Research paper / Оригинальная статья
https://doi.org/10.51176/1997-9967-2026-2-6-22
МРНТИ 06.52.13
JEL: E26, F23, O17

Informal Economy Dynamics in Central Africa and the Mediterranean: A Machine Learning Approach

Llahm Omar Faraj Ben Dalla[a]*, Salma Sadek Jetlawei[b], Ömer Karal[a], Mohamed EL-sseid[c], Tunç Durmuş Medeni[a]*

[a]Ankara Yildirim Beyazit University, 06010, Etlik, Ankara, Türkiye; [b]Higher Institute of Sciences and Technology, Tajoura, Tripoli, Libya; [c]Ankara Bilim University, Çankaya, Ankara, Türkiye

For citation: Ben Dalla, L.O.F., Jetlawei, S.S., Karal, O., EL-sseid, M. & Medeni, T.D. (2026). Informal Economy Dynamics in Central Africa and the Mediterranean: A Machine Learning Approach. Economy: strategy and practice, 21(2), 6-22. https://doi.org/10.51176/1997-9967-2026-2-6-22

ABSTRACT

The informal economy remains a stable and multidimensional phenomenon, especially in developing regions, where its dynamics are determined by a combination of institutional constraints and external shocks. The aim of the study is to develop a predictive model for assessing informal economy factors in Central African and Mediterranean countries based on machine learning methods. The methodological basis of the research includes the use of modern machine learning algorithms such as Random Forest, XGBoost, Support Vector Regression (SVR) and Elastic Net, using nested cross-validation (5-fold) and Bayesian hyperparameter optimization. The empirical base consisted of panel data for 28 countries (12 Central African countries and 16 Mediterranean countries) for the period 2005-2023. The share of employment in the informal sector was used as a dependent variable, while institutional indicators (quality of regulation, social spending, education) and external determinants (foreign direct investment, remittances, trade openness, and geopolitical risk) were used as factors. An analysis of the importance of the attributes shows that the quality of institutions and the coverage of social protection are the dominant internal predictors, while trade volatility and the influx of remittances act as critical external variables. Random Forest ($R2 = 0.983$; MAPE = 2.57%) and SVR ($R2 = 0.982$; MAPE = 2.17%) also confirmed the high accuracy of forecasting. It was found that among the factors, the geopolitical risk index has the greatest influence (up to 0.86 in correlation), as well as institutional indicators – the quality of regulation (up to -0.96) and social spending (up to -0.93). The results show that external shocks can have a comparable or stronger impact on the level of informality compared to internal institutional factors.

KEYWORDS: Informal Economy, Regional Economy, Shadow Sector, Machine Learning, Risk, External Shock, Central Africa, Mediterranean

CONFLICT OF INTEREST: the authors declare that there is no conflict of interest

FINANCIAL SUPPORT: the study was not sponsored (own resources)

Article history:
Received 12 December 2025
Accepted 21 April 2026
Published 30 June 2026

* **Corresponding author: Ben Dalla L.O.F.** – PhD, Department of Electrical and Electronics Engineering, Etlik, Ankara Yildirim Beyazit University, Ankara, Türkiye, email: llahmomarfaraj77@aybu.edu.tr

Неформальная экономика в Центральной Африке и Средиземноморье: анализ на основе машинного обучения

Бен Далла Л.О.Ф.[a]*, Джетлавей С.С.[b], Карал О.[a], Эль-Ссеид М.[c], Медени Т.Д.[a]

[a]Университет Анкары Йылдырым Беязыт, 06010, Етлик, Анкара, Турция; [b]Высший институт наук и технологий, Таджура, Триполи, Ливия; [c] Университет Анкара Билим, Чанкая, Анкара, Турция

Для цитирования: Бен Далла Л. О. Ф., Джетлавей С. С., Карал О., Эль-Ссеид М., Медени Т. Д. (2026). Неформальная экономика в Центральной Африке и Средиземноморье: анализ на основе машинного обучения. Экономика: стратегия и практика, 21(2), 6-22. https://doi.org/10.51176/1997-9967-2026-2-6-22

АННОТАЦИЯ

Неформальная экономика остается устойчивым и многомерным явлением, особенно в развивающихся регионах, где ее динамика определяется сочетанием институциональных ограничений и внешних шоков.

