

Avaliação da aplicação de ensemble learning na predição do percentual de gordura corporal em adolescentes

Assessment of ensemble learning application in predicting body fat percentage in adolescents

Evaluación de la aplicación de ensemble learning en la predicción del porcentaje de grasa corporal en adolescentes

Danilo José dos Santos Costa¹, Ewaldo Eder Carvalho Santana²

1 Bacharel, Laboratório de processamento e aquisição de sinais, Universidade Estadual do Maranhão, São Luís (MA), Brasil.

2 Doutor, Laboratório de processamento e aquisição de sinais, Universidade Estadual do Maranhão, São Luís (MA), Brasil.

Autor correspondente: Danilo José dos Santos Costa

E-mail: danilodalessandro08@gmail.com

Resumo

Objetivo: O presente estudo teve como objetivo estimar o percentual de gordura corporal em adolescentes de São Luís/MA utilizando técnicas de aprendizado de máquina.

Método: Foram usadas técnicas de ensemble com os algoritmos: *Stacking*, *Bagging* e *adaboost*. **Resultado:** Os resultados revelaram que o modelo de *Stacking* apresentou o melhor desempenho, com menor erro quadrático médio (MSE) e maior coeficiente de determinação (R^2), indicando sua eficácia na explicação da variabilidade dos dados.

Conclusão: O *Stacking* é o algoritmo mais indicado para predição de índice de gordura corporal em adolescentes, pois se adaptou muito bem aos dados devido à sua robustez, redução de *overfitting* e alto poder interpretativo.

Descritores: Transição Nutricional; Gordura Corporal; Ensemble Learning

Abstract

Objective: The present study aimed to estimate the percentage of body fat in adolescents from São Luís/MA using machine learning techniques. **Method:** Ensemble techniques with the algorithms Stacking, Bagging, and AdaBoost were employed. **Results:** The findings revealed that the Stacking model demonstrated the best performance, with lower mean squared error (MSE) and higher coefficient of determination (R^2), indicating its effectiveness in explaining the data variability. **Conclusion:** Stacking is the most suitable algorithm for predicting body fat index in adolescents as it adapted well to the data due to its robustness, reduction of overfitting, and high interpretative power.

Keywords: Nutritional Transition; Body Fat; Ensemble Learning

Resumen

Objetivo: El presente estudio tuvo como objetivo estimar el porcentaje de grasa corporal en adolescentes de São Luís/MA utilizando técnicas de aprendizaje automático. **Método:** Se emplearon técnicas de conjunto con los algoritmos Stacking, Bagging y AdaBoost. **Resultados:** Los resultados revelaron que el modelo de Stacking mostró el mejor rendimiento, con un menor error cuadrático medio (MSE) y un mayor coeficiente de determinación (R^2), lo que indica su efectividad en explicar la variabilidad de los datos. **Conclusión:** El Stacking es el algoritmo más adecuado para predecir el índice de grasa corporal en adolescentes, ya que se adaptó bien a los datos debido a su robustez, reducción de sobreajuste y alto poder interpretativo.

Descriptores: Transición Nutricional; Grasa Corporal; Aprendizaje en Conjunto

Introdução

Em 2015, a população mundial de adolescentes era de 1,2 bilhões, representando cerca de 16% do total. Aproximadamente 90% desses adolescentes vivem em países de baixa renda e pouco desenvolvidos, onde são encontrados desafios significativos para se ter uma educação alimentar plena ⁽¹⁾. As diferenças no estado nutricional de um indivíduo podem ser decorrentes de fatores ambientais, genéticos ou uma combinação de ambos. Além disso, existe uma correlação entre o excesso de peso dos pais e o excesso de peso dos filhos, mostrando características de hereditariedade em problemas de obesidade ⁽²⁾.

A adolescência, dos 10 aos 19 anos, é um período de mudanças corporais, incluindo crescimento e possível desenvolvimento da obesidade ⁽¹⁾. A obesidade, causada por excesso de gordura corporal, pode afetar a saúde, pois tem relação direta com doenças como hipertensão e doenças cardiovasculares⁽³⁾. No Brasil, as disparidades sociais refletem-se nos números de obesidade, com regiões mais desenvolvidas geralmente tendo mais casos. Considerada uma epidemia global, a obesidade é um problema de saúde pública, especialmente durante a adolescência, fase de rápido crescimento ⁽⁴⁾.

