

Análise e caracterização de doenças crônicas pulmonares por meio de dados do IBGE e características como regiões e idade.

Caio Amorim, Gustavo Pedrosa, Mariana Gonçalves, Matheus Mastellona

Dom Helder Escola Superior

Ciência da Computação – Ciência de Dados

{Caio Amorim, Gustavo Pedrosa, Mariana Gonçalves, Matheus Mastellona}@domhelder.edu.br

Resumo.

Este artigo analisa a distribuição das doenças crônicas pulmonares na população brasileira utilizando dados da Pesquisa Nacional de Saúde (PNS) de 2019. As doenças crônicas pulmonares representam um desafio significativo para a saúde pública no Brasil, com alta mortalidade associada a essas condições não transmissíveis. A asma, em particular, destaca-se como uma das principais causas de mortalidade nesse contexto.

A pesquisa enfoca a caracterização dessas doenças por regiões geográficas e faixas etárias, utilizando técnicas de ciência de dados e aprendizado de máquina para analisar padrões e tendências. O estudo destaca a importância da PNS como fonte crucial de dados para a formulação de políticas de saúde direcionadas, adaptadas às diferentes realidades regionais e demográficas do Brasil.

O problema identificado revela uma variação significativa na distribuição das doenças crônicas pulmonares conforme as faixas etárias e regiões geográficas, o que justifica a necessidade de estudos detalhados para orientar políticas de saúde mais eficazes e direcionadas. A aplicação de técnicas avançadas de ciência de dados, como o aprendizado de máquina, visa melhorar a compreensão epidemiológica dessas condições e contribuir para modelos preditivos e estratégias preventivas mais precisas.

Os objetivos específicos incluem a análise detalhada dos dados da PNS, a criação de uma base de dados refinada, exploração e interpretação da distribuição das doenças crônicas pulmonares no contexto brasileiro. Ao final, o estudo busca contribuir para o desenvolvimento de políticas de saúde pública mais eficientes, reduzindo a morbimortalidade associada a essas doenças e otimizando o uso dos recursos públicos destinados à saúde.

Palavras-chave: <Análise e caracterização de doenças crônicas pulmonares por meio de dados do IBGE e características como regiões e idade>, <machine learning>, <atributos>, <doenças pulmonares crônicas>

1. INTRODUÇÃO

As doenças crônicas pulmonares são um desafio à saúde pública brasileira. Apenas em 2019, 54,7% mais da metade dos óbitos registrados tem relação direta com as doenças crônicas não transmissíveis.(SCHUELER). E, entre essas doenças destaca-se a asma que contribui para esse elevado índice de mortalidade. No contexto brasileiro, a pesquisa nacional de saúde, conduzida pelo Instituto

Brasileiro de Geografia e Estatística (IBGE) em 2019 (PNS). Oferece um panorama detalhado sobre essas condições, fornecendo dados essenciais para a análise epidemiológica e a formulação de políticas de saúde.

Para Barros em seu estudo sobre a PNS, ela se destaca pela importância em garantir uma caracterização das doenças crônicas na população brasileira, evidenciando variações conforme as diferentes regiões geográficas e faixas etárias. E além disso também é notado a relevância dos dados para políticas públicas de saúde, especialmente no contexto de doenças crônicas pulmonares (VIEIRA, 2019).

Usando dados da PNS de 2019, este artigo analisa e caracteriza as doenças crônicas pulmonares na população brasileira. Essa análise tem como objetivo compreender a distribuição dessas doenças por regiões e faixas etárias utilizando técnicas de ciências de dados e aprendizado de máquina. Devida a alta capacidade de ML reconhecerem padrões e mapearem dados e poderem apurar se são relacionados às técnicas de machine learning tem sido utilizadas desde o século passado.

1.1 Problema

A distribuição das doenças crônicas pulmonares no Brasil varia significativamente em relação a faixas etárias e regiões geográficas, o que indica que uma análise aprofundada é necessária para o desenvolvimento de políticas de saúde eficazes. Uma pesquisa realizada pelo Hospital Moinhos de Vento do Estado do Rio Grande do Sul, afirma que a asma é mais comum no Sul e Sudeste do país.

