Identificação das características relacionadas ao perfil de jovens, adultos e idosos com depressão utilizando da base de dados da PNS 2013



Cinthia Mikaela de Souza Douglas Spencer de Oliveira Lucas Isabela Borlido Barcelo Wenderson Júnio de Souza

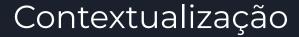
### Sumário



- Introdução
  - Contextualização
  - o Objetivo do trabalho
  - Trabalhos relacionados
- Metodologia
  - o Base de dados
  - Etapas de tratamento da base
  - Tratamento de variáveis
  - Seleção dos casos diagnosticados de depressão por filtro regional
  - Exploração do espaço solução
  - o Treinamento e teste
- Resultados
- Conclusão
- Referências



# Introdução





- Depressão é uma das doenças que mais está presente na população mundial;
- 322 milhões de pessoas;
- 5,8% da população diagnosticada com depressão no Brasil em 2017;
- Sensação de solidão, com sentimento de tristeza e cansaço, falta de interesse;
- Depressão de alterações e desregulação de humor até a depressão maior.



## Objetivo do trabalho

Contribuir na identificação e caracterização do perfil dos diagnosticados com depressão.



#### Trabalhos relacionados

Richter et al. (2020) apresentaram em seu trabalho uma metodologia que testa as diferenças entre os indivíduos ansiosos e deprimidos.

- Técnicas de Machine Learning (ML) para detectar as características;
- Coleta de dados a partir de um questionário;
- Amostra com as respostas do questionário de 125 participantes;
- Baseado em DT;
- Recebe como entrada medidas comportamentais e retorna o perfil sintomático;
- Altos níveis de ansiedade e baixos níveis de depressão (HA), altos níveis de depressão e baixos níveis de ansiedade (HD), altos níveis de ansiedade e depressão (HAD) e baixos níveis de ansiedade e depressão (LAD);
- De acordo com os autores, a associação de métodos de ML com dados de medidas comportamentais possuem resultados promissores.



### Trabalhos relacionados (cont.)

Baek e Chung (2020) propuseram em seu trabalho um modelo para predição do risco da depressão chamado *context*-DNN.

- Utiliza informações de contextos de variáveis e um Deep Neural Network (DNN);
- Korea National Health and Nutrition Examination;
- Pré-processamento e composição dos dados, extração de variáveis e treinamento do modelo context-DNN;
- 52 features extraídas;
- Extração das features utilizando um modelo de regressão múltipla;
- Cada uma das *features* possuem variáveis de contexto;
- Para cada *features*, uma DNN. Os resultados de cada DNN é agregado em um único valor;
- Classificação: "bom", "não-ruim", "perigoso" e "alto risco";
- O melhor resultado foi obtido utilizando 6-fold, com média de erro de classificação de 0,012358.



# Metodologia



#### Base de dados

- Pesquisa Nacional de Saúde (PNS);
- Segundo o site do IBGE: "Visou coletar informações sobre o desempenho do sistema nacional de saúde no que se refere ao acesso e uso dos serviços disponíveis e à continuidade dos cuidados, bem como sobre as condições de saúde da população, a vigilância de doenças crônicas não transmissíveis e os fatores de risco a elas associados";
- Questionários domiciliar, relativo a todos os moradores do domicílio e individual;
- PNS-2013:
  - 205.546 registros;
  - 942 variáveis;



### Base de dados (cont.)

sub-base dados utilizada possui 8470 registros, 50% destes são de pessoas que <u>afirma</u>ram que algum médico ou psiquiatra lhe deu o diagnóstico de 50% depressão afirmaram que não foram diagnosticadas com depressão.

Algum médico ou profissional da saúde mental (como psiquiatra ou psicólogo) já lhe deu diagnóstico de depressão?

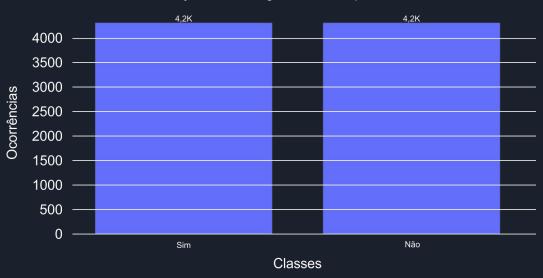


Figura 1 - Divisão de registros



### Etapas de tratamento da base

#### Etapas de tratamento

- 1. Base de dados com 8470 registros
- 2. Padronização dos valores
- 3. Levantamento dos fatores relacionados à depressão
- 4. Análise dos valores faltantes na base
- 5. Análise da incidência de depressão por estado
- 6. Análise causa/consequência
- 7. Filtragem da base pela região escolhida
- 8. Análise individual de cada variável de causa com o agrupamento de variáveis e redução de valores de resposta
- 9. Transformação dos valores das features para categóricos
- 10. Separação da base treino/teste
- 11. Treinamento e teste do modelo



#### Tratamento das variáveis

No levantamento inicial de características relacionadas à depressão, foram identificadas 138 features da base, indicadas no modelo conceitual, relacionadas a:

- Hábitos alimentares;
- Socialização;
- Vícios;
- Desemprego;
- Estresse;
- Trauma;
- Mudanças no apetite;

- Baixa autoestima;
- Problemas de sono;
- Presença de outra doença;
- Dificuldades financeiras;
- Sedentarismo;
- Sentimento de tristeza.

