

# Caracterizando o perfil Artrite-Depressão da População Brasileira - Um Estudo Baseado na Base de Dados PNS 2019

Luís Fernando R. Cancella<sup>1</sup>, Luis E. Zárate<sup>1</sup>

<sup>1</sup>Curso de Ciência de Dados e Inteligência Artificial  
Laboratório de Inteligência Computacional Aplicada - LICAP  
Pontifícia Universidade Católica de Minas Gerais (PUC-Minas)  
CEP — 30140-100 — Belo Horizonte — MG — Brasil

lcancella@sga.pucminas.br, zárate@pucminas.br

**Abstract.** *This study aims to explore the relationship between depression and arthritis in the Brazilian population by identifying patterns using Data Mining techniques. The methodology involves conceptual modeling, exploratory analysis, pre-processing, data preparation, and the transformation and combination of attributes to reduce dimensionality. The models are based on decision trees and random forests. The results indicated the accuracy and recall of the models, reaching values between 76% and 80%. In addition, the results showed that factors such as gender, the presence of diabetes, and self-perception of health are the most relevant for characterizing this profile for people over 60 years old.*

**Resumo.** *Este trabalho tem como objetivo explorar a relação entre depressão e artrite na população brasileira, por meio da identificação de padrões via técnicas de Mineração de Dados. A metodologia envolve modelagem conceitual, análise exploratória, pré-processamento e preparação dos dados, com transformação e combinação de atributos para redução da dimensionalidade. Os modelos são baseados em árvore de decisão e floresta aleatória. Os resultados indicaram a acurácia e o recall dos modelos, atingindo valores entre 76% e 80%. Além disso, a análise de uso dos atributos pelos modelos destacou fatores como sexo, presença de diabetes e autoperccepção de saúde como os mais relevantes para caracterização desse perfil para pessoas acima de 60 anos.*

## 1. Introdução

A depressão é um transtorno mental caracterizado por sintomas como tristeza persistente, fadiga e perda de interesse em atividades, enquanto a artrite é uma condição crônica que afeta as articulações, causando dor e limitações físicas. Ambas impactam significativamente a qualidade de vida, e sua coexistência pode agravar esses efeitos. Segundo o Ministério da Saúde, a depressão atinge cerca de 15% da população brasileira, percentual possivelmente subestimado, pois muitos buscam apoio nas redes sociais ao invés de recorrer a especialistas [Souza et al. 2020]. Já a artrite, doença autoimune crônica, afeta de 0,5% a 1% da população mundial.

De acordo com [Xue et al. 2020], indivíduos com artrite têm de 35% a 50% mais chance de desenvolver depressão crônica que a população geral. Essa relação evidencia a importância de investigar os fatores associados à comorbidade e suas possibilidades

de intervenção, seja por meio de ações clínicas ou políticas públicas que reduzam seus impactos sociais e econômicos.

Este trabalho tem como objetivo descrever o perfil de indivíduos brasileiros que convivem com artrite e depressão como comorbidade. Para isso, aplica-se um processo de descoberta de conhecimento para identificar os fatores mais relevantes dessa associação. Utilizam-se modelos supervisionados baseados em Árvore de Decisão e Floresta Aleatória para distinguir três grupos: indivíduos saudáveis, indivíduos com artrite sem depressão e indivíduos com a comorbidade. Consideram-se dois cenários: a) modelo para as três classes e b) modelo para distinguir saudáveis e indivíduos com a comorbidade. Os dados são provenientes da Pesquisa Nacional de Saúde (PNS) 2019 do IBGE[IBGE 2020], que traça um panorama da saúde e estilo de vida da população brasileira. Ao focar no contexto nacional, este estudo busca identificar padrões específicos e contribuir para estratégias de prevenção, melhorias no diagnóstico e formulação de políticas públicas de saúde.

## 2. Trabalhos Relacionados

A relação entre artrite e depressão tem sido amplamente estudada, evidenciando a coexistência dessas condições e seus impactos na qualidade de vida. A depressão é a comorbidade mais frequente em idosos com artrite reumatoide (AR), com prevalência variando entre 14% e 48% [Fakra and Marotte 2021].

