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

Mnyawami Y.
Department of Computer Science and Engineering
University of Dar es Salaam
Deep Learning Indaba
Accra, Ghana
mnyawamiyuda@gmail.com

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 is 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 the school. Therefore, the enhanced prediction scores indicate an accurate selection of student dropout features that significantly contribute to student dropout, which can be closely examined during the learning process to allow for early interventions.

1 Introduction

In Tanzania, student dropouts in secondary schools increased from 3.8 percent in 2018 to 4.2 percent in 2019 (PO-RALG, 2019; PO-RALG, 2020). Student drop-out has been addressed by the different interventions to reduce a great impact in secondary schools. Many measures to reduce student dropout have been proposed and put in place for education stakeholders (UNESCO-Tanzania, 2015). Faruk (2015) proposed the teachers' training, seminars, and workshops as a measure for students not to drop out of school. Similarly, Bibi (2018) evidenced that parents-teacher meetings contribute 93.5%, and parents workshops 82.3% to control student dropout rate. Likewise, authority in the United Republic of Tanzania (URT) established the Secondary Education Development Programme (SEDP) in 2005 aimed to introduce at least one secondary school in every administrative ward so as to increase the availability of secondary schools and reduce the distance from their homes as a measure to solve school dropout (URT, 2008). The authority also introduced the Big Results Now Initiative (World Bank, 2014) to fast-track quality improvement of education in secondary schools and address student dropout. The sharp decrease in 2016 could be attributed to the government's fee-free basic education policy introduced in January 2016 (UNICEF-Tanzania, 2018).

The studies by Mduma et al. (2019), Lee & Chung (2019), Chareonrat (2016), Aguiar (2015), and Sara et al. (2015) have focused on establishing machine learning (ML) prediction models, as measures to fight against student dropout in secondary schools but unfortunately, the dropout problem still persists. The persisting dropout problem especially in secondary schools attributed to a lack of proper

5th Deep Learning Indaba Conference (DLI 2023).

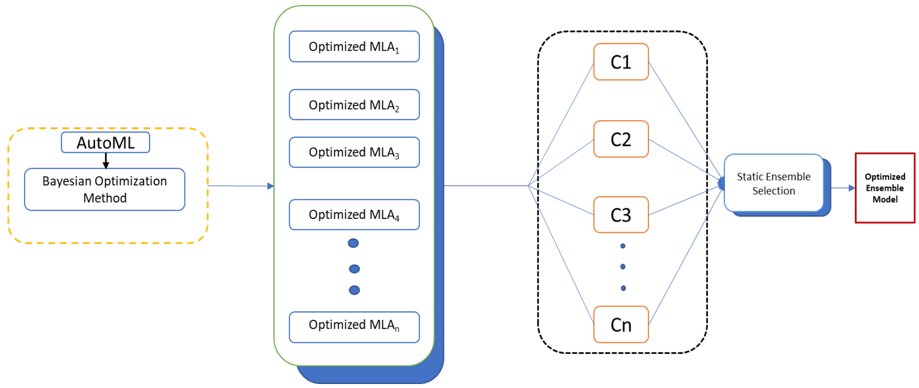

Figure 1: Static optimized machine learning model.

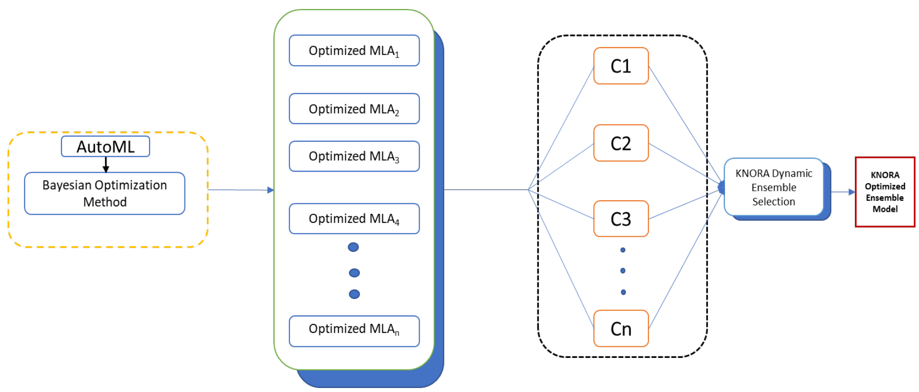

Figure 2: Dynamic optimized machine learning model.

identification of root causes and the unavailability of formal methods that can be used to project the severity of the problem. This study proposed the dynamic model for optimized machine learning algorithms (KNORA-AutoML model).

