

Data-driven prediction of university admission cutoff marks in Sri Lanka using machine learning

Theeksha J.R.N.^{1*}, and Kudagamage U.P.²

¹Department of Computing and Information Systems,
Faculty of Computing, Sabaragamuwa University of Sri Lanka

²Department of Data Science,
Faculty of Computing, Sabaragamuwa University of Sri Lanka

*jrntheeksha@std.appsc.sab.ac.lk

Admissions into universities in Sri Lanka are determined through the cutoff marks published annually in the University Grants Commission (UGC), although the procedure itself is not only unclear but opaque to students as well as counselors. Since the procedure is unclear, students end up making academic choices based on misleading information. This research proposes a machine learning model for accurate determination of the cutoff marks in the UGC university admissions. The proposed approach uses the UGC cutoff dataset (2020-2025), with features like the year, the district quota, the stream, the university, the degree, the intake capacity, and the cutoff marks. The final supervised regression approach using XGBoost with year-related features and lag features for cutoff points is proposed. In this work, the error in the regression method is calculated by the RMSE, MAE, and R^2 values. The regression analysis will be done by the subgroup error analysis for districts and streams, and the regression results will be explained by SHAP values. The RMSE, MAE, and R^2 score of the XGBoost algorithm come out to be 0.2062, 0.1446, and 0.7484, respectively. Among various factors given importance by the algorithm, Z score is given maximum importance, followed by subject stream, district quota, university, intake capacity, and previous year cutoff trends. On checking the equity of the algorithm, it is found that the disparity of errors is very low, which makes it a fair algorithm. This study has shown the viability and accuracy of using machine learning algorithms to predict the admission cutoffs of UGC admissions. This work fills the very significant research gap created by the absence of publications focusing on the educational data-mining domain. This work can form the basis of an effective and transparent decision support system that can help students, counselors, and policymakers make educated decisions regarding educational planning and access to higher education.

Keywords: *Educational Data Mining; Machine Learning; Sri Lanka; University; Admissions*