J. Health Inform. 2024 Janeiro-Dezembro; 16

Artigo Original

DOI: https://doi.org/10.59681/2175-4411.v16.2024.1020

JHI JOURNAL OF HEALTH INFORMATICS

Identificando padrões de depressão em idosos por meio de mineração de dados

Identifying patterns of depression in the elderly through data mining

Identificando patrones de depresión en adultos mayores a través de la minería de datos

Luis Enrique Zárate[1], Arthur Vinicius do Carmo Santos[2], Jefferson Eduardo de Carvalho Camelo[3], Cristiane Neri Nobre[4], Mark Alan Junho Song[5]

Descritores: Depressão; Mineração de Dados; Aprendizado de Máquina.

RESUMO

Objetivo: Identificar padrões de depressão em idosos baseado em variáveis exógenas por meio da mineração de dados. Métodos: O processo aplica técnica de classificação Floresta Aleatória para descrever os padrões de depressão nessa população. Como fonte de dados considera-se a base de dados PNS, IBGE 2013. Resultados: Os resultados evidenciam como fatores relevantes, doenças crônicas pré-existentes, o nível de confiança com amigos e parentes, nível de escolaridade, etc. Para o grupo diagnosticado "Com depressão", a precisão do modelo foi de 68,8%, sensibilidade de 77,2% e medida F1-score de 72,8%. Para o grupo diagnosticado "Sem depressão", a precisão foi de 66,4%, Sensibilidade de 56,2% e a medida F1-score de 60,9%. Conclusão: Dentre os fatores destacam-se, em nível de importância, doença crônica pré-existente, um ou nenhum parente ou amigo em quem confiar, e escolaridade até o ensino médio. A prática de exercícios físicos e manter-se ativo é um aspecto favorável para a não-depressão.

Keywords: Depression; Data Mining; Machine Learning.

ABSTRACT

Objective: Identify patterns of depression in elderly people based on exogenous variables through data mining. Methods: The process applies the Random Forest classification technique to describe the patterns of depression in this population. The PNS, IBGE 2013 database is considered as a data source. Results: The results highlight pre-existing chronic diseases, level of trust with friends and relatives, level of education, etc. as relevant factors. For the group diagnosed "With depression", the accuracy of the model was 68.8%, sensitivity of 77.2% and F1-score measurement of 72.8%. For the group diagnosed "No depression", the accuracy was 66.4%, Sensitivity was 56.2% and the F1-score measure was 60.9%. Conclusion: Among the factors that stand out, in terms of importance, are pre-existing chronic illness, one or no relatives or friends to trust, and education up to high school. Practicing physical exercise and staying active is a favorable aspect for non-depression.

Descriptores: Depresión; Minería de Datos; Aprendizado Automático.

RESUMEN

Objetivo: Identificar patrones de depresión en personas mayores a partir de variables exógenas mediante minería de datos. Métodos: El proceso aplica la técnica de clasificación Floresta Aleatória para describir los patrones en esta población. Se considera como fuente la base de datos PNS, IBGE 2013. Resultados: Los resultados destacan las enfermedades crónicas preexistentes, el nivel de confianza con amigos y familiares, el nivel de educación, etc. como factores relevantes. Para el grupo diagnosticado "Con depresión", la precisión del modelo fue 68,8%, la sensibilidad 77,2. % y medición de puntuación F1 72,8%. Para el grupo diagnosticado "Sin depresión", la precisión fue 66,4%, la sensibilidad fue 56,2% y la medida de puntuación F1 fue 60,9%. Conclusión: Entre los factores, en términos de importancia están enfermedad crónica preexistente, tener o ningún familiar o amigo en quien confiar y la educación hasta la secundaria. Practicar ejercicio físico y mantenerse activo es un aspecto favorable para no deprimir.

[1] Professor do Departamento de Ciência da Computação, Pontifícia Universidade Católica de Minas Gerais – PUC Minas, Belo Horizonte, Minas Gerais, Brasil.
[2] Bacharel em Sistemas de Informação. Pontifícia Universidade Católica de Minas Gerais – PUC Minas, Belo Horizonte (MG), Brasil.
[3] Bacharel em Sistemas de Informação. Pontifícia Universidade Católica de Minas Gerais – PUC Minas, Belo Horizonte (MG), Brasil.
[4] Professor do Departamento de Ciência da Computação, Pontifícia Universidade Católica de Minas Gerais – PUC Minas, Belo Horizonte, Minas Gerais, Brasil.
[5] Professor do Departamento de Ciência da Computação, Pontifícia Universidade Católica de Minas Gerais – PUC Minas, Belo Horizonte, Minas Gerais, Brasil.

Autor Correspondente: **Luis Enrique Zárate**
e-mail: **zarate@pucminas.br**

Artigo recebido: 06/02/2023
Aprovado: 09/02/2024

INTRODUÇÃO

De acordo com a Pesquisa Nacional de Saúde - IBGE de 2019, houve um aumento de 34% de casos de depressão no Brasil em seis anos (tomando como referência a Pesquisa Nacional de Saúde, 2013), atingindo 16,3 milhões de brasileiros. Trata-se de uma doença psiquiátrica crônica que apresenta multicausalidade, podendo ser desencadeada devido a fatores biológicos, socioeconômicos, psicológicos, culturais, emocionais, entre outros[1-2]. Estima-se que em 2030 a depressão seja a segunda maior causa de incapacidade em todo o mundo[3-4].