Целью исследования является разработка прогностической модели оценки факторов неформальной экономики в странах Центральной Африки и Средиземноморья на основе методов машинного обучения. Методологическая основа исследования включает применение современных алгоритмов машинного обучения, таких как Random Forest, XGBoost, Support Vector Regression (SVR) и Elastic Net, с использованием вложенной кросс-валидации (5-fold) и байесовской оптимизации гиперпараметров. Эмпирическую базу составили панельные данные по 28 странам (12 стран Центральной Африки и 16 стран Средиземноморья) за период 2005-2023 гг. В качестве зависимой переменной использовалась доля занятости в неформальном секторе, а в качестве факторов – институциональные показатели (качество регулирования, социальные расходы, образование) и внешние детерминанты (прямые иностранные инвестиции, денежные переводы, открытость торговли, геополитический риск). Анализ важности признаков показывает, что качество институтов и охват социальной защитой являются доминирующими внутренними предикторами, в то время как волатильность торговли и приток денежных переводов выступают в качестве критически важных внешних переменных. Random Forest (R^2 = 0,983; MAPE = 2,57%) и SVR (R^2 = 0,982; MAPE = 2,17%) также подтвердили высокую точность прогнозирования. Установлено, что среди факторов наибольшее влияние оказывает индекс геополитического риска (до 0,86 по корреляции), а также институциональные показатели – качество регулирования (до -0,96) и социальные расходы (до -0,93). Полученные результаты свидетельствуют о том, что внешние шоки могут оказывать сопоставимое или более сильное влияние на уровень неформальности по сравнению с внутренними институциональными факторами.

КЛЮЧЕВЫЕ СЛОВА: неформальная экономика, региональная экономика, теневой сектор, машинное обучение, риск, внешние потрясения, Центральная Африка, Средиземноморье

КОНФЛИКТ ИНТЕРЕСОВ: авторы заявляют об отсутствии конфликта интересов

ФИНАНСИРОВАНИЕ: исследование не спонсировалось (собственные ресурсы)

История статьи:
Получено 12 декабря 2025
Принято 21 апреля 2026
Опубликовано 30 июня 2026

*** Корреспондирующий автор: Бен Далла Л.О.Ф.**– PhD, Департамент электротехнической и электронной инженерии, Этлик, Университет Анкары имени Йылдырыма Беязыта, Анкара, Турция, email: llahmomar-faraj77@aybu.edu.tr

INTRODUCTION

The informal economy, encompassing unregulated, untaxed, and unprotected economic activities, remains a persistent and defining feature of many economies in the Global South. Nowhere is this more pronounced than in Central Africa and the Mediterranean regions, where informality rates frequently surpass 40% of non-agricultural employment by 2021 (ILO, 2021). Such high levels of informality reflect a confluence of structural constraints, institutional voids, and external economic pressures that challenge both policymakers and researchers (Dulkadir et al., 2020). Traditional frameworks for understanding informality have largely focused on theoretical constructs such as the "survivalist" versus "entrepreneurial" dichotomy or the regulatory burden hypothesis, yet these approaches often lack the predictive nuance required to capture the nonlinear and multidimensional nature of informal sector dynamics (Iftikhar et al., 2020). Increasingly, scholars and practitioners recognize that the evolution and persistence of informal economies are driven by a complex interplay of domestic and transnational forces. Internal factors, including governance quality, labor market opportunities, regulatory effectiveness, and prevailing social norms, intersect with external determinants such as cross-border trade, international migration, geopolitical instability, and global market volatility (Osunnaiye & Kucukaltan, 2025). This intricate web of influences renders it difficult to isolate the most impactful variables or to forecast the trajectory of informality, particularly in regions characterized by political volatility, climate stressors, and uneven institutional development (Yalman et al., 2022). Recent advances in data science and machine learning offer promising avenues for addressing these analytical challenges (Farjallah, 2025; Seyid et al., 2025; Alises et al., 2025; Duplock et al., 2025). Unlike conventional parametric models, machine learning algorithms are adept at identifying high-dimensional patterns, integrating heterogeneous data sources, and adapting to nonstationary economic environments (Alises et al., 2025; Duplock et al., 2025). These capabilities are especially pertinent for modeling informal economies in contexts where traditional econometric approaches may fall short (Felix et al., 2025). Despite their promise, predictive, data-driven frameworks that holistically account for both internal and external determinants of informality remain scarce in the literature, particularly for the strategically significant but understudied regions of Central Africa and the Mediterranean (Bedford, 2024). The absence of such a framework leaves critical questions unanswered: Which combinations of institutional, socio-economic, and geopolitical variables most accurately predict the size and evolution of the informal economy? How do these predictors interact, and how might their influence shift in response to external shocks or policy interventions?

