



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

Instituto de Ciências Exatas e de Informática

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

Cinthia M. Souza¹

Douglas Spencer de Oliveira Lucas²

Isabela Borlido Barcelos³

Wenderson Júnio de Souza⁴

Resumo

Este trabalho apresenta o desenvolvimento de um modelo de *Data Mining* para identificar as características da depressão em jovens e adultos utilizando a base de dados PNS de 2013, a partir de um questionário domiciliar de abrangência nacional. Dentre as informações contidas na base, há as informações individuais, com enfoque nas principais doenças crônicas não transmissíveis, aos estilos de vida e ao acesso ao atendimento médico. Neste trabalho foram realizados a seleção e o pré-processamento das variáveis identificadas como preditoras de depressão, a fim de reduzir a quantidade de dados ausente, remover *outliers*, e agregar variáveis que contenham a mesma informação. Para o treinamento do modelo, foi utilizada uma árvore de decisão com 30 *features* obtidas a partir da seleção, agregação e redução de valores das *features* da base original. Apesar do modelo treinado não obter uma avaliação competitiva em relação à literatura, a árvore de decisão trata-se de uma estrutura interpretável e, portanto, nossos resultados podem ser utilizados como *baseline* para outros estudos relacionados. Ademais, os valores obtidos neste estudo podem agregar informação acerca das causas identificadas no modelo.

Palavras-chave: Perfil. Características. Depressão. Árvore de decisão. PNS.

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

¹Graduanda em Engenharia de Computação pela PUC Minas, Brasil – cinthia.mikaela@sga.pucminas.br.

²Graduando em Engenharia de Computação pela PUC Minas, Brasil – douglas.spencer@sga.pucminas.br.

³Graduanda em Engenharia de Computação pela PUC Minas, Brasil – isabela.borlido@sga.pucminas.br.

⁴Graduando em Engenharia de Computação pela PUC Minas, Brasil – wenderson.junio@sga.pucminas.br.

1 INTRODUÇÃO

A depressão é uma das doenças que mais está presente na população mundial, com 322 milhões de pessoas com depressão e a região das Américas é uma das regiões do planeta com maiores incidências de casos de depressão, tendo no Brasil, em 2017, uma presença 5,8% da população diagnosticada com depressão. É uma doença caracterizada pela sensação de solidão, com sentimento de tristeza e cansaço, falta de interesse, foco e prazer e perturbações de sono e apetite (ORGANIZATION, 2017). A depressão pode ser subdividida em diferentes características, passando depressão de alterações e desregulação de humor até uma condição mais crítica, a depressão maior. A depressão maior é caracterizada pela perda de interesse em quase todas as atividades e humor deprimido por períodos de, pelo menos, duas semanas e sua criticidade é percebida pela alta associação com casos de suicídios. A depressão é uma doença que não tem idade para acontecer e pode ser causada e causadora de diferentes fatores, podendo causar a somatização das dores psicológicas, como dores em partes específicas ou generalizadas no corpo (ASSOCIATION, 2014, 2002).

Portanto, este trabalho visa contribuir na identificação e caracterização do perfil dos diagnosticados com depressão, por meio de um modelo computacional, utilizando uma árvore de decisão, capaz de prever a presença de depressão, baseando-se nos dados disponibilizados pelo questionário Pesquisa Nacional de Saúde (PNS), realizado em 2013 (ESTATÍSTICA, 2013).

O trabalho está dividido em cinco seções. A Seção 2 apresenta uma breve descrição dos trabalhos correlatos sobre detecção e previsão do quadro de depressão. A Seção 3 apresenta a metodologia do trabalho e a Seção 4 apresenta os resultados e análises.

2 TRABALHOS RELACIONADOS

Atualmente, a depressão é um problema mental comum na sociedade. Diante disso, o estado-da-arte relacionado ao estudo desse problema vem ganhando notoriedade. Uma das vertentes desses estudos compreende a utilização de modelos computacionais e matemáticos para criação de modelos, utilizando base de dados históricas, a fim de gerar previsões e mapear as características relacionadas a esse problema. Nessa seção, são apresentados alguns dos trabalhos presentes na literatura que utilizam modelos computacionais aplicado ao problema da depressão.