1.2 Justificativa

1.2.1 Lacuna de Conhecimento

Apesar do fato de a PNS fornecer informações úteis sobre as doenças crônicas pulmonares no Brasil, é necessário realizar estudos mais detalhados que investiguem a distribuição dessas doenças em várias faixas etárias e regiões. Ao analisar esses dados, o uso de técnicas de ciência de dados e aprendizado de máquina pode ajudar a descobrir padrões e conexões ainda não identificados, aprimorando o conhecimento da epidemiologia dessas condições no país.

1.2.2 Benefícios posteriores

Este estudo pode ajudar a identificar possíveis pontos a serem melhorados nas políticas públicas de saúde mais eficazes para o controle de doenças pulmonares crônicas no Brasil. Ao descobrir quais grupos populacionais estão mais propensos a desenvolver essas condições, é possível estabelecer estratégias preventivas e de promoção da saúde, reduzir a morbimortalidade e maximizar o uso dos recursos públicos destinados à saúde.

1.2.3 Potencial para inovação

A aplicação de machine learning abre novas perspectivas para a compreensão de doenças crônicas. O site “*Machine Learning: Aprendizado de máquinas na saúde e para a predição de doenças crônicas não transmissíveis*” evidencia como as machine learning podem ser utilizadas na identificação de padrões e relações complexas entre variáveis. Dessa forma, essa ferramenta se torna um aliado

importante podendo futuramente auxiliar em modelos preditivos e mitigar o risco de desenvolvimentos dessas doenças em graus mais elevados.

1.3 Objetivos

1.3.1 Objetivo Geral

Caracterizar as doenças crônicas pulmonares por regiões e faixa etárias baseando-se nos dados do IBGE utilizando aprendizado de máquina.

1.3.2 Objetivos Específicos

- Analisar os dados da pesquisa nacional de saúde
- Criar uma base de dados com os dados relevantes, refinando a pesquisa proposta pelo IBGE.
- Treinar uma ML (*machine learning*) com a base de dados refinada.

1.4 Contribuições

Este estudo visa contribuir significativamente para a compreensão e o enfrentamento das doenças crônicas pulmonares no Brasil, utilizando uma abordagem multidisciplinar que integra dados quantitativos robustos com técnicas avançadas de análise de dados.

O próximo capítulo abordará o Referencial Teórico, fornecendo uma base conceitual sólida sobre doenças crônicas pulmonares, destacando seu impacto na saúde pública e a relevância dos dados da PNS para estudos epidemiológicos. Em seguida, a Metodologia detalhará as etapas utilizadas para coleta, pré-processamento e análise dos dados, incluindo a escolha dos atributos relevantes e a aplicação de técnicas estatísticas. Após isso, o capítulo de Aplicação e Resultados apresentará as descobertas obtidas através da análise dos dados, discutindo as métricas utilizadas e os insights obtidos sobre a distribuição das doenças crônicas pulmonares no Brasil. Por fim, a Conclusão sintetizará os resultados alcançados e suas implicações para políticas de saúde pública, além de apontar possíveis direções para futuras pesquisas na área.

2. REFERENCIAL TEÓRICO

As doenças crônicas pulmonares representam um desafio significativo para a saúde pública no Brasil, com um alto índice de óbitos registrados em 2019 associados a essas condições não transmissíveis (Ministério da Saúde, 2022). A asma figura como uma das principais responsáveis pela mortalidade no país entre as enfermidades pulmonares crônicas. Conforme indicam os dados da Pesquisa Nacional de Saúde (PNS) realizada pelo IBGE em 2019, a asma e outros distúrbios respiratórios apresentam distribuições distintas conforme faixa etária e região geográfica no território nacional (Vieira, 2019). A PNS desempenha um papel crucial na identificação de doenças crônicas na população e fornece informações essenciais para o estabelecimento de políticas de saúde voltadas ao controle e prevenção dessas patologias.