Os efeitos dessa comorbidade são significativos para o curso das doenças. [Fakra and Marotte 2021] demonstraram que a depressão agrava fatores da AR, como dor, remissão, complicações, qualidade de vida e mortalidade. Além disso, a relação é bidirecional: indivíduos com artrite apresentam maior risco de depressão, e indivíduos deprimidos têm maior risco de desenvolver artrite [Fakra and Marotte 2021, Xue et al. 2020]. [Xue et al. 2020]. Além disso, pacientes acometidos por ambas as condições possuem risco 80% maior de mortalidade por todas as causas em comparação com aqueles afetados apenas por uma delas.

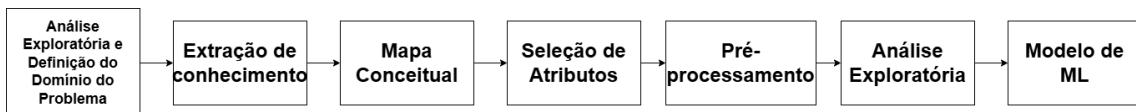
No Brasil, [Mella et al. 2010] investigaram sintomas depressivos e variáveis clínicas e sociodemográficas em pacientes com artrite, encontrando prevalência de 53,2% de sintomas depressivos em pacientes com AR e 28,3% em indivíduos com osteoartrite, além de alta prevalência de ansiedade. Outro estudo relevante, [Campos et al. 2013], analisou qualidade de vida e depressão em indivíduos com AR, evidenciando piores índices de saúde e qualidade de vida em todos os domínios do questionário SF-36 [Laguardia et al. 2013], além de 63,33% de prevalência de sintomas depressivos.

Na área de aprendizado de máquina, [Zárate et al. 2024] identificou padrões de depressão em idosos com base em variáveis exógenas, obtendo 67% de precisão com Floresta Aleatória. Os fatores mais relevantes incluíram doenças crônicas, ausência de rede de apoio e baixa escolaridade. [Imtiaz et al. 2022] revisaram técnicas de aprendizado de máquina no diagnóstico de artrite e destacaram o potencial do aprendizado profundo para prever seu curso. Já [Das and Dhillon 2022] analisaram aplicações de aprendizado de máquina em geriatria, ressaltando a escassez de estudos direcionados à comorbidade entre artrite e depressão.

Esses trabalhos evidenciam a complexidade dessa relação e a necessidade de abordagens integradas para mitigar seus impactos na qualidade de vida dos pacientes.

### 3. Metodologia

Para uma compreensão mais clara das etapas realizadas neste trabalho, é mostrado um fluxograma que detalha o processo de descoberta de conhecimento adotado, ver Figura 1.



**Figura 1. Processo para descoberta de conhecimento adotado**

#### 3.1. Materiais

Neste estudo, foi utilizada a base de dados da Pesquisa Nacional de Saúde (PNS) 2019, [IBGE 2020] conduzida pelo Instituto Brasileiro de Geografia e Estatística (IBGE) em parceria com o Ministério da Saúde. A PNS é uma das principais pesquisas realizadas no Brasil, voltada para a coleta de informações sobre as condições de saúde da população.

A base de dados da PNS 2019 foi selecionada por sua riqueza de informações e escopo nacional, possibilitando uma análise sobre os fatores associados à depressão e à artrite. Essa base contém informações que permitem avaliar tanto aspectos demográficos, como idade, gênero e renda, quanto informações específicas sobre diagnósticos médicos e condições de saúde autorreferidas.

A base de dados contém 293.726 registros e 1.088 atributos, dos quais foram identificados 1620 casos de indivíduos diagnosticados simultaneamente com Artrite e Depressão. Foi considerado o limite de idade de 60 anos, por ter a maior incidência de casos a partir dessa idade.

O objetivo do presente trabalho é investigar a possível relação e similaridade entre essas duas condições, buscando compreender como elas podem estar associadas e quais fatores caracterizam os indivíduos que possuem essa coexistência. Para isso é proposto a construção de modelos de aprendizado baseados em árvore de decisão, por sua capacidade interpretativa, para atender os objetivos.