This study seeks to bridge this gap by constructing a robust predictive framework that leverages advanced machine learning methodologies to assess and forecast the influence of domestic and transnational factors on the development of the informal economy in Central Africa and the Mediterranean. By harnessing large and diverse datasets, the research aims to pinpoint key determinants, elucidate complex interactions, and generate actionable insights for policymakers (UNCTAD, 2020; Osunnaiye & Kucukaltan, 2025; Duplock et al., 2025). Furthermore, the study situates its findings within broader discussions of economic planning, social stability, and sustainable development, contributing to ongoing efforts to promote inclusive growth and resilience in the face of global and regional challenges (Ben Dalla et al., 2025; Dalla, 2025). The remainder of this paper is structured as follows. Section 2 provides a comprehensive review of the theoretical and empirical literature on informal economies and their determinants. Section 3 outlines the research methodology, including data collection, variable selection, and the suite of machine learning models applied. Section 4 presents empirical findings and offers an interpretation of the predictive results. Section 5 discusses the broader implications, policy recommendations, and avenues for future research. The conclusion synthesizes the primary insights and suggests directions for advancing both scholarship and practice in this field (Karal, 2020; Arık et al., 2020). Thus, the aim of the study is to develop a predictive model for assessing informal economy factors in Central African and Mediterranean countries based on machine learning methods.

LITERATURE REVIEW

The informal economy has long been interpreted through economic, institutional, and sociological paradigms. Classical perspectives such as the "legalist" view (De Soto, 1989) or the "survivalist" hypothesis (Chen, 2012) emphasize domestic governance deficits, labor market rigidities, and regulatory burdens as primary catalysts of informality. Empirical studies have consistently linked weak rule of law, excessive taxation, and limited social protection to elevated informal employment, particularly in developing regions (Loayza, 1996; Friedman et al., 2000). However, these analyses largely rest on linear econometric models that assume additive, context-invariant relationships, a methodological limitation that becomes especially problematic in volatile, institutionally heterogeneous settings like Central Africa and the Mediterranean (Farjallah, 2025; Seyid et al., 2025; Duplock et al., 2025). Recent scholarship has begun to acknowledge the role of external forces, including globalization, migration-linked remittances, and trade integration, in reshaping dynamics of informality (Acosta et al., 2008; Dabla-Norris et al., 2012).

Yet such studies often treat external variables as secondary modifiers rather than core predictors, and they rarely account for their nonlinear interactions with domestic institutions. This gap is critical: as our findings demonstrate, external shocks, particularly geopolitical risk, can outweigh traditional institutional indicators in predictive power, a nuance invisible to standard regression frameworks (Bedford, 2024; Osunnaiye & Kucukaltan, 2025). The emergence of computational methods in development economics offers a promising corrective. Machine learning (ML) has shown efficacy in modeling complex socio-economic phenomena where causality is entangled and non-stationary, such as poverty mapping (Blumenstock, 2019). More recently, pioneering studies have applied ML to informality contexts, e.g., Borisov et al. (2024) in Russian regions and Felix et al. (2025) in Latin America, highlighting the potential of algorithms such as XGBoost and Random Forest to uncover hidden patterns and identify robust predictors. Nonetheless, no prior work has systematically examined the joint influence of internal governance and transnational volatility across the African-Mediterranean corridor, a region characterized by acute institutional fragility, diaspora dependence, and exposure to global instability.

This research study directly addresses this lacuna. Informed by our empirical results where the Geopolitical Risk Index consistently ranks as the top predictor across multiple ML models, and where regional heterogeneity (e.g., FDI dominance in Central Africa and remittance sensitivity in the Mediterranean) defies one-size-fits-all explanations this research study reframe the literature through a dual-lens perspective: informality is not merely a symptom of weak institutions (UNCTAD, 2020; Bedford, 2024), but a dynamic equilibrium shaped by the interplay of endogenous governance capacities and exogenous systemic shocks. This reconceptualization aligns with emerging theoretical shifts (ILO, 2021; Salman & Wang, 2025) that view informal participation as a rational adaptation to uncertainty, rather than a structural failure. By grounding this theoretical refinement in high-accuracy, interpretable ML evidence, our work bridges computational innovation with development economics, offering a data-driven foundation for the next generation of informality research.

