

Predição de Evasão em Cursos Universitários: Uma Abordagem Baseada em Stacking

Murilo R. Cândido,¹ Danilo R. Pereira,² Vitor M. Rodrigues,³ Marcus V. M. Souza⁴
DMC-FCT/UNESP, Presidente Prudente, SP

Neste trabalho, utilizando a abordagem proposta por Niyogisubizo et al. [2], desenvolvemos um classificador de risco de evasão para alunos do campus FCT/UNESP que atingiu uma precisão superior a 93%. Os resultados obtidos demonstram que a estratégia baseada em empilhamento (stacking) de dois níveis é eficaz para identificar alunos em risco, mesmo quando aplicada a um conjunto de dados relativamente pequeno. Na primeira camada, previsões temporais foram geradas a partir de algoritmos robustos, como Random Forest, Extreme Gradient Boosting e Gradient Boosting, que processaram dados brutos, caracterizados por variáveis desordenadas e irregulares, extraindo features representativas. Em seguida, as saídas destes modelos foram integradas a um meta-classificador, implementado por meio de um modelo de Feed-forward Neural Network (FNN), que, utilizando validação cruzada, reduziu significativamente os problemas de overfitting e aprimorou a precisão final da predição [1].

A aplicação de técnicas de ensemble, conforme discutido por Xing et al. [4] e fundamentado na metodologia de stacked generalization proposta por Wolpert [3], possibilitou a extração de um conjunto mais rico de informações preditivas, a redução dos erros de generalização por meio da combinação não linear das previsões e a melhoria dos índices de desempenho em relação às abordagens tradicionais. Tais achados sugerem que, mesmo em domínios com dados limitados, a utilização de métodos de ensemble pode oferecer suporte decisivo para intervenções educacionais, permitindo a identificação precoce e o apoio direcionado a alunos em risco de evasão.

Diante dos resultados preliminares, conclui-se que a abordagem de empilhamento em dois níveis se mostra altamente promissora para a predição do abandono em cursos universitários. Entre as principais contribuições deste estudo, destacam-se a inovação na integração de Random Forest, Extreme Gradient Boosting, Gradient Boosting e Feed-forward Neural Networks em um framework de **stacked generalization**, bem como a demonstração empírica de que é possível identificar com alta precisão os alunos em risco mesmo com um conjunto de dados relativamente pequeno.

Agradecimentos

O projeto, intitulado *Políticas Públicas e Ações Afirmativas na Universidade Pública: definindo algoritmo para alerta de evasão de ingressantes universitários via reserva de vagas e cotas raciais*, é realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processo 2023/10240 – 0, e conta com a parceria do INEP e a empresa de soluções acadêmicas UniSoma, elementos fundamentais para o desenvolvimento dos resultados apresentados.

¹mr.candido@unesp.br

²danillo.pereira@unesp.br

³vitor.m.rodrigues@unesp.br

⁴marcus.moraes@unesp.br

Referências

- [1] M. Jiang et al. “An improved Stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms”. Em: **Physica A** 541 (2020), p. 122272. DOI: 10.1016/j.physa.2019.122272.
- [2] J. Niyogisubizo, L. Liao, E. Nziyumva, E. Murwanashyaka e P. C. Nshimyumukiza. “Predicting student’s dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization”. Em: **Computers and Education: Artificial Intelligence** 3 (2022), p. 100066. DOI: 10.1016/j.caeai.2022.100066.
- [3] D. H. Wolpert. “Stacked generalization”. Em: **Neural Networks** 5.2 (1992), pp. 241–259. DOI: 10.1016/S0893-6080(05)80023-1.
- [4] W. Xing, X. Chen, J. Stein e M. Marcinkowski. “Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization”. Em: **Computers in Human Behavior** 58 (2016), pp. 119–129. DOI: 10.1016/j.chb.2015.12.007.