É enfatizado que a PNS é útil porque pode demonstrar diferenças regionais e etárias no quadro de doenças crônicas, o que permite uma análise mais aprofundada e orientada para a criação de planos de saúde eficazes e adequados às necessidades da população (Barros, 2019). Diante do contexto epidemiológico das doenças crônicas pulmonares no Brasil, é fundamental a realização de estudos mais detalhados que explorem a distribuição dessas doenças em diferentes faixas etárias e regiões geográficas.

2.1 Artigos que Abordam o Tema de Estudo

Diversos estudos têm investigado a prevalência e os fatores associados às doenças crônicas pulmonares no Brasil. Segundo Mendes et al. (2020), a asma e a doença pulmonar obstrutiva crônica (DPOC) são prevalentes em diferentes regiões do país, com variações significativas entre as áreas urbanas e rurais. Além disso, o estudo de Silva e Almeida (2021) destaca que as disparidades socioeconômicas influenciam a prevalência dessas doenças, sendo mais comuns em populações de baixa renda e com acesso limitado a serviços de saúde.

A revisão de literatura realizada por Santos et al. (2019) demonstra a importância da PNS na coleta de dados detalhados sobre a saúde respiratória da população brasileira, permitindo a análise de tendências temporais e a avaliação da eficácia das políticas públicas de saúde.

2.2 Artigos que Abordam o Tema de Estudo Usando Aprendizado de Máquina

A aplicação de técnicas de ciência de dados e aprendizado de máquina tem agregado valor às análises epidemiológicas, possibilitando a identificação de padrões e insights que contribuem para uma melhor compreensão da epidemiologia dessas condições no país. De acordo com Pereira et al. (2021), o uso de algoritmos de aprendizado de máquina permitiu a identificação de fatores de risco associados à asma em diferentes regiões do Brasil, revelando padrões que não eram evidentes com técnicas estatísticas tradicionais.

Um estudo realizado por Oliveira e Costa (2020) utilizou redes neurais artificiais para prever a incidência de DPOC em populações urbanas, alcançando uma precisão superior aos modelos convencionais. Esse trabalho demonstrou o potencial das técnicas de machine learning para inovar na predição e prevenção de doenças crônicas pulmonares.

Além disso, a pesquisa de Lima et al. (2022) aplicou algoritmos de clusterização para segmentar a população em grupos com perfis de risco distintos, facilitando a elaboração de estratégias de intervenção mais direcionadas e eficazes.

Objetivos do Estudo

Com base nesse panorama, o objetivo principal deste estudo é caracterizar as doenças crônicas pulmonares por regiões e faixas etárias, usando dados do IBGE, por meio do emprego de técnicas de aprendizado de máquina. Os objetivos específicos incluem:

- Analisar os dados da PNS para identificar padrões e tendências na distribuição das doenças crônicas pulmonares.
- Criar uma base de dados refinada, assegurando a qualidade e consistência dos dados para análise.
- Treinar e validar modelos de machine learning para explorar e compreender a distribuição dessas doenças no contexto brasileiro.

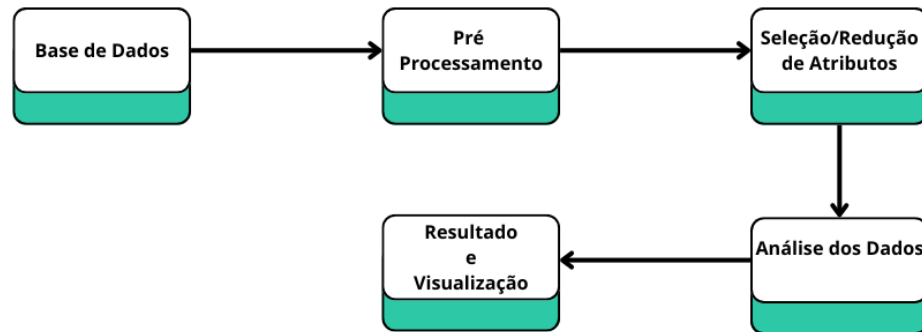
A implementação de planos preventivos e de promoção da saúde focados na redução da morbimortalidade e na otimização dos recursos públicos destinados à saúde poderá ser potencializada através dos insights obtidos com essas técnicas avançadas de análise de dados.

