

Uso da Mineração de Dados na Caracterização da Asma na População Brasileira

Arthur Henrique Chaves Oliveira¹, Sara Alves Martins¹, Luis Enrique Zárate¹

¹Curso Ciência de Dados e Inteligência Artificial
Pontifícia Universidade Católica de Minas Gerais (PUC-Minas)
CEP – 30140-100 – Belo Horizonte – MG – Brasil

Abstract. *Asthma is a chronic respiratory disease with high prevalence in Brazil, influenced by genetic, environmental and lifestyle factors. Understanding the profile of individuals with asthma contributes to improving public health strategies and disease control. This work aims to evaluate data from the 2019 National Health Survey (PNS) to characterize asthma in the Brazilian population. As a main part of the proposed methodology, a conceptual modeling of the problem domain was performed, identifying factors and their multiple dimensions to guide data preparation. As a result, predictive models were developed to classify individuals with or without asthma, seeking to find valuable patterns for future data science research applied to health.*

Resumo. *A asma é uma doença respiratória crônica com alta prevalência no Brasil, influenciada por fatores genéticos, ambientais e relacionados ao estilo de vida. Compreender o perfil dos indivíduos com asma contribui para aprimorar estratégias de saúde pública e o controle da doença. Este trabalho tem como objetivo avaliar dados da Pesquisa Nacional de Saúde (PNS) de 2019 para caracterizar a asma na população brasileira. Como parte principal da metodologia adotada, foi realizada uma modelagem conceitual do domínio do problema, identificando fatores e suas múltiplas dimensões para orientar a preparação dos dados. Como resultado, foram desenvolvidos modelos preditivos para classificar indivíduos com ou sem asma, buscando encontrar padrões valiosos para futuras pesquisas da ciência de dados aplicadas à saúde.*

1. Introdução

A asma é uma condição que apresenta uma alta prevalência no Brasil e no mundo, estando presente em torno de 20 milhões de brasileiros, sendo responsável por aproximadamente 350.000 internações anualmente no Sistema Único de Saúde (SUS). Considerada uma doença imunomediada e multifatorial, a asma apresenta uma clínica variada e complexa, caracterizada por broncoconstricções episódicas e reversíveis, bem como por inflamações das vias aéreas. Esse último caso pode ser desencadeado por fatores como infecções, alérgenos ambientais e irritantes [Rodrigues et al. 2021].

O diagnóstico da asma é baseado em critérios clínicos e funcionais, obtidos por meio de anamnese, exame físico e exames de função pulmonar, também conhecido como a espirometria. Os sintomas característicos incluem tosse, cansaço e aperto no peito. Em crianças de até quatro anos esse diagnóstico é predominantemente clínico, devido à dificuldade da realização de provas funcionais [Guimarães et al. 2024].

Um estudo realizado por [Hisinger-Mölkänen et al. 2022] analisou respostas de uma pesquisa conduzida pela Statistics Finland e identificou, por meio de testes chi-quadrado, com significância de 5%, que indivíduos diagnosticados com asma durante a fase adulta tendem a ser mais sintomáticas do que aquelas diagnosticadas na infância. Essa descoberta sugere que a asma diagnosticada precocemente tem uma resposta mais favorável ao tratamento e maior probabilidade de melhora, enquanto o diagnosticada na idade adulta pode ser associada a uma condição crônica. Além disso, [Menezes et al. 2015] aponta que, na fase adulta, a asma é mais prevalente no sexo feminino, ao contrário da infância, em que o sexo masculino é mais afetado.

No Brasil, a região sul é a que apresentou as maiores prevalências do diagnóstico de asma, independente da faixa etária [Menezes et al. 2015][Ramos et al. 2021]. Esse fenômeno pode estar relacionado as baixas temperaturas da região, que, segundo [Xavier et al. 2022], favorecem as doenças respiratórias, podendo ser notado um aumento dessa doença. Esse aumento pode ser observado em outras regiões do país durante os períodos de outono e inverno.

No que diz respeito ao controle da asma, diversos estudos discutem o assunto [Lual and Awoke 2021][Meltzer et al. 2019][Estanislau et al. 2021]. Estes trabalhos abordam temas como a utilização de medicamentos prescritos, influência da localização de moradia, a prática de exercícios respiratórios e a relação entre o tempo e qualidade do sono do indivíduo. Outro estudo investigou os níveis de vitamina D entre pacientes que possuíam asma e os que não possuíam, e dois subgrupos que praticavam ou não atividades físicas [Lorensia et al. 2019]. Os resultados apresentaram uma diferença nos níveis de vitamina D entre as pessoas que tinha diagnóstico positivo e negativo, além de apresentar uma forte correlação entre os níveis da vitamina D e a prática de atividades físicas.