#### 3.2. Métodos

##### *Etapa 1 - Análise Exploratória e Definição do Domínio Problema:*

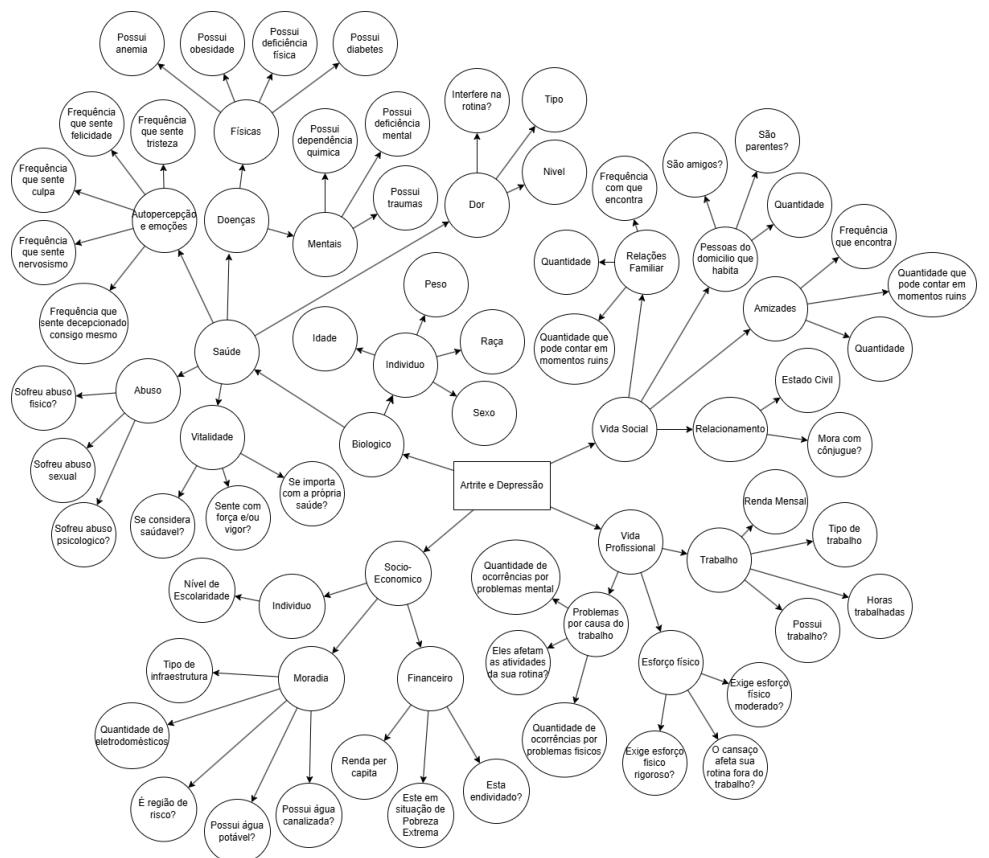
Para definir o escopo do domínio e compreender a distribuição das condições de interesse, foram realizadas análises preliminares na base de dados, utilizando gráficos de distribuições absolutas e relativas de casos de depressão e artrite por estado e região do Brasil. Os resultados indicaram baixa incidência de casos segmentados por unidade federativa. Diante dessa limitação, optou-se por considerar o Brasil como população de estudo, a fim de garantir maior volume de instâncias para as análises e para a construção dos modelos descritivos.

Para este trabalho, foram definidas três populações: indivíduos saudáveis (sem diagnóstico de doenças), indivíduos com artrite sem depressão e indivíduos com a comorbidade artrite com depressão. Registros com respostas “Ignorado” ou ausentes para essas condições foram excluídos do estudo.

##### *Etapa 2 - Entendimento do domínio de problema:*

Para o entendimento do domínio e a seleção conceitual de atributos, aplicou-se o método CAPTO [Gonçalves et al. 2024], que visa capturar conhecimento tácito de especialistas e conhecimento explícito de fontes como literatura científica, relatórios técnicos e dicionários de dados, para estruturar um modelo conceitual que auxilie na seleção de atributos. Neste estudo, utilizou-se principalmente revisão da literatura científica, na ausência de especialistas consultados diretamente.

