

Fatores Determinantes do Estresse Acadêmico: Uma Abordagem Baseada em Dados e Aprendizado de Máquina

Factors Determining Academic Stress: A Data-Driven and Machine Learning Approach

Douglas Baldon Correa*

RESUMO

A previsão dos níveis de estresse em estudantes universitários é um desafio significativo, considerando os múltiplos fatores que influenciam seu bem-estar. Neste estudo, foram analisadas variáveis como qualidade do sono, atividade física e parâmetros fisiológicos para desenvolver um modelo preditivo baseado em aprendizado de máquina. Três modelos foram avaliados: Regressão Logística, Random Forest e Rede Neural. Os resultados demonstraram que o modelo Random Forest obteve o melhor desempenho, com uma acurácia superior a 90%, seguido pela Rede Neural. Em contrapartida, a Regressão Logística apresentou um desempenho inferior, com métricas próximas de 55%. Além disso, verificou-se uma forte correlação entre a qualidade do sono e os níveis de estresse, bem como a relação inversa entre a prática de atividade física e a incidência de estresse elevado. Os achados sugerem que modelos de aprendizado de máquina podem ser utilizados para monitoramento e prevenção do estresse acadêmico, auxiliando na implementação de estratégias de mitigação e promoção da saúde mental entre estudantes universitários.

PALAVRAS-CHAVE: aprendizado de máquina; estresse acadêmico; qualidade do sono; atividade física; previsão de estresse.

ABSTRACT

Predicting stress levels in university students is a significant challenge due to the multiple factors influencing their well-being. This study analyzed variables such as sleep quality, physical activity, and physiological parameters to develop a machine learning-based predictive model. Three models were evaluated: logistic regression, random forest, and neural network. The results showed that the random forest model achieved the best performance, with an accuracy above 90%, followed by the neural network. In contrast, logistic regression performed worse, with metrics around 55%. Additionally, a strong correlation was observed between sleep quality and stress levels, as well as an inverse relationship between physical activity and high stress incidence. The findings suggest that machine learning models can be employed for monitoring and preventing academic stress, assisting in the implementation of mitigation strategies and mental health promotion among university students.

KEYWORDS: machine learning; academic stress; sleep quality; physical activity; stress prediction.

1 INTRODUÇÃO

O estresse acadêmico é reconhecido como um fator predominante que afeta negativamente o bem-estar mental dos estudantes universitários. Estudos indicam que o estresse relacionado às demandas acadêmicas pode reduzir a motivação, prejudicar o desempenho acadêmico e aumentar as taxas de evasão escolar (Pascoe; Hetrick; Parker, 2020). Além disso, há uma associação significativa entre a pressão acadêmica e o surgimento de sintomas de ansiedade e depressão entre os alunos (Eisenberg *et al.*, 2007).

*  Pesquisador em Ciência da Computação.  Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brasil.  douglasbaldon@alunos.utfpr.edu.br.  4704625343652623.  0009-0006-0233-8648.

Estudos apontam que carga de trabalho excessiva, restrições de tempo e dificuldades acadêmicas são fatores predominantes no aumento do estresse acadêmico. Uma revisão recente apontou evidências, em 48 estudos, de uma associação entre a pressão acadêmica e desfechos negativos na saúde mental. (Steare *et al.*, 2023). Especificamente, a qualidade do sono tem sido apontada como um mediador crucial nessa relação, onde padrões de sono inadequados podem exacerbar os níveis de estresse e comprometer o bem-estar psicológico (Musshafen *et al.*, 2021).

A prática regular de atividade física surge como uma estratégia eficaz para mitigar os efeitos do estresse acadêmico. Pesquisas demonstram que o exercício físico regular está associado a uma redução nos níveis de estresse e uma melhoria na saúde mental geral dos estudantes (James *et al.*, 2021). Nesse contexto, técnicas de aprendizado de máquina têm sido exploradas para prever problemas de saúde mental entre estudantes universitários, utilizando múltiplas fontes de dados para identificar padrões e fatores de risco (Madububambachu; Ukpebor; Ihezue, 2024).

