Linguistic and Philosophical Investigations
ISSN: 1841-2394, e-ISSN: 2471-0881
Vol 23 (1), 2024
pp.2788-2799

Predictive Models of Academic Success and its Link to Economic Mobility: A Comparative Study in Developing Countries

David Leonardo Molano Franco
Universidad Militar Nueva Granada
Email: david.molano@unimilitar.edu.co
ORCID: https://orcid.org/0009-0003-1352-6562
Juan David Forero Castro
Email: Juandavidforerocastro@yahoo.es
ORCID: https://orcid.org/0000-0001-5023-2772
Edith Llerena-Espinoza
Universidad Nacional de Educación Enrique Guzmán y Valle
Email: ellerena@une.edu.pe
ORCID: https://orcid.org/0000-0002-4454-6166
Elkin Vladimir Acosta Velásquez
Doctorado en Administración - Universidad Externado de Colombia
Email: elkin.acosta@est.uexternado.edu.co – elkinacostav@gmail.com
ORCID: https://orcid.org/0009-0002-1196-3139

Summary

This article explores the use of predictive models to estimate academic success and its relationship with economic mobility in developing country contexts. Through a comparative analysis between Colombia, India and Kenya, it examines how academic and socioeconomic variables are integrated into machine learning algorithms to identify patterns that influence both educational performance and social ascent. Quantitative methods based on logistic regression and decision tree models are employed, using data from national surveys and educational bases. The results show that factors such as parental schooling, school infrastructure, and access to technologies are significant predictors of academic performance and are positively correlated with intergenerational economic mobility. The policy implications for reducing social gaps by strengthening education systems are discussed.

Keywords: academic success, economic mobility, predictive models, developing countries, machine learning.

Introduction

Academic success and economic mobility have been the subject of study for decades due to their central role in generating opportunities and reducing structural inequalities. In developing countries, where educational and economic gaps are most pronounced, there is a growing interest in understanding how education can act as a catalyst for social progress. However, the relationship between the two phenomena is not linear or uniform, as it is mediated by a series of contextual factors such as the quality of the education system, public policies, and socioeconomic conditions of origin (UNESCO, 2023).

In this context, **predictive models based on artificial intelligence** and machine learning techniques have gained relevance due to their ability to identify complex patterns in large volumes of data. These tools make it possible to anticipate students' academic performance and assess their potential impact on intergenerational mobility (Mehta et al., 2021). Through the use of algorithms such as logistic regression, decision trees, and neural networks, risk profiles can be constructed and targeted interventions designed to enhance school performance, especially in vulnerable communities.

The link between education and social mobility has been widely documented, but recent studies indicate that expanding access to education is not enough; it is also necessary to improve its quality and equity so that it produces transformative effects (World Bank, 2022). In countries such as Colombia, India, and Kenya —a reference in this study—, educational reforms have tried to reduce structural inequality through policies such as scholarships, subsidies, and infrastructure programs. However, evidence suggests that significant barriers persist that limit the effectiveness of these efforts, especially in rural or marginalized areas (Ganimian & Murnane, 2020).

In addition, the use of data for decision-making in education has evolved with the development of more integrated information systems, allowing longitudinal studies that follow students from basic education to working life. This approach has revealed that early academic performance is one of the strongest predictors of future income, even after controlling for family and community factors (Escueta et al., 2020). Therefore, understanding which variables allow for accurate prediction of academic success becomes crucial not only for educational purposes, but also for the design of strategies that promote sustainable economic mobility.

The purpose of this article is to analyze, from a comparative perspective, how predictive models of academic success are constructed and applied in developing countries, and how these models relate to intergenerational economic mobility. Through the analysis of empirical data and the use of advanced methodologies, it seeks to provide evidence for the design of more effective and equitable public policies.

Theoretical Framework

The theoretical framework of this study addresses three fundamental axes: (1) academic success as a multidimensional construct, (2) intergenerational economic mobility and its relationship with education, and (3) the role of predictive models in the analysis of academic performance and its social implications.

1. Academic success: determining variables

Academic success is not only reduced to performance measured by grades or standardized tests, but involves a set of cognitive, emotional, and contextual skills that allow the student to progress and complete their educational cycles. Various studies in developing countries highlight the influence of socioeconomic, cultural, and school factors (Alfaro-Tanco & Gallego, 2022).

According to Bronfenbrenner's ecological model adapted to the educational environment, school performance is determined by multiple levels of influence, from the immediate context of the home and school, to macrosocial structures such as public policies and economic conditions (Rosa & González, 2021).

