

Aprendizado Ensemble e sua Relação com a Interpretabilidade para a Identificação de Crianças e Adolescentes com Depressão

Ludmila Bruna Santos Nascimento¹

¹Pontifícia Universidade Católica de Minas Gerais
Dom José Gaspar, 500, Belo Horizonte - Minas Gerais, Brasil
ludmila.bruna@sga.pucminas.br

Marcelo de Sousa Balbino

Centro Federal de Educação Tecnológica de Minas Gerais
Dezenove de Novembro, 121, Timóteo - Minas Gerais
marcelobalbino@cefetmg.br

Ana Carolina Medeiros Gonçalves¹

ana.medeiros@sga.pucminas.br

Cristiane Neri Nobre¹

nobre@pucminas.br

RESUMO

Algoritmos de aprendizado de máquina têm se mostrado eficientes em problemas de classificação na previsão e identificação de transtornos mentais. Esses métodos podem ser considerados como interpretáveis ou não interpretáveis, sendo a interpretabilidade essencial para a credibilidade dos modelos, principalmente em contextos clínicos. O objetivo deste trabalho foi avaliar o desempenho de três métodos *ensemble* (*Bagging*, *Random Forest* e *XGBoost*) e dois métodos únicos (Árvore de Decisão e Rede Neural) na classificação da depressão em jovens, além de discutir a interpretabilidade desses métodos. Os resultados revelaram que os métodos *ensemble*, especialmente *Random Forest* e *XGBoost*, apresentaram melhor desempenho. No entanto, esses modelos possuem maior necessidade de uma investigação para identificar os atributos importantes no contexto da depressão. Essa discussão ressalta o potencial das abordagens computacionais avançadas em complementar as práticas clínicas tradicionais na área da saúde.

PALAVRAS CHAVE. Ensemble, Aprendizado de Máquina, Interpretabilidade, Depressão em Crianças e adolescentes

Inteligência computacional, Pesquisa Operacional em Saúde.

ABSTRACT

Machine learning algorithms have proven to be efficient in classifying problems and predicting and identifying mental disorders. These methods can be considered interpretable or non-interpretable, with interpretability being essential for the credibility of the models, especially in clinical contexts. The objective of this work was to evaluate the

performance of three *ensemble* methods (*Bagging*, *Random Forest* and *XGBoost*) and two single methods (Decision Tree and Neural Network) in the classification of depression in young people, in addition to discussing the interpretability of these methods. The results revealed that the ensemble methods presented better performance, especially *Random Forest* and *XGBoost*. However, these models have a greater need for investigation to identify the essential variables in the context of depression. This discussion highlights the potential of advanced computational approaches to complement traditional clinical practices in health.

KEYWORDS. Ensemble, Machine Learning, Interpretability.

Computational Intelligence, Operations Research in Health.

1. Introdução

Algoritmos de Aprendizado de Máquina (AM) para problemas de classificação têm sido promissores na previsão de depressão em crianças e adolescentes [Tate e et.al, 2020], assim como em diversos contextos relacionados ao diagnóstico e prognóstico de transtornos mentais, incluindo a depressão em adultos mais velhos, a doença de Alzheimer e a doença de Parkinson [Witten e et.al, 2016]. Na área de AM, duas classificações são comumente encontradas: algoritmos interpretáveis, chamados ‘caixa branca’, e os não interpretáveis, chamados de ‘caixa preta’.

Embora modelos interpretáveis possam ser fáceis de interpretar, como regressão linear e árvores de decisão, os algoritmos mais poderosos com alta precisão, como os métodos *ensemble*, as redes neurais, a máquina de vetores de suporte, são modelos não interpretáveis, que fornecem previsões que não são projetados para serem interpretáveis e não podem ser facilmente interpretados.

No entanto, abordagens interpretáveis de aprendizado de máquina são essenciais para que os pesquisadores entendam, confiem e gerenciem com eficácia as soluções encontradas. Desta forma, a interpretabilidade é um conceito crucial no contexto do AM clínico, pois é improvável que os modelos alcancem um desempenho impecável. Fornecer uma explicação interpretável pode facilitar a tomada de decisões com base nesses modelos [Lu et al., 2023]. A relevância da interpretabilidade para todos os algoritmos de decisão baseados em AM é evidenciada pelo Projeto de Declaração de Direitos de IA do Governo dos Estados Unidos, que inclui o “Aviso e Explicação” como um princípio essencial para modelos de previsão baseados em AM [Office, 2022]. Além disso, as diretrizes da Food and Drug Administration (FDA) dos EUA para sistemas de apoio à decisão clínica (CDSS) sublinham a importância de fornecer a fundamentação das previsões, além de outras regulamentações e padrões aplicáveis aos cuidados de saúde e outras indústrias.

