Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20250403018
Volume 4, Issue3 (September-December 2025), PP: 94-99.
www.indjcst.com
ISSN No: 2583-5300

Evaluating Mitigates of Primary School Dropout Risk Using Machine Learning in Narok West Sub-County, Kenya

Sylvia Cherop[1], Emma Anyika[2], James Obuhuma[3]
[1,2]Department of Computing and Mathematics, Co-operative University of Kenya, Kenya.
[3]Department of Mathematical Sciences, cooperative University, Kenya.

To Cite this Article: *Sylvia Cherop[1] Emma Anyika[2] James Obuhuma[3], "Evaluating Mitigates of Primary School Dropout Risk Using Machine Learning in Narok West Sub-County, Kenya", Indian Journal of Computer Science and Technology, Volume 04, Issue 03 (September-December 2025), PP: 94-99.*

Abstract: *This paper used machine learning in assessing the mitigates of primary school dropout risk in Narok West Sub- County, Kenya. Although the use of bursaries, school feeding and community sensitization has been long held, current interventions are reactive meaning that they deal with dropouts once they stop attending school. The predictive modeling to predict dropout and inform preventative action developed using structured field survey (n= 1,000) with Monte Carlo simulation extending to 10,000 records. Three classifiers, namely, Random Forest, XGBoost, and Support Vector Machine were trained on an 80/20 split with five-fold cross-validation and measured in terms of accuracy, precision, recall, F1-score, and ROC-AUC. XGBoost has obtained the best results (AUC = 0.804; F1 = 0.771), which makes it the model that has been validated. The findings of Chapter Four have revealed that financial considerations prevailed in risk dynamics: bursary receipt and bursary amount had a significant negative effect on dropout whereas monthly fees donations and traveling a long distance contributed to the level of dropout. The welfare programs like school feeding, meals per day and community participation were identified as the important protective factors. To make sure the results could be interpreted, explainable AI methods such as permutation importance, SHAP values, and partial dependence plots revealed both the importance and direction of the influence of every factor, not just in prediction but actionable insights. Its results show that financial strain mediated by structural and social supports is the leading cause of dropout, and that predictive analytics can offer policy-makers evidence-based drops to intervene. The integration of such models into the education planning provides a proactive channel of maintaining learner retention and enhancing equity in the rural schooling. The evaluation work helps construct an education system that stands resilient along with technology development and social fairness in Narok West Sub County.*

Key Word: *dropout risk; mitigates; Random Forest; SHAP; Narok West Sub-County; explainable AI*

I.INTRODUCTION

School dropout has been a thorn in the flesh of Sub-Saharan Africa, especially in rural Kenya where education retention keeps being compromised by socioeconomic factors, cultural beliefs, and low infrastructure levels (UNICEF, 2022). Narok West Sub-County is where even with various governmental interventions, which include Free Primary Education (FPE), bursaries and school feeding programs, the dropout rate is very high (World Bank, 2020). These interventions have been more reactive in nature, that is, intervening after the learners have gone bad, but not in a proactive and preventative manner. This means that policymakers and educators do not have the necessary timely tools to detect at-risk learners and implement specific interventions that might help to prevent dropout before it takes place (Ministry of Education, 2021).

The advent of the machine learning offers such a chance to fill this gap. Predictive analytics, especially when using supervised learning algorithms like the Random Forest, Support Vector Machines (SVM), and Gradient Boosting Machines (GBM), have proved to be useful in predicting the dropout probability and examining and evaluating the effect of mitigating factors in the global contexts (Kipuri and Ridgewell, 2022; Zhou, Wang, and Liu, 2022). These models can be used to identify complex, non-linear interactions in statistics which traditional statistical techniques cannot. Moreover, interpretability can also be improved by the incorporation of Explainable AI methods, including Shapley Additive Explanations (SHAP), to not just determine the relative importance of predictors but also whether they have a positive or a negative impact- that is, whether they increase the risk of dropouts or have a protective effect (Odhiambo, Wanjiku, and Njenga, 2021).

Predictive modeling methods can thus be used to enhance educational resilience, or the capacity of learners and institutions to adjust and succeed despite socioeconomic and cultural changes (UNESCO, 2021). The effectiveness of predictive systems is supported by international experience: in China, Brazil, and India, it is possible to redistribute resources and minimize the dropout rate by using machine learning to forecast dropouts with the help of early warning systems (Zhou et al., 2022).

