KNearest Oracle-AutoML Model for Predicting Student Dropouts in Tanzanian's Secondary Schools

Yuda Mnyawami [2] Hellen Maziku[2] Joseph Mushi[2]

[2]University of Dar es Salaam

Abstract

Secondary school dropout is a major problem in developing countries, particularly in Sub-Saharan Africa. In Tanzania, student dropouts in secondary schools increased from 3.8 percent in 2018 to 4.2 percent in 2019. Student dropout rates increased significantly in secondary schools due to inappropriate identification of the root causes of student dropouts and the method used to project the severity of the problem. In addressing this prevalent problem, machine learning designed to learn from data, revealing previously unknown findings as it discovers historical relationships and trends. The proposed model has done well in addressing secondary school dropouts by accurately identifying the root causes of student dropout. This study discovered that the root causes of student dropout in Tanzanian secondary schools are the number of children, household size, distance, age, household education, student location (area), student gender, and means to school. Therefore, the enhanced prediction scores indicate an accurate selection of student dropout features that have a significant contribution to student dropout, which can be closely examined during the learning process to allow for early interventions.

Introduction

In 2017 indicates student dropout rate in primary school was 1.3% and 3.8% in secondary school when compared to 2016 was 1.0% in primary school and 3.7% in secondary school (PO-RALG, 2018, pp. 53-57). In 2013, the student dropout was very high in both primary and secondary schools by 5.6% and 14%, respectively. Secondary school dropout rates increased by 3.8% in 2018 and by 4.2% in 2019 (PO-RALG, 2019, p. 278; PO-RALG, 2020, p. 258) presented in Figure 1.

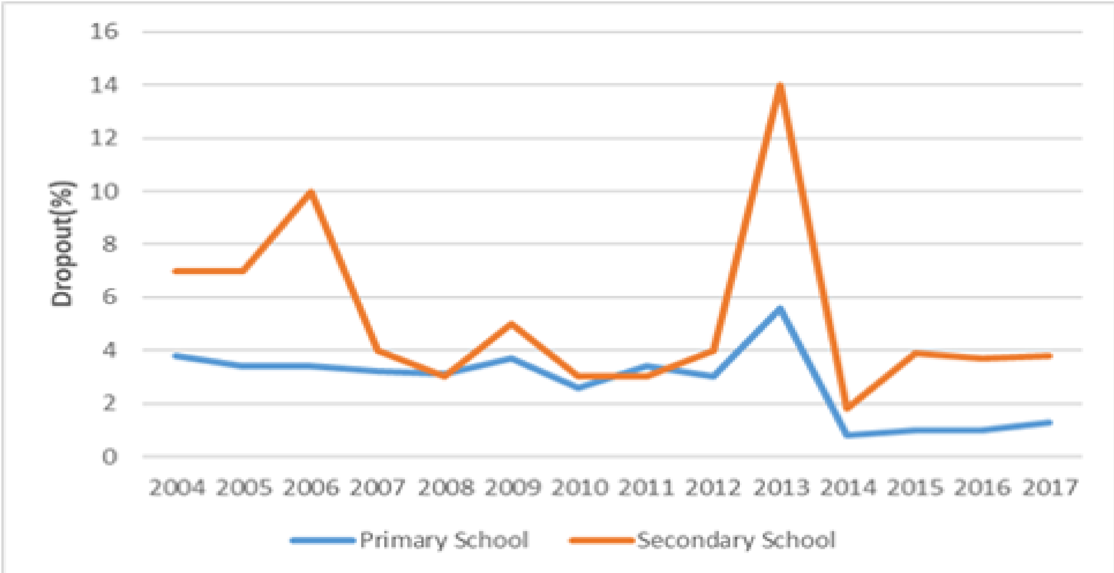

Figure 1. Student Dropout Trends in Tanzania.

Existing Challenges

- Inappropriateness identification of root causes of the student dropouts.
- Methods used cannot effectively project the severity of the problem.
- Using irrelevant and uninformative student dropout features.

Specific Objectives

- To establish factors influencing student dropout in secondary schools in Tanzania context in the order of severity.
- To develop the enhanced model for predicting student dropouts in secondary schools in Tanzania.

