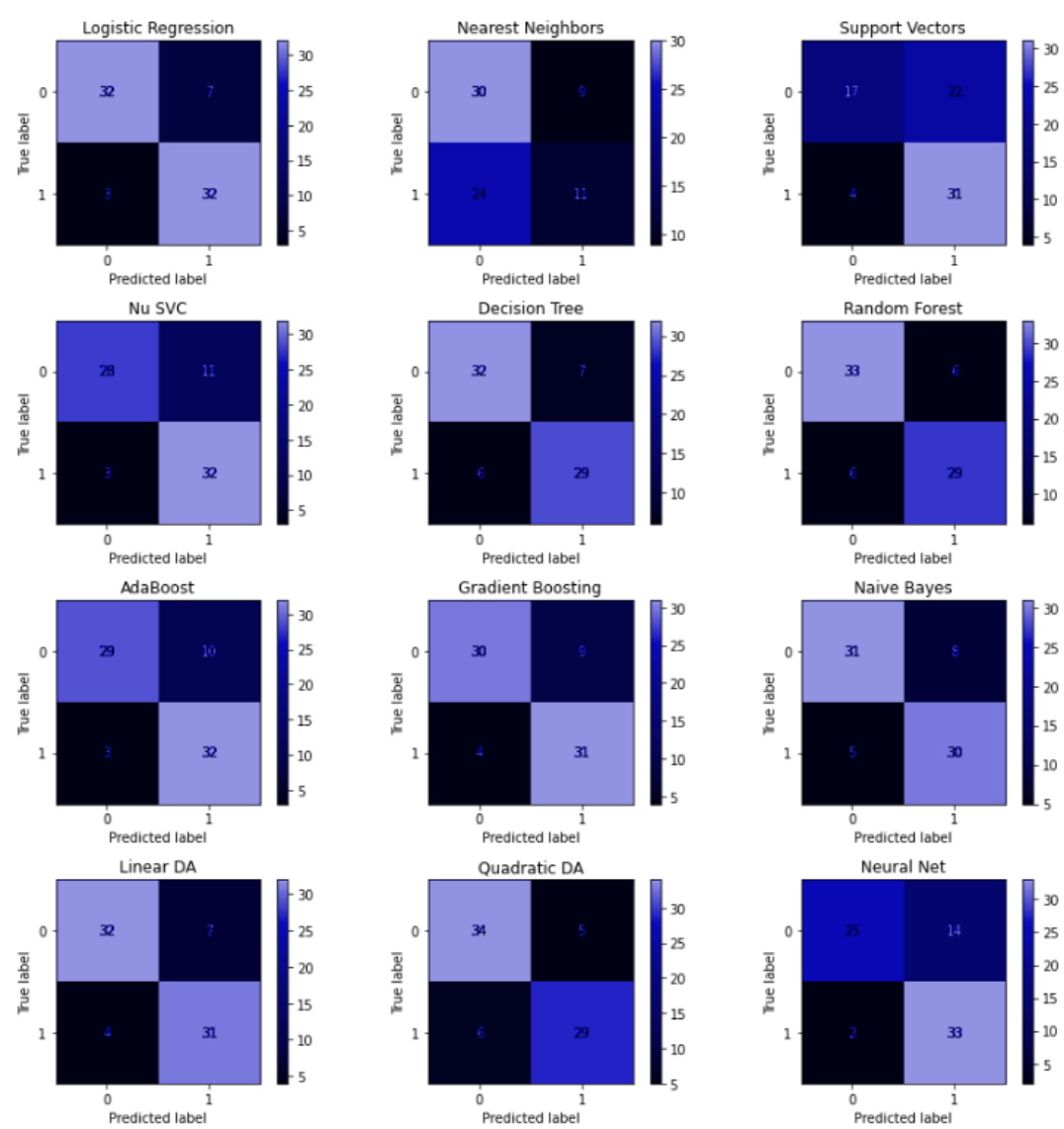
onde valores próximos de 1 indicam um modelo altamente eficiente e valores próximos de 0,5 sugerem um desempenho similar ao acaso.

Figura 9: Curvas ROC



Fonte: O autor

Figura 10: Matriz de confusão modelos Iniciais



Fonte: O autor

Figura 11: Regressão logística

	precision	recall	f1-score	support
0	0.91	0.82	0.86	39
1	0.82	0.91	0.86	35
accuracy			0.86	74
macro avg	0.87	0.87	0.86	74
weighted avg	0.87	0.86	0.86	74

Fonte: O autor

3.3.2 Algoritmos Modernos (Boosted Trees)

- Testados: CatBoost, XGBoost e LightGBM (LGBM).
- Melhor Modelo Tunado:
 - O LGBM ajustado aumentou o recall de 86% para 94%, reduzindo falsos negativos de 5 para 2 no conjunto de validação.