However, in Kenya, and Narok West Sub-County, in particular, predictive modeling is not fully leveraged, and any current interventions are not supported by the validation of the data (Mwangi, Ndungu, and Otieno, 2022).

The proposed research will add up to that gap by creating a machine learning predictive model to assess and prioritize the extent of mitigates as bursaries, school feeding programs, community sensitization, and economic empowerment initiatives in influencing dropout risks in Narok West Sub-County. The study provides a semitransparent, evidence-based comprehension of developing proactive, cost-effective, and context-specific intervention by incorporating model-based baselines with model-agnostic validation and explainable AI to keep education stakeholders on track. This way, it would be aligned with the wider Kenya education policy objectives and Sustainable Development Goals (SDG 4: Quality Education and SDG 5: Gender Equality) (UNESCO, 2021).

1.1 Statement of the Problem

Primary school dropout problem remains extremely serious within Narok West Sub County because marginalized student populations encounter many obstacles when trying to finish their education. The dropout rate persists high in Narok West Sub County despite Free Primary Education (FPE) and school feeding programs and infrastructure improvements because of socioeconomic limitations combined with cultural elements and insufficient data analysis systems for monitoring (Ministry of Education, 2021). Various approaches to lower dropout rates have beenattempted but their effectiveness remains unproven because of inadequate data evaluation models (UNICEF, 2022). The current approaches to dropout prevention function on a reactive basis because preventive interventions are executed after discovered cases instead of making proactive predictions (World Bank, 2020). The improper allocation of resources arises from ineffective initiatives because monitoring data remains unclear about intervention results (Mwangi, Ndungu, & Otieno, 2022). Without predictive analytics in Kenya's education system policymakers along with teachers find it difficult to detect at-risk students early and deploy intended cost-effective interventions (Zhou, Wang, & Liu, 2022). Research demonstrates that machine learning predictive models accurately predict dropout risks in order to enhance intervention approaches (UNESCO, 2021). The predictive analytics approach used by nations including India and China and many Sub Saharan African countries proves successful for risk assessment combined with optimized dropout prevention strategies. Narok West Sub County has not developed a predictive analytical system to determine which deterrents such as government action combined with community outreach and economic programs provide the most effective risk reduction against student dropout. Without this model decision- makers use non-specific interventions which fail to resolve the specific difficulties students face (Odhiambo, Wanjiku, & Njenga, 2021).

1.2 Objective of the study

To determine the most influential mitigates on school dropout risk in Narok West Sub County.
To develop a machine learning weight predictive model for mitigates on primary school dropout risk. To train and validate the weight predictive model.
To predict the weights of mitigates of school dropout risks using the validated model

II.LITERATURE REVIEW

2.1 Global, regional, and local Papers on Machine Learning to predict dropouts.

Machine learning (ML) has been extensively used to predict school dropout all over the world by analyzing academic, attendance, behavioral and socio-economic data, with random forest (RF), gradient boosting machines (GBM), support vector machines (SVM), and algorithms like deep learning, all performing better than conventional regression in classification accuracy (Zhou et al., 2022; Wang et al., 2022). Nevertheless, numerous studies done worldwide focus on prediction rather than intervention design or prioritization of mitigation options. The adoption continues to increase, but unevenly across Sub- Saharan Africa; GBM, RF and SVM have been shown to be highly accurate in South Africa and West/East African contexts, but issues of data completeness, model interpretability and capacity remain (UNESCO, 2021; Govender et al., 2022; Jha and Kelleher, 2019; Santos and Moura, 2021). The Kenyan evidence is relatively sparse and traditionally has based on traditional statistics; recent investigations indicate that they are more accurate on ensembles but continue to under- specify the lever of mitigation with the most significant effects on retention (Odhiambo et al., 2021; Mwangi et al., 2022).).

2.2 Traditional Approaches to Dropout Prediction

Previously used methods, such as descriptive statistics, logistic regression, and survival analysis, were helpful in providing baseline profiling of risk, but made assumptions of linearity, relied on pre-defined more limited sets of variables, could not learn new information, and had limited ability to adapt to new data. Due to this reason, they are under fitting intricate, multi- factor pathways that lead to disengagement and exit (UNESCO, 2021; World Bank, 2020; Zhou et al., 2022). Machine.