Foram analisados artigos sobre as duas doenças para identificar hábitos, características e fatores de risco associados ao surgimento, agravamento e prevenção dessas condições. Esse conhecimento foi sistematizado para embasar as etapas seguintes e serviu de base para a construção de um modelo conceitual (Figura 2), que organiza as informações ao domínio “Perfil de pessoas com diagnóstico de artrite com depressão”.



**Figura 2. Modelo Conceitual para o domínio Artrite e Depressão**

### **Etapa 3 - Seleção conceitual de Atributos:**

O modelo conceitual considera 4 dimensões principais, 14 principais aspectos e 61 atributos identificados durante a modelagem conceitual. No entanto, é importante destacar que nem todas as variáveis identificadas no modelo conceitual estão disponíveis na base de dados da Pesquisa Nacional de Saúde (PNS). Nesse contexto, o modelo conceitual deve ser entendido como um guia conceitual/teórico para orientar a seleção de atributos, onde todos os atributos possíveis de serem coletadas/extraídos da PNS, foram identificados. Note que a partir do modelo conceitual é possível pensar em colocar restrições ao modelo ou ainda abortar o projeto por falta de informação relevante ao domínio. A Tabela 1 detalha cada dimensão.

**Tabela 1. Dimensões, Aspectos e atributos do modelo conceitual - PNS**

<b>Socioeconômico</b>			
Nesta dimensão tenta-se caracterizar a situação socioeconômica da pessoa, incluindo moradia, situação financeira e educacional do morador.			
Variável	Tipo	Valores	Fonte PNS
Ligado a rede de distribuição de água	Categórico	Sim, Não	A005012
Água Canalizada	Categórico	Sim, Não	A00601
Renda Per Capita	Numérico	Mínimo = 0, Máximo = 203.196	VDF003
<b>Biológico</b>			
Nesta dimensão tenta-se caracterizar a pessoa biologicamente, incluindo a saúde e características físicas e mentais.			
Variável	Tipo	Valores	Fonte PNS
Idade	Numérico	Mínimo = 0, Máximo = 112	C008
Sexo	Categórico	Mulher, Homem	C006
Cor/Raça	Categórico	Branco, Preto, Amarelo, Outros	VDF003
Autopercepção de Saúde	Numérico	Mínimo = 0, Máximo = 5	N001, N00101
Obesidade	Categórico	Abaixo do Peso, Peso Normal, Acima do Peso, Obesidade, Obesidade Extrema	W00202, W00203, P00403, P00404,
Possui/possuiu Diabetes	Categórico	Sim, Não	Q03001
Deficiência de Membros Inferiores	Categórico	Não, Pouca/Alguma, Muita	G059, G068, G071, G072
Deficiência Mental	Categórico	Sim, Não	Q11007, Q11008,
Sente Deprimido (últimas 2 semanas)	Categórico	Nenhum dia, Menos da metade dos dias, Mais da metade dos dias, Quase todos os dias	N016
Sente Decepcionado Consigo Mesmo (últimas 2 semanas)	Categórico	Nenhum dia, Menos da metade dos dias, Mais da metade dos dias, Quase todos os dias	N017
Sofreu Abuso Físico/Psicológico ou Agressões	Categórico	Sim, Não	V00202, V00203, V00204, V00205, V01401, V01402,
<b>Vida Profissional</b>			
Nesta dimensão tenta-se caracterizar a situação profissional/financeira da pessoa.			
Variável	Tipo	Valores	Fonte PNS
Possui Trabalho	Categórico	Sim, Não	E001, E003,
Horas Trabalhadas Semanal	Numérico	Mínimo = 1, Máximo = 120	E017, E019
Problemas De Saúde Relacionado a Trabalho	Categórico	Sim, Não	E019, J00404
<b>Vida Social</b>			
Nesta dimensão tenta-se caracterizar a vida social da pessoa, como relacionamentos com parentes, parceiros ou amigos			
Variável	Tipo	Valores	Fonte PNS
Estado Civil	Categórico	Casado(a), Solteiro(a), Viúvo(a), Divorciado(a) ou desquitado(a) ou separado(a) judicialmente	C011
Mora com Cônjuge	Categórico	Sim, Não	C01001
Amigos que Confia	Categórico	Nenhum, Um, Dois, Três ou mais	M01501
Familiares que Confia	Categórico	Nenhum, Um, Dois, Três ou mais	M01401