Figura 12: LGBM ajustado

	precision	recall	f1-score	support
0	0.94	0.79	0.86	39
1	0.80	0.94	0.87	35
accuracy			0.86	74
macro avg	0.87	0.87	0.86	74
weighted avg	0.88	0.86	0.86	74

Fonte: O autor

3.3.3. Explicabilidade do Modelo

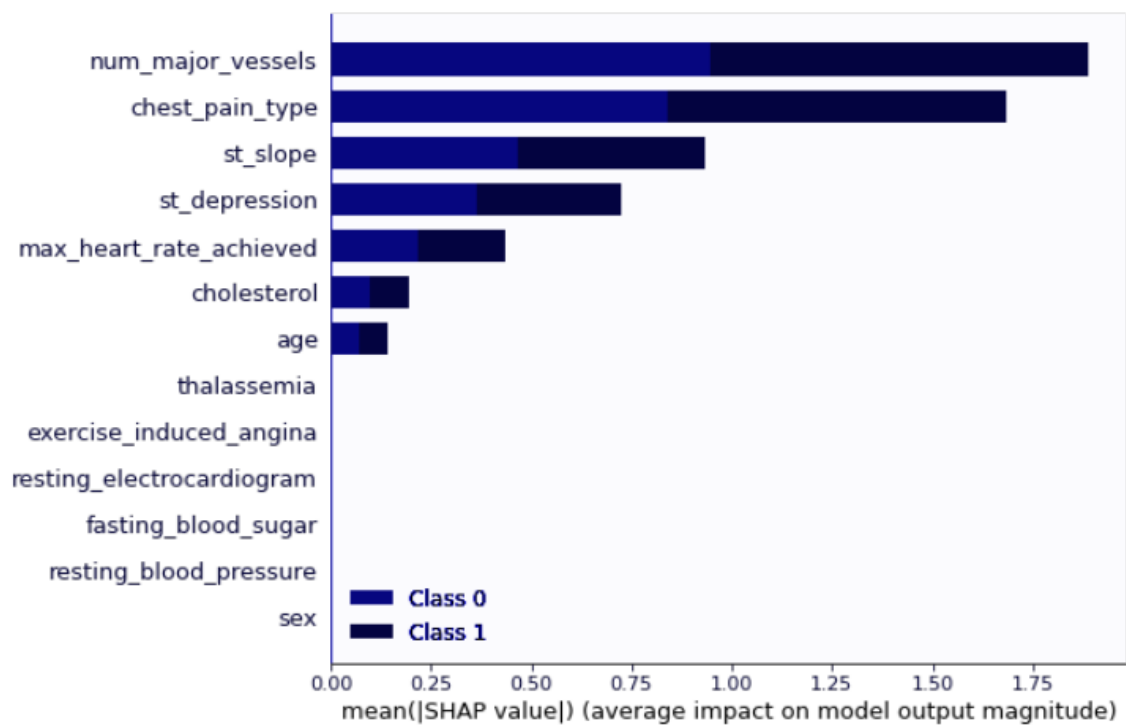
Importância por Permutação

- Mede o impacto de cada variável ao embaralhar seus valores.
- Principais variáveis: número de vasos maiores, tipo de dor no peito, declive ST.

SHAP (*Shapley Additive Explanations*)

- Utiliza valores de Shapley para explicar previsões.
- Confirma as variáveis principais identificadas pelo método de permutação.

Figura 13: variáveis principais identificadas pelo método de permutação



Fonte: O autor

4 MÉTODO

Este trabalho quanto à natureza caracteriza-se como um resumo de assunto, já que se baseia em apenas organizar uma área de conhecimento, indicando sua evolução histórica e estado de arte (Wazlawick, 2014).

Quanto aos objetivos essa pesquisa é exploratória, aquela em que o autor não tem necessariamente uma hipótese ou objetivo definido em mente. Ela pode ser considerada, muitas vezes, como o primeiro estágio de um processo de pesquisa mais longo (Wazlawick, 2014).

Referente aos procedimentos técnicos, trata-se de uma pesquisa bibliográfica e experimental.

A revisão bibliográfica consiste na análise de materiais já publicados, incluindo livros, teses, recursos online e revistas, entre outros. Sua principal vantagem é possibilitar uma abordagem mais ampla de diversos fenômenos, além do que seria viável com a pesquisa direta. (Gil, 2017).

A pesquisa experimental é caracterizada pela manipulação de variáveis independentes para observar seus efeitos sobre variáveis dependentes, geralmente em condições controladas. Sua principal vantagem é permitir o estabelecimento de relações causais entre os fenômenos investigados, fornecendo resultados mais precisos e replicáveis do que outros métodos de pesquisa. Além disso, o controle rigoroso das condições experimentais reduz a influência de fatores externos, aumentando a confiabilidade das conclusões (Gil, 2017).

De acordo com Gil (2017), as etapas da revisão bibliográfica são:

- a) Escolha do tema: Inteligência artificial no monitoramento pós-cirúrgico.
- b) Revisão preliminar da literatura: Fazer o levantamento bibliográfico preliminar de periódicos e artigos relacionados ao assunto de pesquisa. No caso, sobre o impacto da inteligência artificial na saúde - foram feitas na base de dados da CAPES, repositório da PUCGO e Google Acadêmico.
- c) Formulação da pergunta de pesquisa: Como a inteligência artificial pode ajudar no monitoramento pós-cirúrgico.

d) Identificação das fontes: Consultar dissertações, periódicos científicos, obras de referência e outras fontes bibliográficas relevantes para obter informações que respondam ao problema proposto.

e) Análise do conteúdo: Analisar as informações e dados coletados das fontes selecionadas, relacionando-os à pergunta de pesquisa e avaliando a coerência dos argumentos apresentados.

f) Fichamento: O fichamento foi feito com ênfase em trabalhos relevantes para a pesquisa. No entanto, documentos com contribuições médias ou baixas também foram analisados para enriquecer as reflexões do autor. Essa abordagem garantiu uma base sólida para a redação do TCC.

g) Escrita do Trabalho de Conclusão de Curso.