Neste estudo, aplicamos algoritmos de aprendizado de máquina interpretáveis e não interpretáveis para desenvolver e avaliar um modelo destinado ao diagnóstico da depressão em jovens entre 10 e 16 anos. O modelo classifica os dados dos pacientes como exibindo ‘alta’ ou ‘baixa’ sintomatologia de depressão. Nosso objetivo é aprimorar o desempenho do modelo por meio da utilização de conjuntos de classificadores, especificamente empregando e comparando três métodos *ensemble*: *Bagging*, *Random Forest (RF)* e *XGBoost* com dois métodos únicos: Árvore de Decisão (AD) e Rede Neural (RN).

Os métodos *sensemble* de aprendizado tornaram-se uma importante tendência de pesquisa nos últimos anos, o que levou a um aumento no número de pesquisas utilizadas para aprendizagem em conjunto em diversos domínios de aplicações. Neste trabalho os métodos *sensemble* utilizados superaram os métodos únicos utilizados, e, devido à falta de interpretabilidade dos modelos, utilizamos a abordagem SHAP [Lundberg e Lee, 2017], *SHapley Additive exPlanations*, para identificarmos as principais características que colaboram para que uma criança ou adolescente tenha depressão alta ou baixa.

Identificar a depressão de maneira precoce pode ser fundamental para prevenir e aprimorar o tratamento clínico, pois a maioria dos jovens não procura ajuda [Balsters e et.al, 2010]. De acordo com um questionário conduzido pelo Fundo das Nações Unidas para a Infância (UNICEF)¹ em 2020, que envolveu mais de 4 mil adolescentes de todo o Brasil, abordando questões sobre saúde mental durante a pandemia e o acesso a suporte psicológico, constatou-se que 41% dos entrevistados não procuraram assistência para cuidar do seu bem-estar físico e mental durante o período de quarentena.

Na tentativa de obter um diagnóstico mais preciso e o mais cedo possível, muitos estudos buscam evidências verbais, não verbais, comportamentais, linguísticas, gestuais e sociais para auxiliar no diagnóstico precoce e preciso da depressão em adolescentes [Balázs e et.al, 2013].

2. Referencial Teórico

2.1. Depressão em crianças e adolescentes

A Organização Mundial da Saúde (OMS) argumenta que depressão é um transtorno mental comum, afetando mais de 300 milhões de pessoas em todo o mundo, sem distinção de idade. Ela é reconhecida como a principal causa de incapacidade global e desempenha um papel significativo na carga global de doenças. Os distúrbios depressivos incluem o Transtorno Depressivo Maior (TDM) e a Distímia [Pinucci et al., 2023].

Um número considerável de pacientes apresenta as primeiras manifestações de TDM durante a infância e a adolescência, sendo comum que esses indivíduos continuem a enfrentar episódios de TDM ao longo da idade adulta. O TDM é o distúrbio depressivo mais comum em adolescentes, e neste artigo, o termo depressão se refere a essa forma de transtorno depressivo. No Brasil, a depressão atinge aproximadamente 8% a 12% de crianças e adolescentes [WHO, 2017].

O TDM em crianças e adolescentes é caracterizado por diversos sintomas como humor deprimido ou irritabilidade, desânimo, desinteresse, sentimentos inapropriados de culpa, perda ou aumento do apetite e do peso, distúrbios do sono, dificuldades cognitivas e ideação suicida [APA, 2014]. Crianças depressivas podem enfrentar dificuldades significativas em diversas áreas da vida, incluindo relacionamentos prejudicados com familiares e parceiros. Além disso, adolescentes deprimidos estão mais propensos a evasão escolar, gravidez não planejada, abuso de substâncias e problemas comportamentais.

Os critérios clínicos e diagnósticos do TDM, conforme estabelecidos pelo Manual Diagnóstico e Estatístico de Transtornos Mentais (DSM-5), envolvem um conjunto de sintomas somáticos e não somáticos que afetam o funcionamento diário por pelo menos duas semanas [APA, 2014]. Esses sintomas prejudiciais podem comprometer o desempenho escolar, ocupacional e social dos adolescentes.

¹Disponível em <https://www.unicef.org/brazil/comunicados-de-imprensa/72-dos-adolescentes-sentiram-necessidade-de-pedir-ajuda-durante-pandemia>

Pesquisas sugerem que metade dos pacientes que desenvolvem transtornos mentais experimentam os primeiros sintomas antes dos 15 anos de idade. As consequências dos episódios depressivos nesses pacientes incluem prejuízos graves no funcionamento social, ideação suicida e tentativas de suicídio.

2.2. Métodos Ensemble

A abordagem do método *ensemble*, ou aprendizado em conjunto, no campo de aprendizado de máquina consiste na combinação de múltiplos modelos para melhorar a precisão e a capacidade de generalização [Mienye e Sun, 2022]. Esses métodos treinam vários modelos base e os combinam para superar as limitações individuais de cada um. A diversidade e precisão dos modelos base são cruciais para o sucesso dos *ensembles*, influenciando diretamente o desempenho do conjunto [Mienye e et.al, 2020].