A análise da composição corporal destaca a medição do percentual de gordura corporal ⁽⁵⁾. Níveis elevados podem causar resistência à insulina, exigindo acompanhamento rigoroso, pois o descontrole pode prejudicar a saúde ⁽¹⁾. Existem técnicas como a *DEXA (Dual X-ray Absorptiometry)* e a bioimpedância elétrica para avaliação, sendo a *DEXA* mais precisa, apesar dos custos e da radiação necessária ⁽⁶⁾. A bioimpedância elétrica também é utilizada, aplicando correntes elétricas de baixa intensidade ⁽⁷⁾. Embora a *DEXA* seja a preferida, seu alto custo e exigência de especialização são desafios relevantes ⁽¹⁾.

O IMC (Índice de Massa Corporal) é um indicador antropométrico simples e acessível, mas sua eficácia é questionada devido à falta de diferenciação entre tecido adiposo e massa magra. O desenvolvimento de métodos nutricionais mais precisos, especialmente para adolescentes, é crucial. O uso de técnicas de aprendizado de máquina

pode oferecer diagnósticos rápidos e não invasivos para melhorar a avaliação nutricional⁽¹⁾.

Para este trabalho foram empregadas técnicas de *ensemble learning*, uma abordagem robusta em aprendizado de máquina que combina múltiplos modelos de previsão para melhorar a precisão e robustez das previsões. Ao integrar as previsões de diferentes modelos, como árvores de decisão, regressão logística e redes neurais, o *ensemble learning* reduz o viés e a variância do modelo final, resultando em uma melhor generalização para novos dados. Suas vantagens incluem maior precisão, melhor capacidade de generalização e maior robustez em comparação com modelos individuais^(9,10). Para realizar a predição de índice de gordura corporal (IGC) em adolescentes.

Foram exploradas as técnicas de *Stacking*, *Bagging* e *AdaBoost* para regressão. O objetivo é comparar essas abordagens para identificar a mais eficaz na estimativa do percentual de gordura corporal em adolescentes. Além disso, busca-se fornecer uma solução computacional que possa auxiliar no combate à obesidade, agilizando o diagnóstico e facilitando o tratamento. Com base em métricas de desempenho como erro médio absoluto e erro quadrático médio, foi possível determinar o método mais preciso e eficiente, visando intervenções rápidas e eficazes na saúde dos jovens.

Métodos

Nesta seção, são detalhados os materiais utilizados no método desenvolvido, assim como as configurações e os procedimentos empregados em sua avaliação experimental.

Base de dados

Os atributos selecionados para este estudo, como idade, massa corporal, estatura e gênero, são amplamente reconhecidos na literatura e têm diversas aplicações, especialmente na avaliação nutricional e de saúde de adolescentes. Recomendados pela OMS, tais dados fornecem *insights* cruciais para o processo. Índices antropométricos,

como circunferência da cintura, são essenciais para avaliar gordura corporal e risco cardiovascular. Circunferências do braço, quadril e panturrilha são úteis para avaliar perda de massa magra. Esses atributos são fundamentais e têm demonstrado resultados promissores em estudos anteriores ⁽¹⁾.

A base de dados consiste em 772 linhas e nove atributos. Os dados foram coletados por pesquisadores usando como voluntários, alunos da rede pública de São Luís do Maranhão, com idades entre 10 e 19 anos, selecionados de forma não probabilística em 2018. A amostra foi calculada com base na prevalência de excesso de peso em adolescentes (20,5%), uma prevalência sugerida do desfecho de 26,9%, um erro tolerável de 5% (erro tipo I) e um poder do teste de 90% (erro tipo II), com um adicional de 10% para perdas ou recusas, resultando em uma amostra mínima de 513 adolescentes a base contém os seguintes atributos: sexo, peso, altura, circunferência da cintura, idade, medida do quadril, medida do braço e medida da panturrilha. Durante a coleta de dados, foram aplicadas regras de exclusão para indivíduos com os seguintes critérios ⁽¹⁾:

- Adolescentes gestantes, amamentando ou que fazem uso de anticoncepcional;
- Uso de medicamentos que alteram o estado nutricional;
- Incapacidade física que impossibilita ou compromete as medidas antropométricas.
- Não concordância dos responsáveis ou participantes;
- Ausência na coleta de dados.