De acordo com Gil (2017), as etapas da revisão bibliográfica são

a) Delimitação do problema: definir claramente a questão que será investigada. Modelo de predição para doenças cardiovasculares.

b) Formulação de hipóteses: elaborar suposições sobre as relações entre variáveis que serão testadas. Idade (age), Sexo (sex), Dor no peito (cp), Pressão arterial (trestbps), Colesterol, Glicemia jejum (fbs), ECG em repouso (restecg), Frequência cardíaca máxima (thalach), Angina induzida (exang), Depressão ST (oldpeak), Inclinação do ST (slope), Vasos (ca), Talassemia (thal), Alvo (target).

c) Definição das variáveis: identificar e operacionalizar as variáveis independentes (manipuladas) e dependentes (medidas).

d) Elaboração do plano experimental: pesquisar tipos de modelos que mais se adequa ao desenvolvimento do modelo, no caso deste foi utilizado modelo de regressão logística.

e) Execução do experimento: aplicar as condições experimentais e registrar os resultados.

f) Análise dos dados: processar e interpretar os resultados obtidos por meio de métodos estatísticos ou qualitativos.

g) Interpretação e conclusão: avaliar se as hipóteses foram confirmadas e discutir as implicações dos achados.

CONCLUSÃO

A inserção da inteligência artificial na saúde está transformando diagnósticos, tratamentos e a gestão de cuidados, possibilitando uma análise avançada de grandes volumes de dados clínicos. Tecnologias como *machine learning* e *deep learning* facilitam a identificação de padrões complexos, resultando em diagnósticos rápidos e tratamentos personalizados. Essa revolução tecnológica permite uma medicina mais eficiente e adaptada às necessidades individuais, melhorando significativamente os resultados clínicos para os pacientes.

Além de aprimorar a prática médica, a inteligência artificial está impulsionando inovações em dispositivos médicos, como próteses robóticas e sistemas de monitoramento remoto. Essas tecnologias elevam a qualidade de vida dos pacientes, oferecendo maior autonomia e melhor gestão de condições crônicas. Com a capacidade de IA em prever complicações e otimizar a utilização dos recursos médicos, o setor de saúde pode estar vendo uma redução considerável nos custos operacionais e nas internações hospitalares, tornando os cuidados mais acessíveis e eficazes.

O futuro da IA na saúde promete ainda mais avanços, especialmente com a integração de dados de dispositivos vestíveis e aplicativos móveis para um monitoramento mais abrangente e contínuo. À medida que essas tecnologias evoluem, espera-se que a IA continue a transformar o setor, oferecendo soluções inovadoras e centradas no paciente. Com isso, a inteligência artificial não só redefine a prática clínica, mas também molda um sistema de saúde mais eficiente e acessível, melhorando a experiência e os resultados para todos os envolvidos.

Concluindo, o modelo de predição LGBM tunado foi identificado como o mais eficaz entre os avaliados, destacando-se por sua capacidade de priorizar variáveis-chave como o número de vasos maiores, o tipo de dor no peito e o declive ST. De maneira surpreendente, o colesterol, frequentemente considerado um indicador importante em análises cardiovasculares, não apresentou relevância significativa para o modelo. No entanto, é fundamental interpretar esses resultados com cautela, dado o tamanho limitado do conjunto

de dados utilizado, composto por apenas 296 amostras, o que pode influenciar a generalização das conclusões obtidas.

CRONOGRAMA

Ano	2024											
Atividade/Mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Revisão bibliográfica												
Descrever o papel da inteligência artificial na transformação dos cuidados médicos.												
Descrever o papel da inteligência artificial no monitoramento pós-cirúrgico												
Escrever TCC1												
Defender TCC1												
Buscas melhores estratégias para aprimoramento da inteligência artificial no monitoramento pós-cirúrgico												
Analisar resultados												
Escrever TCC2												
Defender TCC2												
Corrigir e entregar TCC2												

APÊNDICE

Código utilizado para o modelo:

```
import numpy as np
import pandas as pd
from scipy import stats
import seaborn as sns
from IPython.core.display import HTML
import matplotlib.pyplot as plt
from scipy.stats import uniform

import warnings
warnings.filterwarnings('ignore')

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

data = pd.read_csv('/content/sample_data/heart.xls')
print('Shape of the data is ', data.shape)

data.head()

data.dtypes

""" Data Dictionary

age: age in years

sex: sex
```

1 = male

0 = female

cp: chest pain type

Value 0: typical angina

Value 1: atypical angina

Value 2: non-anginal pain

Value 3: asymptomatic

restbps: resting blood pressure (in mm Hg on admission to the hospital)

chol: serum cholesterol in mg/dl

fbs: (fasting blood sugar > 120 mg/dl)

1 = true;

0 = false

restecg: resting electrocardiographic results

Value 0: normal

Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)

Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria

thalach: maximum heart rate achieved

exang: exercise induced angina

1 = yes

0 = no

oldpeak = ST depression induced by exercise relative to rest

slope: the slope of the peak exercise ST segment

Value 0: upsloping

Value 1: flat

Value 2: downsloping

ca: number of major vessels (0-3) colored by flourosopy

thal:

0 = error (in the original dataset 0 maps to NaN's)

1 = fixed defect

2 = normal

3 = reversable defect

target (the lable):

0 = no disease,

1 = disease """"

```
data = data[data['ca'] < 4] #drop the wrong ca values data = data[data['thal'] > 0]
# drop the wong thal value print(f'The length of the data now is {len(data)}
instead of 303!')
```

""""Rename columns for the sake of clarity""""

```
data = data.rename( columns = {'cp':'chest_pain_type',
'trestbps':'resting_blood_pressure', 'chol': 'cholesterol', 'fbs':
'fasting_blood_sugar', 'restecg' : 'resting_electrocardiogram', 'thalach':
'max_heart_rate_achieved', 'exang': 'exercise_induced_angina', 'oldpeak':
'st_depression', 'slope': 'st_slope', 'ca':'num_major_vessels', 'thal':
'thalassemia'}, errors="raise")
```

```
data['sex'][data['sex'] == 0] = 'female' data['sex'][data['sex'] == 1] = 'male'
```

```
data['chest_pain_type'][data['chest_pain_type'] == 0] = 'typical angina'
data['chest_pain_type'][data['chest_pain_type'] == 1] = 'atypical angina'
data['chest_pain_type'][data['chest_pain_type'] == 2] = 'non-anginal pain'
data['chest_pain_type'][data['chest_pain_type'] == 3] = 'asymptomatic'
```

```
data['fasting_blood_sugar'][data['fasting_blood_sugar'] == 0] = 'lower than
120mg/ml' data['fasting_blood_sugar'][data['fasting_blood_sugar'] == 1] =
'greater than 120mg/ml'
```

```
data['resting_electrocardiogram'][data['resting_electrocardiogram'] == 0] =
'normal' data['resting_electrocardiogram'][data['resting_electrocardiogram'] ==
1] = 'ST-T wave abnormality'
data['resting_electrocardiogram'][data['resting_electrocardiogram'] == 2] = 'left
ventricular hypertrophy'
```

```
data['exercise_induced_angina'][data['exercise_induced_angina'] == 0] = 'no'
data['exercise_induced_angina'][data['exercise_induced_angina'] == 1] = 'yes'
```

```
data['st_slope'][data['st_slope'] == 0] = 'upsloping'
data['st_slope'][data['st_slope'] == 1] = 'flat' data['st_slope'][data['st_slope'] == 2]
= 'downsloping'
```

```
data['thalassemia'][data['thalassemia'] == 1] = 'fixed defect'
data['thalassemia'][data['thalassemia'] == 2] = 'normal'
data['thalassemia'][data['thalassemia'] == 3] = 'reversible defect'
```

data.dtypes

data.head()

""""Grouping Features (by data type)""""

numerical features 6

```
num_feats = ['age', 'cholesterol', 'resting_blood_pressure',
'max_heart_rate_achieved', 'st_depression', 'num_major_vessels']
```

```

categorical (binary)
bin_feats = ['sex', 'fasting_blood_sugar', 'exercise_induced_angina', 'target']

categorical (multi-)
nom_feats= ['chest_pain_type', 'resting_electrocardiogram', 'st_slope',
'thalassemia'] cat_feats = nom_feats + bin_feats

"""*Exploring Features/Target"""

mypal= ['#FC05FB', '#FEAEFE', '#FCD2FC', '#F3FEFA', '#B4FFE4', '#3FFEBA']

plt.figure(figsize=(7, 5),facecolor='#F6F5F4') total = float(len(data)) ax =
sns.countplot(x=data['target'], palette=mypal[1::4]) ax.set_facecolor('#F6F5F4')

for p in ax.patches:
    height = p.get_height()
    ax.text(p.get_x()+p.get_width()/2.,height + 3,{:1.1f}
    %'.format((height/total)*100), ha="center",
            bbox=dict(facecolor='none', edgecolor='black', boxstyle='round',
linewidth=0.5))

ax.set_title('Target variable distribution', fontsize=20, y=1.05)
sns.despine(right=True) sns.despine(offset=5, trim=True)

"""Numerical Features"""

data[num_feats].describe().T

L = len(num_feats) ncol= 2 nrow= int(np.ceil(L/ncol)) #remove_last= (nrow *
ncol) - L

fig, ax = plt.subplots(nrow, ncol, figsize=(16, 14),facecolor='#F6F5F4')
fig.subplots_adjust(top=0.92)

i = 1 for col in num_feats: plt.subplot(nrow, ncol, i, facecolor='#F6F5F4')

ax = sns.kdeplot(data=data, x=col, hue="target", multiple="stack",
palette=mypal[1::4])
ax.set_xlabel(col, fontsize=20)
ax.set_ylabel("density", fontsize=20)
sns.despine(right=True)
sns.despine(offset=0, trim=False)

if col == 'num_major_vessels':
    sns.countplot(data=data, x=col, hue="target", palette=mypal[1::4])
    for p in ax.patches:
        height = p.get_height()
        ax.text(p.get_x()+p.get_width()/2.,height +
3,{:1.0f}'.format((height)),ha="center",
                bbox=dict(facecolor='none', edgecolor='black', boxstyle='round',
linewidth=0.5))