O processo de combinação dos modelos base pode variar de acordo com o método *ensemble* utilizado, como *bagging*, *boosting* ou *stacking*, cada um com suas próprias técnicas de combinação.

A Figura 1 mostra uma arquitetura de *ensemble* comum, que pode ser homogênea, com modelos do mesmo tipo, ou heterogênea, com modelos de diferentes tipos. A abordagem *ensemble* combina as previsões de múltiplos modelos individuais para produzir uma previsão final mais precisa e robusta. Isso é alcançado treinando-se modelos base independentes, gerando previsões com base nesses modelos e, em seguida, combinando essas previsões usando técnicas como média ou votação.

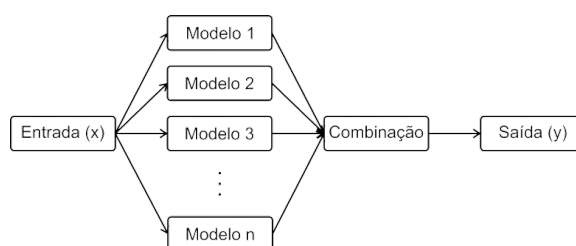


Figura 1: Arquitetura de *ensemble* comum.

Foram desenvolvidos vários métodos para construir conjuntos, incluindo *Bagging*, *Random Forest* e *XGBoost*.

O classificador *Bagging*, também conhecido como *Agregação Bootstrap* [Breiman, 1996], é uma técnica de conjunto que cria diferentes versões de um preditor (ou classificador) para serem aplicadas simultaneamente ao mesmo problema. Cada versão é treinada com conjuntos de dados de treinamento amostrados aleatoriamente. A saída é determinada por uma votação majoritária entre as previsões de cada classificador individual.

O classificador *Random Forest* é um algoritmo de aprendizado de máquina que combina várias árvores de decisão para formar um modelo robusto. Durante o treinamento, cada árvore é gerada com uma amostra aleatória dos dados e uma seleção aleatória de características. As previsões das árvores individuais são combinadas para produzir uma previsão final.

XGBoost (*Extreme Gradient Boosting*) é um algoritmo de aprendizado de máquina que utiliza um conjunto de modelos de árvores de decisão para melhorar a precisão das previsões. Ele otimiza a perda do modelo por meio de um algoritmo de gradiente descendente, produzindo um modelo de *ensemble* eficaz e escalável.

2.3. Interpretabilidade

Muitas vezes, os modelos caixa preta mais complexos (por exemplo, *RF* e Redes Neurais) superam modelos interpretáveis tradicionais (por exemplo, Regressão Linear e Árvore de Decisão) em termos de desempenho preditivo [Arrieta e et.al, 2020]. No entanto, no domínio da saúde, os profissionais acham difícil compreender e confiar nesses modelos complexos devido à falta de intuição e explicação de suas previsões. Por esse motivo cresce a demanda por métodos de interpretabilidade que possam apresentar *insights* sobre o processo de predição dos modelos de aprendizado de máquina [ElShawi e Sherif, 2021].

SHapley Additive exPlanations (SHAP) é uma abordagem proposta por Lundberg e Lee [2017] baseada na teoria dos jogos cooperativos para interpretar modelos de aprendizado de máquina. O SHAP ajuda a explicar modelos de aprendizado, na medida que permite calcular o impacto dos atributos de entrada em previsões individuais em problemas de classificação. Em linhas gerais, a soma dos efeitos de todos atributos de entrada corresponde a uma aproximação da predição que se deseja explicar.

O *Kernel SHAP* consiste em uma versão agnóstica da ferramenta capaz de gerar explicações para qualquer modelo de aprendizagem. Também existem versões específicas para determinados modelos, como o *Linear SHAP*, *Low-Order SHAP*, *Max SHAP*, *Deep SHAP* e o *SHAP Tree Explainer* [Zeng e Davoodi, 2020].

3. Trabalhos Relacionados

Em [Haque e et.al, 2021], os algoritmos *RF*, *XGBoost*, Árvore de Decisão e *Naive Bayes* Gaussiano foram utilizados para prever a depressão em crianças e adolescentes, com base nos dados da Pesquisa Australiana de Saúde Mental e Bem-Estar. Foram identificadas 11 características importantes para a detecção da depressão, e o *RF* superou os demais algoritmos.

Bhakta e et.al [2016] realizaram um estudo comparativo de cinco algoritmos de classificação, em um estudo de caso para prever a depressão em pacientes idosos. O conjunto de dados utilizado foi coletado em um único bairro na cidade de Kolkata, no leste da Índia. Eles coletaram dados de 60 pacientes, todos com 60 anos ou mais, usando a Escala de Depressão Geriátrica. Os autores utilizaram os seguintes classificadores: Rede Bayesiana (BN), Logístico, *Multilayer perceptron* (MLP), Otimização Sequencial Mínima (SMO) e Tabela de Decisão (DT). Em geral, o modelo Bayesiano teve os melhores resultados para o conjunto de dados, com 95% de precisão.