Em Priya et al. (2020), os autores utilizaram algoritmos de aprendizado de máquina para realizar previsões de ansiedade, depressão e estresse. Os dados utilizados nesses algoritmos foram coletados de indivíduos empregados e desempregados em diferentes culturas e comunidades por meio do questionário *Depression, Anxiety and Stress Scale* (DASS 21). Considerando que ansiedade, depressão e estresse ocorrem em cinco níveis de gravidade, os autores aplicaram cinco algoritmos de aprendizado de máquina diferentes – Árvore de Decisão (DT), Floresta Aleatória (RFT), Naïve Bayes, Máquina de Vetores de Suporte (SVM) e K-Vizinhos

Mais Próximo (KNN) – que, segundo eles, são altamente precisos e, conseqüentemente, adequados para prever problemas psicológicos. Após a aplicação dos métodos, pelo fato desse problema produzir classes desequilibradas, a seleção do melhor modelo foi feita com base no *f1-score* – que é usado para casos de particionamento desequilibrado – o que ajudou a identificar o modelo de melhor precisão. Além disso, o parâmetro de especificidade revelou que os algoritmos também eram especialmente sensíveis a resultados negativos. Ao final, a precisão do Naïve Bayes foi considerada a mais alta, embora a RFT tenha sido identificada como o melhor modelo.

Malaquias et al. (2019) apresentaram em seu trabalho um estudo da depressão em crianças e adolescentes brasileiras. Para realizar esse estudo, os autores utilizaram os modelos de predição DT, SVM e uma *Artificial Neural Network* (ANN). A metodologia utilizada pelos autores consiste em três etapas: pré-processamento da base de dados, treinamento do modelo e validação. Na etapa de pré-processamento, os autores realizaram o tratamento dos dados realizando a padronização dos dados, remoção de atributos com poucas informações, tratamento de inconsistências, discretização, dentre outros processos. Os modelos foram treinados utilizando o método *cross-validation* com 10 *folds* e os resultados foram validados utilizando as métricas *precision*, *recall*, *accuracy*, *f-measure* e *roc area*. De acordo com os autores, o trabalho apresentou resultados competitivos com o estado-da-arte.

Baek e Chung (2020) propuseram em seu trabalho um modelo para predição do risco da depressão chamado *context-DNN*. Esse modelo utiliza informações de contextos de variáveis e um *Deep Neural Network* (DNN). Para realizar o treinamento e validação do modelo, os autores utilizaram os dados disponibilizados pelo *Korea National Health and Nutrition Examination*. O modelo proposto possui três etapas: pré-processamento e composição dos dados, extração de variáveis e treinamento do modelo *context-DNN*. Na etapa de pré-processamento e composição dos dados, foram extraídas 52 *features*, para cada *feature* foram removidos os dados nulos e/ou informações como *N/A*, *Not Applicable*, *Unknown* e *No response*. Ao final, todos os dados foram normalizados. Posteriormente, os autores realizaram a extração das *features* utilizando um modelo de regressão múltipla. O resultado da regressão gerou 14 *features*. Cada uma das *features* possuem variáveis de contexto. O modelo de predição proposto treina, para cada uma das *features*, uma DNN. Ao final, os resultados de cada uma das DNN é agregado gerando um único valor. Esse valor é classificado em quatro estágios, sendo “bom”, “não-ruim”, “perigoso” e “alto risco”. O modelo foi treinado utilizando *k-fold cross validation*. Para validar o modelo, foram utilizadas as métricas *goodness-of-fit*, *accuracy* e *recall*. De acordo com os autores, o modelo proposto possui um desempenho superior ao modelo DNN. Sendo que o melhor resultado foi obtido utilizando 6-*fold*, com média de erro de classificação de 0,012358.