#### ***Etapa 4 - Pré-processamento e preparação do conjunto de dados:***

O conjunto de dados inicial contou com 293.726 instâncias e 23 atributos extraídos da PNS 2019. Cabe destacar que nem todos os atributos previstos no modelo conceitual foram identificados na base, sendo possível mapear apenas 23 dos 61 atributos ideais definidos. Isso ocorre porque o modelo conceitual representa a estrutura ideal de conhecimento sobre o domínio, independentemente das restrições impostas pelas bases de dados disponíveis. Como apontado por [Zarate et al. 2023], cabe ao responsável decidir a continuidade ou adaptação do processo diante dessas limitações. Neste estudo, optou-se por prosseguir, considerando como restrição a disponibilidade de atributos agrupados nas dimensões socioeconômicas, biológicas (antropometria), atividade profissional e vida social. A partir desse conjunto reduzido, foram realizadas sub-étapas de pré-processamento, visando preparar os dados para a construção dos modelos de aprendizado, conforme descrito nas etapas seguintes.

- *Remoção de instâncias:* Foram consideradas como comorbidade as pessoas que responderam “Sim” para diagnóstico de Artrite e Depressão. Registros com respostas “Ignorado” ou em branco para essas variáveis foram excluídos. Após essa filtragem, o conjunto de dados foi reduzido para 14.242 instâncias.
- *Remoção de atributos:* Foi realizada uma análise da quantidade de instâncias com dados vazios, nulos e ausentes. Diferente dos dados ausentes, valores vazios ou nulos não permitem imputação direta [Pyle 1999]. Todas as colunas possuíam mais 98% de dados íntegros, com exceção de “Problemas de Saúde Relacionado a Trabalho” e “Horas Trabalhadas Semanal”, que apresentaram, respectivamente, 57% e 80% de dados vazios. Devido ao alto número, qualquer tentativa de imputação ou suposição geraria muito viés, então ambos atributos foram removidos.
- *Eliminação de outliers:* Para o atributo “Idade”, foi realizada uma análise exploratória utilizando gráfico *Box-plot*, a fim de identificar possíveis outliers. O gráfico indicou a presença de valores acima de 103 anos, os quais foram removidos por se tratarem de casos raros e possivelmente inconsistentes, que poderiam distorcer a média e afetar a performance e a interpretação dos modelos. Na variável “Renda Per Capita”, observou-se grande variabilidade e presença de outliers, reflexo das desigualdades socioeconômicas no Brasil. Como valores extremos poderiam influenciar negativamente as análises e enviesar os modelos, adotou-se como critério a exclusão de registros fora do intervalo da média  $\pm 2$  desvios-padrão. Em relação aos atributos categóricos ordinais, diferentes critérios foram aplicados conforme suas características. Por exemplo, o atributo “Autopercepção de Saúde”, com escala de 0 a 5 e boa distribuição de instâncias entre as categorias, foi mantido no conjunto de dados, visto que não apresentou problemas de desbalanceamento ou inconsistência.
- *Imputação para dados ausentes:* atributos contendo valores ausentes apenas entre 1% a 2% foram imputados de acordo com a média (para atributos numéricos) e moda (para atributos categóricos) de acordo com a classe à qual pertencem. Por exemplo, se a moda para quem possui depressão e “Possui Trabalho” é igual a “Sim”, todos os registros que forem condicionados a essa situação recebem o

mesmo valor. Note que o tratamento de dados ausentes, por mais rigoroso e criterioso qualquer processo de imputação altera a realidade do registro, podendo trazer inconsistências e ruídos ao modelo [Rubin 1976].