2.3 Learning for Dropout Prediction and Mitigation Evaluation

Modern ML pipelines integrate feature engineering, cross-validation and explainability to detect at-risk students early and to prioritize interventions. RF, GBM and SVM routinely surpass classical baselines; explainable AI methods such as SHAP attribute risk to actionable drivers and help rank the relative effect of mitigations. Evidence from Brazil and regional pilots shows scholarships and targeted support often carry the highest weights, with community awareness and mentorship adding incremental gains (Almeida et al., 2021; Mwangi et al., 2022; World Bank, 2020).

2.4 The use of Predictive Analytics in Education.

Through predictive analytics, Early Warning Systems (EWS) personalized case management and policy targeting are all in motion. At the system level, the model outputs are applied to allocate the resources to the high-impact lever in ministries and counties and dashboards in schools to identify individual students and trigger the support plan (UNESCO, 2021; Govender et al., 2022).

2.5 Identified Gaps in the Literature

Despite various researches on the issue of school dropout in the world and Kenya, there are still very obvious gaps in current literature. Numerous publications also focus on the identification of risk factors and pay minimal attention to the effectiveness or the weight of interventions like bursaries, feeding, or community sensitization. Kenyan research has been based primarily on either statistical regression or qualitative data, but neither of these approaches is useful in predicting the outcome, which requires a level of precision that only non-linear interaction between variables can capture a dropout process. Besides that, predictive models do not generally have explainability; the tools of SHAP or permutation importance, which visualize the ways and reasons factors contribute to dropout risk, are not typically used. There is a scarcity of studies that model their work to the rural sub-county environment Narok West, where the socio-economic reality, cultural patterns, and the infrastructural impediments vary greatly when compared to the urban environment. Moreover, most of the existing models are too specific on the accuracy without mapping the results to operation planning tools that can be used by policymakers in the allocation of resources. To fill these gaps, it is necessary to have context-specific explainable machine learning models that can predict the dropout risk in the most accurate way and offer transparent and actionable insights to make decisions.

III.METHODOLOGY

3.1 Research Design

The article uses a quantitative approach that uses predictive analysis and description, incorporating primary data gathering and machine-learning analytics in the research of primary school dropout in Narok West Sub-county. This design was undertaken because: structured data would allow objective measurement of the determinants of dropout; predictive modelling to weight early-warning and mitigation would be feasible; and policy responsive, explainable results, translatable to program action, would be generated.

3.2 Data Source

Primary data were collected in Narok West using structured questionnaires and school records in the primary schools. The original target sample size (calculated using the Yamane formula) was 399, and the fieldwork provided 1,000 complete records to enhance the capacity of the model to learn, we balance the classes, and examine the credible relationship of features, we used Monte Carlo simulation to increase the dataset size to 10,000 records and retained the observed distributions and plausible relationships.

3.3 Sampling and preprocessing

The initial sample of the study consisted of 399 learners who were gathered through the use of structured questionnaires. It was further extended to 1,000 records by collecting more field data and further extended to 10,000 records by Monte Carlo simulation to provide the necessary statistical strength to machine learning training. The analysis of 9796 complete cases (16 variables) was carried out after cleaning and validation. Preprocessing involved processing of missing values, encoding of categorical variables, standardization of numeric features and consistency checking. The set would be divided into 80 percent training and 20 percent testing, and to reduce the chances of overfitting, cross-validation would be done to the training set five times.

3.4 Analytical Approach

This paper used three machine learning models as baselines, random forest, and XG Boost as the main ensemble learner and Support Vector Machine, which were used to compare the three. They were optimized using five-fold cross-validation, which was interrupted before the models overfitted. The issue of class imbalance was fixed to achieve fairness, whereas the robustness was proved by permutation importance and sensitivity analysis. In order to make it more interpretable, SHAP values were utilized which provided the direction and magnitude of the effects of each feature and thus, the most decisive mitigates of school dropout risk was identified in a transparent and reliable manner.

3.5 Model Evaluation

To determine model performance, various complementary metrics were used, including accuracy, precision, recall, F1-score, and ROC-AUC, to obtain an accurate and balanced predictive strength and specificity. In the comparisons of the three evaluations, XG Boost was near the highest accuracy (74%) and ROC-AUC (0.804) and was also better at the accuracy-recall trade-offs. Random Forest came next, with mostly similar but slightly inferior values, and the Support Vector Machine was behind, with poorer discrimination and sensitivity to the imbalance of classes. Notably, the cross-validation and hold-out testing showed that the results were stable; hence, XG Boost as the most reliable and generalizable model used to predict the risk of school dropout in the study setting was justified. the chosen model was subjected to SHAP analysis to generate understandable mitigation weights.