```

```
i = i + 1
```

```
plt.suptitle('Distribution of Numerical Features' ,fontsize = 24);
```

```
""""Pair-plots""""
```

```
_ = ['age', 'cholesterol', 'resting_blood_pressure', 'max_heart_rate_achieved',  
'st_depression', 'target'] data_ = data[] g = sns.pairplot(data, hue="target",  
corner=True, diag_kind='hist', palette=mypal[1::4]); plt.suptitle('Pairplot:  
Numerical Features ' ,fontsize = 24);
```

```
""""Selected Features""""
```

```
fig, ax = plt.subplots(1,4, figsize=(20, 4)) sns.regplot(data=data[data['target']  
==1], x='age', y='cholesterol', ax = ax[0], color=mypal[0], label='1')  
sns.regplot(data=data[data['target'] ==0], x='age', y='cholesterol', ax = ax[0],  
color=mypal[5], label='0') sns.regplot(data=data[data['target'] ==1], x='age',  
y='max_heart_rate_achieved', ax = ax[1], color=mypal[0], label='1')  
sns.regplot(data=data[data['target'] ==0], x='age', y='max_heart_rate_achieved',  
ax = ax[1], color=mypal[5], label='0') sns.regplot(data=data[data['target'] ==1],  
x='age', y='resting_blood_pressure', ax = ax[2], color=mypal[0], label='1')  
sns.regplot(data=data[data['target'] ==0], x='age', y='resting_blood_pressure',  
ax = ax[2], color=mypal[5], label='0') sns.regplot(data=data[data['target'] ==1],  
x='age', y='st_depression', ax = ax[3], color=mypal[0], label='1')  
sns.regplot(data=data[data['target'] ==0], x='age', y='st_depression', ax = ax[3],  
color=mypal[5], label='0') plt.suptitle('Reg plots of selected features')  
plt.legend();
```

```
def count_plot(data, cat_feats): L = len(cat_feats) ncol= 2 nrow=  
int(np.ceil(L/ncol)) remove_last= (nrow * ncol) - L
```

```
fig, ax = plt.subplots(nrow, ncol,figsize=(18, 24), facecolor='#F6F5F4')  
fig.subplots_adjust(top=0.92)  
ax.flat[remove_last].set_visible(False)
```

```
i = 1
```

```
for col in cat_feats:
```

```
    plt.subplot(nrow, ncol, i, facecolor='#F6F5F4')  
    ax = sns.countplot(data=data, x=col, hue="target", palette=mypal[1::4])  
    ax.set_xlabel(col, fontsize=20)  
    ax.set_ylabel("count", fontsize=20)  
    sns.despine(right=True)  
    sns.despine(offset=0, trim=False)  
    plt.legend(facecolor='#F6F5F4')
```

```
    for p in ax.patches:
```

```
        height = p.get_height()  
        ax.text(p.get_x()+p.get_width()/2.,height +  
3, '{:1.0f}'.format((height)),ha="center",  
        bbox=dict(facecolor='none', edgecolor='black', boxstyle='round',  
linewidth=0.5))
```

```
i = i +1
```

```
plt.suptitle('Distribution of Categorical Features' ,fontsize = 24)
return 0
```

```
count_plot(data, cat_feats[0:-1]);
```

```
df_ = data[num_feats] corr = df_.corr(method='pearson') mask =
np.triu(np.ones_like(corr, dtype=bool)) f, ax = plt.subplots(figsize=(8, 5),
facecolor=None) cmap = sns.color_palette(mypal, as_cmap=True)
sns.heatmap(corr, mask=mask, cmap=cmap, vmax=1.0, vmin=-1.0, center=0,
annot=True, square=False, linewidths=.5, cbar_kws={"shrink": 0.75})
ax.set_title("Numerical features correlation (Pearson's)", fontsize=20, y= 1.05);
```

```
feats_ = ['age', 'cholesterol', 'resting_blood_pressure',
'max_heart_rate_achieved', 'st_depression', 'num_major_vessels', 'target']
```

```
def point_biserial(x, y): pb = stats.pointbiserialr(x, y) return pb[0]
```

```
rows= [] for x in feats_ : col = [] for y in feats_ : pbs =point_biserial(data[x],
data[y]) col.append(round(pbs,2)) rows.append(col)
```

```
pbs_results = np.array(rows) DF = pd.DataFrame(pbs_results, columns =
data[feats_].columns, index =data[feats_].columns)
```

```
mask = np.triu(np.ones_like(DF, dtype=bool)) corr = DF.mask(mask)
```

```
f, ax = plt.subplots(figsize=(8, 5), facecolor=None) cmap =
sns.color_palette(mypal, as_cmap=True) sns.heatmap(corr, mask=mask,
cmap=cmap, vmax=1.0, vmin=-1, center=0, annot=True, square=False,
linewidths=.5, cbar_kws={"shrink": 0.75}) ax.set_title("Cont feats vs target
correlation (point-biserial)", fontsize=20, y= 1.05);
```

