

Forecasting Students Dropout: A UTAD University Study

O artigo aborda a taxa de abandono nas universidades portuguesas, que é de aproximadamente 29%, com 14% dos alunos restantes não concluindo o curso no tempo estipulado. Essas taxas elevadas são uma grande preocupação para as instituições educacionais, não apenas em Portugal, mas em todo o mundo. As universidades têm adotado várias ações tradicionais para reduzir essas taxas, como monitoramento personalizado dos alunos em risco e reestruturação dos currículos. No entanto, a identificação precoce das razões para o abandono é essencial para qualquer metodologia que visa reduzir essas taxas.

Com o avanço da inteligência artificial (IA) e da ciência de dados, tem-se tornado cada vez mais viável prever o sucesso acadêmico com precisão. Neste estudo, foram aplicados vários modelos de aprendizado de máquina a um conjunto de dados de alunos da Universidade de Trás-os-Montes e Alto Douro (UTAD), focando apenas em dados acadêmicos devido à falta de dados socioeconômicos. O objetivo é construir modelos analíticos capazes de prever com precisão os abandonos escolares usando apenas notas acadêmicas e a idade dos alunos.

O artigo também revisa diversos estudos anteriores que utilizaram técnicas de aprendizado de máquina para prever o abandono escolar em diferentes contextos educacionais. Esses estudos empregaram uma variedade de modelos, como Redes Neurais Artificiais (ANNs), Florestas Aleatórias (RF), e métodos de Ensemble, entre outros, para prever com alta precisão os casos de abandono escolar, utilizando diferentes tipos de dados, desde registros acadêmicos até dados demográficos e de atividades extracurriculares. O estudo em questão aplica essas técnicas a um dataset limitado, buscando minimizar os vieses e melhorar a precisão das previsões.

Por fim, no presente estudo, foi considerada a previsão de abandono acadêmico. Embora o abandono escolar dependa de vários fatores, como condições econômicas, sociais, de formação parental e institucionais, este estudo foi conduzido com dados referentes ao sucesso das unidades curriculares como fonte. Com os resultados alcançados, conclui-se que essa análise é possível mesmo quando os dados dos alunos são escassos. De fato, todos os métodos considerados neste trabalho mostram resultados promissores na previsão de abandono acadêmico, com destaque para Random Forest (RF) e XGBoost, que demonstraram uma precisão de 88% e 90% no conjunto de testes final, respectivamente. Essa previsão é possível porque as notas dos alunos de alguma forma já incorporam suas condições de vida. Por outro lado, o estudo da importância das características revelou

que a conclusão bem-sucedida do curso depende da maturidade dos alunos (idade) e do sucesso em unidades curriculares mais exigentes.

Como estudo futuro, essas novas técnicas de mineração de dados serão aplicadas a outros planos de estudo, o que permitirá a implementação dos modelos mais adequados. Portanto, um classificador será incorporado ao sistema de informação da UTAD para apoiar o corpo docente na previsão de abandono dos alunos. Por outro lado, informações pessoais dos alunos, como dados econômicos e pessoais, serão consideradas quando a UTAD coletar uma quantidade significativa de dados.