Methodology

A Set of Established Student Dropout Factors

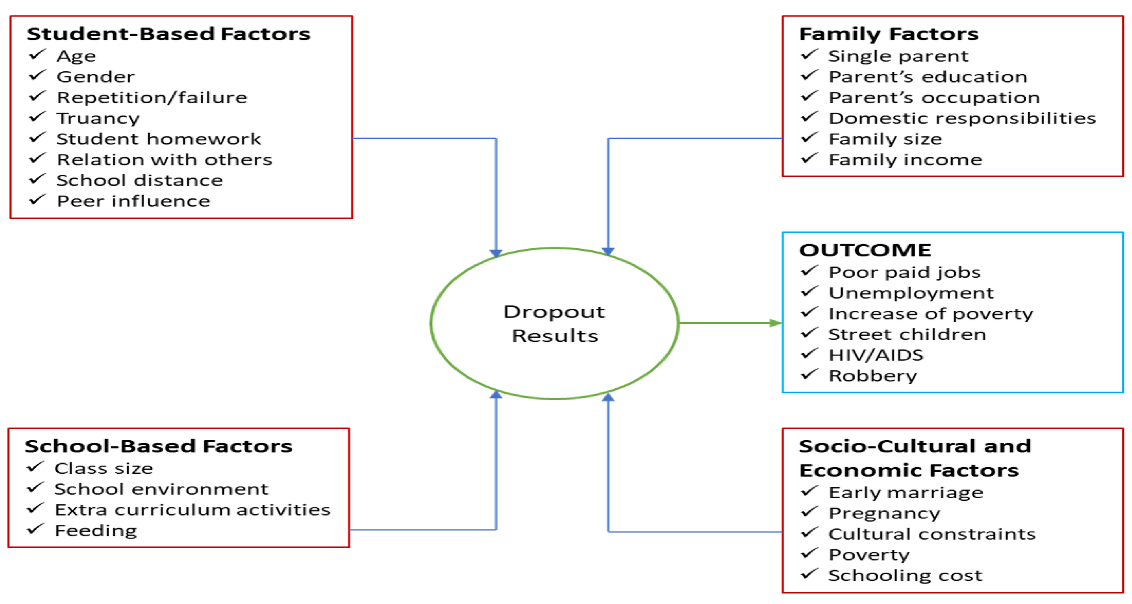

Figure 2. Student Dropout Dataset (Adapted from (Gachungi, 2005)

Technical Implementation for Prediction Model

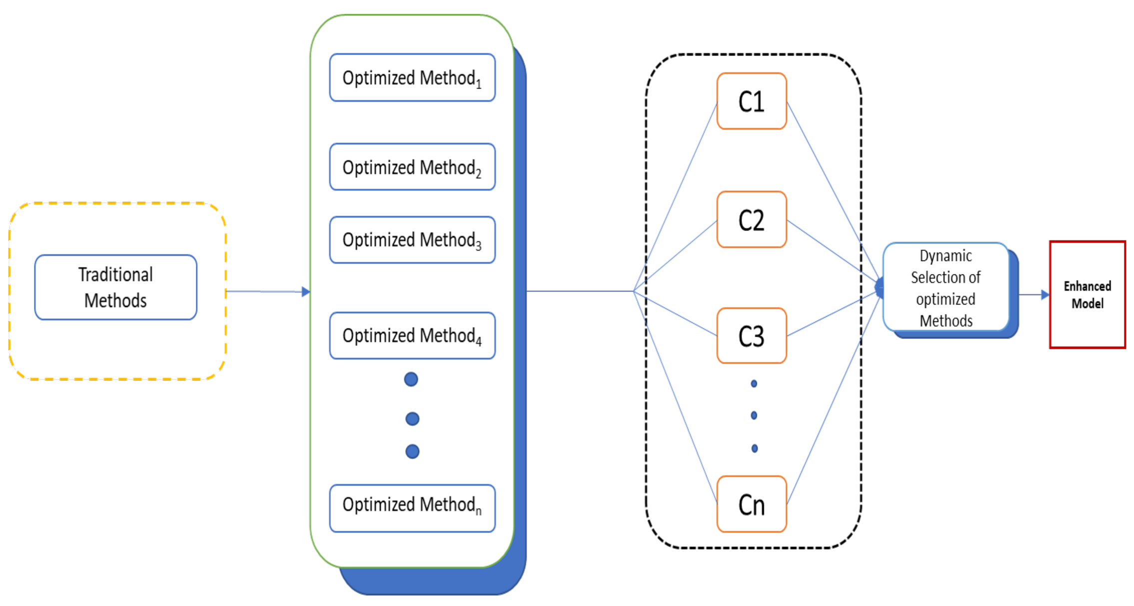

Figure 3. KNORA-AutoML model Prediction Model

Results

Significant Contribution of Combined Features

Combined Features	PredAcc. (%)
(age, school distance, means of transport, household size, household's education, home language, household's occupation, gender, mother's education, grade)	97%
(age, school distance, means of transport, household's education)	95%
(age, mother's education, school distance, means of transport)	95%
(household's education, school distance, means of transport)	95%
(household size, gender, school distance, household children)	94%
(household size, gender, means of transport, household children)	94%
(household size, means of transport, household children)	94%