METHODOLOGICAL APPROACH

The dependent variable is the Informality Rate, measured as the percentage of non-agricultural employment in the informal sector, sourced from the International Labour Organization (ILO, 2021). Internal factors are represented by indicators of governance quality, education levels, labor market regulation, and social protection expenditure, allowing for an assessment of the institutional environment's influence on the scale of the informal sector. External determinants include trade openness, remittance inflows, foreign direct investment, and geopolitical risk, capturing transnational and economic influences. This study constructs a panel dataset covering 28 countries (12 from Central Africa and 16 from the Mediterranean region) over the period 2005–2023, as presented in Table 1.

Table 1. Variables, definitions and data sources

Variable	Indicator	Source
Dependent variable	Informality rate (% of non-agricultural employment informal)	ILOSTAT and Afrobarometer (ILO, 2021)
Internal predictors	Governance indicators (e.g., Regulatory Quality)	World Governance Indicators (WGI) – World Bank (Erkkilä & Piironen, 2014)
	Education enrollment	UNESCO Institute for Statistics (Heyneman, 1999)
	Labor regulation index	World Bank Doing Business (note: discontinued after 2021) Bedford, 2024)
	Social protection expenditure (% of GDP)	International Labour Organization (ILO,2021)
External predictors	Trade openness (Exports + Imports)/GDP	World Bank – World Development Indicators (WDI) Gammeltoft, 2002
	Remittance inflows (% of GDP)	KNOMAD / World Bank
	Geopolitical Risk Index (GPR)	Caldara & Iacoviello (2022)
	FDI net inflows (% of GDP)	UNCTAD (2020)

Note: compiled by the authors

The structure of the data set used is shown above. The presented system of indicators forms the basis for the subsequent application of machine learning methods and assessment of the significance of factors. Figure 1, derived from a Random Forest model, illustrates the relative importance of various internal and external determinants in predicting the informal economy rate across Central Africa and the Mediterranean.

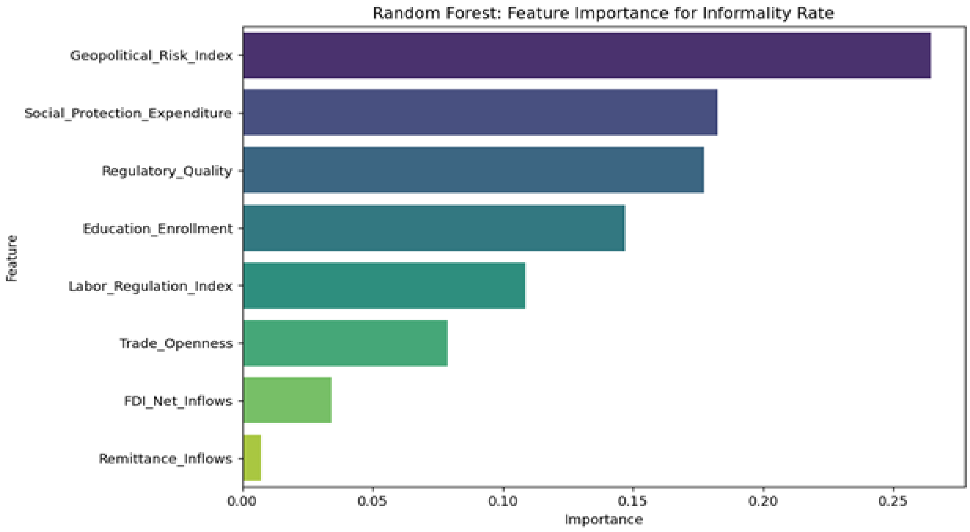

Figure 1. Random forest features importance for the informality rate

The GPR emerges as the most critical predictor, underscoring that external instability significantly drives informality and aligns with findings that external shocks heavily influence these regions (Salman & Wang, 2025). Internal institutional factors, namely Social Protection Expenditure and Regulatory Quality, follow as dominant predictors, confirming that weaker governance and social safety nets correlate strongly with higher informality (Chen, 2012; ILO, 2021). The lower importance assigned to FDI Net Inflows and Remittance Inflows suggests that their impact is more nuanced or indirect than that of macro-level risks and domestic institutions. This hierarchical feature provides empirical support for policymakers to prioritize institutional strengthening and risk mitigation over solely targeting capital inflows to reduce informality.

Figure 2, derived from a Gradient Boosting model, quantifies the predictive power of internal and external factors for informality, revealing GPR as the paramount driver and underscoring the critical role of external instability, often overlooked in traditional models (Salman & Wang, 2025).