Além de ser uma doença muito frequente entre os idosos, a depressão constitui um dos problemas mais comuns e importantes dentro desse grupo, atingindo ao menos 16% dos idosos assistidos na atenção básica[5] - em trabalho recente, foi observado que o número de idosos no Brasil vem crescendo ano após ano, e que cerca de 1 em cada 6 idosos possuem depressão[5]. No entanto, frequentemente esta estatística é negligenciada pois muitas pessoas associam a depressão como um efeito do envelhecimento. Porém, em um estudo[4] foi constatado que fatores biopsicossociais, como religião, práticas de atividades físicas, relações sociais, e o baixo nível de conhecimento sobre a doença, tem influência direta no acometimento da depressão por parte da população idosa.

Dentre os fatores citados anteriormente, a obesidade possui uma grande influência para o aparecimento de sintomas da depressão[6]. Foi observado também que fatores bioquímicos, como a alteração de níveis hormonais, possuem uma relação direta com surgimento de traços depressivos[7]. Além disso, outro fator relevante observado foi que a maioria das pessoas associam os sintomas da depressão como causa emocional ou como consequência de uma saúde mental alterada, levando a concluir, que a depressão pode ser caracterizada por vários outros aspectos[8].

A depressão nas pessoas idosas é uma das doenças psiquiátrica crônicas mais prevalentes e com consequências devastadoras para o indivíduo, atingindo o âmbito pessoal, social e profissional. Destaca-se que as queixas apresentadas pelo idoso devem ser consideradas e investigadas, uma vez que a identificação precoce dos sintomas depressivos, seguido do diagnóstico e início do tratamento é essencial para minimizar os riscos de agravo da doença[9].

Considerada um problema de saúde pública, a depressão em idosos requer atenção dos profissionais de saúde a fim de evitar o sofrimento das pessoas que não recebem um tratamento adequado, bem como de seus familiares, e de reduzir os custos financeiros que a doença impõe à sociedade e ao poder público[10]. De acordo com a literatura[8], é imprescindível que o idoso com depressão seja avaliado considerando aspectos multidimensionais e especificidades do envelhecimento para lidar com a complexidade da doença.

Com o intuito de entender melhor o cenário da depressão, a comunidade científica de Mineração de Dados (do inglês Data Mining) tem aplicado diversas técnicas para descrever, analisar e auxiliar na tomada de decisões para tratamento e prevenção da depressão[12-13]. Neste trabalho, são aplicados conceitos de mineração de dados e aprendizado de máquina para analisar a base de dados da Pesquisa Nacional de Saúde (PNS) do IBGE, estudo que coletou dados sobre a situação de saúde e os estilos de vida da população brasileira no ano de 2013. Para este trabalho propõe-se um modelo conceitual, construído a partir da revisão da literatura, com aspectos que podem influenciar uma pessoa idosa a desenvolver um quadro de depressão. Realizou-se também o tratamento da base de dados PNS, selecionando variáveis relevantes para os aspectos definidos pelo modelo conceitual. Sobre o conjunto de variáveis selecionadas é aplicado o método PICTOREA[14] para definir um processo de descoberta de conhecimento de padrões que caracterizem a depressão em pessoas idosas. A partir do conjunto de dados selecionados, foi utilizado o algoritmo Floresta Aleatória (FA) para descrever os padrões da doença e potenciais perfis de idosos que possam ser diagnosticados como depressão.

Este trabalho pretende contribuir na análise e compreensão do fenômeno da depressão em pessoas idosas e possíveis causas exógenas ao indivíduo que podem estar atreladas, considerando a relação com diversos aspectos que podem servir como catalisadores para o aumento desse fenômeno.

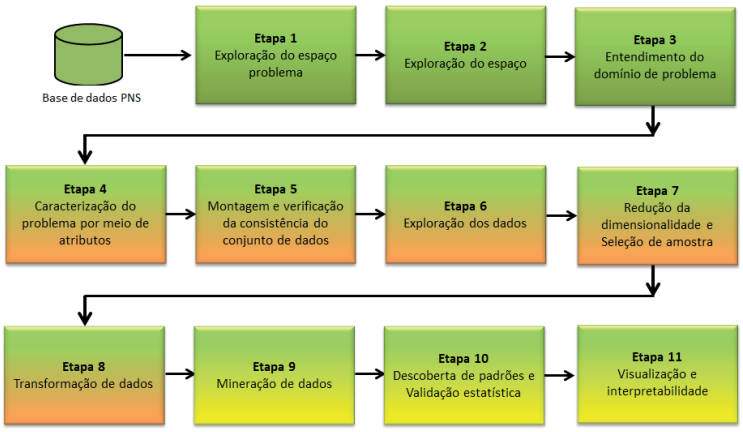
METODOLOGIA

Para desenvolvimento deste trabalho foi aplicado o método PICTOREA que é composto por 13 etapas e que permite a descoberta de conhecimento em base de dados para diferentes domínios de problemas. A Figura 1 ilustra a sequência das etapas da metodologia proposta neste trabalho.