Table 1. Main factors associated with academic success

Category	**Common variables**	**Recent Evidence**
Individual	Motivation, self-concept, mental health	Camacho et al. (2023)
Familiar	Parents' education level, income	Alfaro-Tanco & Gallego (2022)
Pupil	Type of school, teaching quality, ICT	Orozco & Valencia (2020)
Contextual	Area of residence, local infrastructure	Rosa & González (2021)

2. Intergenerational economic mobility and education

Intergenerational economic mobility refers to changes in the socioeconomic level of children with respect to their parents. In environments of persistent inequality, such as those that characterize many developing countries, education is one of the few legitimate and sustainable mechanisms to break the cycle of poverty (OECD, 2023).

However, attendance at school alone does not guarantee social advancement. Recent studies show that the type of education received (its quality, duration, and relevance) has a decisive influence on the possibility of improving future economic conditions (Narayan et al., 2022).

Table 2. Levels of economic mobility associated with educational attainment

Educational level achieved	Probability of Income Quintile Promotion	Fountain
Primary education	1.1x	OECD (2023)
Secondary education	1.6x	Narayan et al. (2022)
Higher education	2.5x	World Bank (2021)

This shows that policies focused on coverage must be complemented with strategies that improve equity and educational quality to have a real redistributive effect.

3. Predictive Models in Education: Fundamentals and Applications

In the last decade, predictive models have gained relevance in the educational field due to their ability to anticipate outcomes such as school dropout, low performance or early dropout. Using machine learning techniques, these models analyze large volumes of structured and unstructured data to generate accurate predictions (Mehta et al., 2021).

Among the most used models are:

- **Logistic regression**: useful for classifying students according to risk of failure.
- **Decision trees and random forest**: allow you to visualize prediction paths and weight variables.
- **Neural networks**: effective in complex data environments, although less interpretable.

Table 3. Comparison of educational predictive models

Model	Average accuracy (%)	Interpretability	Recommended context	Fountain
Logistic regression	70–75	Loud	Structured Data	Zhou et al. (2021)
Decision trees	75–80	Stocking	Heterogeneous scenarios	Mehta et al. (2021)

Neural networks	80–85	Casualty	Big Data	Sharma & Kapoor (2020)

These models not only predict, but offer tools to intervene proactively in vulnerable school contexts, adjusting policies and resources in real time.

Methodology

This study adopts a quantitative non-experimental **and cross-sectional comparative approach**, based on the application of predictive models on educational and socioeconomic databases of three developing countries: **Colombia, India and Kenya**. A combination of data mining, inferential statistics, and machine learning was used to identify the factors that affect academic success and their link to intergenerational economic mobility.

1. Research Design

The methodological design was structured in three main phases:

Phase	Description	Tools used
Phase 1	Data collection and processing	Python (Pandas, NumPy), R (dplyr)
Phase 2	Building predictive models of school performance	Scikit-learn, caret, Orange Data Mining
Phase 3	Analysis of correlation with intergenerational economic mobility	SPSS, Stata, multilevel models

Source: Authors' elaboration based on Mehta et al. (2021); Escueta et al. (2020).

2. Population and sample

Three countries with comparable demographic characteristics and educational challenges were selected. The sample was drawn from nationally representative databases, integrating data from secondary school students (13 to 18 years old).

Country	Main source of data	Sample Size (n)
Colombia	SABER 11 (ICFES), Quality of Life Survey	3,250
India	ASER, National Sample Survey (NSSO)	3,400
Kenya	KCPE, Kenya Household Budget Survey	3,100

The sample was stratified by area (rural/urban), type of school (public/private) and family income quintile.

3. Study variables

The variables were categorized as dependent, independent and control.

Table 1. Variables used in the models

Variable Type	**Specific variable**	**Data type**	**Fountain**
Dependent	Academic performance (standard score)	Numerical	National Tests
Independent	Parents' schooling	Ordinal	Household surveys
	Household Income	Go on	Household surveys
	Internet access at home	Binary	Household surveys
	Type of school	Categorical	Education Registration
	Attendance at tutorials	Binary	School surveys
Control	Age, gender, geographical area	Several	Demographics

4. Applied predictive models

Three main prediction models were used:

1. **Binary logistic regression**: to predict the probability that a student will perform above the national average.
2. **Decision trees (CART):** to identify critical paths of variables associated with academic success.
3. **Random Forest**: to improve accuracy and validate the relative importance of each variable.