Cada medida foi registrada por um único pesquisador utilizando o mesmo instrumento calibrado. As medições foram realizadas em duplicata para permitir a obtenção de uma média, a qual foi então incorporada à base de dados. Para determinar o percentual de gordura corporal, foi empregada a técnica de bioimpedância tetrapolar, conforme detalhado na tabela abaixo:

Tabela 1 – Valores de referência para classificação do %GC ⁽¹⁾

Classificação/Gênero	Masc. (%)	Fem. (%)
Baixo	6 – 10	12 - 15
Otimo	10.1 – 20	15.1 - 25
Moderadamente Alto	20.1 – 25	25.1 - 30



CBIS'24

XX Congresso Brasileiro de Informática em Saúde

08/10 a 11/10 de 2024 - Belo Horizonte/MG - Brasil

Alto	25.1 – 31	30.1 - 35
Muito Alto	>31	>35

Para aplicar os algoritmos selecionados, foi utilizada a plataforma *Google Colab* com a linguagem de programação *Python*, empregando a biblioteca *Scikit-Learn*. Essas ferramentas foram fundamentais por oferecerem grande suporte ao trabalho com dados.

Este estudo foi aprovado pelo comitê de ética em pesquisa da Universidade Federal do Maranhão, conforme parecer CAEE: 83206118.1.0000.5087. Os voluntários foram incluídos somente após receberem informações detalhadas sobre o estudo e assinarem o termo de consentimento livre e esclarecido. A confidencialidade dos dados foi garantida, e os voluntários poderiam desistir de participar do estudo ⁽¹⁾.

Método Desenvolvido

O método proposto neste estudo contém três fases a primeira é a análise da base de dados, a segunda é a aplicação dos algoritmos e a terceira é a extração de métricas.

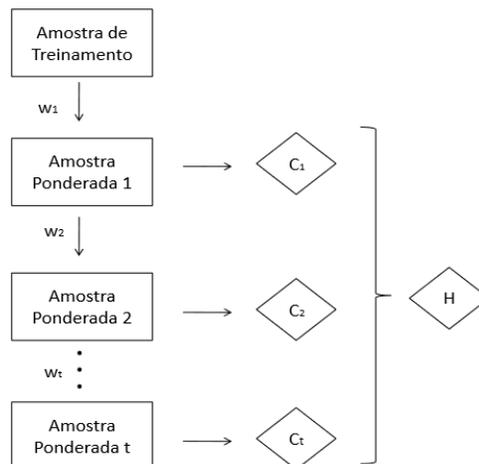
Na primeira fase para compreender melhor a base de dados deste estudo, foi realizada uma análise exploratória. O principal objetivo foi entender as particularidades e características específicas dos dados. A análise mostra que há 9 variáveis, sem dados faltantes ou linhas duplicadas. Isso proporciona uma base sólida para avançar na análise e extrair *insights* valiosos.

Para avaliar a normalidade dos dados, foi empregado o teste *Kolmogorov-Smirnov*, especialmente para a comparação entre os grupos de amostras independentes. No caso de uma distribuição normal das variáveis, utilizou-se o teste "t" de *Student*, enquanto o teste de *Mann-Whitney U* foi aplicado para as variáveis que não apresentaram distribuição normal. Os resultados foram considerados estatisticamente significantes para $p < 0,05$, com 'p' referindo-se ao valor de probabilidade ⁽¹⁾. Além disso, outras informações gerais foram observadas: dos 772 indivíduos, 501 são do sexo masculino e 271 do sexo feminino. Em relação à classificação do percentual de gordura corporal, constatou-se que 539 indivíduos apresentaram valores acima do ideal.