Sunmoo e et.al [2014] analisaram a depressão usando dados do BRFSS dos EUA, selecionando 73 características relevantes de 450 disponíveis. Compararam os algoritmos C4.5, *RF*, MLP e *Support Vector Machine* (SVM). As características mais relevantes incluíram ‘experiência adversa na infância’, ‘conviver com pessoa com doença mental’ e ‘abuso sexual’. Os autores destacaram a importância de incorporar resultados clínicos para melhorar o desempenho dos classificadores.

Adeniji e et.al [2022] propõem um modelo híbrido de *RF* e Rede Neural Artificial (ANN) para prever transtornos mentais em funcionários da Tecnologia da Informação. Utilizando um modelo de ‘*Bagging Ensemble*’, alcançou-se uma precisão, sensibilidade (*Recall*) e acurácia ponderadas de 84.5%. Comparado com os modelos individuais de *RF* e ANN, o modelo híbrido mostrou uma melhoria significativa de desempenho, indicando seu potencial para detecção precoce de transtornos mentais em ambientes de trabalho.

4. Materiais e Métodos

4.1. Descrição da Base de Dados

A base de dados utilizada foi obtida em parceria com a Universidade Federal de Minas Gerais. Trata-se de um conjunto de dados de 377 crianças e adolescentes com idade entre 10 e 16 anos, sendo 158 do sexo masculino e 219 do sexo feminino.

Essa base de dados é composta de 74 características que abrangem diversos aspectos, incluindo dados demográficos, sociais e relacionados à saúde mental. A descrição completa dos atributos encontra-se na Figura 2.

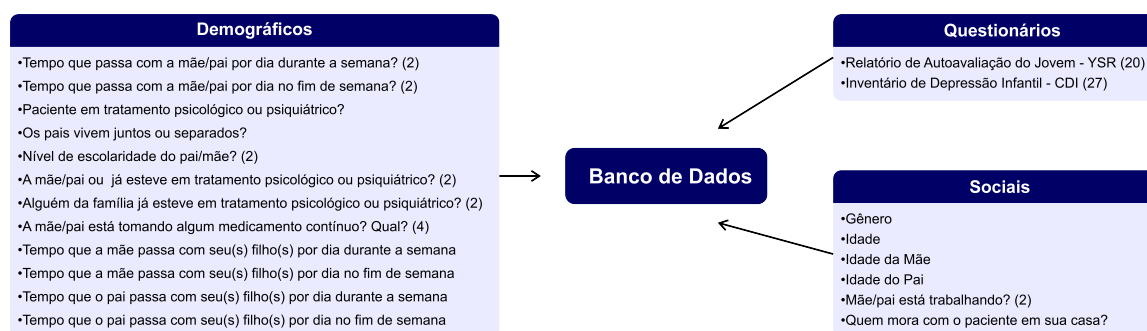


Figura 2: Descrição da base de dados, com indicação dos atributos utilizados. Os números entre parênteses indicam a quantidade de atributos para a característica.

Para a realização desse estudo, a depressão foi categorizada em dois grupos com base em um cálculo do percentil, com ponto de corte definido como percentil 85%, seguindo a recomendação de Kovacs e Staff [2003]. Nesse sentido, 63 indivíduos foram classificados como apresentando ‘depressão alta’, enquanto os demais foram classificados como tendo ‘depressão baixa’.

4.2. Pré-processamento da Base de Dados

Com o objetivo de obter um modelo mais consistente e imparcial, foram adotadas estratégias de pré-processamento antes de aplicar efetivamente os algoritmos de classificação. Todo o pré-processamento de dados foi realizado utilizando a linguagem *Python*². Sequencialmente, as fases de processamento do conjunto de dados foram realizadas.

1. *Remoção de atributos*: os atributos ‘Quais medicações sua mãe faz uso?’ e ‘Quais medicações seu pai faz uso?’ foram removidos devido ao alto valor de dados ausentes. Além desses atributos foram removidos outros 27 atributos relacionados ao questionário CDI, já que esses atributos estão altamente relacionados ao atributo de classe. Restando 45 atributos de entrada.
2. *Tratamento de dados ausentes*: cerca de 17% da informações pertencentes ao conjunto de dados continham dados ausentes, os quais foram imputados através da técnica *KN-Imputer*³, já que alguns algoritmos de aprendizado não suportam dados ausentes. O *KNImputer* é um método de imputação e utiliza os *k* vizinhos mais próximos para estimar os valores ausentes com base nos valores observados dos vizinhos mais próximos. O parâmetro utilizado para o número de vizinhos foi o *default* (igual a 5).