O objetivo deste trabalho é descrever o perfil do indivíduo com asma em regiões brasileiras, aplicando um processo de descoberta de conhecimento que busca identificar os principais fatores que caracterizam a doença dentro da população. Para isso, é considerado o estudo mais recente da Pesquisa Nacional de Saúde (PNS) do IBGE [IBGE 2020]. Esse estudo coletou dados sobre a saúde e os estilos de vida dos brasileiros, no ano de 2019. Por meio de modelos computacionais, busca-se revelar padrões específicos da asma na população brasileira, contribuindo para um entendimento mais aprofundado dessa doença no país. Espera-se que os resultados possam auxiliar no desenvolvimento de estratégias mais eficazes para o diagnóstico, tratamento e controle da asma no Brasil.

2. Trabalhos Relacionados

O trabalho de [Kaplan et al. 2021] realiza uma revisão sobre o uso da inteligência artificial dentro da área da saúde e particularmente nas doenças respiratórias, tendo um foco maior na asma e nas doenças pulmonares obstrutivas crônicas (DPOC). Por meio da revisão de trabalhos relacionados, os autores chegaram a conclusão que o uso de IA na área da saúde tem potencial de ajudar no diagnóstico de pacientes e na diferenciação, entre outras doenças, de doenças respiratórias. É possível destacar então a importância das técnicas de aprendizado de máquina dentro da área da saúde.

Ademais, em [Xavier et al. 2022] é abordado a influência da sazonalidade climática no número de casos de doenças respiratórias em crianças de 0 a 9 anos. Os

autores realizaram essa pesquisa no município de Campina Grande, estado da Paraíba (PB), com dados que abrangem um período de 23 anos (janeiro de 1998 até dezembro de 2020), analisando a frequência da pneumonia, bronquite/bronquiolite e asma durante os períodos sazonais, por meio de modelos Autorregressivos Integrados por Médias Móveis com Sazonalidade (SARIMA). Os resultados do modelo buscavam prever o número de casos para os anos de 2021 e 2022 e, no caso da asma, não houve variação na taxa de ocorrência de internações.

Já no estudo de [Razavi-Termeh et al. 2021] os autores buscaram implementar um modelo de Relação de Frequência (RF) para os dados de pessoas que possuíam asma e dos dados vindo do Sistema de Informação Geográfica (SIG) do Iran. Na pesquisa, foi realizado um total de 5 passos, sendo eles a criação de uma base de dados espaciais utilizando a localização de crianças com asma e treze fatores ambientais que afetam a asma, por exemplo os dados de poluição do ar, trânsito local, produtos agrícolas utilizados na área. O segundo passo envolveu a análise da relação espacial entre pessoas com asma e os fatores ambientais. Já o terceiro passo analisava a correlação da incidência da doença com a localização. No quarto passo, o modelo foi utilizado para determinar as áreas de maior propensão a asma. E, no último passo, o modelo foi avaliado utilizando a curva e análise sensitiva da Característica de Operação do Receptor (ROC). Na conclusão desse estudo foi determinado uma ótima acurácia por parte do modelo, além de determinar as variáveis mais importantes dentre os dados vindos do SIG.

Outro trabalho utilizou de dados de pacientes com asma para tentar prever, por meio de aprendizado de máquina (AM), uma piora da asma em paciente diagnosticados [Finkelstein and Jeong 2017]. Através de modelos de AM, como a rede Bayesiana adaptativa, classificador Naive Bayes e SVM, chegaram a conclusão que técnicas de AM possuem um potencial significativo na predição da piora da asma em pacientes.

3. Metodologia

3.1. A Base de Dados

Neste estudo foi utilizado a base de dados da Pesquisa Nacional de Saúde (PNS) de 2019. Esta pesquisa leva em consideração os dados demográficos dos entrevistados, características da atividade laboral e entre outras características. A base possui 1.088 atributos e 293.726 instâncias, dentre as quais foram identificadas 90.846 pessoas que responderam se possuíam asma ou não, contendo 4.467 respostas de diagnóstico positivo. Utilizando esses dados, o objetivo é classificar se um indivíduo possui ou não asma com base em atributos previamente selecionados conceitualmente.

Para definir o escopo do problema, foi realizada uma análise exploratória da base de dados, utilizando gráficos de distribuição de frequência acumulada por região, estado e faixa etária. Ademais, foi realizado o teste chi-2, para análise de proporcionalidade, para separar as regiões mais semelhantes em relação ao número de casos da classe, obtendo assim uma separação em duas grandes regiões: A primeira composta pelas regiões Norte, Nordeste e Centro-Oeste, e a segunda pelas regiões Sul e Sudeste. Analisando a distribuição acumulada por idade, observou-se uma maior incidência de casos da doença até os 50 anos de idade, possuindo uma maior concentração no período da infância. A partir da análise exploratória, definiu-se que o escopo abrangeia as pessoas com menos de 50 anos e que residiam na segunda região.