Materiais

A base de dados utilizada para este estudo foi extraída da base de dados originada da Pesquisa Nacional de Saúde (PNS) realizada pelo IBGE[15] em 2013. A base original possui informações do domicílio, características gerais e de educação dos moradores, rendimentos financeiros, deficiências, saúde dos indivíduos, emprego, percepção do estado de saúde, estilos de vida, doenças crônicas, informações clínicas, dentre outras. A partir dessa base de dados foram extraídos os dados dos respondentes da pesquisa, com idade acima de 60 anos, que foram diagnosticados com depressão e na mesma quantidade indivíduos sem a doença, de forma a manter um balanceamento no conjunto de dados. A base de dados extraída possui 942 variáveis e 8.470 registros. Os nomes das colunas seguem um padrão de código utilizado pelo IBGE, do questionário aplicado na Pesquisa Nacional de Saúde, que possui um dicionário de variáveis contendo as informações necessárias para análise dos dados.

Figura 1 – Metodologia para análise de depressão de idosos



Métodos

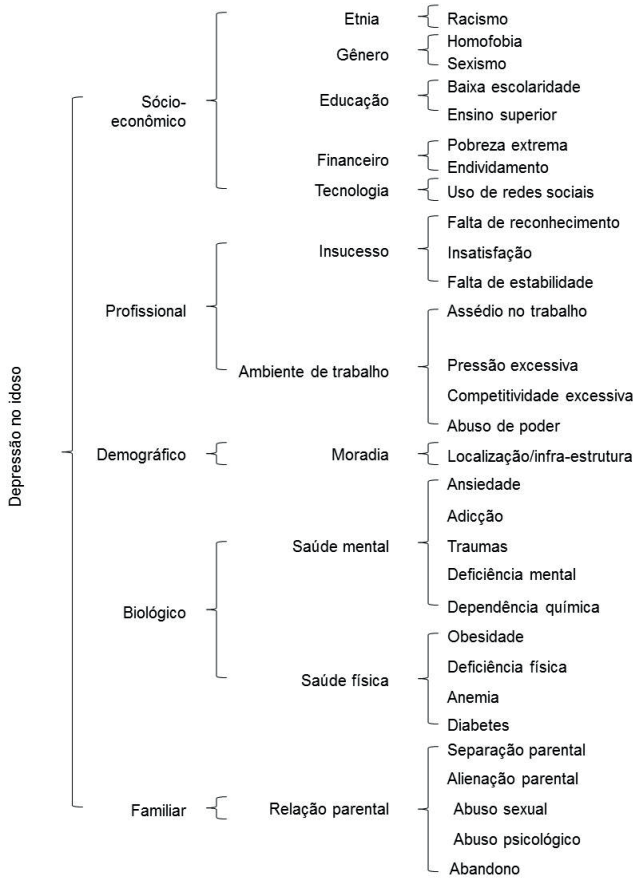
Etapa 1: Exploração do Espaço Problema

Esta etapa corresponde à definição do domínio de problema a ser considerado. Por meio da análise exploratória[19] nos dados da PNS 2013 e 2019, as regiões sul e sudeste do Brasil apresentaram maior incidência de casos de depressão. Sendo o objetivo a população idosa acima de 60 anos, foi considerada para este estudo a região sudeste do Brasil por ser uma das de maior incidência.

Etapa 2: Exploração do Espaço Solução

Foi definido que, por meio do uso de algoritmos de classificação baseados em Floresta Aleatória (FA), seria possível descrever e identificar com certa precisão padrões de potenciais indivíduos idosos que possam ser diagnosticados com depressão. A escolha do algoritmo Floresta Aleatória deve-se ao satisfatório desempenho alcançado para problemas de classificação. O algoritmo FA, quando utilizado com medidas demográficas, clínicas, antropométricas e bioquímicas, mostrou ser uma estratégia eficiente para identificar fatores de risco associados a problemas de saúde como a diabetes tipo-2[16]. De acordo com os autores, tal estratégia pode ser útil para a gestão de políticas de saúde.

Figura 2 – Mapa conceitual - fatores que podem afetar a depressão em idosos



Etapa 3: Entendimento do Domínio de Problema

A partir da revisão da literatura, foi elaborado um modelo conceitual que buscou identificar os principais dimensões e aspectos que corroboram para as causas de casos de depressão, como apresentado na Figura 2. Para construção do modelo foi aplicado o método CAPTO18 que considera conhecimento explícito obtido a partir da literatura. O período de busca foi de 2006-2021 em repositórios brasileiros (Scielo, Rbone, RSR), e os descritores foram: (Idoso AND depressão) OR (Idoso AND saúde).

Etapa 4: Caracterização do Problema por meio de Atributos

O dicionário de dados disponível para a base de dados do PNS 2013 foi utilizado junto com o Mapa conceitual para realizar uma seleção conceitual por meio de uma análise interpretativa, e identificar as principais variáveis que podem representar o domínio do problema. Inicialmente, a base possui 942 variáveis, onde após análise foram reduzidos para 118. Ainda nessa etapa, ficou definido que seriam analisados apenas os dados de pessoas com idade igual ou superior a 60 anos e que moram na região Sudeste do Brasil.

Etapa 5: Montagem e verificação da consistência do conjunto de dados

Com as variáveis selecionados na etapa anterior foi montado um subconjunto de dados preliminar para analisar a consistência e coerência nos valores dos atributos selecionados. Porém tratando-se de um estudo verificado (pelo IBGE) não foram detectados valores inconsistentes.