Richter et al. (2020) apresentaram em seu trabalho uma metodologia que testa as diferenças entre os indivíduos ansiosos e deprimidos. Para desenvolver o trabalho, os autores utilizaram técnicas de Machine Learning (ML) para detectar as características únicas relacionadas ao diagnóstico de ansiedade e da depressão. Para realizar os experimentos, os autores coletaram os dados a partir de um questionário. Com isso, foi criada uma amostra com as

respostas do questionário de 125 participantes. O classificador implementado pelos autores é baseado em DT e recebe como entrada medidas comportamentais e retorna o perfil sintomático. O classificador foi projetado para identificar quatro grupo de participantes, os com altos níveis de ansiedade e baixos níveis de depressão (HA), os com altos níveis de depressão e baixos níveis de ansiedade (HD), os com altos níveis de ansiedade e depressão (HAD) e os com 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.

Com base nos trabalhos apresentados, verifica-se que a tarefa de classificação do risco da depressão e da identificação das características relacionadas a depressão vem sendo bastante estudada atualmente. Contudo, essa área ainda possui melhorias a serem exploradas, principalmente considerando a interpretação dos modelos de classificação criados para essa tarefa. Diante disso, entende-se que o diferencial desse trabalho está na identificação das características relacionadas a depressão.

3 METODOLOGIA

Nesta seção são apresentados os métodos e materiais utilizados para a realização deste trabalho.

3.1 Materiais

Para o desenvolvimento desse trabalho, utilizou-se a linguagem de programação Python e a base de dados da Pesquisa Nacional de Saúde de 2013 (PNS-2013). Essa base possui 205.546 registros e para cada registro existem 942 variáveis (representadas por um código), referente à perguntas feitas no questionário da PNS. Entretanto, a base de dados utilizada neste trabalho é uma sub base gerada a partir da PNS e possui 8470 registros, 50% destes são de pessoas que afirmaram que algum médico ou psiquiatra lhe deu o diagnóstico de depressão e 50% afirmaram que não foram diagnosticadas com depressão (variável de código 'Q092').

3.2 Métodos

A metodologia utilizada para o desenvolvimento do trabalho passou por algumas etapas. A primeira etapa constitui na elaboração de um modelo conceitual que engloba os fatores relacionados à depressão. Conforme apontado por organizações de saúde e psiquiatria (ASSOCIATION, 2002, 2014; ORGANIZATION, 2017), são avaliados como preditores os fatores relacionados à hábitos alimentares, tabagismo, alcoolismo, traumas psicológicos, diagnósticos de outras doenças, baixa socialização e atividade física, além dos comportamentos avaliados no

questionário PHQ-9, cujas percepções do estado mental avaliadas pelo questionário realiza um pré-diagnóstico de saúde. Todas as variáveis da base relacionadas aos fatores apontados foram adicionadas ao modelo conceitual.

Na segunda etapa foram removidas as variáveis apontadas pelos estudos com consequência (ou seja, sintoma) da depressão. A partir das selecionadas como preditoras de depressão, foi realizado um estudo do percentual de dados ausentes na base para cada uma dessas variáveis. A fim de evitar analisar regiões com menor incidência de depressão, foi realizada uma avaliação da ocorrência de depressão por Estado, selecionando para a próxima etapa as regiões com maiores ocorrências.

A etapa seguinte consiste no tratamento das variáveis. Nesta etapa, buscou-se agregar variáveis que continham a mesma informação (por exemplo, *Pratica exercício físico?* e *Quantas vezes por semana pratica exercícios físicos?*), e na união de respostas a fim de obter uma relação mais balanceada entre a quantidade de registros para cada resposta. Mesmo na impossibilidade de agregação de variáveis ou união de respostas, foi realizado um estudo individual para cada uma a fim de utilizar sua informação agregada ou preencher os dados ausentes, quando possível.