Diante disso, o presente estudo objetiva analisar os principais fatores determinantes do estresse acadêmico e desenvolver um modelo preditivo, baseado em aprendizado de máquina, capaz de identificar níveis de estresse em estudantes universitários. Serão exploradas variáveis como atividade física, qualidade do sono, estado emocional, carga horária de estudos e parâmetros fisiológicos. Com os resultados obtidos, espera-se contribuir para a compreensão dos fatores que impactam a saúde mental dos alunos e propor estratégias baseadas em dados para a mitigação do estresse no ambiente acadêmico.

2 MATERIAL E MÉTODOS

Nesta seção, são apresentados os procedimentos adotados para a coleta, processamento e análise dos dados utilizados no estudo. Primeiramente, descreve-se a base de dados e as variáveis analisadas. Em seguida, detalham-se as etapas de pré-processamento, análise estatística e o desenvolvimento do modelo preditivo baseado em aprendizado de máquina. O objetivo é garantir a reprodutibilidade do experimento e a confiabilidade dos resultados obtidos.

2.1 Base de Dados

O presente estudo utiliza um conjunto de dados coletado de estudantes universitários, contendo informações relacionadas a fatores que podem influenciar os níveis de estresse, tais como atividade física, qualidade do sono, estado emocional, carga horária de estudos e parâmetros fisiológicos, incluindo frequência cardíaca e pressão arterial. Pesquisas anteriores apontam que essas variáveis possuem correlação significativa com a saúde mental dos estudantes (Musshafen *et al.*, 2021; Pascoe; Hetrick; Parker, 2020).

Os dados foram anonimizados para garantir a privacidade dos participantes e foram tratados seguindo princípios de ética na pesquisa, conforme diretrizes estabelecidas por estudos anteriores na área (Eisenberg *et al.*, 2007). A análise dos dados foi realizada utilizando a linguagem Python, com as bibliotecas pandas para manipulação dos dados e matplotlib e seaborn para visualização.

2.2 Processamento dos Dados

Inicialmente, os dados passaram por um processo de limpeza e tratamento para remoção de valores ausentes e inconsistências. Para variáveis categóricas, foram aplicadas técnicas de codificação (*one-hot encoding*) a fim de facilitar a utilização dos modelos de aprendizado de máquina (Madububambachu; Ukpebor; Ihezue, 2024). Além disso, a normalização dos dados numéricos foi realizada utilizando a técnica *Min-Max Scaling*, que mantém os valores dentro de um intervalo compreendido entre 0 e 1, garantindo que variáveis com escalas diferentes não prejudiquem o desempenho dos modelos (Zhu *et al.*, 2021).

2.3 Análise Estatística

A análise exploratória incluiu o cálculo de estatísticas descritivas e a verificação de correlações entre as variáveis. As métricas de correlação de *Pearson* e *Spearman* foram aplicadas para identificar relações entre os níveis de estresse e fatores como sono e atividade física. Estudos prévios apontam que essas métricas são eficazes na identificação de padrões em dados de saúde mental (Stearne *et al.*, 2023).

2.4 Desenvolvimento do Modelo de Aprendizado de Máquina

Para a predição dos níveis de estresse, serão testados diferentes algoritmos de aprendizado de máquina, incluindo **Regressão Logística**, **Random Forest**, e **Redes Neurais Artificiais (RNA)**. Modelos baseados em árvores de decisão, como o *Random Forest*, são amplamente utilizados para classificação e predição em saúde mental devido à sua interpretabilidade e robustez em relação a dados com múltiplas variáveis (James *et al.*, 2021). Além disso, modelos de redes neurais têm sido empregados para detectar padrões complexos em dados multidimensionais, sendo uma abordagem promissora para a predição de estresse (Sharma; Kapoor; Kang, 2020).

A seleção do melhor modelo será baseada nas métricas **Acurácia**, **Precisão**, **Recall** e **F1-Score**. A técnica de validação cruzada *k-fold* será utilizada para evitar o *overfitting* e garantir que o modelo tenha um bom desempenho em dados não vistos (Baba; Bunji, 2023).

2.5 Implementação

A implementação do modelo será realizada utilizando a biblioteca *scikit-learn*, e as redes neurais serão testadas com o *TensorFlow/Keras*. A divisão dos dados será feita em 80% para treinamento e 20% para teste. Os hiperparâmetros dos modelos serão ajustados utilizando a busca em grade (*Grid Search*) para otimizar o desempenho preditivo (Yang; Shami, 2020).