Etapa 6: Exploração dos Dados

A partir do subconjunto de dados, as variáveis foram analisadas para identificar os tipos de dados, intervalos de valores e frequência de valores. O objetivo foi explorar a informação contida nos dados. Atributos com valores de alta frequência de repetição (baixa informação e entropia) foram desconsiderados.

Etapa 7: Redução da Dimensionalidade e Seleção de Amostra

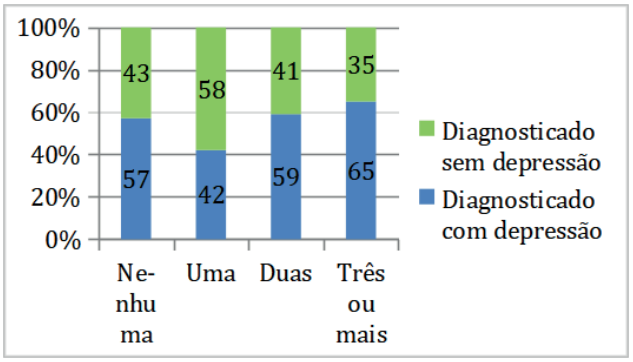
Foi analisada a representatividade dos atributos e seus respectivos valores presentes para saber se seriam relevantes ou não para os resultados esperados. Nesta etapa 99 variáveis foram removidos pela presença de excessivos dados ausentes, resultando em 19 variáveis, sendo um deles o atributo de classificação da base, tendo o valor 1 para "diagnosticados com depressão" e o valor 2 para os que "não foram diagnosticados com depressão". Registros que apresentavam excessivos dados ausentes também foram retirados do conjunto de dados. O número de instâncias foi reduzido de 8.470 para 1.861. Note que o tratamento de dados ausentes por imputação não foi considerado neste trabalho, desde que qualquer método, por mais criterioso, distorce a realidade do dado, o que pode ser prejudicial na área da saúde. A Tabela 1 mostra a relação de variáveis, resultantes da aplicação das etapas anteriores, consideradas para construção dos modelos de aprendizado de máquina.

Etapa 8: Transformação de dados

Os atributos numéricos do subconjunto de dados foram transformados (categorizados) de forma a facilitar a interpretabilidade no processo de mineração de dados.

Primeiramente, todos os atributos numéricos foram discretizados e convertidos em variáveis categóricas de escala ordinal. Em seguida foi realizada uma análise da capacidade classificatória de cada variável categórica referente a cada classe (Diagnosticado com depressão-1 e sem depressão-2). Se a porcentagem dos valores das variáveis categóricas ordinais, vinculados ao rótulo de cada classe, apresentassem uma diferença considerada baixa entre as classes, as escalas adjacentes da variável seriam agrupadas em nova categoria unindo dois ou mais valores de escalas adjacentes. Isso objetivou uma maior relevância da contribuição da variável, em relação a sua capacidade classificatória.

Figura 3 – Representatividade do atributo NÚMERO_DOENCAS_CRÔNICAS



A Figura 3 mostra um exemplo da variável <Número de doenças crônicas> diagnosticadas para o idoso. Essa variável foi transformada para obter maior ganho discriminativo por meio de uma análise das porcentagens de valores da variável que trata o número de doenças crônicas em relação à classificação. Primeiramente a variável <Número de doenças crônicas> foi obtida utilizando a resposta a 5 questões diferentes, sendo elas <Hipertensão, Diabetes, Colesterol Alto, Artrite/Reumatismo e Problemas na Coluna>. Ao analisar o gráfico é possível perceber que o valor "Uma" possui uma relação peculiar entre as classificações, sendo maior a quantidade de não diagnosticados com depressão do que no valor "Nenhuma", "Duas" e "Três ou mais". Note que as instâncias com número de doenças crônicas correspondentes a "Nenhuma" e "Duas" apresentam porcentagens próximas, mas não podem ser agrupadas, pois não pertencem a valores conceitualmente adjacentes.

Como o problema da depressão é um problema complexo multicausas, e como a base possui outras variáveis, o processo de mineração de dados procura pelas inter-relações de variáveis até a classificação.

A Tabela 1 mostra o conjunto de atributos selecionados com a descrição dos valores que compõem a base de dados que foi utilizada na fase seguinte de aprendizado de máquina. O conjunto de dados ficou composto de 747 registros de indivíduos diagnosticados com depressão e de 934 registros de indivíduos sem a doença. O leve desbalanceamento não é considerado um problema de balanceamento que deva ser tratado para evitar viés[17].

Etapa 9: Mineração de Dados

O conjunto de dados resultante da etapa anterior (1.681 registros) foi dividido nos conjuntos de treinamento (70%) e teste (30%) pela técnica hold-out. Utilizando o software WEKA[1], foi escolhido o algoritmo Floresta Aleatória, usando o método de validação cruzada com 10 dobras. Para o treinamento, o modelo obteve uma taxa de acerto global (acurácia) de 67,8%. Para o conjunto de teste a acurácia foi de 65,9%. Para este estudo foi considerada a acurácia por ser a medida global mais adequada para avaliar o poder descritivo e não o poder preditivo do modelo para discriminar ambas as populações (de depressão e não-depressão). A Tabela 2 mostra a matriz de confusão obtida no teste a qual mostra resultados aceitáveis para a complexidade do problema considerado. A partir do modelo de aprendizado, foi iniciada a análise dos possíveis perfis e padrões extraídos a partir dos dados disponíveis.