4 RESULTADOS

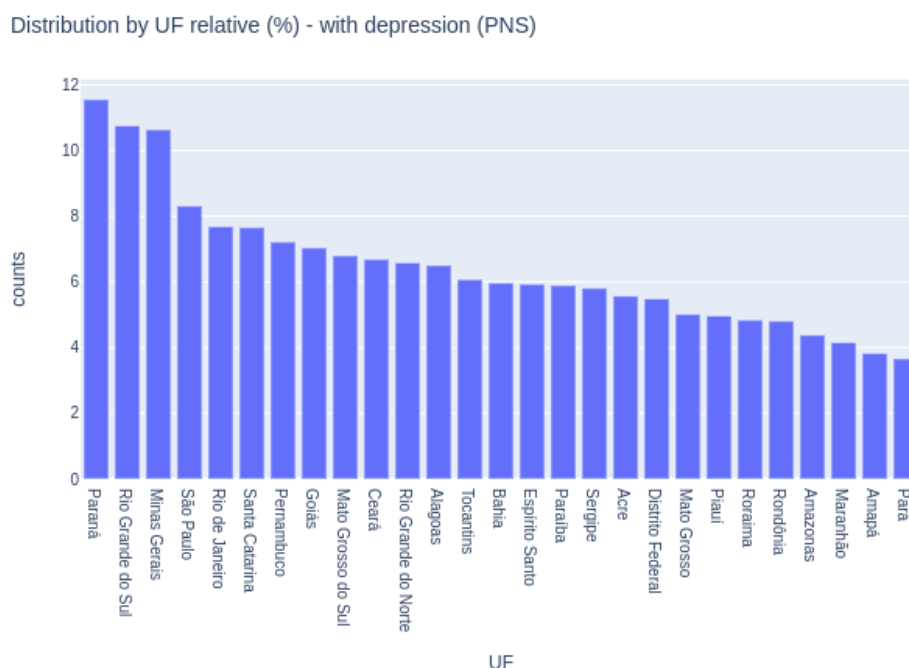
Nesta seção são apresentados os resultados obtidos e análises realizadas neste trabalho.

4.1 Resultados iniciais

Para avaliar os estados ou Unidade Federativa (UF) com maior presença de diagnósticos de depressão, foi avaliado a base original PNS, com 205.546 registros; Foram removidos os registros que não haviam respondido a questão sobre terem ou não o diagnóstico de depressão, reduzindo a base para 60.202 registros; Convertido o tipo do dado referente a UF de numérico para a sigla da UF; e feito o somatório dos registros de cada UF. Para a Figura 1, foram contabilizados a porcentagem de casos de depressão em relação ao total de pessoas que responderam ao questionário. A Figura 2 contabiliza os valores absolutos da quantidade de casos de depressão por UF.

Avaliando a Figura 2, o estado de São Paulo possui a maior quantidade absoluta de casos de depressão. Em contrapartida, a 1 demonstra que a UF com maior índice é o estado do Paraná. Realizando uma análise mais ampla, é possível identificar que os estados pertencentes às regiões do Sul e Sudeste, possuem altos níveis de índice de depressão. Portanto, ao ampliar a análise, agrupando os estados por regiões (Figura 3), os estados que integram as regiões do Sul e Sudeste possuem o maior e o terceiro maior índices de depressão, respectivamente. A região do Nordeste possui o segundo maior índice, mas vale ressaltar que é uma região que possui 9 estados, ao contrário do Sul e Sudeste que possuem, respectivamente, 3 e 4 estados.