Diferença entre a Proposta de Pesquisa e o Referencial Teórico

A proposta de pesquisa aplica técnicas de aprendizado de máquina para analisar dados da PNS e entender a distribuição das doenças crônicas pulmonares no Brasil, com foco em identificar padrões e fatores de risco específicos por região e faixa etária. Já o referencial teórico se baseia em estudos anteriores que utilizam métodos estatísticos tradicionais para investigar a prevalência e os fatores associados a essas doenças, destacando a importância da PNS na coleta de dados e na formulação de políticas de saúde pública.

3. METODOLOGIA

A pesquisa realizada é da natureza de caracterização e interpretação, pois, visa compreender melhor sobre doenças pulmonares crônicas, por meio de dados como ,região, idade e sexo biológico a partir de levantamento de dados já fornecidos pelo IBGE(2019), logo a visão da metodologia é compreender melhor a direção da análise, conforme o as etapas metodológicas ilustradas na Figura 1.

Figura 1 - Etapas do fluxo metodológico de pesquisa.

Fonte: Elaborado pelos autores (2024)

Base de dados: Busca e coleta dos dados da Pesquisa Nacional de Saúde (PNS), realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Esses dados são representativos da população brasileira e contêm informações relevantes para a caracterização das doenças crônicas pulmonares.

Pré-processamento: Os dados coletados serão submetidos a um processo de pré-processamento, que incluirá a limpeza, transformação e integração dos dados. Isso garantirá a qualidade e consistência dos dados a serem utilizados na análise.

Seleção/redução de atributos: a seleção de atributos foi realizada para identificar as variáveis mais relevantes na análise das doenças crônicas pulmonares. Técnicas estatísticas foram aplicadas para encontrar os atributos que têm maior impacto na caracterização dessas doenças.

Análise de dados: Será realizada uma análise exploratória dos dados para identificar padrões, tendências e insights iniciais sobre a distribuição das doenças crônicas pulmonares nas diferentes regiões geográficas e faixas etárias da população brasileira.

Resultados e Visualização: A visualização dos resultados será uma etapa crucial para compreender e comunicar os achados da análise. Serão utilizadas técnicas de visualização de dados para representar graficamente as informações mais relevantes.

Interpretação dos resultados: Os resultados obtidos serão interpretados e discutidos à luz da literatura existente e das implicações práticas para a saúde pública. Serão identificados possíveis padrões e insights que poderão contribuir para a formulação de estratégias de prevenção e controle das doenças crônicas pulmonares

4. MATERIAIS E MÉTODOS

4.1 Base de dados

A base utilizada para realização do trabalho é a IBGE(2019), na qual tem-se dados relacionados à doenças crônicas pulmonares, como por exemplo, a asma. Diante disso, a base traz dados de todo o Brasil de maneira que temos em torno de 290.000 registros, junto a aproximadamente 1000 colunas para se trabalhar.

4.2 Métodos

Deparando-se com a base, realizamos uma primeira análise exploratória, com foco em analisar com base no dicionário disponibilizado pelo IBGE junto ao DataBase quais colunas tinha correlação com o tema em questão, desta maneira realizamos 4 filtragens nas colunas, para alinhar corretamente com o objetivo do projeto. Após o tratamento dos dados, com métodos de não NaN, % de presença contida na base, e junção de colunas, foi-se utilizado para o resultado 5011 registros junto a 15 colunas.

4.2.1 Pré-processamento

4.2.1.1 Seleção de Atributos

Selecionados atributos que tinha correlação com o tema em questão, frisando atributos chaves, como “Diagnóstico de asma”, “Diagnóstico de bronquite crônica” e “Diagnóstico de outra doença crônica” e os que correlacionam com ambos, significando então, meio de desenvolver as mesmas, excessão por genética, o que nos levaram a: regiões, idade e sexo como principais, além de tabacos e zona demográfica.