Tabela 1 – Atributos para caracterizar a depressão em idosos e respectivos valores

Variável Original	Atributo	Valor: Descrição
C006	Sexo	1: Masculino; 2: Feminino
C008	Idade	1:6 0 a 65; 2: 66-75; 3: 76-80; 4: 81-85
C010	Cônjuge	1: Sim; 2: Não
D009	Escolaridade	1: Sem escolaridade ou somente ensino básico; 2: Ensino médio; 3: Ensino Superior
I001	Plano_Saúde	1: Sim; 2: Não
J001	Estado_Saúde	1: Bom; 2: Regular; 3: Ruim
J002	Dx_Atv_Habitual_Saude	1: Sim; 2: Não
J007 + J008	Diagn_Doen_Limit	1: Sim, limita atividades; 2: Sim, não limita atividades; 3: Não
M009	Trabalha	1: Sim; 2: Não
M014	Parentes_Confia	1: Nenhum ou apenas 1; 2: 2 a 4; 3: 5 ou mais
M015	Amigos_Confia	1: Nenhum; 2: 1 a 3; 3: 4 ou mais
N001	Auto-Av_Saúde	1: Bom; 2: Regular; 3: Ruim
P027	Bebida_Alcoolica	1: Sim; 2: Não
P034 + P035	Exercício_Físico	1: Sim; 2: Não
P045	Televisão	1: Não assiste ou assiste menos de 1 hora; 2: De 1 a 4 horas; 3: Mais de 4 horas
	Número_Doenças_Crônicas	1: Nenhum; 2: Uma; 3: Duas; 4: 3 ou mais
Q132	Medicamento_Dormir	1: Sim; 2: Não
W00103	Peso	1: 30 a 69,9; 2: 70 a 89,9; 3: 90 ou mais

Etapa 10: Descoberta de Padrões e Validação Estatística

O algoritmo de Floresta Aleatória faz parte do grupo de algoritmos denominados "Caixa Preta", que não possuem fácil interpretação, sendo necessário utilizar uma combinação de técnicas para realizar a extração de conhecimento. Uma das técnicas abordadas foi a utilização de um ranqueamento com a relevância das variáveis, calculada de acordo com o ganho de informação obtido pela variável em cada nodo da árvore. Além disso, a partir das árvores de decisão geradas pela floresta aleatória, foi realizada uma análise das regras presentes nas mesmas, considerando sua frequência e tamanho. A frequência pode ser descrita como a quantidade de vezes em que a regra aparece em relação à quantidade de regras. Após essa avaliação, as regras foram ordenadas, sendo o tamanho o critério de desempate em caso de frequência e erros iguais.

Etapa 11: Visualização

Para visualização dos padrões encontrados, foram utilizadas a descrição das regras encontradas nas árvores de decisão geradas na Floresta Aleatória, por meio de comparação entre árvores e análise dos perfis relacionados com as regras extraídas.

EXPERIMENTOS E ANÁLISE DOS RESULTADOS

O algoritmo Floresta Aleatória, implementado na ferramenta WEKA, foi parametrizado para gerar 100 árvores considerando 4 atributos na seleção aleatória. Para validação foi aplicado o método de validação cruzada com 10 dobras. Essa combinação obteve uma medida de Acurácia de 67,8%, Precisão de 68,8%, Sensibilidade de 77,2% e F1-score de 72,8% para a classe de "diagnosticados com depressão". Para a classe dos "diagnosticados sem depressão", a medida de Precisão foi de 66,4%, a Sensibilidade de 56,2% e a medida F1-score foi de 60,9%. Essa diferença talvez possa ser explicada pelo fato da base de dados possuir atributos com mais valores associados a fatores responsáveis pelo quadro de depressão.

[1] Disponível em https://www.cs.waikato.ac.nz/ml/weka/

Tabela 2 – Matriz de confusão para o conjunto de teste

Matriz de Confusão	Classificado para Depressão	Classificação para não-depressão
Real para depressão	211	88
Real para não-depressão	84	121

Tabela 3 – Relevância de atributo

ATRIBUTO	IMPORTÂNCIA
NUMERO_DOENCAO_CRITICAS	0,101599
IDADE	0,083481
AMIGOS_CONFIA	0,074003
TELEVISAO	0,072817
PESO	0,072755
PARENTES_CONFIA	0,072231
DIAGN_DOEN_LIMIT	0,060845
ESCOLARIDADE	0,051997
PLANO_SAUDE	0,045051
ESTADO_SAUDE	0,044419
AUTO_AV_SAUDE	0,044057
CONJUGE	0,042987
MEDICAMENTO_DORMIR	0,037904
EXERCICIO_FISICO	0,035778
SEXO	0,035576
TRABALHA	0,032109
BEBIDA_ALCOOLICA	0,031664
LER_ESCREVER	0,030671
DX_ATV_HABITUAL_SAUDE	0,030055