3 RESULTADOS E DISCUSSÃO

Nesta seção, são apresentados e discutidos os resultados obtidos na avaliação dos modelos de aprendizado de máquina para a predição dos níveis de estresse. Os modelos testados incluem **Regressão Logística**, **Random Forest** e **Rede Neural**. A Tabela 1 apresenta as métricas de desempenho dos modelos, considerando **precisão**,

revocação e F1-score.

Tabela 1 – Métricas de desempenho dos modelos de aprendizado de máquina.

Modelo	Precisão	Revocação	F1-Score
Regressão Logística	0.556	0.556	0.556
Random Forest	0.913	0.913	0.913
Rede Neural	0.909	0.909	0.909

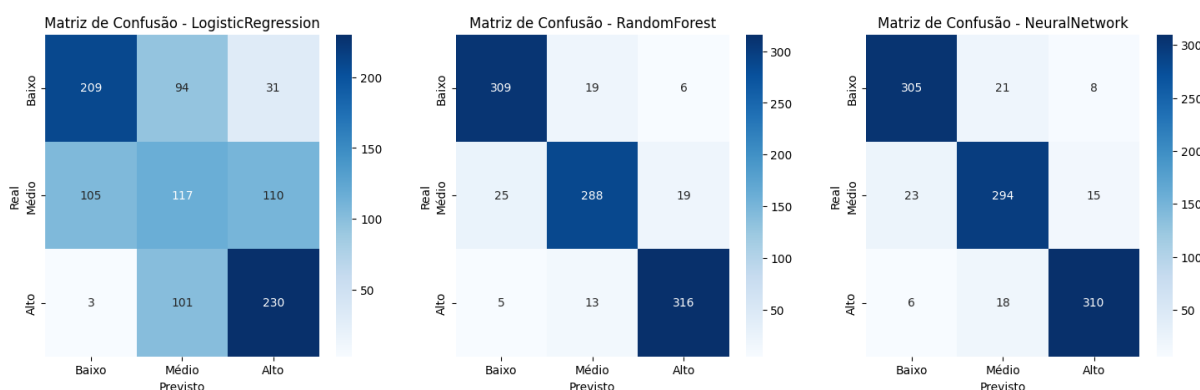
Fonte: Autoria própria (2025)

O modelo **Random Forest** apresentou o melhor desempenho geral, seguido pela **Rede Neural**, ambos com acurácia superior a 90%. Em contrapartida, a **Regressão Logística** demonstrou desempenho inferior, com valores de precisão e revocação próximos de 55%. Isso sugere que abordagens não lineares são mais eficazes para capturar as complexidades dos dados.

3.1 Matrizes de Confusão dos Modelos

A Figura 1 apresenta as matrizes de confusão para os três modelos avaliados. Observa-se que o modelo de **Random Forest** apresenta maior concordância entre valores reais e previstos, com menor taxa de erros de classificação em comparação com os demais modelos. A **Rede Neural** também apresenta um desempenho satisfatório, enquanto a **Regressão Logística** exibe um maior número de previsões incorretas, particularmente nas classes de estresse moderado.

Figura 1 – Matrizes de confusão para os modelos testados.

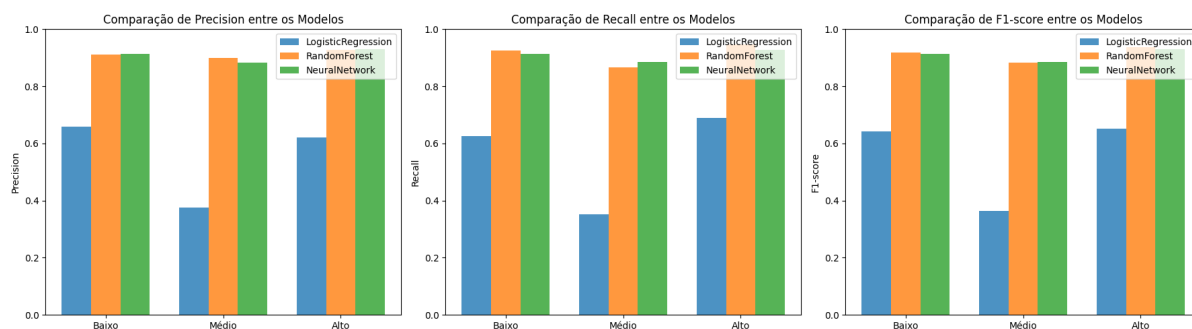


Fonte: Autoria própria (2025)