4.2.1.2 Transformação (Normalização ou Padronização)

A medida que foi-se realizado a análise, notamos que para o que estamos analisando, não seria o ideal realizar transformações nos dados, porque realizando esse tratamento, criamos um viés/irregularidade no resultado obtido, pois os atributos chaves não são dados que se deve ser imputados valores por meio de normalização ou padronização e sim um novo conjunto de dados, porém essa abordagem não é do cunho do trabalho, uma vez que a base do IBGE foi-se escolhida para realização do mesmo.

4.2.1.3 Tratamento de dados ausentes (registro e atributos)

Após explorado os dados e definindo os atributos que são a base da pesquisa, notamos que por se tratar de uma caracterização, analisamos somente dados que tenham resposta “sim” a doenças crônicas pulmonares. Logo, tratamos com cláusulas “ou” se havia pelo menos uma resposta “sim” para os 3 atributos principais, o que nos resultou em registros de pessoas que poderiam ter ao menos 1 delas, mesmo que a outra não, diminuindo a perda de dados no tratamento .

4.2.1.4 Tratamento de outliers

Devido ao foco do estudo, não foi-se necessário o tratamento de outliers, pois o que poderia haver de diferente nos dados seria o atributo idade, com número distorcido, contudo, como o resultado em interesse é o todo, dados mínimos como estes não afetam o resultado.

4.2.2 Análise dos Dados

4.2.2.1 Parâmetros para a Visualização dos dados

Após o processamento dos dados, unificamos atributos de regiões para que facilitem o manuseio dos dados para exibição dos resultados, então criamos faixas como “Grandes Regiões”, assim também foi feito para idade, separada em 5 faixas. Com isto, estava pronto a base para a caracterização e análise com base em gráficos.

4.2.2.2 Visualização dos dados

Posteriormente aos parâmetros, construímos gráficos em pizza, para fácil visualização e exibição dos resultados. Dito isto, o foco principal exibido é doenças pulmonares crônicas, com os demais atributos, como: grandes regiões, faixa da idade e sexo.

4.2.3 Porque não utilizado a machine learning

Nesta pesquisa, o resultado principal não teve auxílio de Machine Learning, pois após utilizar Decision Tree e Random Forest, que são ótimos algoritmos de classificação, notamos que o resultado não estava de fácil interpretação e o esperado do trabalho pode ser obtido de outra maneira. Contudo testamos esses dois códigos pois ambos tem funcionamento baseado em uma estrutura hierárquica que se assemelha a uma árvore, onde cada nó interno representa uma decisão sobre um atributo, cada ramo representa o resultado dessa decisão e cada nó folha representa uma classe ou valor de saída (Quinlan,1993).

5. APLICAÇÃO E RESULTADOS

5.1 Definição do conjunto analisado

O conjunto de dados utilizado neste estudo foi direcionado aos atributos específicos "Q074, Q11606, Q11607" da base do IBGE 2019. Estes atributos foram selecionados com o objetivo de caracterizar as diferentes grandes regiões brasileiras, considerando também idade e sexo dos indivíduos pesquisados.

5.2 Métricas utilizadas para montagem do gráfico

As métricas aplicadas incluíram a resposta "sim" para os principais atributos do questionário, a unificação das regiões para a criação de macro-regiões geográficas, e faixa etária da idade. Essas medidas foram adotadas para facilitar a análise comparativa entre as diferentes regiões do Brasil.

5.3 Resultados e Discussões

Os resultados indicam uma maior incidência de diagnósticos de doenças pulmonares crônicas na região Nordeste, especialmente entre indivíduos com idade entre 30 e 44 anos. Surpreendentemente, constatou-se que o não uso de tabaco ou produtos similares não corresponde a uma menor prevalência dessas condições de saúde na população estudada.