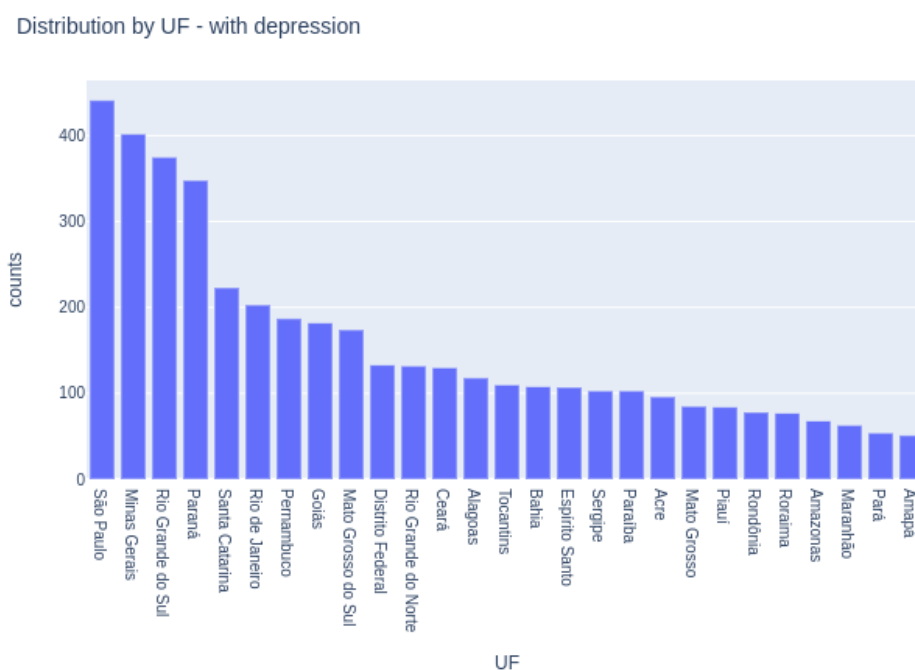
Figura 1 – Casos de depressão em relação à quantidade total de pessoas por UF



Fonte: Elaborada pelos autores.

Após esta análise, decidiu-se filtrar os registros da base de dados regiões Sul e Sudeste, reduzindo o número de registros para 3.604, de uma base com 8.470 registros, sendo 2.099 registros de casos com depressão e 1505 de casos sem depressão.

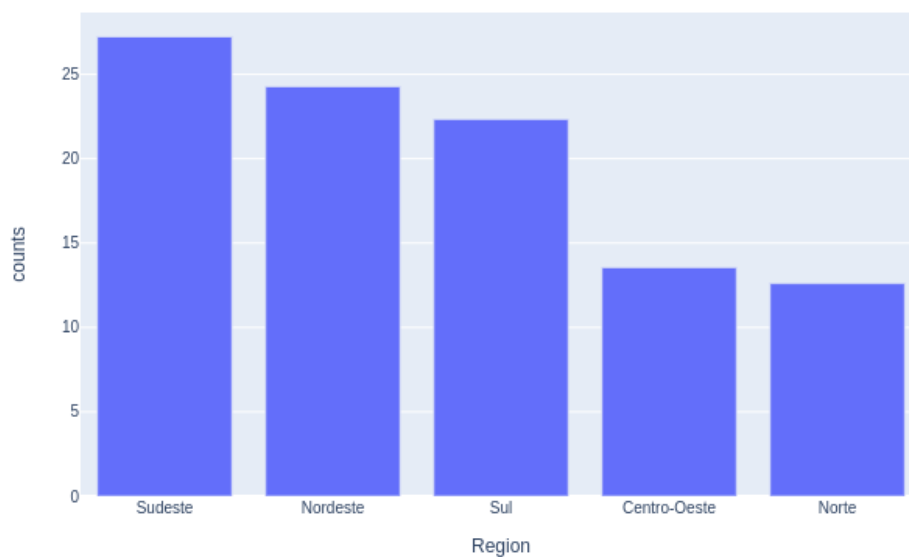
Figura 2 – Casos absolutos de depressão por UF



Fonte: Elaborada pelos autores.

Figura 3 – Distribuição relativa de casos de depressão por região

Relative Distribution by Region (%) (PNS) - with depression



Fonte: Elaborada pelos autores.

4.2 Tratamento das Variáveis

O tratamento das variáveis foi realizado visando utilizar as variáveis que possam melhor caracterizar ou identificar prováveis causas para a depressão e não foram utilizadas variáveis categorizadas como consequências da depressão. Além da análise individual das variáveis de causa, consequência e relevância, também foi observado e analisado a presença da quantidade de valores omissos e a possibilidade de inferência destes. Entretanto, foram descartadas as variáveis que não podiam ser inferidas, pois dependiam exclusivamente da pessoa que respondeu o questionário ou não haviam dados suficientes para realizar a inferência.