3.2. Modelagem Conceitual do problema

Para a criação de um modelo conceitual do problema, que nos ajude a entendê-lo e identificar seus atributos relacionados, foi-se utilizado o método CAPTO, proposto em [Gonçalves et al. 2024]. Este método busca reunir conhecimento tácito, vindo do especialista de domínio, com o conhecimento explícito, que vem de diversas fontes técnicas, para criar um modelo conceitual que ajude na seleção de atributos para projetos na área de ciência de dados. Estes modelos são divididos em dimensões, a maior abstração do problema, que, posteriormente, são divididas em aspectos — um entendimento mais específico da situação — e, por último, em atributos potenciais.

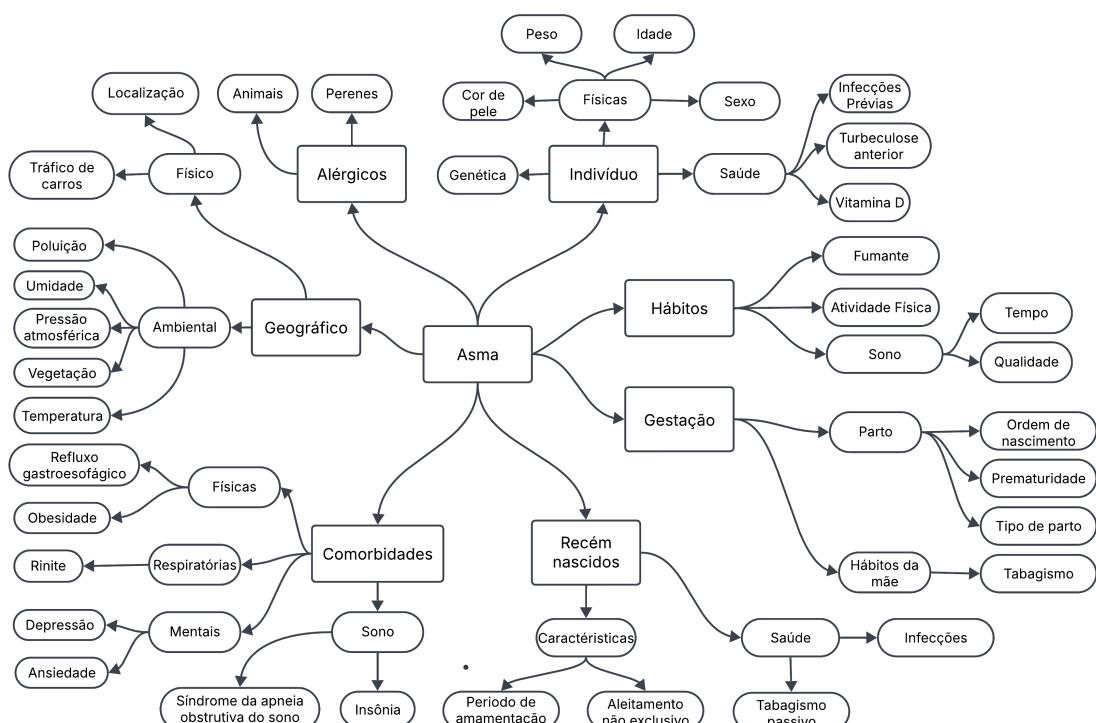


Figura 1. Modelo Conceitual

As dimensões identificadas no domínio do problema foram: Indivíduos (1), Hábitos (2), Gestação (3), Recém nascidos (4), Comorbidades (5), Geográfico (6), Alérgicos (7). Os Aspectos que foram identificados na literatura, associados as suas respectivas dimensões, podem ser visualizados na Tabela 1.