A partir da execução do algoritmo, Floresta Aleatória, foi extraída a lista de relevância das variáveis. A Tabela 3 mostra essas variáveis e seus respectivos valores de relevância. De posse da lista de importância dos atributos, foram avaliadas as regras extraídas das árvores de decisão geradas pelo algoritmo da Floresta Aleatória. Foram observados os critérios de frequência e tamanho, que dizem respeito ao número de vezes que a regra apareceu nas árvores geradas e a quantos nós a regra possui até chegar à classificação, respectivamente. As Tabelas 4 e 5 mostram a descrição das regras com melhor desempenho e suas análises. Ao analisar as regras extraídas com melhor desempenho, é possível identificar alguns padrões de perfis de pessoas idosas com depressão e sem depressão:

Por exemplo, a regra <*ESCOLARIDADE = "Sem escolaridade ou somente ensino básico" & TRABALHA = "Sim" & NÚMERO_DOENCAS_ CRÔNICAS = "3 ou mais" & PARENTES_CONFIA = "Nenhum ou apenas 1" -> "Depressivo"*>, indica uma pessoa com baixa escolaridade, que potencialmente precisa trabalhar para manter seu sustento apesar de seu alto número de doenças crônicas, tendo em vista que o mesmo não possui nenhum parente em que possa confiar. Esse perfil pode indicar que o fator de solidão ou abandono familiar possa causar os sintomas de depressão em uma pessoa idosa que precisa se manter por conta própria.

Tabela 4 – Regras extraídas da Floresta Aleatória

Tamanho	Frequência %	Condições	Predição
4	25	NÚMERO_DOENCAS_CRÔNICAS = "Nenhum" & PESO = [70-90> & IDADE = [66-75] & TELEVISAO = "Não assiste ou assiste menos de 1 hora"	Não Depressivo
3	25	PESO = [70-90> & IDADE = [76-80] & NÚMERO_DOENCAS_ CRÔNICAS = "Uma"	Depressivo
4	23	TRABALHA = "Não" & PESO = "90 ou mais" > & EXERCICIO_FISICO = "Não" & IDADE = [81-85]	Depressivo
4	23	ESCOLARIDADE = "Sem escolaridade ou somente ensino básico" & TRABALHA = "Sim" & NÚMERO_DOENCAS_ CRÔNICAS = "3 ou mais" & PARENTES_CONFIA = "Nenhum ou apenas 1"	Depressivo
4	20	NÚMERO_DOENCAS_ CRÔNICAS = "Uma" & EXERCICIO_FISICO = "Sim" & IDADE = [66-75] & AMIGOS_CONFIA = "1 a 3"	Não Depressivo
3	16	ESCOLARIDADE = "Ensino médio" & IDADE = [66-75] & DIAGN_DOEN_LIMIT = "Sim"	Depressivo
7	13,2	IDADE = [66-75] & DIAGN_DOEN_LIMIT = "Não" & MEDICAMENTO_DORMIR = "Não" & TELEVISAO = "1 a 4 horas" & TRABALHA = "Não" & PESO = [70-90>	Não Depressivo
4	9	IDADE = [66-75] & TELEVISAO = "1 a 4 horas" & PARENTES_CONFIA = "Nenhum ou apenas 1" & NÚMERO_DOENCAS_ CRÔNICAS = "3 ou mais"	Depressivo
4	8	PLANO_SAUDE = "Não" & ESCOLARIDADE = "Sem escolaridade ou somente ensino básico" & TRABALHA = "Sim" & PARENTES_CONFIA = "5 ou mais"	Não Depressivo
4	5	TRABALHA = "Sim" & ESCOLARIDADE = "Ensino médio" & SEXO = "Feminino" & NÚMERO_DOENCAS_ CRÔNICAS = "Uma"	Depressivo

Tabela 5 – Análise das regras extraídas da Floresta Aleatória

REGRAS	ANÁLISE
ESCOLARIDADE = "Sem escolaridade ou somente ensino básico" & TRABALHA = "Sim" & NÚMERO_DOENCAS_ CRÔNICAS = "3 ou mais" & PARENTES_CONFIA = "Nenhum ou apenas 1"	O indivíduo possui baixa escolaridade, trabalha, possui muitas doenças crônicas e tem apenas um ou nenhum parente em que possa confiar.
NÚMERO_DOENCAS_CRÔNICAS = "Nenhum" & PESO = [70-90> & IDADE = [66-75] & TELEVISAO = "Não assiste ou assiste menos de 1 hora"	O indivíduo não possui nenhuma doença crônica, peso moderado, idade entre 66 e 75 anos e não assiste ou assiste menos de 1 hora de televisão por dia.
NÚMERO_DOENCAS_ CRÔNICAS = "Uma" & EXERCICIO_FISICO = "Sim" & IDADE = [66-75] & AMIGOS_CONFIA = "1 a 3"	Indivíduo com alguma doença crônica, mas que faz exercícios físicos, com idade entre 66 a 75 anos e tem até 3 amigos que pode confiar.
ESCOLARIDADE = "Ensino médio" & IDADE = [66-75] & DIAGN_DOEN_LIMIT = "Sim"	Indivíduo com escolaridade média, idade entre 66 e 75 anos e que foi diagnosticado com doença que limita suas atividades.
TRABALHA = "Não" & PESO = "90 ou mais" > & EXERCICIO_FISICO = "Não" & IDADE = [81-85]	Indivíduo que não trabalha, possui peso acima dos 90 quilos, não pratica exercícios físicos e possui idade entre 81 e 85 anos.
IDADE = [66-75] & DIAGN_DOEN_LIMIT = "Não" & MEDICAMENTO_DORMIR = "Não" & TELEVISAO = "1 a 4 horas" & TRABALHA = "Não" & PESO = [70-90>	Indivíduo que possui entre 66 a 75 anos de idade, não foi diagnosticado com doença que limita as atividades, não toma medicamentos para dormir, assiste de 1 a 4 horas de televisão diariamente, não trabalha e possui peso moderado.
PESO = [70-90> & IDADE = [76-80] & NÚMERO_DOENCAS_ CRÔNICAS = "Uma"	Indivíduo com peso entre 70 a 89 quilos, idade entre 76 a 80 anos e que possui 2 doenças crônicas.
IDADE = [66-75] & TELEVISAO = "1 a 4 horas" & PARENTES_ CONFIA = "Nenhum ou apenas 1" & NÚMERO_DOENCAS_ CRÔNICAS = "3 ou mais"	Indivíduo com idade entre 66 a 75 anos, que assiste de 1 a 4 horas de televisão diariamente, possui apenas um ou nenhum parente que confia, possuindo 3 ou mais doenças crônicas.
PLANO_SAUDE = "Não" & ESCOLARIDADE = "Sem escolaridade ou somente ensino básico" & TRABALHA = "Sim" & PARENTES_ CONFIA = "5 ou mais"	Indivíduo não possui plano de saúde, escolaridade baixa, mas trabalha e possui 5 ou mais parentes que confia.
TRABALHA = "Sim" & ESCOLARIDADE = "Ensino médio" & SEXO = "Feminino" & NÚMERO_DOENCAS_ CRÔNICAS = "Uma"	Indivíduo do sexo feminino que não trabalha, possui escolaridade média e possui alguma doença crônica.