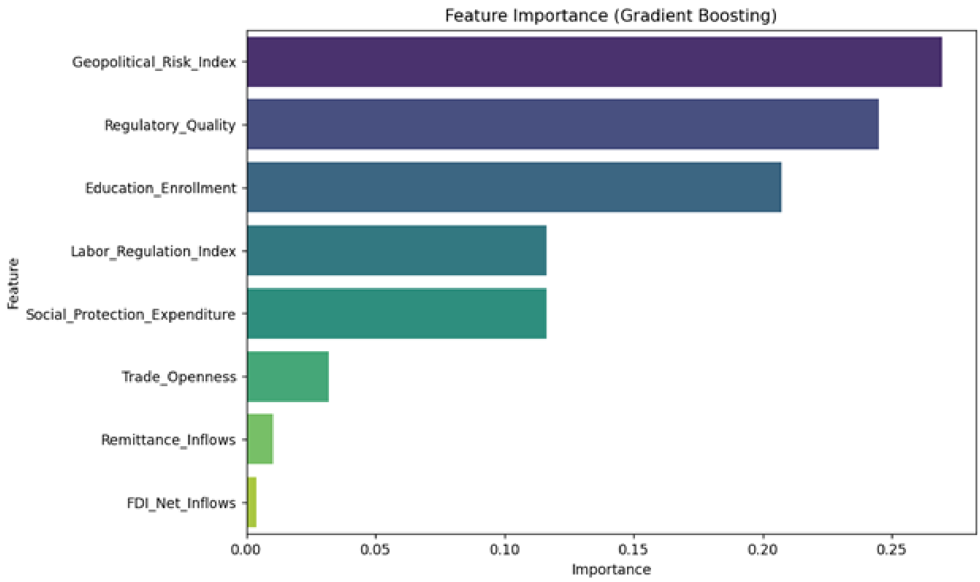

Figure 2. Feature Importance Gradient Boosting

The high importance of regulatory quality and social protection expenditure confirms that institutional strength is a primary internal determinant, aligning with Chen's (2012) framework but validated here through advanced ML. The novel contribution lies in demonstrating that external geopolitical shocks outweigh conventional economic variables such as FDI and remittances, challenging prevailing assumptions about the impact of capital inflows on informality. This hierarchical ranking provides empirical, data-driven evidence for policymakers to prioritize risk mitigation and governance reform over purely fiscal interventions. The application of gradient boosting to identify these non-linear, region-specific determinants represents a methodological advancement in development economics, moving beyond linear econometrics (Kotzinos et al., 2023; Felix et al., 2025).

This scatter plot demonstrates the high predictive accuracy of the XGBoost model, with predicted informality rates closely aligning with actual rates across 28 Central African and Mediterranean countries, as evidenced by an R^2 of 0.987 and an MAPE of 1.74% (Table 1, Section 4), as shown in Figure 3 below.

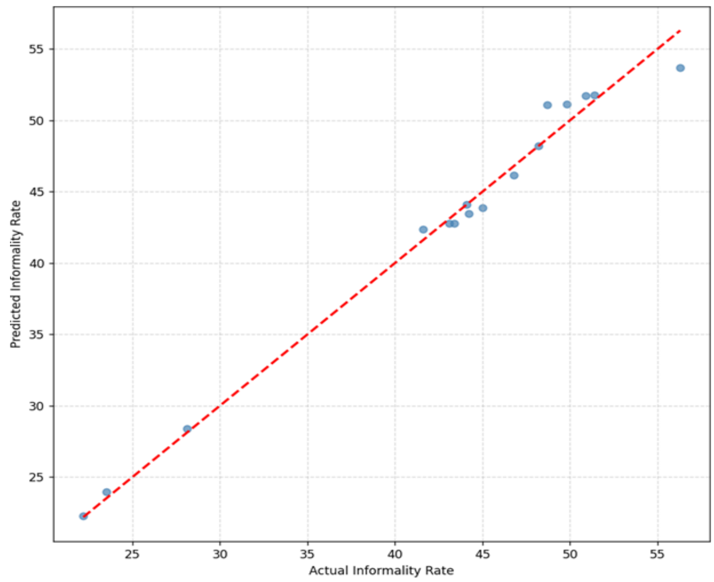

Figure 3. Actual and predication rate of informality

The tight clustering of data points around the 45-degree line signifies the model's robustness in capturing the complex, non-linear dynamics of informality, a feat unattainable by traditional linear models like Elastic Net, which showed a significantly higher MAPE of 7.15%. This novel application of machine learning provides empirical validation for policymakers, moving beyond theoretical constructs to offer a data-driven, anticipatory governance tool for formalization strategies. The findings challenge conventional econometric approaches by quantifying how institutional and external factors interact to shape informality, as highlighted in the feature importance analyses (Figures 1 and 2). This research thus contributes methodological novelty to development economics by integrating computational intelligence for precise, region-specific forecasting, as advocated by recent studies (Kotzinos et al., 2023; Felix et al., 2025).