- *Categorização de atributos:* Alguns atributos foram categorizados para facilitar a análise. A variável “Renda Per Capita” foi transformada em “Vulnerabilidade-RPC”, variável dicotômica com valores [Sim, Não]. O critério adotado considerou como vulneráveis indivíduos com renda per capita inferior a  $\frac{1}{4}$  do salário mínimo, conforme definição do Ministério da Previdência Social (MPS).
- *Fusão de Atributos:* Para reduzir a dimensionalidade e enriquecer as variáveis do estudo, foi aplicado um processo de fusão de atributos. Por exemplo, o atributo “grau de obesidade” foi derivado a partir do IMC, calculado com os dados de peso e altura, categorizado segundo os intervalos estabelecidos pela *World Obesity Federation*. A variável “autopercepção de saúde” foi definida pela média das respostas a duas perguntas: uma sobre a avaliação geral da saúde e outra considerando saúde como bem-estar físico e mental, ambas em escala de 1 (pior) a 5 (melhor). Da mesma forma, o atributo “Horas Trabalhadas Semanal” foi obtido pela soma das horas dedicadas ao emprego principal e subempregos. Outros atributos listados na Tabela 1, com múltiplas origens na PNS, também foram combinados. Um exemplo é “Sofreu Abuso Físico/Psicológico ou Agressões”, construído a partir de respostas a perguntas sobre situações de violência física, psicológica ou sexual nos últimos 12 meses. Caso o entrevistado respondesse “sim” a qualquer uma dessas situações, o atributo era considerado positivo. Essa fusão de atributos visou captar informações relevantes de forma mais sintética e consistente, preservando a variabilidade explicativa e reduzindo a complexidade do modelo.
- *Análise da capacidade discriminatória dos atributos:* Para avaliar a capacidade discriminatória dos atributos, foi calculada a entropia em relação à variável de classe *Saudável, com Artrite, com Artrite-e-Depressão*. Os resultados indicaram entre 1,551 e 1,595, valores de entropia próximos entre os atributos, sugerindo que todos apresentam relevância potencial para a classificação.

O conjunto final possui 4.200 registros e 19 atributos, com 3.136 casos de artrite, 752 de comorbidade (artrite e depressão) e 312 saudáveis. Para lidar com o desequilíbrio entre classes, aplicou-se *undersampling*, que reduz a quantidade de registros das classes majoritárias para evitar que o modelo favoreça essas classes. O conjunto de dados está disponível em: <https://github.com/licapLaboratory/ DataBase-PNS-Artrite-Depressao>.

### 3.3. Modelos de aprendizado

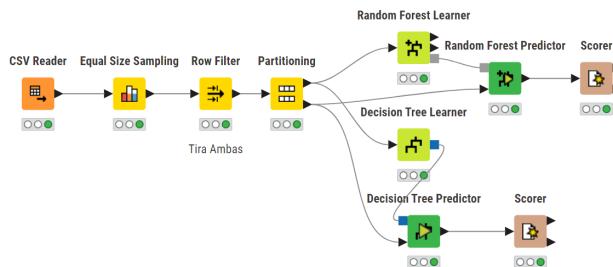
Os modelos de aprendizado são construídos objetivando distinguir por meio de modelos supervisionados, baseados em Árvore de decisão (caixa branca de maior interpretabilidade), e Floresta aleatória (ensemble de maior precisão), três populações (classes) distintas: indivíduos saudáveis, indivíduos que sofrem de artrite sem depressão; e indivíduos com presença da comorbidade de artrite com depressão. Para construção dos modelos foram considerados dois cenários: A) Modelo para representar as três classes de populações; e B) Modelo para representar duas classes, de pessoas saudáveis e as que possuem a comorbidade. Os modelos, baseados em árvore de decisão e floresta aleatória foram construídos no ambiente *KNIME*

	Árvore de Decisão		Floresta Aleatória	
Classe Alvo	Recall	Precisão	Recall	Precisão
Saudável	54,70%	43,90%	74,10%	50,60%
Artrite	39,40%	41,30%	31,50%	45,10%
Artrite + Depressão	47,80%	55,90%	54,10%	56,90%