Na segunda fase foram escolhidos três algoritmos com potencial para entregar um resultado satisfatório, na aplicação dos algoritmos selecionados, empregamos o método de validação cruzada *k-fold*, começando com 3 e posteriormente, ampliando para 5 *folds*. Essa abordagem estratégica divide os dados em conjuntos distintos para treinamento e teste, garantindo uma avaliação robusta do desempenho do modelo. Com *k-fold* 3 e 5, cada parte dos dados é utilizada como conjunto de teste em rotação, aumentando a confiabilidade dos resultados e minimizando viés.

O primeiro algoritmo usado foi o *bagging* que é como criar várias cópias de um modelo de previsão, usando o mesmo algoritmo, para reunir opiniões diferentes. Essas opiniões são então combinadas para chegar a uma decisão final. Em vez de depender apenas de uma previsão, várias versões do modelo dão seu palpite e por meio de uma votação, a opinião mais comum é escolhida. Isso cria uma espécie de consenso entre os modelos. Essas diferentes versões são criadas para aprender juntas e melhorar o desempenho geral do modelo. Os resultados mostram que essa abordagem pode aumentar significativamente a precisão das previsões ⁽⁸⁾.

Figura 1 - Funcionamento do algoritmo *bagging* ⁽¹⁴⁾



Matematicamente o *bagging* é um algoritmo onde dado um conjunto de dados de treinamento D com N amostras, seleciona aleatoriamente N amostras com substituição para criar T conjuntos de dados de treinamento D_i , onde $i=1, 2, \dots, T$. onde para cada conjunto D_i é treinado um modelo de $h_i(x)$ onde para previsão feita para um novo exemplo

x , a previsão é determinada pela média das previsões dos modelos individuais. Para um

conjunto de T modelos ^(9,10), a previsão final H(x) é calculada como:

$$H(x) = \sum_{i=1}^T h_i(x) \tag{1}$$

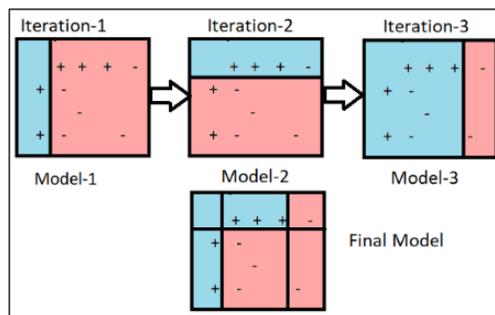
Onde:

- H(x) é a previsão final para a entrada x.
- h_i(x) é a previsão do i-ésimo modelo para a entrada x.

No algoritmo de *bagging* aplicado neste trabalho foi utilizado o parâmetro *random_state=42*, com ele se garante que os resultados do modelo sejam reproduzíveis e consistentes entre diferentes execuções do código. Com isso se controla aleatoriedade durante o processo de amostragem dos dados. O valor específico de 42 é arbitrário e poderia ter sido qualquer outro número. O importante é que seja fixado para que possamos reproduzir os resultados exatos sempre que necessário. O *bagging* não apenas reduz a variação, mas também diminui o erro médio quadrado, o que é ótimo para tornar as previsões mais confiáveis ⁽⁸⁾.

O segundo algoritmo utilizado foi o *AdaBoost*. A ideia central do *AdaBoost* é treinar uma sequência de classificadores fracos de forma iterativa, dando mais peso às amostras mal classificadas em cada iteração subsequente. Isso permite que o algoritmo se concentre nas amostras mais difíceis de classificar, melhorando gradualmente o desempenho do modelo final ^(11,12).

Figura 2 - Funcionamento do algoritmo *Adaboost* ⁽¹⁵⁾





CBIS'24

XX Congresso Brasileiro de Informática em Saúde

08/10 a 11/10 de 2024 - Belo Horizonte/MG - Brasil



CBIS'24

XX Congresso Brasileiro de Informática em Saúde

08/10 a 11/10 de 2024 - Belo Horizonte/MG - Brasil

Suponha que temos um conjunto de dados de treinamento $D=\{(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)\}$ onde x_i é o vetor de características da i -ésima amostra e y_i é o rótulo correspondente. O objetivo é aprender uma função de regressão $F(x)$ que minimize o erro médio quadrático ^(10,11):

(2)

Durante cada iteração, o *AdaBoost* ajusta um classificador fraco à distribuição atualizada das amostras, onde a importância de cada amostra é ajustada com base em sua dificuldade de classificação nas iterações anteriores. No final, os classificadores fracos são combinados de forma ponderada para formar um classificador forte, que é capaz de fazer previsões precisas sobre novos dados ⁽¹¹⁾.