Tabela 1. Descrição do Mapa Conceitual – Domínio de problema: Asma

Descrição do Mapa Conceitual – Domínio de problema: Asma		
Dimensão: Indivíduos		
Aspectos	Atributos Associados	Bases de Dados
Genético: A fisiopatogenia da asma, que acomete as vias aéreas inferiores, resulta de uma complexa interação genética e ambiental[Ramos et al. 2021]	- Atopia - Asma nos genitores	Não foram encontrados registros para esse atributo
Físicos: Diagnósticos de asma após a fase adulta fazem com que os fatores de risco da doença sejam mais sintomáticos [Hisinger-Mölkänen et al. 2022]. Na idade adulta há uma maior prevalência de asma no sexo feminino [Menezes et al. 2015]. Uma associação entre o excesso de peso e asma foi achado no estudo de [Estanislau et al. 2021].	- Idade - Sexo - Peso	Módulo C - Características gerais dos moradores e Módulo W - Antropometria: C6, C8, C9, W00103; Fonte: BD-PNS

Continua na próxima página

Aspectos	Atributos Associados	Bases de Dados
Saúde: Pacientes que tenham um diagnóstico anterior de tuberculose possuem um fator de risco maior para a asma [Ramos et al. 2021]. Uma relação entre vitamina D e pessoas com asma foi apontada por [Lorensia et al. 2019]	- Diagnóstico anterior de tuberculose - Vitamina D	Não foram encontrados registros para esse atributo
Dimensão: Hábitos		
Sono: O tempo de sono possui um impacto objetivo e subjetivo nos sintomas diários da asma [Meltzer et al. 2019] A qualidade do sono, uma medida subjetiva, foi relacionada a asma em [Estanislau et al. 2021]	- Tempo de sono	Módulo N - Percepção do estado de saúde: N10; Fonte: BD-PNS
Atividade física: Existe um diferença significativa entre praticantes de atividade física entre asmáticos e não asmáticos [Lorensia et al. 2019]	- Prática de atividade física - Tempo de prática	Módulo P - Estilos de Vida e Módulo M - Características do trabalho e apoio social: M16a, P34, P35, P37; Fonte: BD-PNS
Fumante: O hábito de fumar foi considerado em [Lual and Awoke 2021]	- Hábito de fumar	Módulo P - Estilos de vida: P50, P52, P53, P59; Fonte: BD-PNS
Dimensão: Gestação		
Parto: A prematuridade também se configura como um fator de risco para asma, sendo bem prevalente nesse grupo [Rodrigues et al. 2021]. Há uma íntima relação entre a idade gestacional, ordem de nascimento e tipo de parto influenciam na asma [Ramos et al. 2021]	- Prematuridade - Idade gestacional - Ordem de nascimento - Tipo de parto	Não foram encontrados registros para esse atributo
Hábitos da mãe: Fatores como exposição precoce à fumaça do tabaco (mães que fumaram na gestação ou passiva) nos primeiros 6 meses de vida também estão diretamente relacionados como um fator de risco [Rodrigues et al. 2021].	- Tabagismo	Não foram encontrados registros para esse atributo
Dimensão: Recém nascidos		
Saúde: Infecções de via aérea superior (entre elas as mais frequentes são as por vírus sincicial respiratório, adenovírus e por influenzae) nos primeiros 6 meses de vida estão relacionados como um fator de risco [Rodrigues et al. 2021].	- Infecções (recém nascidos)	Não foram encontrados registros para esse atributo
Características: aleitamento materno não exclusivo nos primeiros 6 meses de vida também é ligado aos fatores de risco [Rodrigues et al. 2021]. A duração da amamentação possui uma relação com o desenvolvimento da asma [Ramos et al. 2021]	- Aleitamento materno não exclusivo. - Duração da amamentação	Módulo L - Crianças com menos de dois anos de idade: L01701, L18; Fonte: BD-PNS
Dimensão: Comorbidades		
Físicas: Várias condições consideradas no diagnóstico diferencial da asma podem ocorrer como comorbidades, como rinossinusites e doença do refluxo gastroesofágico e contribuem para o agravamento dos sintomas da asma [Guimarães et al. 2024]. A obesidade é um fator de risco independente para a asma, ou seja, pode contribuir para o desenvolvimento da mesma [Estanislau et al. 2021]	- Refluxo gastroesofágico - Obesidade	Não foram encontrados registros para esse atributo
Respiratórias: Rinossinusites contribuem para o agravamento dos sintomas da asma e piora na qualidade de vida dos asmáticos [Guimarães et al. 2024]. O estudo de [Ramos et al. 2021] aponta uma íntima relação da rinite e a asma	- Rinossinusites - Rinite	Módulo Q - Doenças crônicas: Q116a; Fonte: BD-PNS
Mentais: Há uma associação significante entre asma e problemas mentais comuns. Ansiedade, depressão, presença de conflitos entre a família foram associados com a baixa busca de tratamento e a maior morbidade e mortalidade da doença [Estanislau et al. 2021]	- Depressão - Ansiedade	Módulo Q - Doenças crônicas: Q92, Q110a; Fonte: BD-PNS
Sono: Em [Lual and Awoke 2021] distúrbios do sono e apneia do sono são categorizados como fatores relacionados a asma	- Apneia do sono - Insonia	Não foram encontrados registros para esse atributo
Dimensão: Geográfico		
Ambiental: A região Sul foi a que mostrou prevalências mais altas de diagnóstico médico de asma, um dos fatores responsável é o clima [Menezes et al. 2015]	- Temperatura - Umidade	Não foram encontrados registros para esse atributo
Físico: Alguns estudos realizados pelo mundo detectaram maiores prevalências de asma na região urbana comparada à rural [Menezes et al. 2015]	- Localização	Parte 1 - Identificação e Controle: V0001; Fonte: BD-PNS
Dimensão: Alérgicos		
Perenes: Exposição a alérgenos e infecções virais da via aérea superior agem em sinergia ativando essa hiper-responsividade brônquica e contribuindo pro desenvolvimento da asma. [Rodrigues et al. 2021].	- Poeira - Mofo - Camundongo - Rato - Barata	Não foram encontrados registros para esse atributo
Animais: No estudo de [Lual and Awoke 2021] a presença de animais na vida do paciente foi considerado relacionado a asma.	- Gatos - Cachorros - Outros	Módulo A - Informações do Domicílio: A22a, A23a; Fonte: BD-PNS