As Tabelas 1,2 e 3 apresentam as *features* mantidas neste estudo como causadoras de depressão. As variáveis das Tabelas 1,2 e 3 foram, então, agrupadas conforme o seu tipo de informação. Cada grupo de *features* foi analisado de maneira a identificar quais informações eram mais relevantes e buscando unir àquelas com mais dados ausentes, reduzindo a quantidade de *features*. Também foi reduzida a quantidade de valores de algumas das *features*. Ao final do processo, foram obtidas 30 *features*.

5 TREINAMENTO E TESTE

Para o 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.

Tabela 1 – Variáveis relacionadas à apoio emocional, violência e acidentes.

Código	Pergunta
M014	Quantidade de familiares que se sente à vontade para falar sobre quase tudo
M015	Quantidade de quantos amigos que se pode falar sobre quase tudo
M016	Frequência de atividades esportivas ou artísticas em grupo
M018	Frequência de trabalho voluntário não remunerado
M019	Frequência a cultos ou atividades religiosas
O009	Se envolveu em algum acidente de trânsito com lesões corporais
O014	Para o acidente mais grave, deixou de realizar atividades habituais
O020	Teve ou tem alguma sequela e/ou incapacidade decorrente deste acidente
O022	Para o acidente de trabalho mais grave, deixou de realizar atividades habituais
O024	Teve ou tem alguma sequela e/ou incapacidade decorrente deste acidente
O025	Sofreu alguma violência ou agressão de pessoa desconhecida
O037	Sofreu alguma violência ou agressão de pessoa conhecida

Tabela 2 – Variáveis relacionadas à alimentação e exercícios físicos

Código	Pergunta
P008	Quantas vezes por dia come salada
P010	Quantas vezes por dia come verdura ou legume cozido
P017	Quantos copos por dia toma de suco de frutas natural
P019	Quantas vezes por dia come frutas
P021	Tipo de refrigerante ou suco artificial que costuma tomar
P025	Em quantos dias da semana come alimentos doces
P026	Em quantos dias da semana substitui o almoço ou jantar
P02601	Consumo de sal
P028	Quantos dias por semana costuma tomar alguma bebida alcoólica
P034	Nos últimos três meses, praticou algum tipo de exercício físico ou esporte
P035	Quantos dias por semana costuma praticar exercício físico ou esporte
P042	Quantos dias por semana faz deslocamento a pé ou bicicleta
P045	Em média, quantas horas por dia assiste televisão
P050	Atualmente, fuma algum produto do tabaco?
P068	Com que frequência alguém fuma dentro do seu domicílio

1. Separação da base em treino e teste utilizando 20% para teste e 80% para treino;
2. Transformação das variáveis categóricas utilizando One-hotting 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 *Undersampling* chamada NearMiss;
4. Validação do modelo utilizando as métricas *accuracy*, *precision*, *recall* e *f1-score*;
5. Poda da árvore chegando nos seguintes hiperparâmetros: $max_depth = 10$, $min_samples_split = 100$, $min_samples_leaf = 50$;
6. O treinamento foi realizado utilizando as mesmas técnicas apresentadas etapa 3, contudo, foram utilizados os valores de hiper-parâmetros obtidos ao realizar a poda;

Tabela 3 – Variáveis relacionadas à outras doenças e aborto.

Código	Pergunta
Q002	Algum médico já lhe deu o diagnóstico de hipertensão arterial
Q030	Algum médico já lhe deu o diagnóstico de diabetes
Q05501	Tem ou teve problemas na vista por causa do diabetes
Q05502	Tem ou teve infarto por causa do diabetes
Q05503	Tem ou teve AVC ou derrame por causa do diabetes
Q05504	Tem ou teve outro problema circulatório por causa do diabetes
Q05505	Tem ou teve problemas problema nos rins por causa do diabetes
Q05506	Tem ou teve Úlcera/ferida nos pés por causa do diabetes
Q05507	Tem ou teve que fazer amputação
Q05508	Tem ou teve coma diabético
Q05509	Outro
Q063	Algum médico já lhe deu o diagnóstico de uma doença do coração?
Q068	Algum médico já lhe deu o diagnóstico de AVC ou derrame?
Q079	Algum médico já lhe deu o diagnóstico de artrite ou reumatismo?
Q084	O(a) Sr(a) tem algum problema crônico de coluna?
Q116	Algum médico já lhe deu o diagnóstico de alguma doença no pulmão ou DPOC?
Q120	Algum médico já lhe deu diagnóstico de câncer?
Q124	Algum médico já lhe deu o diagnóstico de insuficiência renal crônica?
R041	A Sra já teve algum aborto espontâneo?
R042	A Sra já teve algum aborto provocado?

