

Predição de Evasão na UNESP: Engenharia de Features e Análise Exploratória

Murilo R. Cândido¹, Danillo R. Pereira², Marcus V. M. Souza³, Vitor M. Rodrigues⁴
DMC-FCT/UNESP, Presidente Prudente, SP

A evasão no ensino superior configura um desafio crítico para o cenário educacional brasileiro, impactando a formação de capital humano qualificado e comprometendo os investimentos públicos e privados. Conforme evidenciado no *Mapa do Ensino Superior no Brasil 2024* [3], a evasão é um fenômeno multifatorial, cuja incidência varia conforme a natureza da instituição, a região e o curso oferecido. Fatores econômicos, acadêmicos e socioeconômicos interagem para determinar os índices de abandono, destacando-se dificuldades financeiras, insuficiência de apoio pedagógico e a necessidade de conciliar atividades profissionais com os estudos, condições que sobrecarregam os alunos e elevam os índices de evasão. Ademais, a crescente demanda do mercado de trabalho por formações especializadas impõe uma pressão adicional, ocasionando profundas implicações econômicas, como a redução da produtividade, a diminuição do potencial inovador e o comprometimento de indicadores de crescimento, tais como o PIB e a arrecadação tributária [3].

Este trabalho objetiva analisar os marcadores preditivos do risco de evasão entre universitários ingressantes por meio de reserva de vagas e cotas na UNESP, utilizando técnicas de machine learning para desenvolver um sistema de alerta precoce baseado em modelagem preditiva. No âmbito da literatura, diversos estudos demonstram a eficácia de algoritmos supervisionados para previsão da evasão: a aplicação do algoritmo Naïve Bayes na Hellenic Open University resultou em acurácia elevada [1], enquanto árvores de decisão superaram modelos baseados em regras de indução em estudo realizado no México [5]. Em outros trabalhos, Support Vector Machines, foram empregadas em um estudo na Universidade de Bari Aldo Moro, apresentando desempenho superior em relação à regressão logística e ao Naïve Bayes Gaussiano [4], e redes neurais têm sido utilizadas para identificar padrões de risco de evasão a partir de características dos estudantes [2]. A abordagem proposta compreende quatro etapas: extração e pré-processamento dos dados brutos, seleção de variáveis representativas, divisão do conjunto de dados e previsão inicial do risco de evasão, sendo que os resultados preliminares indicam uma previsão correta de mais de 90% dos alunos desistentes após análise do desempenho no primeiro e segundo semestres. Estes resultados superam benchmarks presentes na literatura especializada e demonstram o potencial do sistema de Aprendizagem de Máquina para subsidiar o aperfeiçoamento das políticas públicas de promoção da equidade no ensino superior.

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),

¹mr.candido@unesp.br

²danillo.pereira@unesp.br

³marcus.moraes@unesp.br

⁴vitor.m.rodrigues@unesp.br

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.

Referências

- [1] S. B. Kotsiantis, C. Pierrakeas e P. E. Pintelas. “Preventing student dropout in distance learning using machine learning techniques”. Em: **International Conference on Knowledge-Based and Intelligent Information and Engineering Systems**. Vol. 2774. Lecture Notes in Computer Science. Springer, 2003, pp. 267–274. DOI: 10.1007/978-3-540-45226-3_37.
- [2] V. R. C. Martinho, C. Nunes e C. R. Minussi. “Prediction of school dropout risk group using neural network”. Em: **2013 Federated Conference on Computer Science and Information Systems**. IEEE, 2013, pp. 111–114. URL: <http://hdl.handle.net/11449/227468>.
- [3] SEMSP. **Mapa do Ensino Superior no Brasil 2024**. Online. Acessado em 13/03/2025. URL: <https://www.semesp.org.br/wp-content/uploads/2024/04/mapa-do-ensino-superior-no-brasil-2024.pdf>.
- [4] A. Serra, P. Perchinunno e M. Bilancia. “Predicting student dropouts in higher education using supervised classification algorithms”. Em: **International Conference on Computational Science and Its Applications**. Vol. 10962. Lecture Notes in Computer Science. Springer, 2018, pp. 18–33. DOI: 10.1007/978-3-319-95168-3_2.
- [5] C. M. Vera, A. Cano, C. Romero, A. Y. M. Noaman, H. M. Fardoun e S. Ventura. “Early dropout prediction using data mining: A case study with high school students”. Em: **Expert Systems** 33.1 (2016), pp. 107–124. DOI: 10.1111/exsy.12135.