2 Methodology

Figure 1 demonstrates the ensemble of the optimized MLAs by the traditional ensemble selection strategy.

Moreover, Figure 2 shows the implementation of the KNORA to combine the optimized MLAs. All optimized machine-learning algorithms by the Bayesian Optimization method combined to give out a single improved prediction model.

3 Results

3.1 Student Dropout Features

Figure 3 shows that each feature has an impact on student dropout prediction, such as school distance (28%), household size (17%), means to school (16%), household children (12%), age (7.5%), parents' occupation (4.5%), mother's education (4%), father's education (3.5%), gender (3%), and mother tongue language (2.5%). A higher feature score indicates a significant contribution to predicting student dropouts.

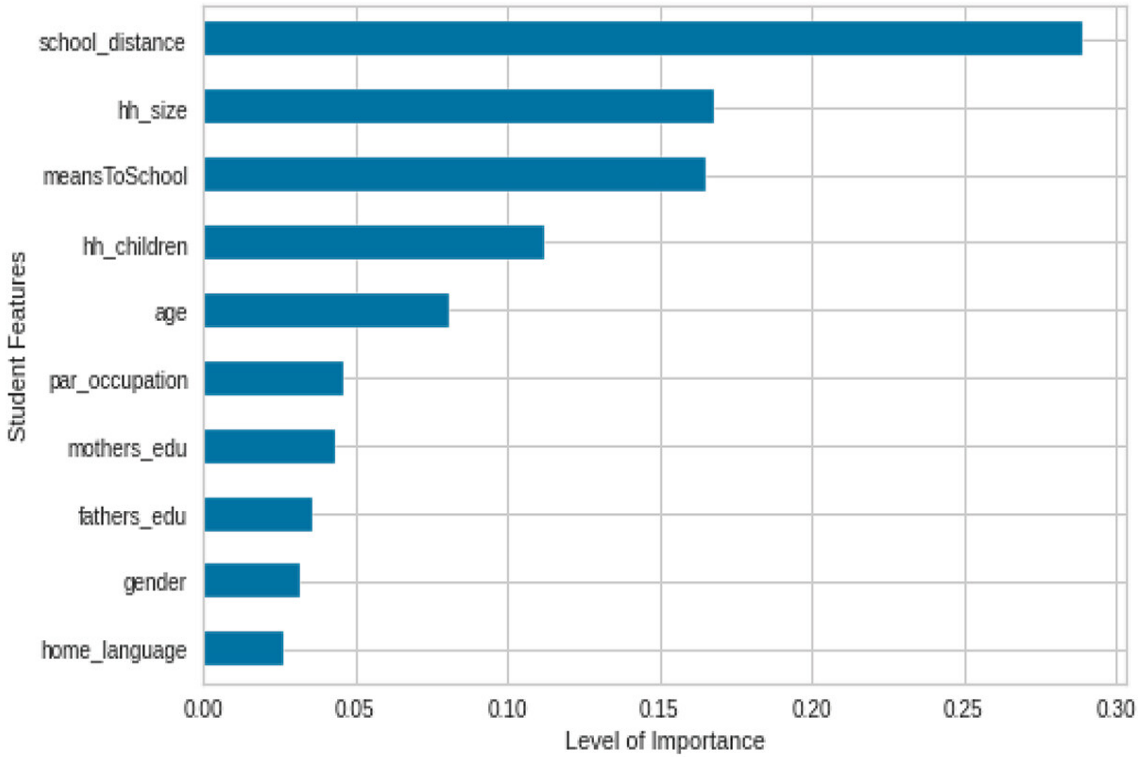

Figure 3: Feature Importance Selection.

3.2 Static Optimized Ensemble Model

Figure 4 shows the prediction results of the optimized machine learning algorithms by the Bayesian Optimization Technique. The ensemble of classifiers applied a traditional selection strategy using the majority-voting rule. The classifiers voted by members to obtain the best performing single classifier applied for predicting student dropouts in developing countries.

3.3 Dynamic Optimized Ensemble Model

Figure 5 shows that the KNORA Optimized model outperforms the static-optimized model ensemble. The KNORA-AutoML model scored 97% in accuracy, precision = 71%, recall = 76%, F1 = 74%, and AUC = 87% when compared to the conventional ensemble of optimized ML models with accuracy = 96%, precision = 70%, recall = 58%, F1 = 64%, and AUC = 78%. KNORA-AutoML model performance increased by 0.6% accuracy, 0.8% precision, 17.8% recall, 9.9% F1, and 8.7% AUC. The proposed approach shows better prediction results compared to the traditional ensemble of optimized machine learning algorithms/models.