7. Validação utilizando a base de teste. Novamente, foram utilizadas as métricas *accuracy*, *precision*, *recall* e *f1-score*;
8. Seleção e análise das *features* mais importantes;
9. *Plot* da árvore, análise e extração de regras.

5.1 Métricas

Nesta seção são apresentados as métricas resultantes do treinamento e teste do modelo. Analisando a Tabela 4 Analisando a Tabela 5, podemos confirmar

Tabela 4 – Métricas para a Base de Treinamento

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

A Tabela 6 apresenta as variáveis com importância (peso) maior que zero para a construção da árvore, com a variável "Q063_Sim" sendo a mais importante para a árvore, sendo, portanto, a sua raiz.

Tabela 5 – Métricas para a Base de Testes

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

Tabela 6 – Importância das *Features*

<i>Feature</i>	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
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

Visto que o modelo busca classificar pessoas com depressão, entende-se que, neste caso, uma melhor classificação é feita para um maior *recall*, uma vez que esta métrica leva em consideração apenas as classificações positivas da base, ou seja, das pessoas com depressão, quantas o modelo classificou como depressivas. Contudo, o modelo obteve maior precisão, indicando que uma maior porcentagem de classificações positivas foi obtida em relação à quantidade de registros com depressão na base. Ou seja, a taxa de acerto é maior em relação ao que o modelo classificou do que em relação à todos os registros positivos.