²Documentação disponível em: <https://www.python.org/doc/>

³Documentação disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.impute.KNNImputer.html>

3. *Discretização*: os atributos que indicam o tempo que a criança ou adolescente passam com seus pais por dia e o tempo que os pais passam com os filhos, foram divididos faixas de valores, que incluem os intervalos ‘0 a 3’, ‘4 a 7’, ‘8 a 11’, ‘12 a 15’, ‘16 a 19’ e ‘mais de 19’.
4. *Codificação*: foram realizadas duas etapas para transformar os atributos nominais em numéricos, uma etapa para o dado nominal ordinal e outra para o nominal não ordinal. A base de dados possui dados nominais ordinais, os relacionados ao tempo que foram discretizados na fase anterior e também os de escolaridade. Esses atributos foram transformados em numéricos através da codificação ordinal. Já em relação aos dados nominais não ordinais, o conjunto de dados possui apenas um dado desse tipo, ‘Mora com quem?’, sabendo que a maioria dos algoritmos de aprendizado requer dados numéricos, convertemos esse atributo em números através da codificação *Dummy*. Essa codificação é semelhante a *One Hot*, porém, quando aplicada a um atributo categórico com N categorias distintas, a codificação *Dummy* cria $N - 1$ colunas, transformando o atributo nominal em uma matriz binária, usando 0 e 1 como valores. No caso desse atributo, ele possui 17 categorias que foram transformadas em 16 novas colunas.
5. *Balanceamento*: tendo em vista que o banco de dados possuía 314 e 63 indivíduos pertencentes às classes ‘baixa’ e ‘alta’, respectivamente, optou-se por balancear as classes com o objetivo de evitar que os algoritmos aprendam melhor a classe majoritária e, com isso, a precisão da classe minoritária seja prejudicada. Foi utilizado o método de balanceamento conhecido como *oversampling*, com a técnica SMOTE, gerando novas instâncias de forma sintética para a classe minoritária, aproximando-o ao número de instâncias da classe majoritária.

4.3. Algoritmos de Aprendizagem

Esta pesquisa empregou cinco algoritmos distintos, sendo dois deles técnicas individuais e os outros três pertencentes ao tipo *ensemble*. A seguir, são apresentados os algoritmos e seus respectivos métodos da biblioteca *Scikit-learn*⁴ utilizados: Árvore de Decisão (*DecisionTreeClassifier*), *Random Forest* (*RandomForestClassifier*), *XGBoost* (*XGBClassifier*) e Rede Neural (*MLPClassifier*).

Esses algoritmos foram implementados e avaliados para comparação em *Python*, utilizando os parâmetros *default*. Para avaliar a capacidade de generalização do modelo, utilizamos uma validação cruzada de 10 dobras. Todas as técnicas de pré-processamento, algoritmos de aprendizado de máquina e gráficos foram realizados no *Python 3.10.9*.

4.4. Métricas de Avaliação

As métricas utilizadas para avaliar a qualidade dos modelos foram: Sensibilidade ($S = \frac{VP}{VP+FN}$), Precisão ($P = \frac{VP}{VP+FP}$) e *F1-Score* ($F1 = \frac{2 \times \text{Precisão} \times \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}}$); onde Verdadeiro Positivo (VP), Falso Positivo (FP) e Falso Negativo (FN).

5. Resultados e Discussões

A Figura 3 apresenta os resultados dos classificadores utilizados neste trabalho: Árvore de Decisão, Rede Neural, *Bagging*, *XGBoost* e *Random Forest*, juntamente com suas pontuações nas métricas de precisão, sensibilidade, *F1-Score*, destacando as pontuações de cada classe para cada algoritmo avaliado.

⁴Documentação disponível em: https://scikit-learn.org/stable/getting_started.html

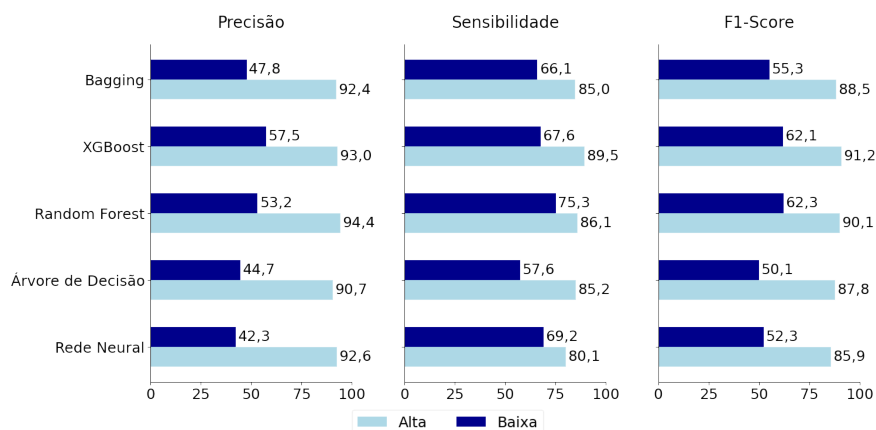


Figura 3: Avaliação da qualidade dos cinco algoritmos de aprendizado.