3.2 Comparação de Desempenho entre os Modelos

A Figura 2 ilustra uma comparação entre os modelos avaliados considerando as métricas de desempenho. A **Random Forest** e a **Rede Neural** apresentam desempenho superior em todas as métricas quando comparadas à **Regressão Logística**, demonstrando que técnicas mais avançadas são mais adequadas para esse tipo de previsão.

Figura 2 – Comparação das métricas de desempenho entre os modelos avaliados.

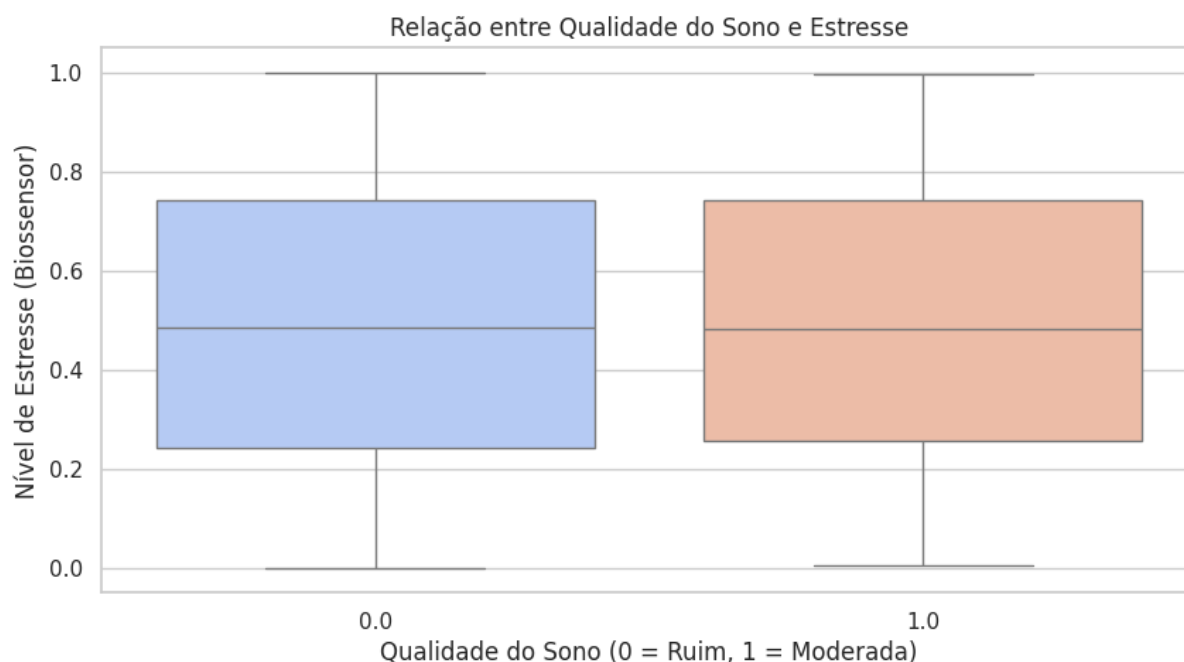


Fonte: Autoria própria (2025)

3.3 Relação entre Qualidade do Sono e Estresse

A Figura 3 apresenta a distribuição dos níveis de estresse em relação à qualidade do sono. Indivíduos que relataram baixa qualidade de sono demonstram uma maior incidência de estresse elevado. Esses achados são coerentes com estudos que indicam que a privação de sono pode aumentar os níveis de cortisol e diminuir a resiliência ao estresse.

Figura 3 – Distribuição dos níveis de estresse em relação à qualidade do sono.