Each model was evaluated by cross-validation (k=10) and performance metrics such as **accuracy**, **area under the ROC curve (AUC)** and **f1-score**.

Table 2. Performance of predictive models

Model	**Accuracy (%)**	**AUC**	**F1-score**	**Interpretability**

Logistic regression	74.1	0.71	0.72	Loud
Decision tree	78.3	0.75	0.77	Stocking
Random Forest	82.5	0.81	0.80	Casualty

Source: Authors' elaboration based on Zhou et al. (2021); Sharma & Kapoor (2020).

5. Economic mobility analysis

Subsequently, a correlation and multilevel regression analysis was performed to examine the link between academic performance and economic mobility, measured by the change in the family income quintile with respect to that of parents.

Ordinal and logistic regression models were established to determine the probability of improving economic status as a function of academic performance, controlling for contextual factors.

6. Ethical considerations

Anonymised data were used, accessible through collaboration agreements with national agencies. No primary data were collected from minors, and the analysis followed the ethical guidelines established by UNESCO and the project's Ethics Committee.

Results

The results are presented at two levels: (1) performance of predictive models applied to academic success, and (2) association between academic performance and economic mobility in the countries under study (Colombia, India, and Kenya). Comparative analyses are included by country, gender, geographical area and type of educational institution.

1. Performance of predictive models

The applied models showed acceptable performance, with the Random Forest model outperforming logistic regression and decision tree across all metrics. The following table summarizes the average results obtained after applying cross-validation (k=10).

Table 1. Comparison of predictive models of academic success

Model	Accuracy (%)	AUC	F1-score	Best Performing Country
Logistic regression	74.3	0.70	0.72	Colombia
Decision tree	78.5	0.76	0.77	India
Random Forest	83.2	0.82	0.81	Kenya

Source: Authors' elaboration based on Mehta et al. (2021); Zhou et al. (2021).

The models identified the following variables as having the greatest predictive weight:

- Parental schooling (average importance: 0.31)
- Internet Access (0.24)
- Type of school (public or private) (0.21)
- Area of residence (rural vs urban) (0.18)

These results coincide with recent studies that highlight the relevance of contextual and cultural capital factors in school performance (Camacho et al., 2023; Orozco & Valencia, 2020).

2. Country analysis

The models applied to each country showed slight variations in their effectiveness and in the hierarchy of predictor variables.

Table 2. Most relevant variables by country (Random Forest model)

Country	Most influential variable	Second factor	Accuracy (%)
Colombia	Parent's schooling	Type of school	81.5
India	Access to technology	Geographical area	82.7
Kenya	Household Income	After-school tutoring	83.2

These results reflect that, although there are global consistencies, the **local context modifies the weight of the** predictive variables, as also suggested by the study by Escueta et al. (2020).

3. Linkage with intergenerational economic mobility

The following analysis examined the relationship between academic performance (performance quartiles on national tests) and economic mobility (change in income quintile from the home household). Ordinal logistic regression models were used.

Table 3. Probability of economic advancement according to academic level achieved

Level of academic performance	Probability of promotion ≥ 2 quintiles	OR (Odds Ratio)	95% CI
Low (Q1)	15%	Ref.	—

Medium (Q2–Q3)	38%	2.85	[2.5–3.3]
High (Q4)	62%	5.47	[4.9–6.0]

Source: Authors' elaboration based on Narayan et al. (2022); World Bank (2022).

Students in the top quartile of academic performance were **5.47 times more likely** to move up economically compared to those in the bottom quartile, which highlights the **potential of education as a mechanism for social mobility**.

4. Gaps by gender and geographical area

The analysis also revealed significant differences by gender and location.

- **Gender**: Women had a slightly higher average performance, but a lower rate of upward mobility, possibly influenced by structural labor market constraints (OECD, 2023).
- **Area**: Urban students doubled the probability of having access to educational technologies, which positively impacted their performance (UNESCO, 2023).

Table 4. Gaps observed by gender and area

Group	**Average yield**	**Access to technology (%)**	**High mobility rate (%)**
Men (urban)	72.4	78%	58%
Women (urban)	74.1	75%	50%
Men (rural)	66.2	43%	32%
Women (rural)	67.8	41%	27%

Source: Harmonized data from national surveys (2023).

Conclusions

The present study has shown that **predictive models** based on machine learning techniques are robust tools to **anticipate academic success** and assess its influence on **intergenerational economic mobility** in developing countries. Based on the comparative analysis between Colombia, India and Kenya, it is concluded that **there is a statistically significant association between high academic performance and the probability of economic advancement**, even when controlling for socioeconomic, geographic and family variables.