Considerando todos os classificadores investigados, é possível observar que dentre eles, os métodos *ensembles* *XGBoost*, *Bagging*, e *Random Forest* obtiveram um desempenho em geral melhor, com um valor de F1-score acima de 88% e uma sensibilidade acima de 66%. No entanto, o valor da precisão ficou baixo para todos os algoritmos investigados para a classe ‘alta’. Uma precisão baixa para a classe ‘alta’ indica uma alta taxa de falsos positivos para esta classe. Ou seja, há pessoas da base de dados que foram classificadas como ‘depressão alta’ tendo depressão ‘baixa’. Estes casos precisam de uma maior investigação a fim de investigar se realmente estes casos são um falso positivo ou se estes indivíduos foram rotulados como tendo depressão baixa, mas têm comportamento de depressão alta.

No entanto, na detecção de alta sintomatologia de depressão, uma alta sensibilidade é fundamental para garantir intervenção e suporte adequados aos indivíduos afetados. Isso significa que o modelo é capaz de identificar a maioria dos casos verdadeiros de depressão grave entre todos os casos positivos, minimizando assim os falsos negativos. Ressalta-se aqui que a alta taxa de falsos positivos pode ser um problema para os pais de uma criança ou adolescente que tenha depressão baixa e o sistema o classifique como tendo alta depressão. A preocupação dos pais pode ser intensificada. No entanto, acreditamos que o caso do falso negativo é pior. Ou seja, o indivíduo ter depressão alta e o sistema indicar o contrário.

Assim, identificar corretamente estes casos de depressão alta permite um tratamento precoce e eficaz, incluindo uma triagem mais eficiente, encaminhamento para serviços especializados e acompanhamento mais próximo durante o tratamento. Essas medidas não apenas ajudam a evitar complicações adicionais, mas também melhoram os resultados do tratamento.

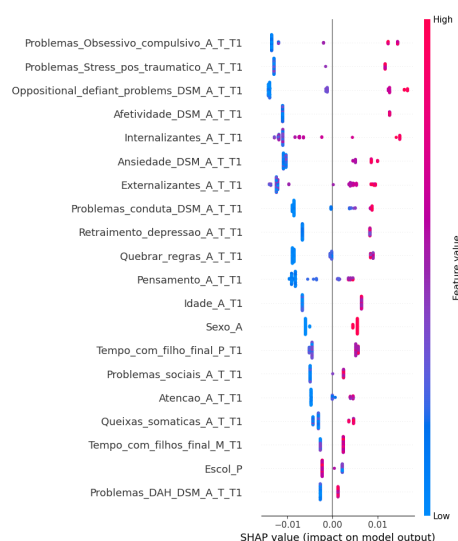
Levando isso em consideração, o melhor classificador considerado neste trabalho é aquele que possui valores altos na métrica de sensibilidade, nesse caso, o classificador *RF* destacou-se com valores da classe ‘Alta’ igual a 75,32%.

Portanto, uma alta sensibilidade na detecção de ambos os casos é fundamental para garantir que esses indivíduos recebam o tratamento adequado e proporcional às suas necessidades específicas. Essa abordagem incentiva uma gestão mais eficiente da saúde mental, priorizando o bem-estar e o tratamento adequado dos indivíduos afetados.

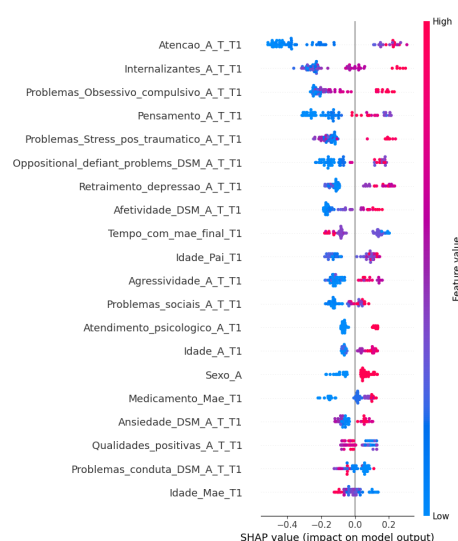
Em resumo, os métodos *ensembles* obtiveram os melhores resultados. O *Random Forest* é uma escolha eficaz para ambas as categorias de sintomatologia, enquanto o *XGBoost* pode ser preferível para a detecção de alta sintomatologia devido à sua sensibilidade ligeiramente superior. Essa metodologia é vantajosa por reduzir variância e bias, compensar deficiências individuais dos modelos e combinar decisões independentes para previsões mais robustas.

Sabendo que os métodos *ensembles* não são interpretáveis e que neste contexto a interpretabilidade é relevante, utilizamos a técnica SHAP para compreender melhor os modelos gerados.