**Tabela 2. Resultado do treinamento dos modelos (A)**

No modelo de árvore de decisão com três classes, a Tabela 2 mostra baixo desempenho, com recall abaixo de 54%, e precisão de 56%. Isso indica que o modelo tem dificuldade em distinguir pessoas com apenas artrite daquelas com ambas as doenças, devido à similaridade das características e à generalidade dos dados da PNS. No modelo (B) de Floresta Aleatória (Tabela 2), apesar do recall da classe saudável atingir 74%, ambas as métricas atingiram o valor máximo de 56%, evidenciando a dificuldade dos algoritmos em classificar adequadamente as três classes.

Para melhorar a capacidade discriminatória optou-se por eliminar a classe correspondente aos indivíduos somente com artrite, considerando somente as populações de indivíduos saudáveis e com comorbidade, correspondente ao cenário (B). A Figura 3 mostra o *dataflow* para o cenário (B), e na Tabela 3 estão os resultados desse novo cenário, sendo possível observar um aumento significativo.

**Figura 3. Dataflow para o cenário (B) com duas classes**

	Árvore de Decisão		Floresta Aleatória	
Classe Alvo	Recall	Precisão	Recall	Precisão
Saudável	80,60%	83,30%	83,90%	77,60%
Artrite + Depressão	84,10%	81,50%	76,20%	82,20%

**Tabela 3. Resultado do treinamento dos modelos (B)**

A Tabela 3, mostra o desempenho do modelo baseado em Floresta aleatória. O modelo obteve um aumento significativo em suas métricas. Apesar disso, a árvore de decisão mostrou ainda uma média superior. As principais regras, extraídas a partir da árvore de decisão, são dadas a seguir:

*Regra 1: SE deficiência mental = sim E possui/possuiu diabetes = não ENTÃO classe = "Artrite + Depressão" .*

Regra direta, indicando que a deficiência mental é um fator fortemente associado à comorbidade. A simplicidade da regra sugere que a deficiência mental pode ser um fator de risco significativo para essas condições. Alcançou cobertura de 90/90 de precisão, sendo altamente confiável para esse grupo. Vale lembrar que a regra não possui

um caráter preditivo e sim descritivo para um grupo da população brasileira. É importante ressaltar que embora o modelo tenha alcançado um desempenho aceitável, a regra dentro do ambiente clínico, deve ser considerada como uma hipótese a ser investigada clinicamente.

*Regra 2: SE deprimido (últimas 2 semanas) = menos da metade dos dias E 1,25 < autopercepção\_de\_saúde ≤ 4,75 E sexo = mulher E deficiência\_mental = não E possui/possuiu\_diabetes = não ENTÃO classe = "Artrite + Depressão".*

Regra mais complexa, com cobertura de 31/36 (86%). Associa sintomas depressivos moderados, autopercepção de saúde regular e sexo feminino à presença da como

*Regra 3: SE deficiência\_de\_membros\_inferiores = alguma/pouca E 1,25 < autopercepção\_de\_saúde ≤ 3,25 E deprimido (últimas 2 semanas) = nenhum dia E sexo = mulher E deficiência\_mental = não E possui/possuiu\_diabetes = não ENTÃO classe = "Artrite + Depressão".*

Alcançando uma cobertura de 21/36 (58%), a *Regra 3* destaca a deficiência física combinada a uma autopercepção de saúde reduzida como indício para a comorbidade. Apesar da regra não associar sentimentos depressivos, a presença de uma deficiência física e uma autopercepção de saúde reduzida podem corroborar a prevalência das doenças. Apesar da baixa cobertura da regra, ela mostra ser relevante, especialmente para mulheres que não apresentam deficiência mental ou diabetes.

