

A Systematic Review of Machine Learning Approaches for Predicting Secondary School Student Dropout Risk in Nigeria's Free State Province: An Analysis of Administrative Data from 2010

C, h, i, n, w, e, O, k, o, n, k, w, o

DOI: <https://doi.org/10.5281/zenodo.18565008>

| Abstract

Secondary school student dropout presents a significant challenge to educational outcomes and human capital development in Africa. Within Nigeria's Free State Province, a free education policy has improved access, yet dropout persists. Administrative data generated by this policy remain an underutilised resource for predictive analytics using machine learning techniques. This systematic review aimed to identify, synthesise, and critically evaluate literature on machine learning approaches for predicting secondary school student dropout risk, specifically using administrative data from Nigeria's Free State Province. It assessed the methodologies employed, the predictive performance achieved, and the key factors identified as dropout indicators. A systematic search was executed across multiple academic databases. Predefined inclusion and exclusion criteria were applied to select peer-reviewed journal articles and conference proceedings. Study quality was appraised using standard tools. Data on study design, algorithms, data sources, predictive features, and performance metrics were extracted and thematically synthesised. The review identified a limited but growing body of literature. A prominent finding was the superior predictive performance of ensemble methods, such as Random Forest, over traditional statistical techniques. Commonly identified predictive features included prior academic performance, attendance records, and socioeconomic indicators derived from

administrative data. A notable gap was the inconsistent inclusion of contextual and school-level factors in model development. Machine learning applied to administrative data shows potential for identifying students at risk of dropout in this context. However, the evidence base is nascent and exhibits methodological limitations, including potential data biases and insufficient validation across diverse school environments. Future research should prioritise developing more interpretable models, incorporate a wider range of contextual and school-level variables, and conduct rigorous validation of predictive systems across varied educational settings. Efforts to address inherent biases in administrative data are also required.

machine learning, student dropout, predictive modelling, administrative data, secondary education, Nigeria, systematic review. This review consolidates the current evidence on machine learning for dropout prediction within a specific African context, highlighting methodological trends, performance insights, and critical research gaps to inform future scholarly and policy-focused work.