Outra regra a ser analisada no perfil de um idoso com depressão é a <*ESCOLARIDADE = "Ensino médio" & IDADE = [66-75] & DIAGN_DOEN_LIMIT = "Sim" -> "Depressivo"*>, que mostra um indivíduo com 66 a 75 anos que possui escolaridade média, mas foi diagnosticado com alguma doença que limita suas atividades habituais. Esse perfil pode indicar que idosos nessa faixa de idade com doenças limitantes podem sofrer de depressão por dependerem de outras pessoas para concluírem atividades básicas ou a inabilidade total de executar outras, causando uma dependência de ajuda e sentimento de incapacidade.

Pode-se observar também regras relacionadas aos indivíduos que não foram diagnosticados com depressão. Tem-se, por exemplo, a regra: <*NÚMERO_DOENCAS_ CRÔNICAS = "Uma" & EXERCICIO_FISICO = "Sim" & IDADE = [66-75] & AMIGOS_CONFIA = "1 a 3" -> "Não depressivo"*>, que indica indivíduos que mesmo possuindo uma doença crônica, pratica exercícios físicos para se manter ativo na idade em que se encontra, e possui amigos em que pode confiar. Outro exemplo de regra que possui características parecidas é a <*PLANO_ SAUDE = "Não" & ESCOLARIDADE = "Sem escolaridade ou somente ensino básico" & TRABALHA = "Sim" & PARENTES_CONFIA = "5 ou mais" -> "Não depressivo"*>, que mostra um indivíduo que não possui plano de saúde, possui uma escolaridade baixa, mas trabalha e possui parentes em que pode confiar. Essa regra mostra uma forte relação entre os idosos que se mantêm ativos realizando atividades, sejam laborais ou físicas, com o fato de não serem diagnosticados com depressão, pelo fato de que podem se sentir ainda dispostos, com energia e sem o risco de não se sentirem dependentes ou inúteis. Outro ponto a se observar nessa regra é a relação com os parentes que esse idoso considera que pode confiar, pois mostra uma conexão com o aspecto familiar por possuir pessoas ao qual ele tem um bom convívio, vivendo em um ambiente potencialmente saudável sem abandono familiar.

CONCLUSÕES

A abordagem apresentada teve como objetivo identificar padrões de perfis de pessoas idosas com depressão, usando algoritmo Floresta Aleatória. Embora esse algoritmo seja considerado "caixa-preta", foi possível extrair os valores de importância dos atributos para as árvores geradas e identificar regras criadas pelas mesmas, considerando seu tamanho, sua frequência e sua taxa de erro. A partir dessas regras extraídas e analisadas, foi possível identificar alguns padrões comportamentais e de estilo de vida das pessoas idosas que podem contribuir para o diagnóstico de depressão, assim como também foram identificados perfis que podem contribuir para que se possa evitar que uma pessoa idosa possa desenvolver um quadro depressivo.

Dentre as principais variáveis relacionadas com a depressão do idoso destacam-se em nível de importância minimamente

uma doença crônica pré-existente, um ou nenhum parente ou amigo em quem confiar, e escolaridade até o ensino médio. A prática de exercícios físicos é um aspecto favorável para a não-depressão, porém se manter ativo trabalhando não é determinante para o estado da doença.

Tendo em vista esses perfis percebidos, espera-se que possa ser possível criar propostas de estudos pela área da medicina e da psicologia para entender os fenômenos comportamentais e contribuir para uma melhoria da qualidade de vida dos idosos no país, além de políticas públicas que possam incentivar ainda mais o convívio em sociedade, ajudar que o idoso se mantenha ativo fisicamente e mentalmente e evitar ou auxiliar idosos em situação de abandono familiar.