*Regra 4. SE deprimido (últimas 2 semanas) = quase todos os dias E 1,25 < autopercepção de saude ≤ 4,75 E sexo = mulher E deficiência mental = não E possui/possuiu diabetes = não ENTÃO classe = "Artrite + Depressao".*

Essa regra alcançou uma cobertura de 15/17 (88%). Ela foca em um pequeno grupo de mulheres que se sentem predominantemente sentem deprimidas, autopercepção de saúde moderada e sem deficiência mental ou diabetes. A frequência elevada de sentimentos depressivos (*quase todos os dias*) parece ser o fator mais determinante para a ocorrência de Artrite + depressão nesse caso.

No geral, as regras evidenciam a complexidade das interações entre saúde física e mental, indicando que múltiplos fatores contribuem para a comorbidade Artrite + Depressão. Deficiência mental e frequência de sintomas depressivos surgem como os preditores mais fortes, com destaque também para a autopercepção de saúde, fator recorrente nas regras e alinhado à literatura [Campos et al. 2013].

#### 4. Conclusões

Este estudo caracterizou o perfil de indivíduos com depressão e artrite a partir da base PNS 2019. Utilizando técnicas de Mineração de Dados, foram extraídos insights relevantes e construídos modelos descritivos que, apesar das limitações iniciais, apresentaram melhor desempenho ao considerar apenas indivíduos saudáveis e com comorbidade. Os resultados indicam o potencial desses modelos para auxiliar na identificação precoce dessas condições, contribuindo para análises mais eficazes e apoio a políticas públicas voltadas a idosos e populações vulneráveis.

Como trabalhos futuros, sugere-se explorar algoritmos mais avançados, como redes neurais, integrar variáveis externas à PNS e envolver especialistas clínicos para apri-

morar a aplicabilidade e o impacto dos modelos na prática médica e social.

### 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

- Campos, A., Silva, C., de Castro, S., and Graminha, C. (2013). Depressão e qualidade de vida em indivíduos com artrite reumatoide e indivíduos com saúde estável: um estudo comparativo. *Fisioterapia e Pesquisa*.
- Das, A. and Dhillon, P. (2022). Aplicação de aprendizado de máquina na medição do envelhecimento e doenças geriátricas: uma revisão sistemática. *BMC Geriatr*, 23(841).
- Fakra, E. and Marotte, H. (2021). Rheumatoid arthritis and depression. *Joint Bone Spine*.
- 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.
- IBGE (2020). Pesquisa nacional de saúde 2019 - instituto brasileiro de geografia e estatística. <https://www.ibge.gov.br/estatisticas/sociais/saude/9160-pesquisa-nacional-de-saude.html?edicao=25921t=resultados>. Acesso em: 2024-07-15.
- Imtiaz, M., Shah, S. A. A., and ur Rehman, Z. (2022). A review of arthritis diagnosis techniques in artificial intelligence era: Current trends and research challenges. *Neuroscience Informatics*, 2(4):100079.
- Laguardia, J., Campos, M. R., Travassos, C., Najar, A. L., dos Anjos, L. A., and Vasconcellos, M. M. (2013). Dados normativos brasileiros do questionário short form-36 versão 2. *Rev. bras. epidemiol.*, (4).
- Mella, L., Bértolo, M., and Dalgalarrodo, P. (2010). Depressive symptoms in rheumatoid arthritis. *Brazilian Journal of Psychiatry*.
- Pyle, D. (1999). *Data Preparation for Data Mining*. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, (3):581–592.
- Souza, V., Nobre, J., and Becker, K. (2020). Characterization of anxiety, depression, and their comorbidity from texts of social networks. In *Anais do XXXV Simpósio Brasileiro de Bancos de Dados*, pages 121–132, Porto Alegre, RS, Brasil. SBC.
- Xue, Q., Pan, A., Gong, J., Wen, Y., Peng, X., Pan, J., and Pan, X. (2020). Association between arthritis and depression risk: a prospective study and meta-analysis. *J Affect Disord*.
- Zarate, L., Petrocchi, B., Maia, C. D., Felix, C., and Gomes, M. P. (2023). Capto - a method for understanding problem domains for data science projects. *Concilium*.
- Zárate, L., Santos, A., Camelo, J., Nobre, C., and Song, M. (2024). Identificando padrões de depressão em idosos por meio de mineração de dados. *Journal of Health Informatics*, (1):1–8.