3.3. Pré-Processamento e preparação de dados

Após a seleção conceitual dos atributos, o conjunto de dados passou por um processo de preparação. Esse processo consistia em realizar a remoção de outliers, tratamento de dados vazios, imputação de valores ausentes e a codificação dos atributos.

Etapa 1: Análise dos dados e tratamento de outliers

A análise dos atributos foi feita por meio de um relatório gerado por código que informava, para cada valor numérico, o mínimo, o máximo, a média, o desvio padrão, a quantidade de dados faltantes e a porcentagem dos mesmos. Para os valores categóricos

era exibido, para cada possibilidade de valores no atributo, o número de vezes que ele apareceu na base, a porcentagem dele na base e a quantidade de instâncias por classe.

Com base nessa análise, retiramos atributos categóricos que possuíam uma discrepância muito grande entre seus valores, como os atributos *Q11604* e *Q11006*.

Já na análise de outliers foi realizada por meio de gráfico boxplot e pelo método IQR, filtrando todos valores acima e abaixo do intervalo, que revelou a existência de diversas instâncias outliers. Estes valores foram excluídos da tabela, com exceção daqueles presentes nos atributos *P00104* que possuía valores significativos, pois essas instâncias poderiam representar pessoas acima do peso e obesas, e *P059*, que indicava a quanto tempo o entrevistado não fumava.

Etapa 2: Tratamento de dados vazios

Essa etapa se trata da junção ou preenchimento de atributos relacionados, buscando reduzir a dimensionalidade da base de dados e os atributos ausentes. Os atributos agrupados foram: *A02201*, *A02305* e *A02306*; *P034* e *P035*; *P03701* e *P03702*; *P050* e *P052*; *P059*; *P053*.

Os atributos dependentes de outros para serem respondidos, como os atributos "Quantos destes animais são gatos"(*A02305*) e "Quantos destes animais são cachorros"(*A02306*) que estavam condicionados à resposta do atributo "Em seu domicílio, há algum animal de estimação"(*A02201*), foram preenchidos com o valor zero caso o entrevistado informasse que não possuía animais de estimação. Posteriormente, esses atributos foram agrupados em um único valor, excluindo os atributos originais. Este processo foi repetido para cada um dos atributos informados inicialmente, com poucas modificações no processo de concatenação.

Etapa 4: Imputação de valores ausentes, criação de atributos e discretização

Após todos os outros processos, as únicas variáveis que possuíam dados ausentes eram numéricas. As imputações de valores foram feitas por meio da função *KNNImputer*. Antes de imputar os valores, separou-se as instâncias das pessoas que tinham asma das que não possuíam, para que possíveis ruídos não aparecessem na base de dados.

Ademais, houve a criação do atributo *IMC*, que indica o nível de massa corporal do indivíduo, categorizando-o entre normal, sobre peso ou obeso, essas faixas foram definidas por [World Health Organization 2025] e [BRASIL 2020]. Outro atributo que foi discretizado foi *P053*, o qual indica com quantos anos o entrevistado começou a fumar, onde as faixas foram retiradas de [Vineis et al. 1984]. O atributo *P059*, que informa há quanto tempo a pessoa parou de fumar, teve suas faixas definidas em [Nunes 2006]. Alguns outros atributos, como idade, foram discretizados de forma estatística para que dessa forma fosse mantida uma distribuição de tamanho igual entre as faixas.