Figure 4. Combination of student dropout features

Static Optimized Ensemble Model

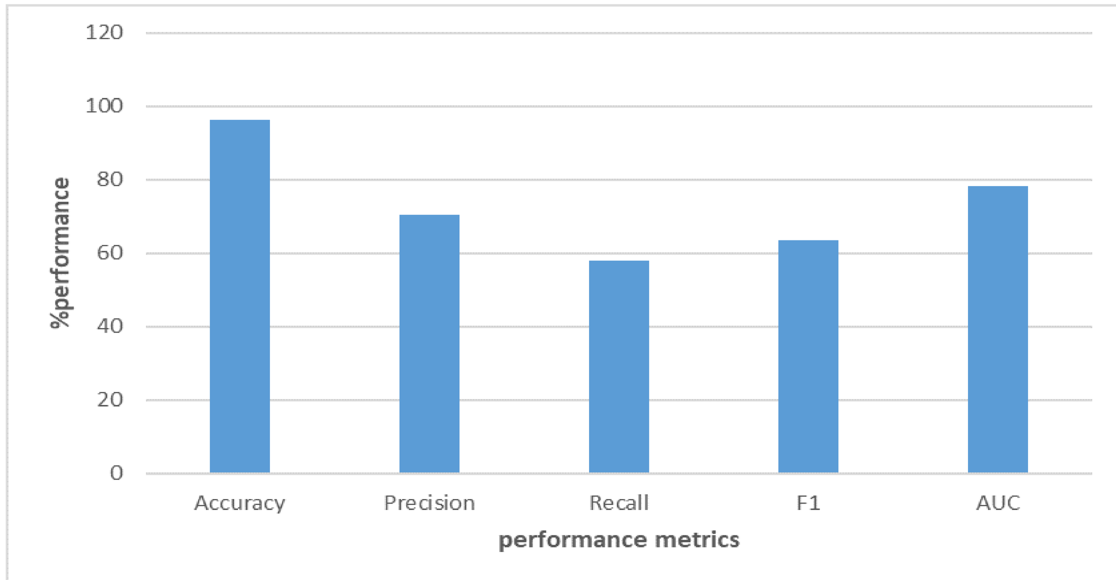

Figure 5. Results for Static Optimized Ensemble Model

Dynamic Optimized Ensemble Model

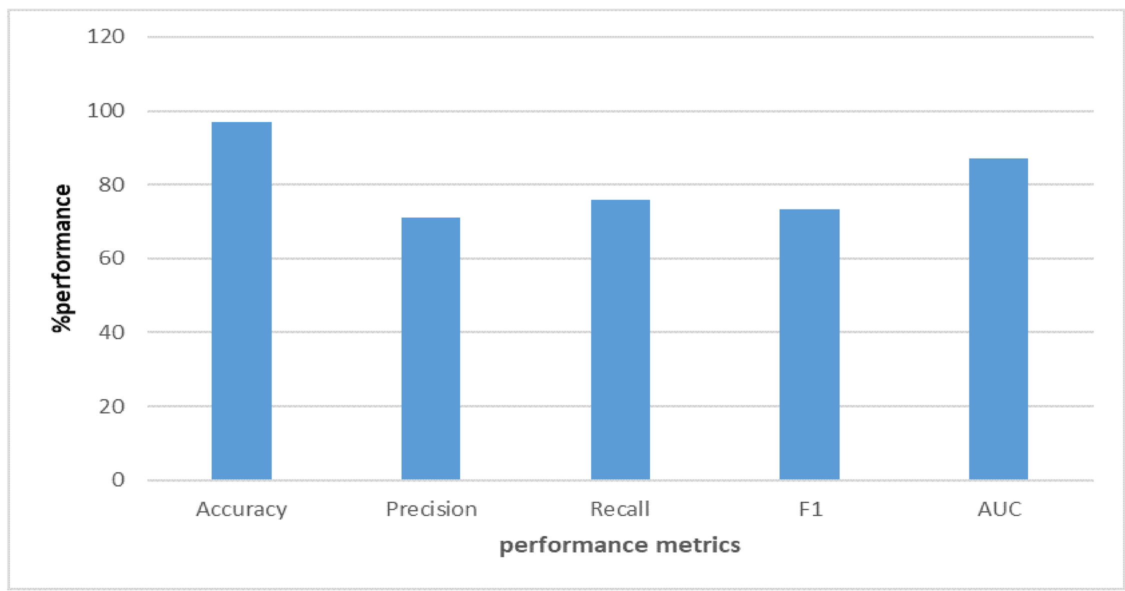

Figure 6. Results for Dynamic Optimized Ensemble Model

Conclusion

Student dropout in developing countries such as Tanzania is an important issue with promising solutions for policymakers and education stakeholders. The previous studies used inappropriate formal methods for identifying root causes that led to student dropouts in Tanzania which compromised prediction accuracy. To address this problem, researchers proposed an improved machine-learning model for predicting student dropouts. The proposed model aided in establishing relevant and informative features such as school-based, student-based, parental, socio-cultural, economic, and demographic factors.

References

[1] PO-RALG. (2018). Basic Education Statistical Abstract 2004-2017.

[2] PO-RALG. (2019). Pre-Primary, Primary, Adult and Non-Formal Education Statistics. In President's Office Regional Administration and Local Government.

[3] PO-RALG. (2020). Pre-Primary, Primary, Adult and Non-Formal Education Statistics (Vol. 1).