4 Conclusion

In many secondary schools in developing countries, children leave school for various reasons that are hard to pinpoint precisely. The severity of the issue has been lessened in developing countries through several programs, including the Big Results Now projects, Free Education for All, No Child Left Behind, the Secondary Education Development Programme, and machine learning prediction models. Due to incorrect root cause identification and a lack of formal methods that may be used to gauge the severity of the problem, dropout rates continue to be a problem, especially in secondary schools. The KNORA-AutoML model outperforms the conventional ensemble learning model and the static ensemble of optimized models using AutoML. Moreover, the findings show that each feature has a significant impact on student dropout prediction; for example, school distance is 28%, household size is 17%, means to school is 16%, household children is 12%, age is 7.5%, parents' occupation is 4.5%,

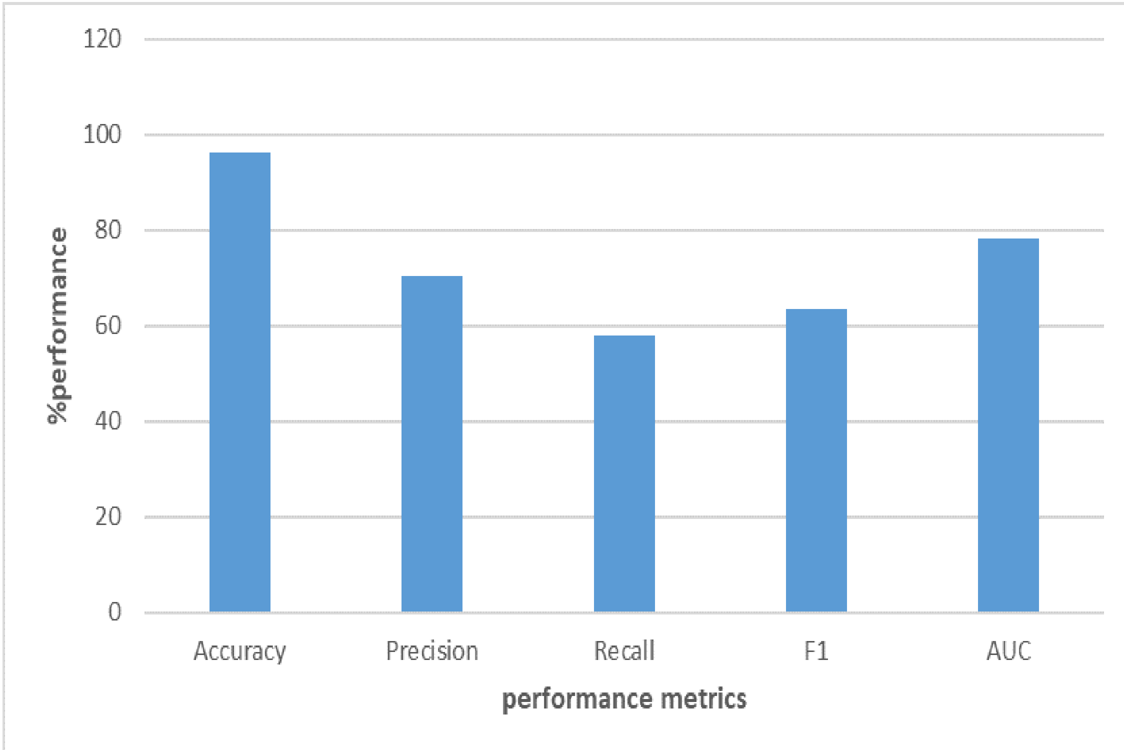

Figure 4: Results for Static Optimized Ensemble Model.