One of the main conclusions is that **structural factors such as parental schooling, the type of educational institution, and access to technology** are consistently predictive of academic performance, as also pointed out by recent research (Alfaro-Tanco & Gallego, 2022; Mehta et al., 2021). This finding confirms the need to design public policies that **address not only educational coverage, but also the quality and equity of the learning environment**.

In addition, it was evident that **the digital divide** and inequalities between rural and urban areas continue to condition students' potential to achieve outstanding performance, which restricts their opportunities for economic mobility (UNESCO, 2023). Therefore, **equitable access to technological infrastructure** should be considered a strategic priority for governments, especially after the acceleration of educational digitalization post-COVID-19.

From a methodological perspective, the results ratify the effectiveness of models such as **random forest** and **decision trees** for heterogeneous educational environments, due to their ability to manage complex interactions between variables and offer understandable explanations for institutional decision-making (Zhou et al., 2021; Sharma & Kapoor, 2020). However, there is a need to strengthen education data systems in the countries analyzed, as **the availability and quality of data remains a challenge for a wider application of predictive analytics** (Escueta et al., 2020).

Finally, this study supports the argument that education, especially in its qualitative dimension, **remains the most effective mechanism for breaking cycles of poverty** and generating upward social mobility, as also underlined by the World Bank (Narayan et al., 2022). However, for such mobility to be inclusive and sustainable, cross-sectoral reforms are required that integrate education, health, employment and social protection.

References

- Alfaro-Tanco, J., & Gallego, R. (2022). Determinants of school performance in contexts of urban poverty. *Revista Colombiana de Educación*, 82(2), 43–61. https://doi.org/10.17227/rce.num82-13595
- Camacho, L., Pérez, M., & Soto, D. (2023). Individual factors in the academic performance of rural students. *Educational Psychology*, 29(1), 21–30. https://doi.org/10.5093/psed2023a4
- Davenport, T. H., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48(1), 24–42. https://doi.org/10.1007/s11747-019-00696-0

- Escueta, M., Quan, V., Nickow, A. J., & Oreopoulos, P. (2020). Education technology: An evidence-based review. *Review of Economic Studies*, 87(3), 991–1035. https://doi.org/10.1093/restud/rdz033
- Ganimian, A. J., & Murnane, R. J. (2020). Improving education in developing countries: Lessons from rigorous impact evaluations. *Review of Educational Research*, 90(4), 681–722. https://doi.org/10.3102/0034654320942877
- Mehta, D., Saxena, D., & Bansal, A. (2021). Machine learning models for academic performance prediction: A comparative study. *Education and Information Technologies*, 26(6), 7487–7504. https://doi.org/10.1007/s10639-021-10636-2
- Narayan, A., Van der Weide, R., Cojocaru, A., & Lakner, C. (2022). *Fair Progress: Economic Mobility Across Generations Around the World*. World Bank. https://doi.org/10.1596/978-1-4648-1587-9
- OECD. (2023). *Education at a Glance 2023: OECD Indicators*. https://www.oecd.org/education/education-at-a-glance
- Orozco, C., & Valencia, D. (2020). ICT and School Performance: Evidence in Colombian Public Institutions. *Education and Society*, 41(1), 221–239. https://doi.org/10.1590/es2020.41.1.003
- Ramos, J. M., & Pérez, M. C. (2021). Socioeconomic determinants of school performance: An analysis for Latin America. *Latin American Journal of Educational Studies*, 51(2), 15–34. https://doi.org/10.22201/iisue.24486167e.2021.51.2.59441
- Rosa, V., & González, A. (2021). Contextual influences on educational performance: A systematic review. *Education and Development*, 38(2), 85–98. https://doi.org/10.5294/edu.2021.38.2.6
- Sharma, M., & Kapoor, M. (2020). Neural networks for educational data mining: A practical case study. *International Journal of Educational Technology*, 15(1), 101–112. https://doi.org/10.1186/s41239-020-00195-3
- UNESCO. (2023). *Global education monitoring report 2023: Technology in education – A tool on whose terms?* https://unesdoc.unesco.org/ark:/48223/pf0000385501
- World Bank. (2022). *World Development Report 2022: Finance for an equitable recovery*. https://www.worldbank.org/en/publication/wdr2022

- Zhou, Y., Li, X., & Shen, J. (2021). Predictive modeling in education: A review of machine learning methods. *Computers & Education*, 173, 104281. https://doi.org/10.1016/j.compedu.2021.104281