Das 138 features identificadas, foram removidas aquelas reconhecidas como consequência e aquelas posteriormente identificadas como não relacionadas à depressão. Após esta remoção, restaram 86 features.



### Tratamento das variáveis (cont.)

A partir das 86 features identificadas como causadoras de depressão, as features foram separadas em grupos de informações semelhantes para análise, sendo identificados os seguintes grupos de informações:

- Estresse no Trabalho;
- Desemprego;
- Amizades;
- Hábitos alimentares;
- Participação de atividades em grupo;
- Consumo de bebidas alcoólicas;

- Exercícios físicos;
- Horas de TV;
- Presença de outras doenças;
- Trauma psicológico;
- Vício em cigarros, fumante passivo ou ex-fumante.

A distribuição dos valores das features de cada grupo foram analisadas buscando unir as que continham a mesma informação, reduzindo a quantidade de features. Também foi reduzida a quantidade de valores de algumas das features. Ao final do processo, foram obtidas 30 features.



# Seleção dos casos diagnosticados de depressão por filtro regional

Baseado na base disponibilizada pela pesquisa da PNS, foram feitas as análises dos valores absolutos e proporcionais (relativos) por UF (Unidade Federativa) e região nacional da quantidade de casos de depressão, em relação ao total de indivíduos daquela localização analisada.

Utilizando a sub-base com 8470 registros, foi realizado o filtro para as regiões Sul e Sudeste dos indivíduos com casos de depressão (Q092), resultando em uma sub-base com 3604 registros, sendo 2099 registros de casos **com** depressão e 1505 de casos **sem** depressão.

Quantidade absoluta de casos de depressão, em relação ao total de indivíduos por UF.

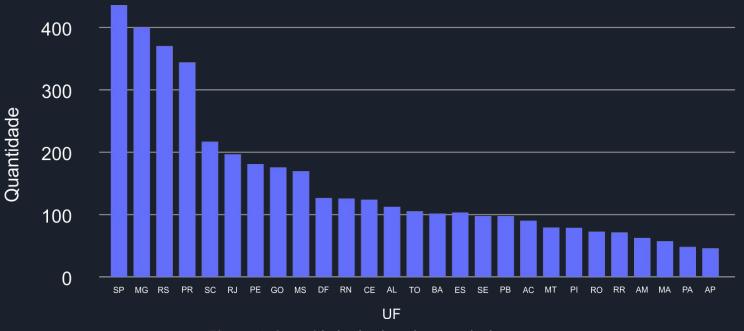
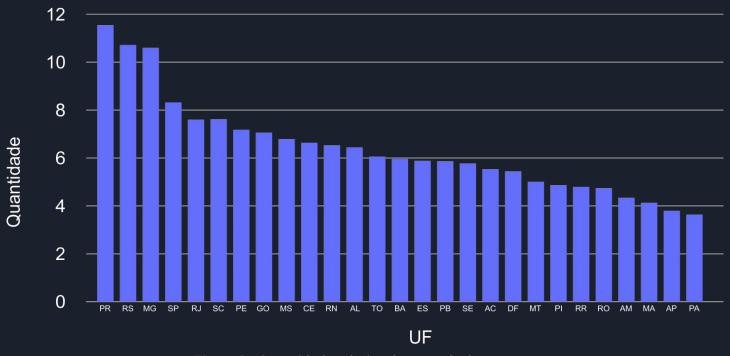




Figura 1 - Quantidade absoluta de casos de depressão.

# Quantidade (%) de casos de depressão, em relação ao total de indivíduos por UF.



PUC Mina

Figura 2 - Quantidade relativa de casos de depressão.

# Quantidade (%) de casos de depressão, em relação ao total de indivíduos por região.

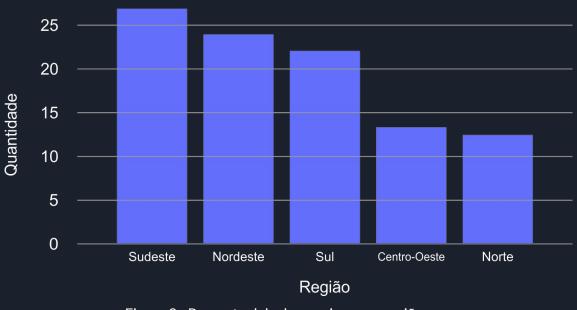


Figura 3 - Percentual de depressivos por região.