Etapa 5: Seleção de atributos pela análise de entropia

A análise de entropia foi feita com o objetivo de realizar uma redução da dimensionalidade da base, removendo os atributos que não agregavam no modelo (alta entropia em relação à classe) ou aqueles que eram uma classificação direta (baixa entropia) do nosso atributo alvo.

Os resultados apontaram que não há nenhum atributo dominante, aqueles que re-

alizam classificação direta, ou que o grau de confusão é muito alto, ou seja, aqueles que não agregariam no resultado. Os valores, para todos os atributos, ficaram próximos de 0,33, o que pode indicar uma seleção conceitual adequada.

4. Experimentos e análise dos resultados

Foram construídos dois tipos de modelos, um modelo caixa-branca, interpretável, e um baseado em Árvore de decisão, um modelo Ensemble, baseado em Random Forest. Para o treinamento destes modelos foi aplicada para o processo de treinamento e teste, a técnica *hold out*, sendo 70% para treino e 30% para teste. A base de dados está disponível em <https://github.com/licapLaboratory/Database-Asma>

Para a construção dos modelos, utilizou-se o python no desenvolvimento. Foi realizado uma etapa adicional de balanceamento, pelo método de Random Undersampling(RUS), do conjunto de treino e teste. A base então, agora completamente preparada, foi utilizada no treinamento dos modelos que foram otimizados pelo algoritmo *BayesSearchCV*. Os melhores hiperparâmetros encontrados podem ser observados na Tabela 2. Após o treinamento e teste dos modelos, é possível analisar as métricas de avaliação pela Imagem 2 e observar as matrizes de confusão dos modelos pela tabela 3.