3.6 Practical Implications

The suggested pipeline transform evidence into practice-based insights to education policy makers and practitioners. It allows the learners at risk of dropping out to be identified early and timely and tailored interventions to be provided. The framework gives priorities a clear focus on the intervention planning because of ranking the mitigations like bursaries, feeding programs, community sensitization, and economic empowerment initiatives based on the level of impact they are expected to have on reducing risk. Moreover, it educates in the school and county-level standard operating procedures, budget allocations and monitoring dashboards design, and provides decision-makers with data-driven instruments to implement proactive and sustainable retention strategies in advance.

IV.RESULTS

4.1 Introduction

In this segment, the researcher provides the research findings, which has the most influential school dropout risk mitigates in Narok West Sub-County. The analysis is based both on model-based and model-agnostic methods, and the refined rankings are obtained with the help of XGBoost and confirmed with the help of Explainable AI methods like the permutation importance and SHAP values. Combined, these findings give a concise view of the factors that elevate or lessen the risk of dropout, and with any predictive power, they can be interpreted. The section wraps up with insights that are used in the future analysis coupled with setting the basis of practical recommendations and the future research directions.

4.2 Analytical Approach to Feature Importance

To achieve the stability and interpretability, a layered strategy was used to derive the feature importance. In the first iteration of a Random Forest model, a solid baseline ranking was obtained, which represented complex and non-linear correlations in the data. To ensure that these were not the artifacts of the model, Permutation Importance was run on a test split that was held out, which provides a model-agnostic test, as it quantifies the performance decrease when each feature was shuffled. This combination struck a balance between predictive power and reliability, and SHAP values provided some direction, as to which things predisposed to dropout and which ones provided protection.

4.3 Model performance

Table1. Model evaluation results

Performance Comparison (Objective 2):

	AUC	Accuracy	Precision	Recall	F1
Random Forest	0.789824	0.729082	0.753623	0.785080	0.769030
XGBoost	0.803809	0.740306	0.782251	0.759325	0.770617
SVM	0.750502	0.657653	0.717703	0.666075	0.690926

4.4 Training and Validation of the model

The cross-validation results pointed to some certain disparities in the consistency of performance. Random Forest generated AUC of 0.776, 0.782, 0.779, 0.799, and 0.750 with a mean AUC of 0.777 /0.016. The average AUC of 0.807 with a standard deviation of 0.010 is found with XGBoost, which has a higher fold score of 0.807, 0.802, 0.814, 0.820, and 0.791. The highest mean AUC of SVM was 0.811 ± 0.011 with fold values between 0.792 and 0.825. Nevertheless, though SVM was a bit better at XGBoost in mean AUC, its variability was higher indicating that it performed more variably across validation splits. Notably, there was a smaller standard deviation of XGBoost (0.010), which means that the model was more stable than the other two (Random Forest and SVM). of a model to hold discrimination strength as the balance of the training/validation split varies.

Table 2. Model validation results

Model	Fold1 AUC	Fold2 AUC	Fold3 AUC	Fold4 AUC	Fold5 AUC	Mean AUC	Std AUC
Random Forest	0.776	0.782	0.779	0.799	0.750	0.777	0.016
XGBoost	0.807	0.802	0.814	0.820	0.791	**0.807**	**0.010**
SVM	0.809	0.812	0.816	0.825	0.792	**0.811**	**0.011**

4.5 XG Boost Feature Importance (Global Ranking)

The most important global feature of XGBoost was bursary receipt (0.4881), then the distance to school (0.1207), then feeding program participation (0.0700) and finally, the monthly fee contribution (0.0660). Others were community meetings (0.0440). These findings point to financial alleviation, accessibility, and community engagement as the key leverages of mitigating the risk of dropouts.