the cramers_v function is copied from <https://towardsdatascience.com/the-search-for-categorical-correlation-a1cf7f1888c9>

```
def cramers_v(x, y): confusion_matrix = pd.crosstab(x,y) chi2 =
stats.chi2_contingency(confusion_matrix)[0] n = confusion_matrix.sum().sum()
phi2 = chi2/n r,k = confusion_matrix.shape phi2corr = max(0, phi2-((k-1)*(r-
1))/(n-1)) rcorr = r-((r-1)**2)/(n-1) kcorr = k-((k-1)**2)/(n-1) return
np.sqrt(phi2corr/min((kcorr-1),(rcorr-1)))
```

calculate the correlation coefficients using the above function

```
data_ = data[cat_feats] rows= [] for x in data_ : col = [] for y in data_ : cramers
=cramers_v(data_[x], data_[y]) col.append(round(cramers,2)) rows.append(col)
```

```
cramers_results = np.array(rows) df = pd.DataFrame(cramers_results, columns
= data_.columns, index = data_.columns)
```

color palette

```
mypal_1= ['#FC05FB', '#FEAEFE', '#FCD2FC', '#F3FEFA',
'#B4FFE4', '#3FFEBA', '#FC05FB', '#FEAEFE', '#FCD2FC']
```

plot the heat map

```

mask = np.triu(np.ones_like(df, dtype=bool)) corr = df.mask(mask) f, ax =
plt.subplots(figsize=(10, 6), facecolor=None) cmap =
sns.color_palette(mypal_1, as_cmap=True) sns.heatmap(corr, mask=mask,
cmap=cmap, vmax=1.0, vmin=0, center=0, annot=True, square=False,
linewidths=.01, cbar_kws={"shrink": 0.75}) ax.set_title("Categorical Features
Correlation (Cramer's V)", fontsize=20, y= 1.05);

from sklearn.linear_model import LogisticRegression from sklearn.neighbors
import KNeighborsClassifier from sklearn.svm import SVC, LinearSVC, NuSVC
from sklearn.tree import DecisionTreeClassifier from sklearn.ensemble import
RandomForestClassifier, AdaBoostClassifier, GradientBoostingClassifier from
sklearn.naive_bayes import GaussianNB from sklearn.discriminant_analysis
import LinearDiscriminantAnalysis from sklearn.discriminant_analysis import
QuadraticDiscriminantAnalysis

from sklearn.metrics import confusion_matrix, plot_confusion_matrix,
classification_report from sklearn.metrics import recall_score,
accuracy_score, roc_curve, auc from sklearn.model_selection import
train_test_split from sklearn.ensemble import RandomForestClassifier from
sklearn.neural_network import MLPClassifier

from sklearn.preprocessing import LabelEncoder import shap

def label_encode_cat_features(data, cat_features): ''' Given a dataframe and its
categorical features, this function returns label-encoded dataframe '''

label_encoder = LabelEncoder()
data_encoded = data.copy()

for col in cat_features:
    data_encoded[col] = label_encoder.fit_transform(data[col])

data = data_encoded

return data

def score_summary(names, classifiers): ''' Given a list of classifiers, this function
calculates the accuracy, ROC_AUC and Recall and returns the values in a
dataframe '''

cols=["Classifier", "Accuracy", "ROC_AUC", "Recall", "Precision", "F1"]
data_table = pd.DataFrame(columns=cols)

for name, clf in zip(names, classifiers):
    clf.fit(X_train, y_train)

    pred = clf.predict(X_val)
    accuracy = accuracy_score(y_val, pred)

    pred_proba = clf.predict_proba(X_val)[:, 1]

    fpr, tpr, thresholds = roc_curve(y_val, pred_proba)

```

```

roc_auc = auc(fpr, tpr)

# confusion matrix, cm
cm = confusion_matrix(y_val, pred)

# recall: TP/(TP+FN)
recall = cm[1,1]/(cm[1,1] +cm[1,0])

# precision: TP/(TP+FP)
precision = cm[1,1]/(cm[1,1] +cm[0,1])

# F1 score: TP/(TP+FP)
f1 = 2*recall*precision/(recall + precision)

df = pd.DataFrame([[name, accuracy*100, roc_auc, recall, precision, f1]],
columns=cols)
data_table = data_table.append(df)

return(np.round(data_table.reset_index(drop=True), 2))

def plot_conf_matrix(names, classifiers, nrows, ncols, fig_a, fig_b): """ Plots
confusion matrices in a subplots.

Args:
    names : list of names of the classifier
    classifiers : list of classification algorithms
    nrows, ncols : number of rows and rows in the subplots
    fig_a, fig_b : dimensions of the figure size
    """

fig, axes = plt.subplots(nrows=nrows, ncols=ncols, figsize=(fig_a, fig_b))

i = 0
for clf, ax in zip(classifiers, axes.flatten()):

    clf.fit(X_train, y_train)
    plot_confusion_matrix(clf, X_val, y_val, ax=ax)
    ax.title.set_text(names[i])
    i = i + 1

plt.tight_layout()
plt.show()

def roc_auc_curve(names, classifiers): """ Given a list of classifiers, this function
plots the ROC curves
    """
plt.figure(figsize=(12, 8))