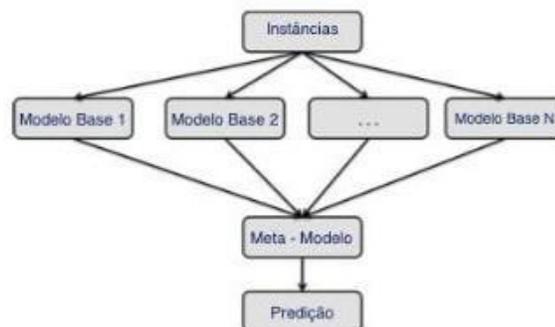
Na implementação deste algoritmo, uma árvore de decisão foi utilizada como o regressor base, funcionando como o estimador fraco padrão. No entanto, é importante destacar que o algoritmo oferece a flexibilidade de ser configurado com outros estimadores, proporcionando a adaptação a diferentes cenários e necessidades específicas. Essa característica confere ao algoritmo uma versatilidade que pode resultar em desempenho otimizado, especialmente quando combinado com modelos de aprendizado diversificados.

O terceiro algoritmo, foi *stacking* que é uma técnica que combina as previsões de vários modelos de aprendizado de máquina para melhorar o desempenho preditivo. Ao contrário do *bagging* e *boosting*, que usam métodos paralelos para combinar os modelos,

o *stacking* utiliza uma abordagem de meta-aprendizagem, onde um modelo final, chamado de meta-modelo, é treinado para fazer previsões com base nas saídas dos modelos base ⁽⁹⁾.

O algoritmo de *stacking* seleciona vários modelos de aprendizado de máquina, treinados individualmente com o conjunto de dados original. Após fazerem previsões no conjunto de validação, essas previsões são usadas como *features* adicionais, combinadas com as originais, formando um novo conjunto de dados de nível 2. Um meta-modelo é então treinado com esse conjunto expandido, recebendo as previsões dos modelos base e a variável alvo original. Seu objetivo é combinar essas previsões de forma ponderada para otimizar o desempenho preditivo geral. O meta-modelo, uma vez treinado, faz previsões sobre novos dados, combinando as previsões dos modelos base para produzir a previsão final ^(9,13).

Figura 3 - Funcionamento do algoritmo *Stacking* ⁽¹⁶⁾



Naturalmente o *stacking* usa vários estimadores para este algoritmo foram utilizados a regressão linear árvore de decisão e o *random forest*, Esses estimadores são escolhidos para representar uma variedade de abordagens de modelagem de dados, Podemos fornecer esses estimadores como entrada para o modelo técnico dando a possibilidade para que ele possa combinar suas previsões de maneira eficaz cada estimador contribui para um modelo final com uma perspectiva diferente fazendo com que a tomada de decisão seja mais robusta ao seu final.

Resultados e Discussões

Neste estudo, foram avaliados três modelos de regressão: *Bagging Regressor*, *Stacking Regressor* e *AdaBoost Regressor* em termos de desempenho preditivo utilizando validação cruzada. Foram utilizadas métricas comuns de avaliação de regressão, tais como Erro Médio Quadrático (MSE), Erro Médio Absoluto (MAE), coeficiente de determinação (R^2) e Raiz do Erro Médio Quadrático (RMSE). A seguir, são apresentados os resultados das métricas usadas para avaliar o desempenho de cada algoritmo. Essas métricas serão exibidas na tabela 1 e 2, juntamente com os resultados de cada *k-fold* aplicado.