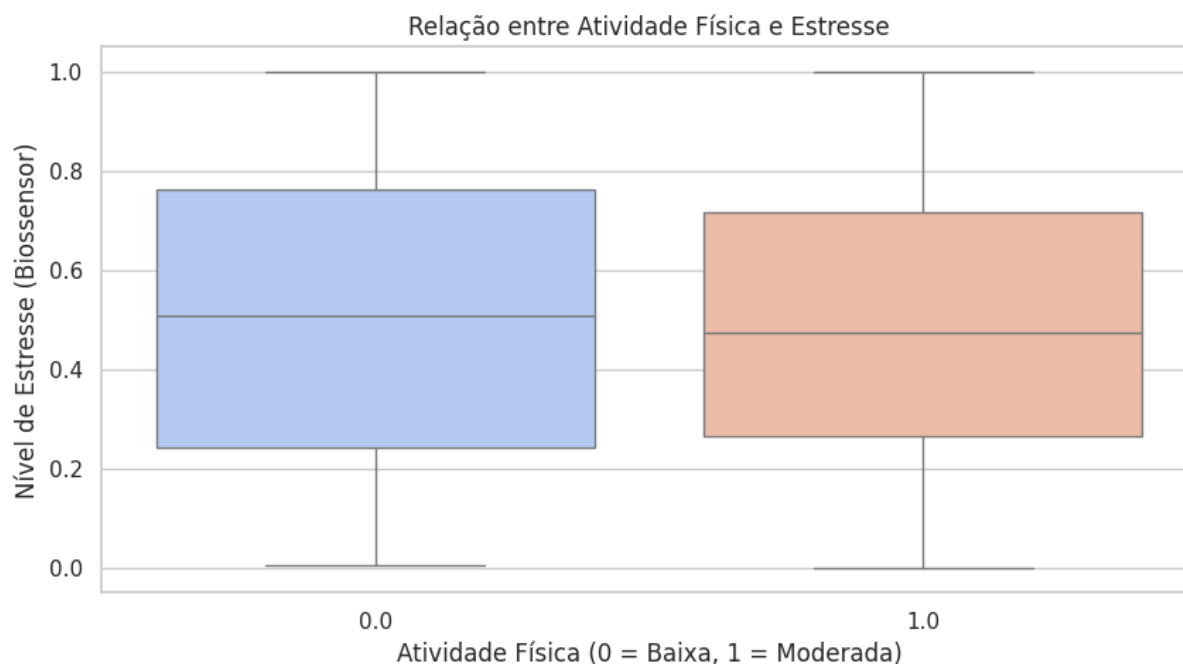
Fonte: Autoria própria (2025)

3.4 Relação entre Atividade Física e Estresse

A Figura 4 apresenta a relação entre os níveis de estresse e a prática de atividade física. Verifica-se que indivíduos com níveis moderados e altos de atividade física tendem

a apresentar menores níveis de estresse elevado. Esse padrão reforça evidências de que a prática regular de exercícios físicos atua como fator protetor contra o estresse, devido à liberação de neurotransmissores como endorfina e serotonina.

Figura 4 – Distribuição dos níveis de estresse conforme os níveis de atividade física.



Fonte: Autoria própria (2025)