Uma análise mais aprofundada sugere que fatores como o clima predominante na região e aspectos socioeconômicos, como a presença de usinas e outros poluentes ambientais, podem estar influenciando os índices observados. Notavelmente, os registros indicam uma concentração significativa de casos nas regiões Nordeste e Sudeste, destacando a importância de considerar esses elementos no contexto da saúde pública e na formulação de políticas adequadas de prevenção e tratamento (REVIEW, 2021).

6. CONCLUSÃO

Durante todo o progresso do trabalho, tínhamos como objetivos principais, uma caracterização detalhada das doenças crônicas pulmonares por regiões e faixa etária. Esperávamos identificar onde a asma se fazia mais predominante, como por exemplo: grandes regiões, idade e até mesmo o gênero mais atingido. Inicialmente, estávamos com dúvidas sobre a base de dados utilizada, pois não estava boa como o esperado, porque notamos um baixo número de registros de causas, causas essas que tinham correlação com doenças pulmonares crônicas, o que nos levou a mudança do tema, voltado-se mais a regiões, idade e sexo. Após a mudança, obtivemos uma quantidade ok para se trabalhar, além da questão numérica, lidamos com dados nulos, os quais foram tratados da melhor maneira para não se perder mais registros, o que levou a atingirmos nosso objetivo com bastante precisão e mesmo com pouco número dados.

Com os resultados obtidos, podemos afirmar que, com a base de dados do IBGE(2019), as doenças pulmonares crônicas tem-se sido predominantes, em regiões que detenham climas mais secos, que tem um grande número de poluições no ar, onde a população tem a inalação de materiais tóxicos ao sistema respiratório, questões socioeconômicas e idade. Com isso damos mais um passo na maneira de tratamento e prevenção das mesmas, fazendo assim com que a saúde seja cada dia mais eficaz.

6.1 Pontos fortes e fracos do Modelo

Por conta dos métodos utilizados, tivemos uma capacidade de trabalhar com um grande volume de dados e identificar os padrões complexos que não seriam identificados com as técnicas de estatística mais tradicionais. Embora não tenhamos tido sucesso com os modelos de Decision Tree e Randon Forest, que foram inicialmente testados, a abordagem permitiu que a gente testasse diferentes possibilidades analíticas.

Por outro lado, temos desvantagens em relação aos métodos usados, como por exemplo, a complexidade na leitura dos resultados obtidos. Além disso, a necessidade de um pré-processamento extenso e a seleção um pouco mais criteriosa dos atributos que fossem relevantes, foi algo mais desafiador porque poderiam limitar a aplicabilidade desses métodos em alguns contextos.

Contudo o que conseguimos levar desta pesquisa foi, o aprendizado de, o que um cientista de dados realiza com suas análises e aplicações de Machine Learning, visando ajudar o meio pesquisado de diversas maneiras e quem sabe, resolver problemas.

6.2 Trabalhos Futuros

Como trabalho futuro, planeja-se que com o aumento da base de dados ou base nova e dados mais enriquecidos que envolvam as possíveis causas para essas enfermidades, seja possível realizar uma predição, a fim de diminuir os casos das mesmas, ou melhorar a maneira de combater e preveni-las atualmente.

REFERÊNCIAS

BARROS, M. B. ET AL. RELATÓRIO FINAL DA PESQUISA NACIONAL DE SAÚDE 2019. [S.L: S.N.].

BARROS, M.B.A. (2019). PESQUISA NACIONAL DE SAÚDE: CARACTERIZAÇÃO DAS DOENÇAS CRÔNICAS NA POPULAÇÃO BRASILEIRA. [HTTPS://WWW.SCIELO.BR/J/CSC/A/RYSFFTQRWZPZNGHSQ5CJHSG/]

DOENÇAS CRÔNICAS SÃO RESPONSÁVEIS POR 54,7% DOS ÓBITOS NO BRASIL. PORTAL DO GOVERNO. [HTTPS://WWW.GOV.BR/SAUDE/PT-BR/ASSUNTOS/NOTICIAS/2021/SETEMBRO/SAUDE-APRESENTA-ATUAL-CENARIO-DAS-DOENCAS-NAO-TRANSMISSIVEIS-NO-BRASIL]