## Exploração do espaço solução

- Decision Tree Classifier (DTC)
  - Técnica simples e interpretável;
  - Classificadores fracos, que podem não apresentar um bom resultado

#### Treinamento e teste

Para treinamento do modelo foi utilizada uma base de dados com 3604 registros e 30 features. As etapas metodológicas utilizadas para treinamento e validação do modelo são descritas a seguir.

- 1. Separação da base em treino e teste utilizando 20% para teste e 80% para treino;
- 2. Transformação das variáveis categóricas utilizando One-hoting encode;
- 3. Treinamento do modelo utilizando k-fold cross-validation com k = 10, utilizando as configurações default do modelo. Durante o treinamento, as dobras de treinamento foram balanceadas usando uma técnica de Undersamplig chamada NearMiss;
- 4. Validação do modelo utilizando as métricas accuracy, precision, recall e f1-score;
- 5. Poda da árvore chegando nos seguintes hiperparâmetros: max\_depth = 10, min\_samples\_split = 100, min\_samples\_leaf = 50;

### Treinamento e teste (Cont.)

- 6. O treinamento foi realizado utilizando as mesmas técnicas apresentadas etapa 3, contudo, foram utilizados os valores de hiperparâmetros obtidos ao realizar a poda;
- 7. Validação utilizando a base de teste. Novamente, foram utilizadas as métricas *accuracy*, *precision*, *recall* e *f*1-*score*;
- 8. Seleção e análise das features mais importantes;
- 9. Plot da árvore, análise e extração de regras.



## Resultados



## Métricas para a Base de Treinamento

	Base de Treinamento	Base de Teste
Accuracy	0.61	0.63
Precision	0.69	0.71
Recall	0.63	0.62
F1-score	0.66	0.67



## Métricas para a Base de Testes

	Precision	Recall	F1-Score	Support	Depressão
Sim	0.73	0.61	0.66	421	Sim
Não	0.55	0.68	0.61	300	Não
Accuracy			0.64	721	-
Macro AVG	0.64	0.64	0.63	721	-
Weighted AVG	0.65	0.64	0.64	721	-



## Importância das Features

Feature	Importância	
Q063_Sim	0.443708	
P027_Não bebo nunca	0.088212	
P015	0.068467	
M016_Nunca	0.047630	
M01301_Não	0.042387	
P045_Menos de 2h	0.035718	
P018	0.035173	

Feature	Importância	
P006	0.029448	
M019_Pouco	0.025989	
M007_Não	0.023658	
M019_Nunca	0.023146	
R041_Não se aplica	0.022805	
O009_Sim	0.020235	
P016	0.018659	

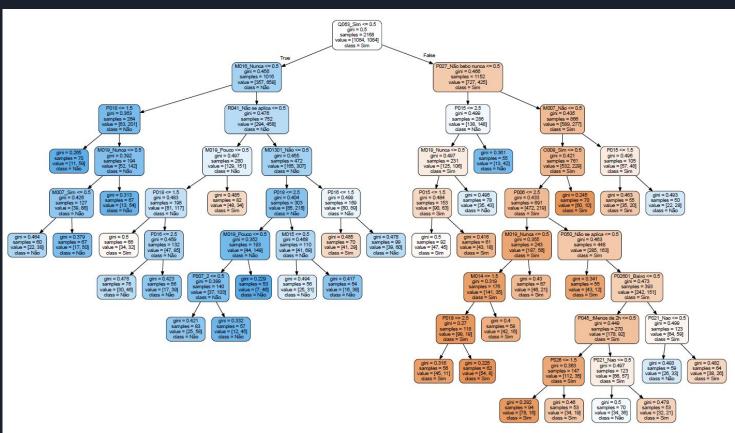
### Resultados - Árvore



- M007\_Não <= 0.5 : (TRUE) 0: está exposto, (FALSE) 1: não está exposto;</li>
- Gini: incidência de registros que utilizam um dado nó para uma dada classe. Se 0, há muita variação nas classes que usam aquele nó e, se 1, o nó é puro e apenas uma classe utiliza aquele nó;
- Azul: Não prevê diagnóstico de depressão;
- Laranja: Prevê diagnóstico de depressão;
- Branco: Não há uma decisão, apesar de prever uma classe;