Tabela 2 – Métricas para *k-fold* = 3

Algoritmo	MSE	MAE	R^2	RMSE
Bagging	0.16	0.28	0.21	0.40
Stacking	0.15	0.32	0.24	0.39
Adaboost	0.16	0.35	0.20	0.40

Tabela 3 – Métricas para *k-fold* = 5

Algoritmo	MSE	MAE	R^2	RMSE
Bagging	0.16	0.28	0.19	0.40
Stacking	0.15	0.31	0.25	0.39
Adaboost	0.15	0.34	0.22	0.39

A validação cruzada foi realizada com diferentes números de folds (3 e 5), onde os dados foram divididos em conjuntos de treino e teste de forma iterativa. Em cada iteração, os modelos foram treinados nos conjuntos de treino e avaliados nos conjuntos de teste. Os resultados indicam que, em média, o *Stacking Regressor* obteve o menor MSE e RMSE em comparação com os outros modelos, tanto para validação cruzada com 3 folds quanto com 5 folds. Isso indica que a combinação de diferentes tipos de regressores no *Stacking Regressor* levou a uma melhor capacidade de generalização e predição em relação aos outros modelos.

No entanto, é importante observar que o *Stacking Regressor* teve um desempenho um pouco inferior em termos de MAE em comparação com o *Bagging Regressor* em ambas as configurações de validação cruzada. Isso pode ser atribuído à natureza do

MAE, que é menos sensível a valores discrepantes em comparação com o MSE, e pode refletir diferenças na distribuição dos erros de predição entre os modelos.

O *AdaBoost Regressor*, por sua vez, apresentou resultados intermediários em termos de todas as métricas avaliadas. Embora tenha mostrado um desempenho ligeiramente melhor do que o *Bagging Regressor* em termos de R^2 , ficou atrás do *Stacking Regressor* em relação ao MSE e RMSE.

Para este trabalho também foi utilizado o teste de Friedman utilizando a soma dos quadrados das diferenças entre as classificações de cada modelo para cada métrica e cada configuração de validação cruzada. o resultado do teste de Friedman é um valor de teste de aproximadamente 0.0952 para $\alpha = 0.05$. Comparando este valor com o valor crítico da distribuição qui-quadrado com 2 graus de liberdade ($k - 1$, onde k é o número de modelos, que é 3), o valor crítico é aproximadamente 5.991.

Como o valor calculado do teste de Friedman (0.0952) é menor que o valor crítico (5.991), não se pode rejeitar a hipótese nula. E também a possibilidade de que os algoritmos usados tenham estatisticamente desempenhos similares.

Conclusão

Com base nos resultados apresentados, se identificou que o *Stacking Regressor* se destacou como o melhor algoritmo para prever o índice de gordura corporal em adolescentes. Este modelo apresentou as métricas mais favoráveis, incluindo o menor MSE e RMSE em ambas as configurações de validação cruzada, bem como um R^2 mais alto, indicando melhor capacidade de explicar a variabilidade nos dados.

A precisão e consistência demonstradas pelo *Stacking Regressor* sugerem que ele pode ser uma ferramenta valiosa para profissionais de saúde na identificação precoce de adolescentes em risco de obesidade e complicações associadas. Com uma estimativa mais precisa do índice de gordura corporal, os profissionais de saúde podem implementar intervenções preventivas e personalizadas, como programas de educação nutricional e atividade física, direcionados a indivíduos específicos.

O diagnóstico preciso do IGC em adolescentes é crucial para avaliar a saúde e bem-estar dessa população em rápido desenvolvimento. Neste estudo, foram exploradas

J. Health Inform. 2024, Vol. 16 Especial - ISSN: 2175-4411 - jhi.sbis.org.br

DOI: 10.59681/2175-4411.v16.iEspecial.2024.1282

várias métricas de avaliação de modelos de regressão para prever o IGC, utilizando métodos como o Erro Médio Quadrático (MSE), o Erro Médio Absoluto (MAE), o coeficiente de determinação (R^2) e a Raiz do Erro Médio Quadrático (RMSE). A análise dessas métricas forneceu *insights* valiosos sobre o desempenho dos modelos de regressão testados.

Os resultados destacaram a importância de métodos de avaliação abrangentes, pois cada métrica fornece uma perspectiva única sobre a qualidade das previsões. O MSE e o RMSE são particularmente úteis para avaliar a precisão das previsões, enquanto o MAE oferece uma medida robusta da magnitude dos erros de previsão. Por outro lado, o R^2 fornece uma indicação da variação explicada pelo modelo em relação à variabilidade total dos dados.