```

```

for name, clf in zip(names, classifiers):
    clf.fit(X_train, y_train)

    pred_proba = clf.predict_proba(X_val)[:, 1]

    fpr, tpr, thresholds = roc_curve(y_val, pred_proba)
    roc_auc = auc(fpr, tpr)

    plt.plot(fpr, tpr, lw=3, label= name + ' ROC curve (area = %0.2f)' % (roc_auc))
    plt.plot([0, 1], [0, 1], color='navy', lw=1, linestyle='--')
    plt.xlim([0.0, 1.0])
    plt.ylim([0.0, 1.0])
    plt.xlabel('False Positive Rate')
    plt.ylabel('True Positive Rate')
    plt.title('Receiver operating characteristic (ROC) curves', fontsize=20)
    plt.legend(loc="lower right")

```

split the data into train and test sets

```
cat_features = cat_feats data = label_encode_cat_features(data, cat_features)
```

```
seed = 0 test_size = 0.25
```

```
features = data.columns[:-1]
```

```
X = data[features] y = data['target']
```

```
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size = test_size,
random_state=seed)
```

classifier algorithms with default parameters

```
names = [ 'Logistic Regression', 'Nearest Neighbors', 'Support Vectors', 'Nu
SVC', 'Decision Tree', 'Random Forest', 'AdaBoost', 'Gradient Boosting', 'Naive
Bayes', 'Linear DA', 'Quadratic DA', "Neural Net" ]
```

```
classifiers = [ LogisticRegression(solver="liblinear", random_state=seed),
KNeighborsClassifier(2), SVC(probability=True, random_state=seed),
NuSVC(probability=True, random_state=seed),
DecisionTreeClassifier(random_state=seed),
RandomForestClassifier(random_state=seed),
AdaBoostClassifier(random_state=seed),
GradientBoostingClassifier(random_state=seed), GaussianNB(),
LinearDiscriminantAnalysis(), QuadraticDiscriminantAnalysis(),
MLPClassifier(random_state=seed), ]
```

```
score_summary(names, classifiers).sort_values(by='Accuracy' , ascending =
False)
```

```
.style.background_gradient(cmap='coolwarm')
```

```
.bar(subset=["ROC_AUC",], color='#6495ED')
```

```
.bar(subset=["Recall"], color='#ff355d')
```

```

.bar(subset=["Precision"], color='lightseagreen')
.bar(subset=["F1"], color='gold')
"""### 2.1.3 ROC curves"""
roc_auc_curve(names, classifiers)
"""### 2.1.4 Confusion matrix"""
plot_conf_matrix(names, classifiers, nrows=4, ncols=3, fig_a=12, fig_b=12)
from sklearn.model_selection import RandomizedSearchCV from scipy.stats
import uniform from sklearn.model_selection import RepeatedStratifiedKFold
lr = LogisticRegression(tol=1e-4, max_iter=1000, random_state=seed)
space = dict(C=uniform(loc=0, scale=5), penalty=['l2', 'l1'], solver= ['liblinear'])
search = RandomizedSearchCV(lr, space, random_state=seed, cv = 5,
scoring='f1')
rand_search = search.fit(X_train, y_train)
print('Best Hyperparameters: %s' % rand_search.best_params_)
params = rand_search.best_params_ lr = LogisticRegression(**params)
lr.fit(X_train, y_train) print(classification_report(y_val, lr.predict(X_val)))
plot_confusion_matrix(lr, X_val, y_val)
from catboost import CatBoostClassifier from xgboost import XGBClassifier
from lightgbm import LGBMClassifier
names_boost = [ 'Catboost', 'xgboost', 'light GBM' ] classifiers = [
CatBoostClassifier(random_state=seed, verbose=0), XGBClassifier(objective=
'binary:logistic', random_state=seed), LGBMClassifier(random_state=seed) ]
"""### 2.2.1 Performance metrics summary table"""
score_summary(names_boost, classifiers).sort_values(by='Accuracy' ,
ascending = False)
.style.background_gradient(cmap='coolwarm')
.bar(subset=["ROC_AUC",], color='#6495ED')
.bar(subset=["Recall"], color='#ff355d')
.bar(subset=["Precision"], color='lightseagreen')
.bar(subset=["F1"], color='gold')
"""### 2.2.2 Confusion matrix"""
plot_conf_matrix(names=names_boost, classifiers=classifiers, nrows=1,
ncols=3, fig_a=12, fig_b=3);
from sklearn.model_selection import GridSearchCV rs_params = {
'num_leaves': [20, 100], 'max_depth': [5, 15], 'min_data_in_leaf': [80, 120], }

```

```

rs_cv = GridSearchCV(estimator=LGBMClassifier(random_state=seed,
verbose=-1), param_grid=rs_params, cv = 5)

rs_cv.fit(X_train, y_train) params = rs_cv.best_params_

lgbm = LGBMClassifier(**params);

lgbm.fit(X_train, y_train, eval_set=(X_val, y_val), verbose=False, );

print(classification_report(y_val, lgbm.predict(X_val)))

plot_confusion_matrix(lgbm, X_val, y_val);

import eli5 from eli5.sklearn import PermutationImportance

perm_imp = PermutationImportance(lgbm, random_state=seed).fit(X_train,
y_train) eli5.show_weights(perm_imp, feature_names = X_val.columns.tolist())

import shap shap.initjs() explainer = shap.TreeExplainer(lgbm) shap_values =
explainer.shap_values(X_val) shap.summary_plot(shap_values, X_val,
feature_names=features, plot_type="bar", )

shap.summary_plot(shap_values[0], X_val)

```

REFERÊNCIAS

AWRAHMAN, B. J.; AZIZ FATAH, C.; HAMAAMIN, M. Y. A review of the role and challenges of big data in healthcare informatics and analytics. **Computational Intelligence and Neuroscience**, v. 2022, n. 5317760, p. 1–10, 29 set. 2022.