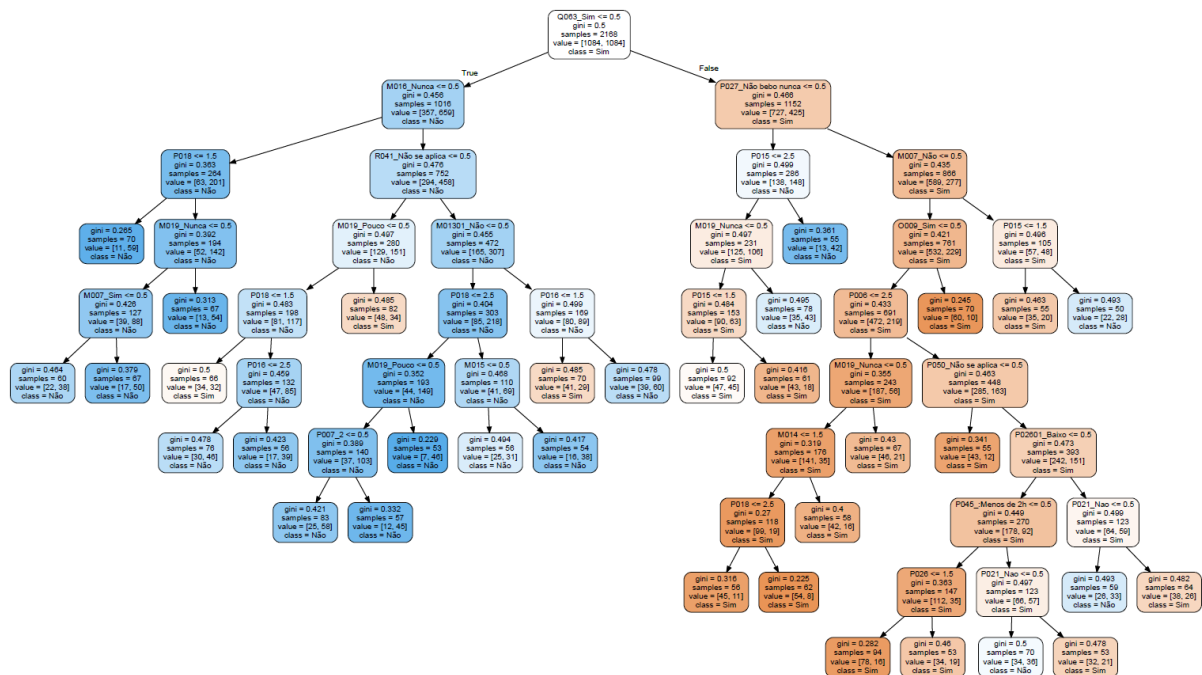
5.2 Árvore de Decisão

A árvore de decisão gerada pelo treinado é apresentada na Figura 4, com com 65 nós e possui o índice de Gini e qual a classe atribuída para cada nó, como métricas mais importantes. Para o índice Gini, quanto menor o valor, mais pura é a decisão da árvore para a escolha da classe naquele nó. Para os nós com a cor azul, temos a classe 'não', referente à decisão da árvore que o registro analisado não possui depressão e se laranja, se possuir depressão. Para cada nó da árvore, até as folhas, há duas arestas, sendo a da esquerda referente à condição

verdadeira do nó e a da direita, se falso. Um exemplo de análise da árvore é: "M007_Não <= 0.5 : (Se verdade é 0, logo está exposto. Se falso, será 1, então não está exposto."

A principal informação que se tira da árvore é a raiz da árvore dividir e começar sua análise com base na variável "Doença cardíaca"(Q063), que é uma doença que pode ser impeditiva. Analisando a árvore, têm-se, de modo geral, que caso a pessoa possua alguma doença (relativo a condição falsa do nó raiz), ou seja, a análise vá para o lado direito da árvore, há uma grande probabilidade de haver depressão, dado a quantidade de folhas que são alcançadas, que associam a depressão, por este nó.

Figura 4 – Árvore de decisão gerada pelo treinamento



Fonte: Elaborada pelos autores.

Uma das análises que podem ser feitas, e que se destacam, da árvore é analisar os nós que possuem a variável M007, referente a atividades ou trabalhos estressantes. Do lado direito da árvore, podemos ver que se a pessoa realiza esta atividade, muito provavelmente ela terá depressão, pois a maioria das folhas abaixo deste nó são da classe "sim"(possui depressão). Curiosamente, somente não terá depressão se comer peixe várias vezes na semana (P015). Porém, do lado esquerdo da árvore, no nó que possui a variável M007, vemos que o indivíduo provavelmente não terá depressão. Porém, este fato contradiz o lado direito da árvore, como apresentado anteriormente. Portanto, analisando os nós superiores, temos os nós de variável M019 (frequentar eventos religiosos). Isso pode indicar que a fé dos indivíduos podem amenizar seus sintomas de depressão. Voltando a análise para o lado direito da árvore, ainda na variável M007, a árvore pode ter gerados uma informação probabilisticamente correta, mas que pode não se confirmar em correlação, que seria que o consumo de refrigerante ou suco artificial (P021) não está associado a depressão ou que consumir tais produtos não levem a depressão.

5.3 Insight

Analisando os resultados da árvore, pode-se observar julgamentos que a árvore mostrou serem adequados, baseado nos registros recebidos. Baseado neste conhecimento formado, tiramos o *insights* de que estudar os trabalhadores que estão envolvidos em situação de risco e estresse, junto com os seus hábitos, podem ser descobertos padrões que identifiquem o que ameniza e/ou o que aumentam os riscos causadores de depressão, para essa parcela de indivíduos e possam ser aplicados em ambientes perigosos e estressantes, de forma a melhorar a qualidade de vida destes trabalhadores.

6 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.

REFERÊNCIAS

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: <<https://www.ibge.gov.br/estatisticas/sociais/saude/9160-pesquisa-nacional-de-saude.html>>.

MALAKUIAS, 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.l.], 2019. p. 534–539.

ORGANIZATION, World Health. **Depression and other common mental disorders: global health estimates**. [S.l.], 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.