Além disso, é preciso olhar para a necessidade de métodos de diagnóstico rápidos e acessíveis para avaliar o IGC em adolescentes. A obesidade e o sobrepeso na adolescência estão associados a uma série de riscos à saúde, incluindo diabetes tipo 2, hipertensão e doenças cardiovasculares. Portanto, é essencial identificar e intervir precocemente para prevenir complicações de saúde a longo prazo.

Os métodos rápidos e acessíveis têm o potencial de facilitar a triagem e o monitoramento do IGC em ambientes clínicos e escolares. Modelos de regressão baseados em dados antropométricos, como medidas de altura, peso e circunferência da cintura, podem fornecer estimativas razoáveis do IGC com um custo relativamente baixo e equipamentos simples. Essa abordagem permite a avaliação em larga escala de adolescentes em diferentes contextos, sem a necessidade de equipamentos sofisticados ou procedimentos invasivos.

Em suma, este estudo destaca o algoritmo *Stacking* Regressor como o mais promissor com base nas métricas de desempenho avaliadas, como erro médio absoluto e erro quadrático médio. Ressalta-se a importância de utilizar métodos rápidos e acessíveis para o diagnóstico do percentual de gordura corporal (IGC) em adolescentes. Ao empregar abordagens abrangentes e computacionalmente eficientes, pode-se melhorar a identificação precoce de indivíduos em risco e promover intervenções eficazes para prevenir problemas de saúde relacionados à obesidade na adolescência.

Referências

1. Sousa Junior CM. Desenvolvimento de um sistema para triagem de adolescentes obesos utilizando variáveis clínicas [Tese]. São Luís: UFMA; 2019. 68 p.
2. Fonseca VM, Sichieri R, Veiga GV. Fatores associados à obesidade em adolescentes. Rev Saúde Pública. 1998;32:541-9.
3. Mattar R, et al. Obesidade e gravidez. Rev Bras Ginecol Obstet. 2009;31:107-110.
4. Wanderley EN, Ferreira VA. Obesidade: uma perspectiva plural. Ciência Saúde Colet. 2010;15:185-194.
5. Rezende FAC, et al. Aplicabilidade do índice de massa corporal na avaliação da gordura corporal. Rev Bras Med Esporte. 2010;16:90-94.
6. Zeballos L, et al. Avaliação da composição corporal total e segmentar de alunos do curso de nutrição pela densitometria por dupla emissão de raios-x. RBONE-Rev Bras Obes Nutr Emagrecimento. 2020;14(89):914-920.
7. Britto EP, Mesquita ET. Bioimpedância elétrica aplicada à insuficiência cardíaca. 2008.
8. Lopes NS. Modelos de classificação de risco de crédito para financiamentos imobiliários: regressão logística, análise discriminante, árvores de decisão, bagging e boosting. 2011.
9. Breiman L. Bagging predictors. Mach Learn. 1996;24:123-140.
10. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001;29(5):1189-1232.
11. Freund Y, Schapire RE. Experiments with a new boosting algorithm. In: International Conference on Machine Learning. 1996. pp. 148-156.
12. Schapire RE. The Boosting Approach to Machine Learning: An Overview. In: Nonlinear Estimation and Classification. Springer, New York, NY; 2012. pp. 149-171.
13. Ting KM, Witten IH. Issues in stacked generalization. J Artif Intell Res. 1999;10:271-289.
14. Oliveira LM. Classificação de dados sensoriais de cafés especiais com resposta multiclasse via Algoritmo Boosting e Bagging. 2016.

15. Pinho CMA, et al. Análise de textos com aplicação de técnicas de inteligência artificial: estudo comparativo para classificação de fuga ao tema em redações. 2021.
16. Oliveira Filho IL. Algoritmo papílio como método de proteção de templates para aumentar a segurança em sistemas de identificação biométricos. 2014.



CBIS'24

XX Congresso Brasileiro de Informática em Saúde

08/10 a 11/10 de 2024 - Belo Horizonte/MG - Brasil



CBIS'24

XX Congresso Brasileiro de Informática em Saúde

08/10 a 11/10 de 2024 - Belo Horizonte/MG - Brasil



CBIS'24

XX Congresso Brasileiro de Informática em Saúde

08/10 a 11/10 de 2024 - Belo Horizonte/MG - Brasil