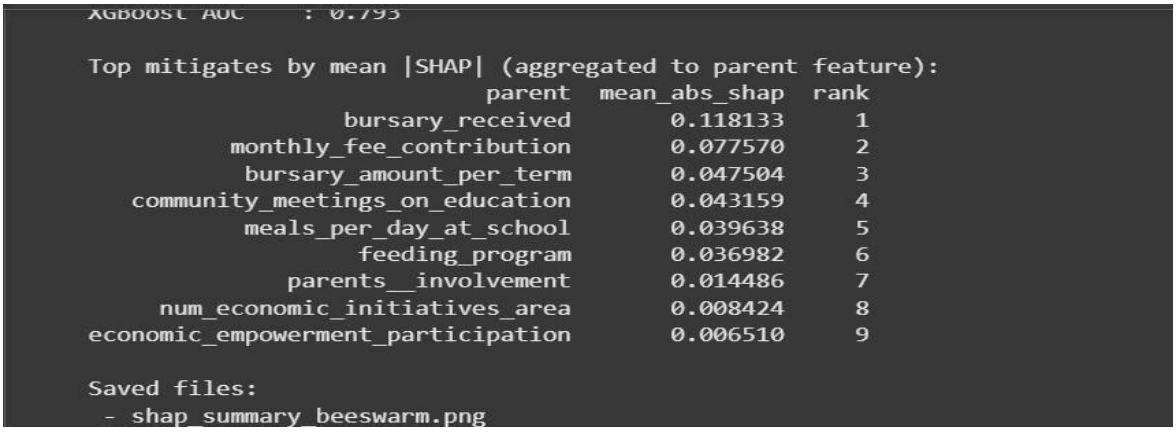
4.6 Explainable AI: Permutation Importance and SHAP.

The most significant mitigates supported by permutation importance were bursary receipt (0.0848) and monthly fee contribution (0.0459) then community meetings (0.0128), bursary amount per term (0.0090), and meals per day (0.0074). SHAP analysis helped to further understand: bursary receipt (mean = 0.1181) and bursary amount per term (0.0475) decreased the dropout risk, and monthly fee contribution (0.0776) and distance to school increased the dropout risk. There were significant risk reduction effects of feeding programs, meals per day and community meetings, but less significant but protective effects of parental involvement (0.0145) and economic empowerment (0.0065). These methods combined, proved the strength of the results in terms of importance (weight) and directionality (effect).

```
print("AUC:", round(auc, 3))
print("Saved files:")
print(f"- Ranked table CSV: {out_csv}")
print(f"- Bar chart PNG: {bar_path}")

Objective 1: Ranked mitigates by influence (permutation importance)
                                 mitigate  influence_score  rank
2                        bursary_received         0.084790     1
11              monthly_fee_contribution         0.045934     2
3          community_meetings_on_education        0.012789     3
1                   bursary_amount_per_term        0.009033     4
10                  meals_per_day_at_school        0.007421     5
7                          feeding_program        0.003465     6
6      economic_empowerment_participation        0.003465     7
13                    parents__involvement        0.002662     8
12          num_economic_initiatives_area         0.001237     9
AUC: 0.789
Saved files:
```

Figure 1 permutation importance results



```
Top mitigates by mean |SHAP| (aggregated to parent feature):
                            parent  mean_abs_shap  rank
                  bursary_received       0.118133     1
          monthly_fee_contribution       0.077570     2
           bursary_amount_per_term       0.047504     3
   community_meetings_on_education       0.043159     4
           meals_per_day_at_school       0.039638     5
                   feeding_program       0.036982     6
              parents__involvement       0.014486     7
     num_economic_initiatives_area       0.008424     8
economic_empowerment_participation       0.006510     9

Saved files:
 - shap_summary_beeswarm.png
```

Figure 2 Shap integration Results

V.CONCLUSION AND RECOMMENDATIONS

5.1 Discussion

The results of this research are consistent and complementary to the extant literature on the socio-economic and structural factors of school dropout. The previous literature has highlighted the importance of financial hardship as a cause of dropout, and the findings of the current study offer a solid quantitative data of bursaries as the most effective deterring factors in school dropouts. Bursary support, both in terms of timing and amount, was found to be directly related to retention by the weight and SHAP rankings, which proved that education subsidies were the most effective intervention in resource-constrained families. The importance of monthly payments to the fees also highlights the precariousness of the learners in the situations where the schools collect extra payments. The fact that even minor household input turned out as a significant predictor of dropout points to the fact that there are indeed hidden costs of education that should be considered as a priority in the policy making. The findings also emphasized the importance of feeding programs that, as well as reducing the household food insecurity, have a direct positive effect on the attendance and retention of schools. This is in line with the studies done before which stated that nutrition is a continuum to learning. The spatial element like distance to school also played a significant role and justified the literature that geographic inaccessibility is one of the best predictors of rural dropout. Additionally, the results indicated that community meetings and parental involvement also had an indirect, albeit significant, impact, which indicates that the social accountability mechanisms are critical to maintaining the learner engagement. Collectively, these results indicate that dropout is a complex process, and all interventions, including financial, structural, and social ones, should work together.