É importante ressaltar que na área de aprendizado de máquina, a partir de um conjunto de dados, busca-se por padrões ou regras, as quais são potenciais hipóteses. Essas hipóteses podem ser avaliadas e validadas por meio de estudos de campo complementares. Em trabalhos futuros, poderiam ser utilizadas técnicas de transformação de dados para os atributos da base de dados utilizada a fim de melhorar sua capacidade de prover informação, além de realizar a comparação com outras abordagens de extração de regras da Floresta Aleatória, provendo melhor interpretabilidade dos resultados.

AGRADECIMENTOS

Os autores agradecem à PUC Minas, CNPq e CAPES (Grant PROAP 88887.842889/2023-00-PUC/MG,Grant PDPG88887.708960/2022-00 PUC/MG, INFORMÁTICA Code 001).

REFERÊNCIAS

1. Maier A, Riedel-Heller SG, Pabst A, Luppa M. Risk factors and protective factors of depression in older people 65+. A systematic review. PLoS One. 2021 May 13;16(5):e0251326.
2. Lelis KCG, Brito RVNE, Pinho S, Pinho L. Sintomas de depressão, ansiedade e uso de medicamentos em universitários. Revista Portuguesa de Enfermagem de Saúde Mental. 2020; 23:9-14.
3. Lopez AD, Mathers CD. Measuring the global burden of disease and epidemiological transitions: 2002-2030. Ann Trop Med Parasitol. 2006 Jul-Sep;100(5-6):481-99.
4. Andrade ABCA, Ferreira AA, Aguiar MJ. Conhecimento dos idosos sobre os sinais e sintomas da depressão. Saúde Redes. 2016;2(2):157-166.
5. Sousa KA, Freitas FFQ, Castro AP, Oliveira CDB, Almeida AAB, Sousa KA. Prevalência de sintomas de depressão em idosos assistidos pela Estratégia de Saúde da Família. Rev Mineira de Enferm. 2017;21:e-1018.
6. Soares ThD, Peroza LR, Cerezer M, Nedel Sh.S, Branco JC. Efeitos do exercício físico na obesidade e depressão: uma revisão. Rev. Bras. de Obesidade, Nutrição e Emagrecimento. 2020;14(86):511-518.
7. Zwolińska W, Dmitrzak-Węglarz M, Słopień A. Biomarkers in Child and Adolescent Depression. Child Psychiatry Hum Dev. 2023 Feb;54(1):266-281. Epub 2021 Sep 29. PMID: 34590201; PMCID: PMC9867683.
8. Peluso ETP, Blay SL. Percepção da depressão pela população da cidade de São Paulo. Rev. Saúde Pública. 2008;42(1):41-48.
9. Ferreira PCS, Martins NPF, Rodrigues LR, Ferreira LA. Características sociodemográficas e hábitos de vida de idosos com e sem indicativo de depressão. Rev Eletrônica de Enfermagem. 2013;15(1):197-204.
10. Benedetti TRB, Borges LJ, Petroski EL, Gonçaçves LHT. Atividade física e estado de saúde mental de idosos. Rev Saúde Pública. 2008;42(2):302-307.
11. Tier CG, Lunardi VL, Santos SSC. Cuidado ao idoso deprimido e institucionalizado à luz da Complexidade. Rev. Elet. Enferm. 2008;10(2):530-536.
12. Dipnall JF, Pasco JA, Berk M, Williams LJ, Dodd S, Jacka FN. Fusing Data Mining, Machine Learning and Traditional Statistics to Detect Biomarkers Associated with Depression. PLoSONE. 2016;11(2): e0148195.
13. Oh J, Yun K, Maoz U, Kim T-S, Chae J-H. Identifying depression in the National Health and Nutrition Examination Survey data using a deep learning algorithm. Journal of Affective Disorders. 2019;257:623-631.
14. Montevecchi ALD, Zárate LE. PICTOREA: Um método para descoberta de conhecimento em banco de dados convencionais. Novas Edições Acadêmicas; 2014, 104 p.
15. Brasil. Instituto Brasileiro de Geografia e Estatística -IBGE. PNS – Pesquisa Nacional de Saúde; 2013. [Citado 2024 jun 24]. Disponível em: https://www.ibge.gov.br/estatisticas/sociais/justica-e-seguranca/29540-2013-pesquisa-nacional-de-saude.html?=&t=resultados.
16. Esmaily H, Tayefi M, Doosti H, Ghayour-Mobarhan M, Nezami H, Amirabadizadeh A. A Comparison between Decision Tree and Random Forest in Determining the Risk Factors Associated with Type 2 Diabetes. J Res Health Sci. 2018 Apr 24;18(2):412.
17. Batista GEAPA, Prati RC, Monard MC. 2004. A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor. Newsl. 6, 1 (June 2004), 20–29.
18. Zarate L, Petrocchi B, Maia CD, Felix C, Gomes MP. CAPTO - A method for understanding problem domains for data science projects. 23(15):922-41.
19. Brito VCA, Bello-Corassa R, Stopa SR, Sardinha LMV, Dahl CM, Viana MC. Prevalência de depressão autorreferida no Brasil: Pesquisa Nacional de Saúde 2019 e 2013. Epidemiologia e Serviços de Saúde v. 31, n. spe1, e2021384.