### Resultados - Árvore

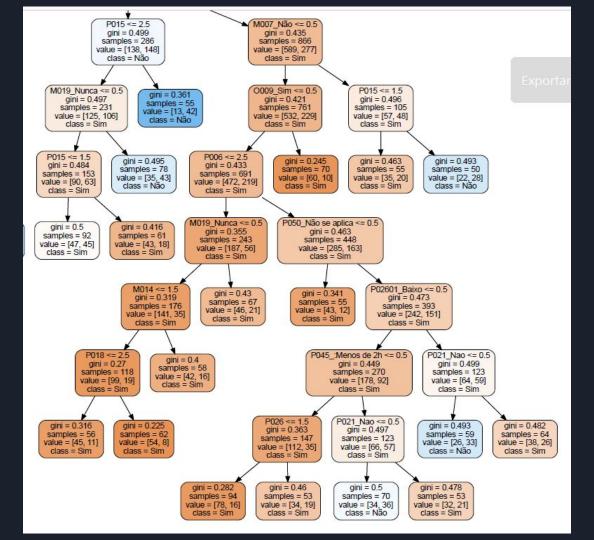


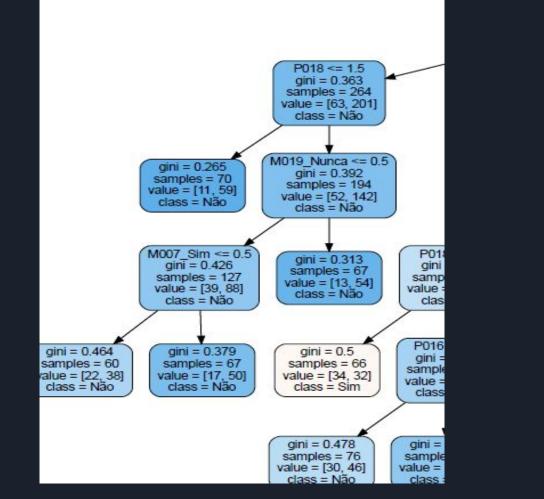






- M007:No trabalho, está exposto a alguma substância ou atividade que lhe cause estresse?
- P021: Consumo de refrigerante ou suco artificial;
- P015: Consumo de peixe na semana;
- M019: Frequência que frequenta cultos ou eventos religiosos;
- Focar na análise das exposições dos trabalhadores expostos a riscos de estresse e seus hábitos diários, como causadores e amenizadores da depressão;





#### Conclusão



Diferentes aspectos podem impactar no quadro de depressão e a essa possibilidade de variação impacta em sua detecção e diagnóstico. Pelo trabalho realizado é possível notar que o senso comum ou preconceitos sobre a doença podem não se traduzir, diretamente, no mesmo conhecimento gerado pelo aprendizado da máquina, mas que serve como base para avaliar as características que podem ser melhor estudadas e avaliadas em trabalhos futuros.

Com o trabalho, também foi possível notar a importância e dificuldade da análise de variáveis em uma base tão ampla e com a presença de variáveis muito específicas ou que não tenham relação com o problema de depressão.

Assim, o trabalho ajudou e percepção das capacidades da mineração de dados, aplicando-se a análise e tratamento de variáveis e, posteriormente, aprendizado de máquina e análise dos resultados.





ASSOCIATION, American Psychiatric. Manual diagnóstico e estatístico de transtornos mentais: DSM-IV-TR. 4. ed. Lisboa: CLIMEPSI, 2002.

ASSOCIATION, American Psychiatric. Manual diagnóstico e estatístico de transtornos mentais: DSM-5. 5. ed. Porto Alegre: artmed, 2014.

BAEK, Ji-Won; CHUNG, Kyungyong. Context deep neural network model for predicting depression risk using multiple regression. IEEE Access, IEEE, v. 8, p. 18171–18181, 2020.

ESTATÍSTICA, Instituto Brasileiro de Geografia e. Pesquisa Nacional de Saúde. 2013. Disponível em: <a href="https://www.ibge.gov.br/estatisticas/sociais/saude/9160-pesquisa-nacional-de-saude.html">https://www.ibge.gov.br/estatisticas/sociais/saude/9160-pesquisa-nacional-de-saude.html</a>.

MALAQUIAS, Kelly et al. Classification and characterization of children and adolescents with depressive symptomatology using machine learning. In: IEEE. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). [S.I.], 2019. p. 534–539.

ORGANIZATION, World Health. Depression and other common mental disorders: global health estimates. [S.I.], 2017. 24 p.

PRIYA, Anu; GARG, Shruti; TIGGA, Neha Prerna. Predicting anxiety, depression and stress in modern life using machine learning algorithms. Procedia Computer Science, Elsevier, v. 167,p. 1258–1267, 2020.

RICHTER, Thalia et al. Using machine learning-based analysis for behavioral differentiation between anxiety and depression. Scientific Reports, Nature Publishing Group, v. 10, n. 1, p.1–12, 2020.