3.5 Distribuição dos Níveis de Estresse

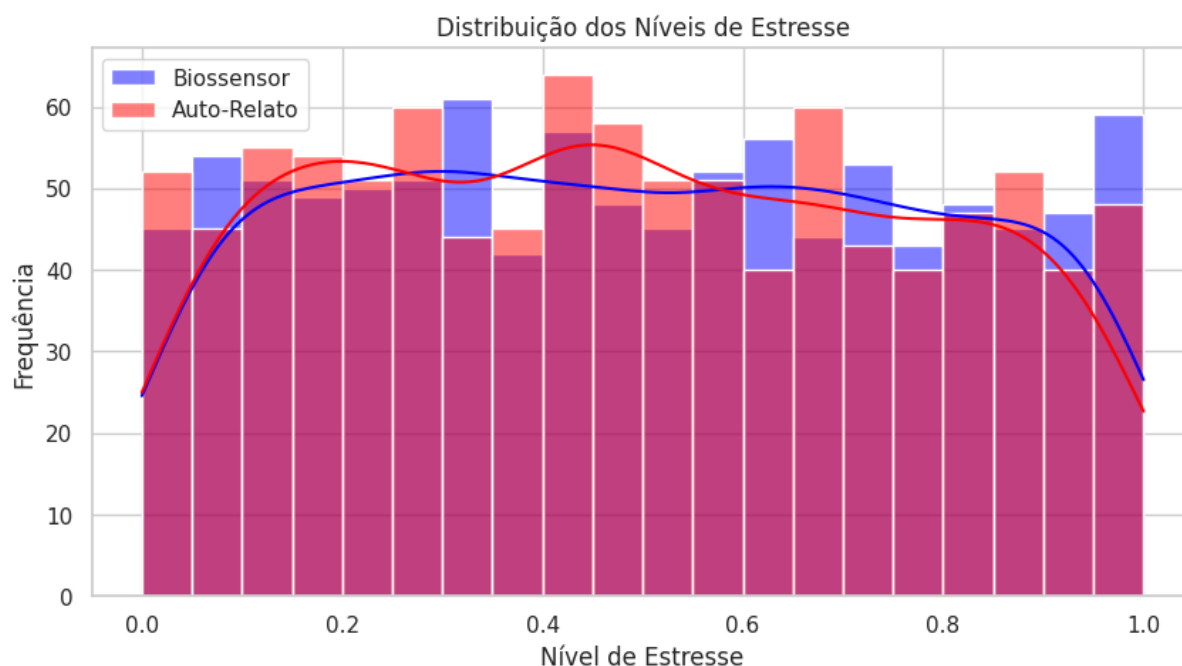
A Figura 5 apresenta a distribuição dos níveis de estresse na base de dados. A maior parte dos indivíduos encontra-se na categoria de estresse moderado, com uma distribuição relativamente equilibrada entre os diferentes níveis. Essa distribuição sugere que a amostra utilizada contempla uma ampla gama de condições de estresse, permitindo uma análise mais aprofundada dos fatores associados.

Os resultados confirmam a influência da qualidade do sono e da prática de atividades físicas na variação dos níveis de estresse. Além disso, os resultados indicam que os modelos de aprendizado de máquina **Random Forest** e **Rede Neural** apresentam maior eficácia na predição dos níveis de estresse. Esses modelos demonstram potencial para aplicações futuras no monitoramento e predição do estresse em diferentes populações, podendo ser empregados para o desenvolvimento de sistemas preditivos personalizados para a gestão da saúde mental.

4 CONCLUSÕES

O presente estudo teve como objetivo avaliar a eficácia de diferentes modelos de aprendizado de máquina na predição dos níveis de estresse, considerando fatores como qualidade do sono, atividade física e características individuais. Os modelos de **Random Forest** e **Rede Neural** apresentaram os melhores desempenhos, com valores

Figura 5 – Distribuição dos níveis de estresse na base de dados.



Fonte: Autoria própria (2025)

de acurácia superiores a 90%, evidenciando sua capacidade de modelar padrões complexos nos dados. Em contrapartida, a **Regressão Logística** demonstrou desempenho inferior, com métricas em torno de 55%, sugerindo limitações na representação da variabilidade dos dados.

Os achados reforçam a influência de fatores como qualidade do sono e prática de atividade física na regulação dos níveis de estresse. Indivíduos que relataram baixa qualidade de sono apresentaram maior incidência de estresse elevado, enquanto aqueles com prática regular de exercícios físicos tiveram menores níveis de estresse. Esses resultados corroboram estudos prévios, que indicam que o equilíbrio entre descanso e atividade física é um componente essencial para a manutenção da saúde mental e do bem-estar.

Além disso, a análise da distribuição dos níveis de estresse revelou uma predominância de indivíduos na categoria de estresse moderado, destacando a importância de estratégias preventivas para mitigar o impacto do estresse no cotidiano. O uso de aprendizado de máquina na predição do estresse pode viabilizar o desenvolvimento de ferramentas automatizadas de monitoramento, auxiliando na identificação precoce de padrões de estresse e na adoção de estratégias preventivas.