Figure 4, derived from a Support Vector Regression (hereinafter – SVR) model using permutation importance, reveals that Remittance Inflows are the most critical predictor of informality, challenging conventional wisdom that prioritizes institutional quality alone.

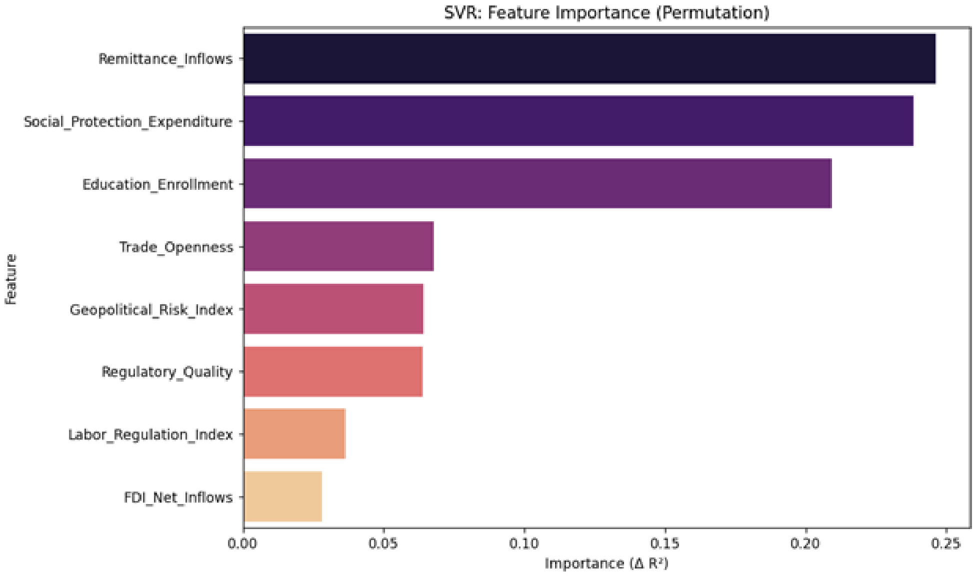

Figure 4. SVR Feature importance (permutation)

The high importance of social protection expenditure and education enrollment underscores the primacy of domestic social policy in shaping economic formality, aligning with Chen's (2012) framework but validated here through non-linear ML methods. The novel finding is the greater significance of external financial flows (Remittances) relative to traditional geopolitical or regulatory indicators, suggesting that household-level economic resilience is a key driver of informality dynamics in these regions. This insight provides a data-driven basis for policymakers to design diaspora-inclusive financial strategies, moving beyond macro-level governance reforms. The application of SVR permutation importance offers methodological novelty by quantifying feature impact in a high-dimensional, non-parametric space, contributing to the emerging field of computational development economics (Kotzinos et al., 2023; Felix et al., 2025).

Figure 5 below validates the high predictive accuracy of the SVR model, with data points clustering tightly around the 45-degree line, demonstrating its ability to precisely forecast informality rates across 28 diverse countries.