A Figura 4 apresenta um *Summary Plot* do SHAP para os modelos com *Random Forest* (Figura 4a) e *XGBoost* (Figura 4b). Estes gráficos destacam os atributos mais importantes na previsão da sintomatologia da depressão, ordenados por importância. Utilizando os *SHAP values*, eles mostram o impacto dos atributos na predição. Cada ponto na linha de um atributo representa um indivíduo afetado por esse atributo. Se um ponto está à esquerda do eixo central, o atributo contribuiu para a classificação como ‘depressão baixa’; se está à direita, contribuiu para ‘depressão alta’. Quanto mais afastado do eixo central, mais significativo é o impacto do atributo. No entanto, é importante ressaltar que a influência de um atributo isoladamente não determina a classe do indivíduo. A cor dos pontos indica o valor do atributo na instância, sendo vermelho para valores altos e azul para valores baixos.



(a) *Summary Plot* do SHAP para o modelo *Random Forest*.



(b) *Summary Plot* do SHAP para o modelo *XGBoost*.

Figura 4: Impacto dos atributos na detecção de depressão nos modelos.

Após a utilização do SHAP, foram identificados os atributos mais relevantes associados à depressão em jovens de 10 a 16 anos. Dentre eles, nos dois modelos que obtiveram os melhores resultados, *Random Forest* e *XGBoost*. Observando os gráficos, percebe-se que os atributos referentes a problemas afetivos, ansiedade, atenção, idade, internalizantes, oposição e desafio, pensamento, conduta, obsessivo compulsivo, estresse pós traumático, sociais, reatamento e depressão, além do sexo do indivíduo, apresentam importância destacada por ambos os modelos.

Apresentamos alguns atributos mais ressaltados pela literatura, comprovando a eficácia dos modelos:

- *Afetividade*: surge como um elemento relevante, demonstrando um aumento nos afetos negativos e uma diminuição nos afetos positivos. Em termos gerais, os afetos negativos refletem o nível de estados emocionais adversos, como medo, tristeza, raiva e culpa [Watson e et.al, 1995].
- *Ansiedade*: os sintomas frequentemente se sobrepõem aos da depressão. Segundo a OPAS (2023)⁵, muitas pessoas com esse transtorno também sofrem com sintomas de ansiedade.
- *Problemas de atenção*: também emergem como sintomas diagnósticos da depressão, uma vez que a falta de concentração interfere no desempenho acadêmico [Baptista et al., 2006].
- *Problemas sociais*: indivíduos que enfrentam episódios depressivos frequentemente manifestam dificuldades sociais como um dos principais sintomas, OPAS (2023) ⁵.
- *Sexo*: torna-se evidente que o gênero é um dos fatores de destaque. Conforme destacado por Albert [2015], mulheres tendem a apresentar uma prevalência maior de depressão em relação aos homens, influenciadas por uma complexa interação de fatores biológicos, socioeconômicos e psicossociais.
- *Stress pós traumático*: aqueles que experimentaram um evento traumático, como o estresse pós-traumático, têm uma probabilidade aumentada de desenvolver e serem diagnosticados com depressão [Berlim et al., 2003].
- *Transtornos internalizantes*: caracterizados por manifestações voltadas para o próprio indivíduo, como tristeza, retraimento, queixas somáticas e medo, estão frequentemente associados aos transtornos de humor, como a depressão [Hess e et.al, 2013].
- *Transtorno obsessivo-compulsivo*: é comum observar uma comorbidade entre o transtorno obsessivo-compulsivo (TOC) e a depressão, uma vez que os sintomas dessas condições se entrelaçam [IPP, 2024].

6. Considerações Finais

Este estudo investigou o uso de técnicas de aprendizado de máquina, juntamente com a análise do SHAP (*SHapley Additive exPlanations*), para analisar o perfil de crianças e adolescentes com depressão com idade de 10 a 16 anos. Utilizando um conjunto de dados fornecido pela Universidade Federal de Minas Gerais e cinco algoritmos de classificação, incluindo métodos individuais e *ensemble*, avaliamos o desempenho de cada modelo na detecção de alta e baixa sintomatologia de depressão.

Os resultados confirmaram o que a literatura tem mostrado: que os métodos *ensemble* geralmente produzem soluções mais precisas do que um único modelo. Mas por outro lado, as suas soluções são não interpretáveis.

Desenvolver modelos precisos e interpretáveis para a identificação precoce da depressão em jovens pode melhorar resultados clínicos e promover bem-estar mental. Integrar métodos computacionais avançados com práticas clínicas tradicionais, juntamente com interpretabilidade fornecida pelo SHAP, pode ser ferramenta valiosa para profissionais de saúde e pesquisadores na identificação e tratamento precoce da depressão em adolescentes.

Os resultados mostraram que os métodos *ensemble*, especialmente *Random Forest* e *XGBoost*, apresentaram desempenho eficaz em ambas as categorias de sintomatologia.

⁵Disponível em: <https://www.paho.org/pt/topicos/depressao>

Além disso, a análise do SHAP permitiu uma compreensão mais profunda dos principais fatores contribuintes para o diagnóstico de depressão, ajudando a identificar padrões e insights importantes para intervenção e tratamento adequado.