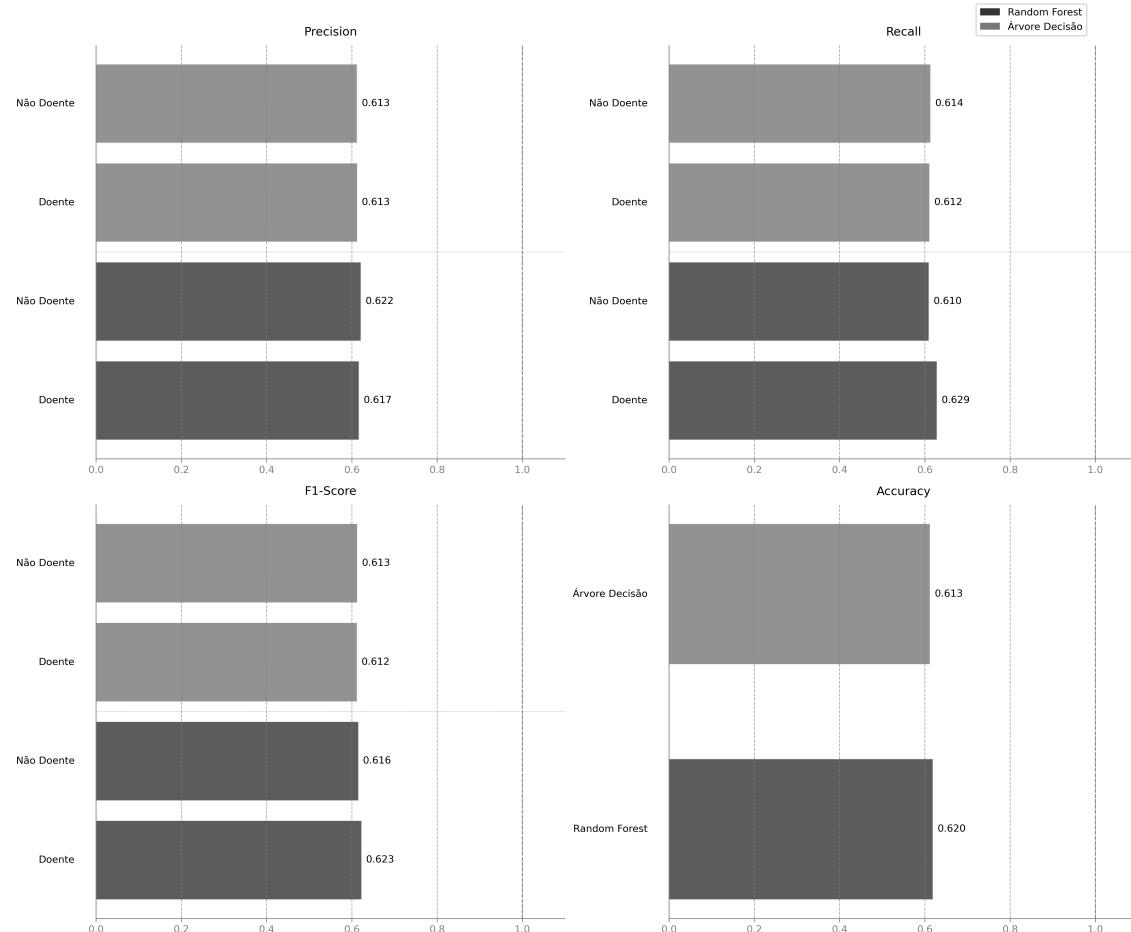


Figura 2. Métricas dos modelos

Tabela 2. Valores de hiperparâmetros

modelo	critério	max_depth	min_samples_leaf	min_samples_split	max_features	n_estimators
Árvore de decisão	entropy	4	1	20	x	x
Random Forest	entropy	6	3	2	sqr	50

Tabela 3. Matrizes de confusão: Árvore vs Random Forest

		Árvore		Random Forest	
		Doente	Não Doente	Doente	Não Doente
Doente	320	203	329	194	
Não Doente	202	321	204	319	

Os modelos de árvore utilizados apresentaram um desempenho geral aceitável, com a Árvore de decisão se destacando levemente (F1-score médio de 61,2%) em termos de métricas de desempenho. Devido ao nosso problema alvo, a métrica mais importante se trata do recall, que teve um resultado consideravelmente aceitável para ambos os modelos.

A árvore de decisão gerou regras bem condizentes com o esperado proposto pela revisão conceitual, tendo como os atributos mais importantes: *P050—P052*(Pessoa fumante ou não, e caso fume a frequência), *Q092*(Diagnóstico de depressão) e *A02201 |A02305 |A02306*(Quantos animais existem na residência). Foram geradas 15 regras no total, sendo as 3 com maior cobertura as seguintes:

Regra 1: SE já fumou ou fuma atualmente E não foi diagnosticado com depressão E Não possui problemas com sono quase todos os dias E faixa de idade entre 15 e 24 anos → Classificação: Não tem ama. Possuindo uma cobertura de 41,32%, representando 486 registros.

Regra 2: SE nunca fumou E possui 1 ou menos animais em casa E menos de 5 pessoas residem na casa E o tempo de treino médio é menor que 90 minutos → Classificação: Tem Asma. Tendo uma cobertura de 39,37% com 463 registros.

Regra 3: SE nunca fumou E possui mais de 4 animais em casa E pratica atividades físicas → Classificação: Não tem asma. Tendo uma cobertura de 8,5% com 100 registros.

5. Conclusão

Este estudo buscou analisar o perfil dos brasileiros que possuem asma, diferenciando-as daquelas que não possuem. Utilizando dados oriundos da Pesquisa Nacional de Saúde (PNS) de 2019, foram aplicadas técnicas de aprendizado de máquina, desenvolvendo modelos preditivos capazes de identificar características dos dois grupos de indivíduos. A análise exploratória identificou um número maior na incidência de asma nos indivíduos com menos de 50 anos de idade, tendo seu ápice na população mais nova, o que indica a importância de uma pesquisa futura buscando analisar este grupo específico.

Dois modelos foram analisados, sendo a Árvore de Decisão, e Floresta aleatória. Os modelos derivados de árvore apresentaram um desempenho satisfatório, o atributo *P050—P052*(Pessoa fuma ou não, e caso fume a frequência) foi o que mais se destacou em relação aos demais, sendo seguido por *C006*(Sexo do indivíduo) e *P034—P035*(Se pratica ou não atividade física). Os resultados sugerem que hábitos que impactam a saúde

pulmonar, assim como o sexo da pessoa, podem ser fatores relevantes na categorização de indivíduos com asma para a população estudada.

Como trabalhos futuros seria necessário a inclusão de atributos adicionais relacionados as condições ambientais no qual os indivíduos vivem, como índice de poluição, umidade, vegetação, categorização do município em que vive e a temperatura. Ademais, a aplicação de uma seleção das instâncias mais representativas seria relevante para este contexto.

Agradecimentos

Os autores agradecem o apoio recebido do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Processo No 303133/2021-0, e da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Processo PCE-00349-25.