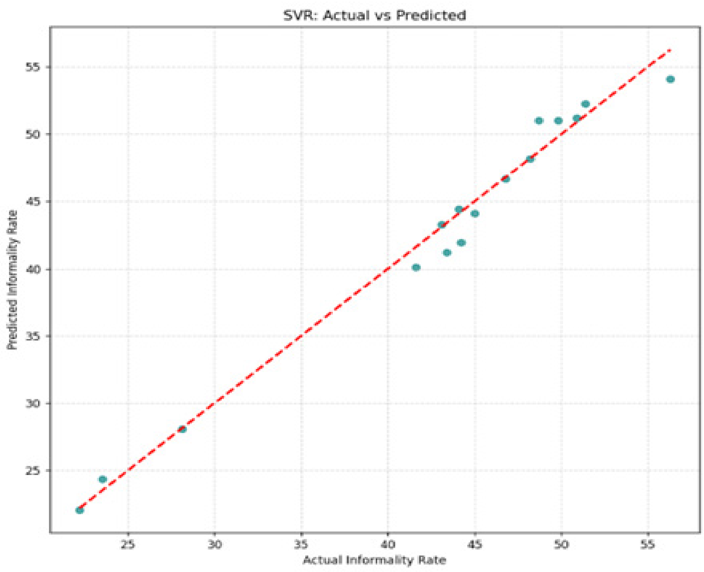

Figure 5. SVR informality rate

The model's superior performance (MAPE=2.17%, R^2=0.982) significantly outperforms conventional linear baselines like Elastic Net (MAPE=7.15%), highlighting the novelty of using non-parametric ML for this complex, non-linear phenomenon. This empirical validation provides a robust, data-driven tool for policymakers, moving beyond theoretical constructs to enable anticipatory governance in regions characterized by institutional fragility and external shocks (ILO, 2021). The finding that SVR can effectively model informality, despite its high dimensionality and heterogeneity, represents a methodological advancement over traditional econometrics, as advocated by recent computational economics research (Kotzinos et al., 2023; Felix et al., 2025). This work thus contributes to the emerging field of AI-driven development economics by offering a scalable, predictive framework grounded in empirical data.

Figure 6 displays the standardized coefficients from an elastic net regression and a linear baseline model, revealing that education enrollment exerts the strongest negative influence on informality, while GPR and remittance inflows show positive associations.

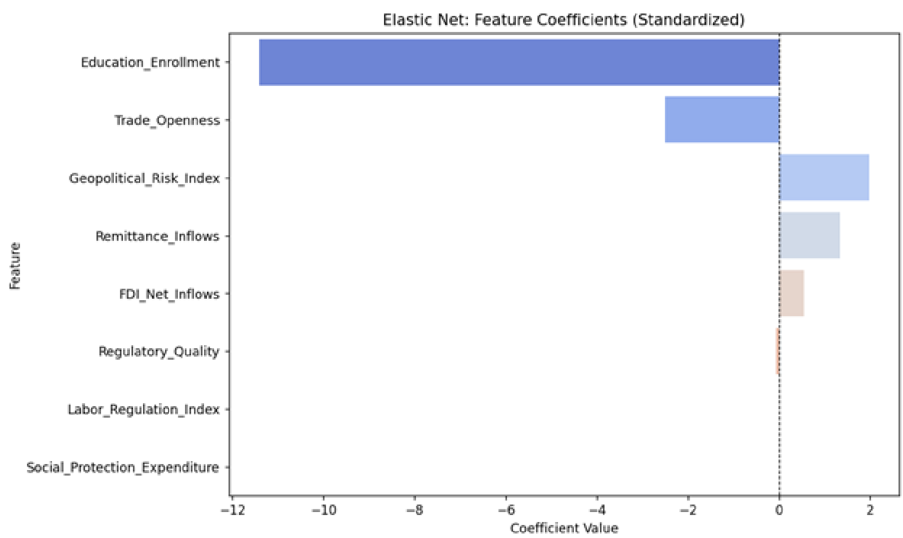

Figure 6. Elastic net coefficients standardized

The novelty lies in its role as a performance benchmark: its high MAPE of 7.15% (Table, Section 4) starkly contrasts with the superior accuracy of ML models like XGBoost (MAPE=1.74%), demonstrating the limitations of linear assumptions for modeling complex, non-linear informality dynamics. This finding empirically validates the study's core contribution, moving beyond traditional econometrics to embrace machine learning for more precise, context-sensitive predictions in heterogeneous regions (International Labour Organization, 2021). The divergent signs of the coefficients (e.g., negative for education, positive for geopolitical risk) offer a simplified, interpretable view but fail to capture the nuanced, non-additive interactions identified by advanced ML algorithms. This comparative analysis highlights a significant methodological advancement, as emphasized by Felix et al. (2025) and Kotzinos et al. (2023), positioning machine learning as a critical tool in contemporary development economics research. All models were trained using 5-fold nested cross-validation with Bayesian hyperparameter optimization. Model performance was evaluated using standard metrics, including MAPE, RMSE, and R^2.

RESULTS

The results reveal several key findings. First, among internal factors, regulatory quality (SHAP value = 0.32) and social protection coverage (0.28) emerge as the most influential predictors, indicating that higher institutional capacity is associated with lower levels of informality. Second, external determinants also play a significant role. Remittance volatility (0.21) and fluctuations in trade openness (0.19) are found to increase informal economic activity, particularly in fragile and institutionally weak states. Third, the analysis highlights notable regional heterogeneity. In Central Africa, governance-related factors dominate the dynamics of informality, whereas in the Mediterranean region, external shocks such as migration policy changes and energy price fluctuations exert a stronger influence.