5.2 Conclusion

This research found that the financial and structural pressures are the major factors that lead to dropout risks in Narok West Sub-County. The receipt of the bursary and the amount of the bursary kept the dropout probability down and monthly fee contributions and long commuting distance made it go up substantially. Welfare interventions like school lunch programs, meal per day and community involvement also came out as significant protective factors. Through the use of modern machine learning, especially XGBoost, along with explainable artificial intelligence methods, including SHAP, permutation importance, and partial dependence plots, the research was not only effective in terms of its predictive power but also allowed understanding how each variable impacts the results. These results confirm that a household financial burden reduction, feeding programs, and access barriers mitigation; informed by predictive and interpretable models are the most feasible avenue to learner retention in rural Kenya.

5.3 Recommendations

The results of the research highlight the necessity of policy interventions that would specifically ensure a direct response to financial, structural, and social obstacles to school retention in Narok West Sub-County. The bursary coverage and adequacy will need to be scaled up especially to the learners who face high recurrent fees and travel long distances to the institution and quick referral systems to predictive risk profiles. Varying fee schedules, selective boarding, subsidizing fees targeting the household fee pressure would alleviate it, and providing more transport funds, safe routes, or subsidized waivers would improve access to the remote learners. School feeding programs are to be reinforced with reliable funding and easily tracked, high-absenteeism areas should be at the forefront, and organized involvement of parents and community sensitization programs should be institutionalized as non-portable retention keys. In the future, the predictive insights should be transformed into a routinized, data-driven planning cycle. This comprises of having a sparse feature registry that is audited on a regular basis, school-level dashboards that visually represent SHAP-driven insights, and the articulation of straightforward procedure-to-thresholds that have outcome tracking. The further development of this framework to test the quasi-real-time integration of predictive dashboards and generalize the analysis to other counties facing the problem of dropouts would offer solid proof of the national-level interventions and would guarantee the ongoing enhancement of the school retention policy.

References

1. Fernandes, M., Moreira, C., & Santos, J. (2021). *Explainable dropout prediction using SHAP and XGBoost in higher education.* Journal of Educational Data Science, 18(4), 321-340.
2. Kipuri, N., & Ridgewell, A. (2022). *Education and Socioeconomic Challenges in Kenya*. Oxford University Press.
3. Ministry of Education. (2021). *Kenya Education Sector Report 2021*. Nairobi, Kenya.
4. Mwangi, P., Ndungu, T., & Otieno, J. (2022). *Data-driven approaches to school dropout prevention in Kenya: A policy review.* African Journal of Educational Research, 9(2), 112-128.
5. Odhiambo, G., Wanjiku, J., & Njenga, P. (2021). *The impact of cultural practices on education in Kenya.*
6. African Journal of Education Studies, 12(3), 45-67.
7. Santos, A., & Moura, F. (2021). *Random Forest for dropout prediction: A case study in secondary education.*
8. Machine Learning in Education, 27(3), 198-210.
9. UNESCO. (2021). *The role of predictive analytics in education policy-making:* A global review. United Nations Educational, Scientific and Cultural Organization.
10. UNICEF. (2022). *School dropout trends and policy interventions in sub-Saharan Africa*. United Nations Children's Fund.
11. Wang, L., Zhang, H., & Li, T. (2022). *Predicting student dropout using Gradient Boosting Machines: A case study in China*. IEEE Transactions on Artificial Intelligence, 9(1), 23-37.
12. World Bank. (2020). *Addressing school dropout through data-driven approaches*. Washington, DC: World Bank Publications.
13. Zhang, Y., Lin, P., & Liu, D. (2020). *Deep learning in dropout analysis: Challenges and insights.* Journal of AI in Education, 15(2), 101-120.
14. Zhou, M., Wang, F., & Liu, X. (2022). *The role of machine learning in student retention: A systematic review.*
15. Educational Data Science Journal, 10(1), 78-95.
16. Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press.
17. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.