BIOINFO. **Métricas de avaliação em machine learning**. Disponível em: <https://www.bioinfo.com.br>. Acesso em: 05 dez. 2024.

CHEN, P.-F. et al. Predicting Postoperative Mortality With Deep Neural Networks and Natural Language Processing: Model Development and Validation. **JMIR Medical Informatics**, v. 10, n. 5, p. e38241, 10 maio 2022.

DATA GEEKS. **Curva ROC: Entenda Como Avaliar Modelos de Classificação!** Disponível em: <https://www.datageeks.com.br>. Acesso em: 05 dez. 2024.

DATA GEEKS. **Matriz de Confusão: Como Utilizar para Análise de Modelos?** Disponível em: <https://www.datageeks.com.br>. Acesso em: 05 dez. 2024.

DES. **Heart Disease Predictions**. Disponível em: <https://www.kaggle.com/code/desalegngeb/heart-disease-predictions>. Acesso em: 05 jun. 2024.

FAGUNDES, Robson. **O papel da inteligência artificial na detecção precoce de doenças**. RDICOM, 2023. Disponível em: <https://rdicom.com.br/blog/o-papel->

da-inteligencia-artificial-na-deteccao-precoce-de-doencas/. Acesso em: 05 jun. 2024.

GIL, Antônio Carlos. Como Elaborar Projetos de Pesquisa. 6. ed. São Paulo: Editora Atlas Ltda., 2017

HOSPITAL 9 DE JULHO. **Inteligência artificial é usada para evitar quedas.** H9J, 2018. Disponível em: <https://www.h9j.com.br/pt/sobre-nos/blog/inteligencia-artificial-e-usada-para-evitar-quedas>. Acesso em: 05 jun. 2024.

KIM, J.-H. et al. Harnessing Machine Learning for Prediction of Postoperative Pulmonary Complications: Retrospective Cohort Design. **Journal of Clinical Medicine**, v. 12, n. 17, p. 5681–5681, 31 ago. 2023.

MICROSOFT. **LightGBM: A Fast, Distributed, High Performance Gradient Boosting (GBDT, GBRT or GBM) Framework.** Disponível em: <https://github.com/microsoft/LightGBM>. Acesso em: 05 jun. 2024

PEREIRA, S. et al. Impactos de dispositivos tecnológicos avançados para o autocuidado e monitoramento de pacientes ostomizados. **Revista Eletrônica Acervo Saúde**, v. 24, n. 3, p. e15952–e15952, 8 mar. 2024.

PORTAL TELEMEDICINA. **Inteligência Artificial na Medicina. Como TensorFlow é usado.** Portal telemecina, 2019. Disponível em: <https://portaltelemedicina.com.br/inteligencia-artificial-na-medicina-tensorflow>. Acesso em: 05 jun. 2024.

SCHULZE, Clay. **Como as Big techs estão mudando o mercado da saúde.** Linkedin, 2024. Disponível em: <https://www.linkedin.com/pulse/como-big-techs-est%C3%A3o-mudando-o-mercado-de-sa%C3%BAdade-clay-schulze-os7cf>. Acesso em: 05 jun. 2024.

UFPE. **HC desenvolve projeto de saúde digital para monitorar feridas de pacientes no pós-cirurgia.** Ufpe Notícias, 2023. Disponível em: https://www.ufpe.br/agencia/noticias/-/asset_publisher/dlhi8nsrz4hK/content/hc-desenvolve-projeto-de-saude-digital-para-monitorar-feridas-de-pacientes-no-pos-cirurgia/40615#:~:text=Monitorar%20a%20evolu%C3%A7%C3%A3o%20da%20cicatriz%C3%A7%C3%A3o,Facepe%20%E2%80%9CPernambucanas%20Inovadoras%202023%E2%80%9D. Acesso em: 05 jun. 2024.

WAZLAWICK, R. S. Metodologia de Pesquisa para Ciência da Computação. 2º ed. Rio de Janeiro: Elsevier, 2014. cap.4, p. 21 – 26.

WU, G. et al. Performance of machine learning algorithms for surgical site infection case detection and prediction: A systematic review and meta-analysis. **Annals of Medicine and Surgery**, v. 84, 1 dez. 2022.

ZAPAROLLI, Domingos. **A inteligência artificial chega a saúde.** Revista Pesquisa Fapesp, 2022. Disponível em: <https://revistapesquisa.fapesp.br/a-inteligencia-artificial-chega-a-saude/>. Acesso em: 05 jun. 2024.