No entanto, é importante destacar que ainda há espaço para melhorias e refinamentos. Para pesquisas futuras, recomenda-se explorar outras técnicas *ensembles*, e investigar a eficácia de diferentes conjuntos de características na identificação da depressão em jovens. Além disso, sugere-se a investigação de métodos de extração de interpretabilidade específicos dos métodos *ensemble*, diferentemente do SHAP que é um método agnóstico, ou seja, extrai conhecimento de qualquer modelo de aprendizado.

Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico do Brasil (CNPq – Código: 311573/2022-3), à Pontifícia Universidade Católica de Minas Gerais (PUC-Minas - Código: PIBIC/PIBIT-2023/29487), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES - Grant PROAP 88887.842 889/2023-00 - PUC/MG, Grant PDPG 88887.708960/2022-00 - PUC/MG - Informática and Finance Code 001), à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG – Códigos: PCE-00437-24, APQ-03076-18 e APQ-05058-23). O trabalho foi desenvolvido na Pontifícia Universidade Católica de Minas Gerais, PUC Minas, no laboratório de Inteligência Computacional Aplicada – LICAP.

Referências

- Adeniji, O. D. e et.al (2022). An improved bagging ensemble in predicting mental disorder using hybridized random forest-artificial neural network model. *Informatica*.
- Albert, P. (2015). Why is depression more prevalent in women? *Journal of psychiatry neuroscience : JPN*.
- APA (2014). *DSM-5: Manual diagnóstico e estatístico de transtornos mentais*.
- Arrieta, A. B. e et.al (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Infor. fus*.
- Balázs, J. e et.al (2013). Adolescent subthreshold-depression and anxiety: Psychopathology, functional impairment and increased suicide risk. *Journal of child psychology and psychiatry*.
- Balsters, M. e et.al (2010). Measuring potential cues for depression in adolescents.
- Baptista, N., de Lima, M., e et.al (2006). Sintomatologia depressiva, atenção sustentada e desempenho escolar em estudantes do ensino médio. *Psicologia Escolar e Educacional*.
- Berlim, M. T., Perizzolo, J., e Fleck, M. P. A. (2003). Transtorno de estresse pós-traumático e depressão maior. *Brazilian Journal of Psychiatry*, 25.
- Bhakta, I. e et.al (2016). Prediction of depression among senior citizens using machine learning classifiers. *International Journal of Computer Applications*.
- Breiman, L. (1996). Bagging predictors. *Machine learning*.

- ElShawi, R. e Sherif, e. (2021). Interpretability in healthcare: A comparative study of local machine learning interpretability techniques. *Computational Intelligence*.
- Haque, U. M. e et.al (2021). Detection of child depression using machine learning methods. *PLoS one*.
- Hess, A. R. B. e et.al (2013). Sintomas internalizantes na adolescência e as relações familiares: uma revisão sistemática da literatura. *Psico-USF*.
- IPP (2024). Transtorno obsessivo-compulsivo (TOC) e depressão: qual a relação? *Instituto de Psiquiatria do Paraná Blog*.
- Kovacs, M. e Staff, M. (2003). Children's depression inventory (cdi): Technical manual update. multi-health systems. Inc.: North Tonawanda, NY, USA.
- Lu, S.-C., Swisher, C. L., Chung, C., Jaffray, D., e Sidey-Gibbons, C. (2023). On the importance of interpretable machine learning predictions to inform clinical decision making in oncology. *Front. Oncol.*, 13:1129380.
- Lundberg, S. M. e Lee, S.-I. (2017). A unified approach to interpreting model predictions. Curran Associates, Inc.
- Mienye, I. D. e et.al (2020). Improved predictive sparse decomposition method with denset for prediction of lung cancer. *Int. J. Comput.*
- Mienye, I. D. e Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access*.
- Office, W. H. (2022). The blueprint for an ai bill of rights: Making automated systems work for the american people. disponível em: <https://www.whitehouse.gov/wp-content/uploads/2022/10/blueprint-for-an-ai-bill-of-rights.pdf>.
- Pinucci, I., Pasquini, M., e Longhi, E. V. (2023). Persistent depressive disorder (dysthymia) and recurrent unipolar major depressive disorder. Springer.
- Sunmoo, Y. e et.al (2014). Using a data mining approach to discover behavior correlates of chronic disease: A case study of depression. *Studies in Health Technology and Informatics*.
- Tate, A. E. e et.al (2020). Predicting mental health problems in adolescence using machine learning techniques. *PLOS ONE*.
- Watson, D. e et.al (1995). Testing a tripartite model: I. evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *J. of abno. psych.*
- WHO (2017). Depression and other common mental disorders: global health estimates. Technical documents.
- Witten, I. H. e et.al (2016). *Data Mining: practical machine learning tools and techniques with Java implementations*. Morgan Kaufmann.
- Zeng, W. e Davoodi, e. (2020). Explainable DRC hotspot prediction with random forest and SHAP tree explainer. IEEE.