Como trabalhos futuros, sugere-se a inclusão de variáveis adicionais, como alimentação e uso de dispositivos eletrônicos antes do sono, para refinar as análises preditivas. Além disso, a incorporação de técnicas de aprendizado profundo e a utilização de conjuntos de dados mais amplos podem aprimorar ainda mais a capacidade dos modelos em capturar padrões sutis e melhorar sua generalização. Dessa forma, espera-se que pesquisas futuras possam consolidar o uso de inteligência artificial como um recurso valioso para a promoção da saúde mental e a prevenção do estresse.

Agradecimentos

Agradece-se à Universidade Tecnológica Federal do Paraná (UTFPR) pelo suporte acadêmico e estrutural na realização deste trabalho, bem como ao professor José Alves de Faria Filho, pela orientação e contribuições ao estudo.

Material Suplementar

O material suplementar deste estudo inclui os códigos-fonte utilizados para a análise dos dados e o treinamento dos modelos de aprendizado de máquina, além dos arquivos contendo os resultados das matrizes de confusão e os gráficos gerados. Todo esse material está disponível no repositório vinculado a este trabalho.

Disponibilidade de Código

Os códigos desenvolvidos para a análise dos dados, treinamento dos modelos de aprendizado de máquina e geração dos resultados estão disponibilizados publicamente. O repositório contendo o código-fonte pode ser acessado no seguinte endereço:

<https://github.com/oBaldon/stress-prediction>

Conflito de interesse

Não há conflito de interesse.

Objetivos de Desenvolvimento Sustentável (ODS)

O presente estudo contribui para a promoção da saúde mental e do bem-estar de estudantes universitários ao desenvolver modelos preditivos capazes de identificar precocemente os níveis de estresse acadêmico. A detecção antecipada permite a implementação de intervenções eficazes para mitigar os impactos do estresse, promovendo um ambiente educacional mais saudável e acessível. Dessa forma, ao analisar os fatores associados ao estresse e propor estratégias para sua redução, este trabalho auxilia na melhoria da qualidade de vida dos estudantes e na sua permanência nas instituições de ensino superior.



REFERÊNCIAS

- BABA, Ayako; BUNJI, Kyosuke. Prediction of mental health problem using annual student health survey: machine learning approach. **JMIR Mental Health**, JMIR Publications Toronto, Canada, v. 10, e42420, 2023.
- EISENBERG, Daniel *et al.* Prevalence and correlates of depression, anxiety, and suicidality among university students. **American journal of orthopsychiatry**, Wiley Online Library, v. 77, n. 4, p. 534–542, 2007.
- JAMES, Michaela *et al.* Impact of school closures on the health and well-being of primary school children in Wales UK: a routine data linkage study using the HAPPEN Survey (2018–2020). **BMJ open**, British Medical Journal Publishing Group, v. 11, n. 10, e051574, 2021.
- MADUBUBAMBACHU, Ujunwa; UKPEBOR, Augustine; IHEZUE, Urenna. Machine learning techniques to predict mental health diagnoses: A systematic literature review. **Clinical Practice and Epidemiology in Mental Health: CP & EMH**, v. 20, e17450179315688, 2024.
- MUSSHAFEN, Leslie A *et al.* Associations between sleep and academic performance in US adolescents: a systematic review and meta-analysis. **Sleep medicine**, Elsevier, v. 83, p. 71–82, 2021.
- PASCOE, Michaela C; HETRICK, Sarah E; PARKER, Alexandra G. The impact of stress on students in secondary school and higher education. **International journal of adolescence and youth**, Taylor & Francis, v. 25, n. 1, p. 104–112, 2020.
- SHARMA, Disha; KAPOOR, Nitika; KANG, Sandeep Singh. Stress prediction of students using machine learning. **International Journal of Mechanical and Production Engineering Research and Development**, v. 10, n. 3, 2020.
- STEARE, Thomas *et al.* The association between academic pressure and adolescent mental health problems: A systematic review. **Journal of affective disorders**, Elsevier, 2023.
- YANG, Li; SHAMI, Abdallah. On hyperparameter optimization of machine learning algorithms: Theory and practice. **Neurocomputing**, Elsevier, v. 415, p. 295–316, 2020.
- ZHU, Xihe *et al.* Academic stress, physical activity, sleep, and mental health among Chinese adolescents. **International Journal of Environmental Research and Public Health**, MDPI, v. 18, n. 14, p. 7257, 2021.