EPIDEMIOLOGY AND BURDEN OF CHRONIC RESPIRATORY DISEASES IN BRAZIL FROM 1990 TO 2017: ANALYSIS FOR THE GLOBAL BURDEN OF DISEASE 2017 STUDY[[HTTPS://WWW.SCIELOSP.ORG/ARTICLE/RBEPID/2020.V23/E200031/](https://www.scielosp.org/article/rbepid/2020.v23/e200031/)]

HOSPITAL MOINHOS DE VENTO. A ASMA E O INVERNO. DISPONÍVEL EM: <HTTPS://WWW.HOSPITALMOINHOS.ORG.BR/INSTITUCIONAL/BLOGSAUDEEVOCE/A-ASMA-E-O-INVERNO>. ACESSO EM: 16 JUN. 2024.

MACHINE LEARNING: APRENDIZADO DE MÁQUINAS NA SAÚDE E PARA A PREDIÇÃO DE DOENÇAS CRÔNICAS NÃO TRANSMISSÍVEIS. DISPONÍVEL EM: <HTTPS://WP.UFPEL.EDU.BR/EAPELOTAS/2022/10/13/MACHINE-LEARNING-APRENDIZADO-DE-MAQUINAS-NA-SAUDE-E-PARA-A-PREDICAO-DE-DOENCAS-CRONICAS-NAO-TRANSMISSIVEIS/>. ACESSO EM: 16 JUN. 2024.

MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA EM SAÚDE DEPARTAMENTO DE ANÁLISE EM SAÚDE E VIGILÂNCIA DE DOENÇAS NÃO TRANSMISSÍVEIS.2022. DISPONÍVEL EM: [HTTPS://WWW.GOV.BR/SAUDE/PT-BR/CENTRAIS-DE-CONTEUDO/PUBLICACOES/SVSA/DOENCAS-CRONICAS-NAO-TRANSMISSIVEIS-DCNT/09-PLANO-DE-DANT-2022_2030.PDF](https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/svsa/doencas-cronicas-nao-transmissiveis-dcnt/09-plano-de-dant-2022_2030.pdf). ACESSO EM: 16 DE JUNHO DE 2024.

PAIXÃO, G. M. DE M. ET AL. MACHINE LEARNING NA MEDICINA: REVISÃO E APLICABILIDADE. ARQUIVOS BRASILEIROS DE CARDIOLOGIA, V. 118, N. 1, P. 95–102, 2022.

QUINLAN, J. R. C4.5: PROGRAMS FOR MACHINE LEARNING. MORGAN KAUFMANN PUBLISHERS, 1993.

REVIEW, BEATRIZ; THAYNÁ; MARIA, PREVALÊNCIA DA ASMA NAS REGIÕES DO BRASIL: UMA REVISÃO SISTEMÁTICA.UNIVERSIDADE TIRADENTES - UNIT, 2021 JUN;11341(1):2-15. DOI: 10.34119/BJHRV4N3-133.

SCHUELER, P. MINISTÉRIO DA SAÚDE APRESENTA CENÁRIO DAS DOENÇAS NÃO TRANSMISSÍVEIS NO BRASIL. DISPONÍVEL EM: <HTTPS://WWW.BIO.FIOCRUZ.BR/INDEX.PHP/BR/NOTICIAS/2604-MINISTERIO-DA-SAUDE-APRESENTA-CENARIO-DAS-DOENCAS-NAO-TRANSMISSIVEIS-NO-BRASIL>. ACESSO EM: 16 JUN. 2024.

VIEIRA, M. L. F. P. ET AL. DISTRIBUIÇÃO GEOGRÁFICA E FATORES ASSOCIADOS À PREVALÊNCIA DE DOENÇAS CRÔNICAS NA POPULAÇÃO BRASILEIRA: RESULTADOS DA PESQUISA NACIONAL DE SAÚDE 2019. REVISTA BRASILEIRA DE EPIDEMIOLOGIA, 2019..