Referências

- BRASIL, M. d. S. (2020). Sobrepeso e obesidade em adultos. *PORTARIA SCTIE/MS*, (53):1–3.
- Estanislau, N. R. d. A., Jordão, E. A. d. O. C., Abreu, G. d. A., Bloch, K. V., Kuschnir, M. C. C., Felix, M. M., and Kuschnir, F. C. (2021). Association between asthma and sleep hours in brazilian adolescents: Erica. *Jornal de Pediatria*, 97(4):396–401.
- Finkelstein, J. and Jeong, I. c. (2017). Machine learning approaches to personalize early prediction of asthma exacerbations. *Annals of the New York Academy of Sciences*, 1387(1):153–165.
- Gonçalves, L., Franca, D., and Zarate, L. (2024). Relevância do entendimento do domínio de problema na construção de modelos computacionais de aprendizado. In *Anais do XVIII Brazilian e-Science Workshop*, pages 135–142, Porto Alegre, RS, Brasil. SBC.
- Guimarães, A. C. C. M., Ribeiro Filho, M. N., Silva, A. C. P., de Almeida Cunha, J., da Silva Neto, J. T., de Paula, L. R. F., da Cunha, L. C., Martins, G. H. C., Ribeiro, A. L., Araújo, I. V. G., et al. (2024). Diagnóstico e tratamento da asma: uma revisão de literatura. *Brazilian Journal of Implantology and Health Sciences*, 6(8):5230–5240.
- Hisinger-Mölkänen, H., Honkämäki, J., Kankaanranta, H., Tuomisto, L., Backman, H., Andersén, H., Lindqvist, A., Lehtimäki, L., Sovijärvi, A., Rönmark, E., Pallasaho, P., Ilmarinen, P., and Piirilä, P. (2022). Age at asthma diagnosis is related to prevalence and characteristics of asthma symptoms. *World Allergy Organization Journal*, 15:100675.
- IBGE (2020). Pesquisa nacional de saúde 2019. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/saude/9160-pesquisa-nacional-de-saude.html>.
- Kaplan, A., Cao, H., FitzGerald, J. M., Iannotti, N., Yang, E., Kocks, J. W., Kostikas, K., Price, D., Reddel, H. K., Tsiligianni, I., Vogelmeier, C. F., Pfister, P., and Mastoridis, P. (2021). Artificial intelligence/machine learning in respiratory medicine and potential role in asthma and copd diagnosis. *The Journal of Allergy and Clinical Immunology: In Practice*, 9(6):2255–2261.

- Lorensia, A., Suryadinata, R., and Saputra, R. (2019). Physical activity and vitamin d level in asthma and non-asthma. *Jurnal Farmasi Indonesia*, 11:454–465.
- Lual, K. and Awoke, M. (2021). Factors associated with suboptimal control of asthma among adult asthma patients: A cross-sectional study. *The Open Respiratory Medicine Journal*, 15:35–42.
- Meltzer, L., Beebe, D., Jump, S., Flewelling, K., Sundström, D., White, M., Zeitlin, P., and Strand, M. (2019). Impact of sleep opportunity on asthma outcomes in adolescents. *Sleep Medicine*, 65.
- Menezes, A. M. B., Wehrmeister, F. C., Horta, B., Szwarcwald, C. L., Vieira, M. L., and Malta, D. C. (2015). Prevalência de diagnóstico médico de asma em adultos brasileiros: Pesquisa nacional de saúde, 2013. *Revista Brasileira de Epidemiologia*, 18:204–213.
- Nunes, E. (2006). Consumo de tabaco. efeitos na saúde. *Revista Portuguesa de Medicina Geral e Familiar*, 22(2):225–44.
- Ramos, B. G., Martins, T. B. D., and de Castro, M. E. P. C. (2021). Prevalência da asma nas regiões do brasil: uma revisão sistemática/prevalence of asthma in brazil's five geographic regions: a systematic review+. *Brazilian Journal of Health Review*, 4(3):11341–11359.
- Razavi-Termeh, S. V., Sadeghi-Niaraki, A., and Choi, S.-M. (2021). Asthma-prone areas modeling using a machine learning model. *Scientific Reports*, 11(1):1912.
- Rodrigues, A. S., Sobrinho, L. A., Ferreira, B. D., Mota, S. M., Cardoso, I. C., Rahal, M. R., Melchiori, B. R., de Lima Rossi, A. L., Moreira, L. S., and Miura, F. K. (2021). Abordagem geral da asma: uma revisão narrativa. *Revista Eletrônica Acervo Médico*, 1(2):e9129–e9129.
- Vineis, P., Estève, J., and Terracini, B. (1984). Bladder cancer and smoking in males: Types of cigarettes, age at start, effect of stopping and interaction with occupation. *International Journal of Cancer*, 34(2):165–170.
- World Health Organization (2025). Obesity and overweight. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>.
- Xavier, J. M. d. V., Silva, F. D. d. S., Olinda, R. A. d., Querino, L. A. L., Araujo, P. S. B., Lima, L. F. C., Sousa, R. S. d., and Rosado, B. N. C. L. (2022). Climate seasonality and lower respiratory tract diseases: a predictive model for pediatric hospitalizations. *Revista Brasileira de Enfermagem*, 75(2):e20210680.