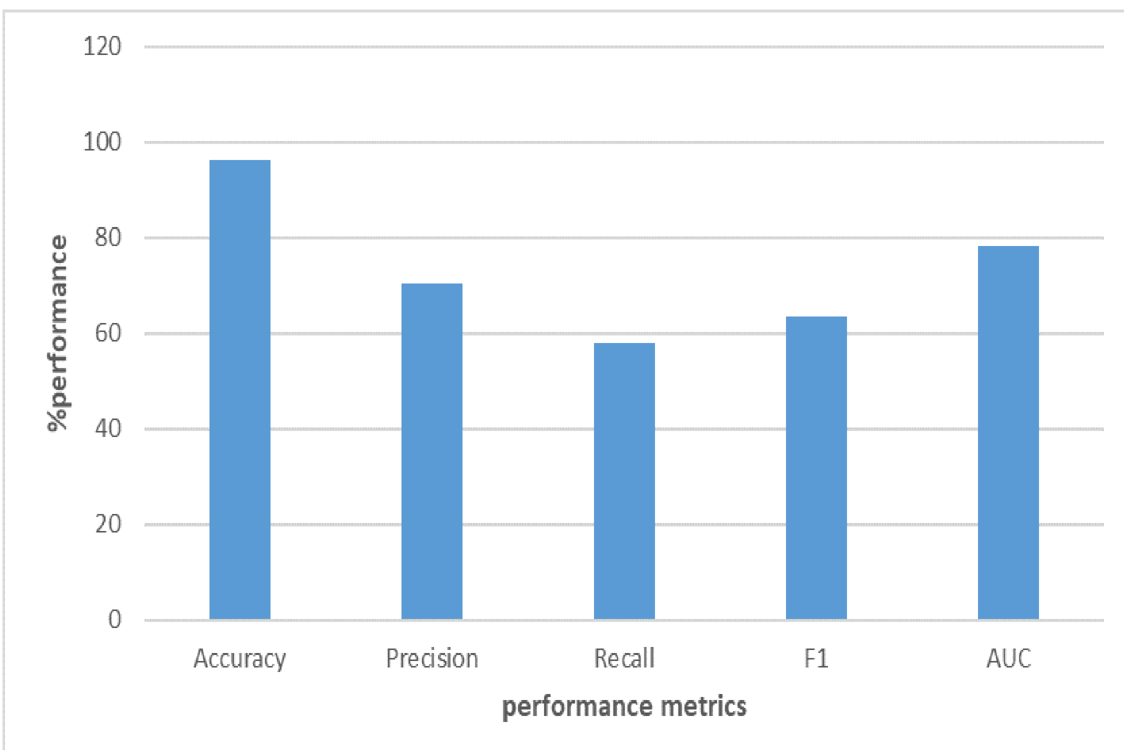

Figure 5: Results for Static Optimized Ensemble Model.

mother's education is 4%, father's education is 3.5%, gender is 3%, and mother tongue language is 2.5%. Therefore, the proposed prediction model emphasizes closely following up on the suggested features and proper planning for early intervention.

References

Aguiar, E. (2015). Identifying Students at Risk and Beyond: A Machine Learning Approach (Issue July) [University of Notre Dame]. https://curate.nd.edu/downloads/w089280479b.

Bibi, T. (2018). Factors Affecting Dropout Rate at Secondary School Level in Private Schools of Punjab, Pakistan. *International Journal of Management Sciences and Business Research,* **7***(4), 1–7.*

Chareonrat, J. (2016). Student Dropout Factor Analysis and Trend Prediction using Decision Tree. *Journal of Science and Technology,* **23***(2), 187–193.*

Lee, S., & Chung, J. Y. (2019). The Machine Learning-Based Dropout Early Warning System for Improving the Performance of Dropout Prediction. *Applied Sciences,* **9***(15), 3093. https://doi.org/10.3390/app9153093.*

Mduma, N., Kalegele, K., & Machuve, D. (2019). Machine Learning Approach for Reducing Students Dropout Rates. *International Journal of Advanced Computer Research,* **9***(42), 156–169.* https //doi.org/10.19101/ijacr.2018.839045.

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

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

Sara, N. B., Halland, R., Igel, C., & Alstrup, S. (2015). High-School Dropout Prediction using Machine Learning: A Danish Large-scale Study. ESANN 2015 Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 22–24.

UNESCO-Tanzania. (2015). National Commissions for United Nations Educational, Scientific and Cultural Organizations (UNESCO). https://unesdoc.unesco.org/ark:/48223/pf0000244391_eng.

UNICEF-Tanzania. (2018). Global Initiative on Out of School Children: Tanzania Verification of the Out of School Children Study (Issue March).

URT. (2008). The United Republic of Tanzania. Education Sector Development Programme 2008-2017.

World Bank. (2014). How Tanzania Plans to Achieve "Big Results Now" in Education. http://www.worldbank.org/en/news/feature/2014/07/10/how-tanzania-plans-to-achieve-big-reforms-now-in-education.