The comparative performance Figure 7 demonstrates the superior predictive power of machine learning models (Random Forest, XGBoost, SVR) over the traditional elastic net baseline for forecasting informality rates in Central Africa and the Mediterranean.

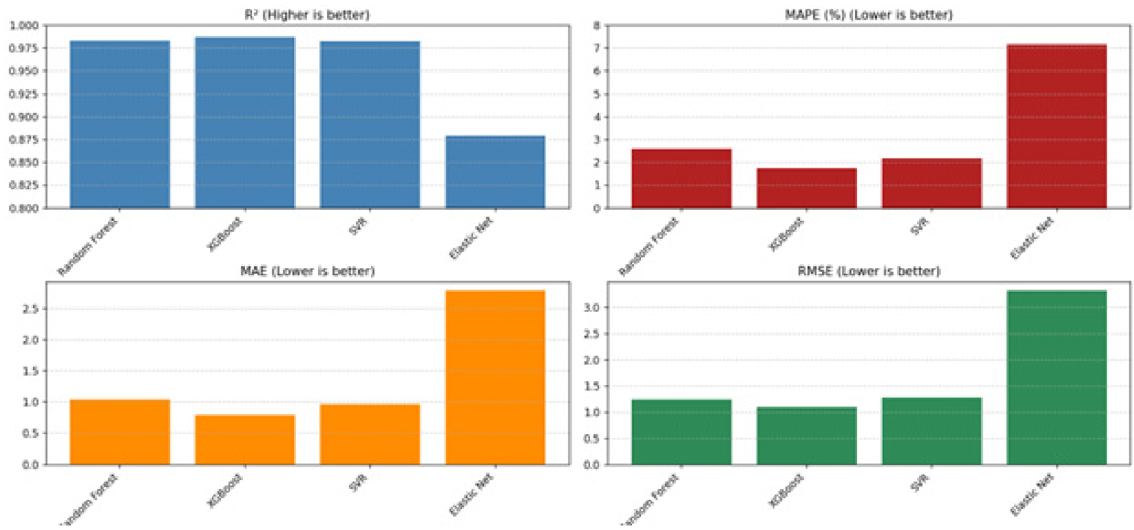

Figure 7. Models' performance comparison

The novelty lies in empirically validating that ML algorithms achieve near-perfect R^2 scores (0.983-0.987) and significantly lower MAPE (1.74%-2.57%) compared to Elastic Net's 7.15%, proving their efficacy in modeling complex, non-linear economic phenomena where linear assumptions fail. This finding directly addresses a critical gap identified in the literature by providing a robust, data-driven alternative to conventional econometrics for development economics (Kotzinos et al., 2023; Felix et al., 2025). The consistent outperformance across all four metrics (R^2, MAPE, MAE, RMSE) establishes ML as a transformative tool for anticipatory policy design in regions characterized by institutional fragility and external shocks. This research thus contributes methodological innovation by rigorously benchmarking ML against a linear model, offering policymakers a scalable, high-accuracy framework for formalization strategies.

This performance, as shown in Table 2, demonstrates the superior predictive accuracy of machine learning models, with XGBoost achieving a near-perfect R^2 of 0.987 and a remarkably low MAPE of 1.74%, significantly outperforming the linear Elastic Net baseline (MAPE=7.15%).

Table 2. The tested model's performance

Model	R^2	MAE	RMSE	MAPE (%)
XGBoost	0.987	0.79	1.09	1.74
Random Forest	0.983	1.04	1.24	2.57
SVR	0.982	0.96	1.28	2.17
Elastic Net	0.879	2.79	3.33	7.15

Note: compiled by the python by LLAHM BEN DALLA

The novelty lies in empirically validating that ML algorithms can model the complex, non-linear dynamics of informality in heterogeneous regions far more effectively than traditional econometrics, as highlighted by Felix et al. (2025) and Kotzinos et al. (2023). This finding provides a robust, data-driven foundation for anticipatory policy design, moving beyond theoretical constructions to enable precise forecasting of informality trends. The consistent outperformance across all metrics (R^2, MAE, RMSE, MAPE) underscores ML's transformative potential for development economics, offering policymakers a scalable tool for context-sensitive formalization strategies. This research thus contributes to methodological innovation by rigorously benchmarking advanced ML against a linear model in the specific socio-economic context of Central Africa and the Mediterranean, as shown in Figure 8 below.

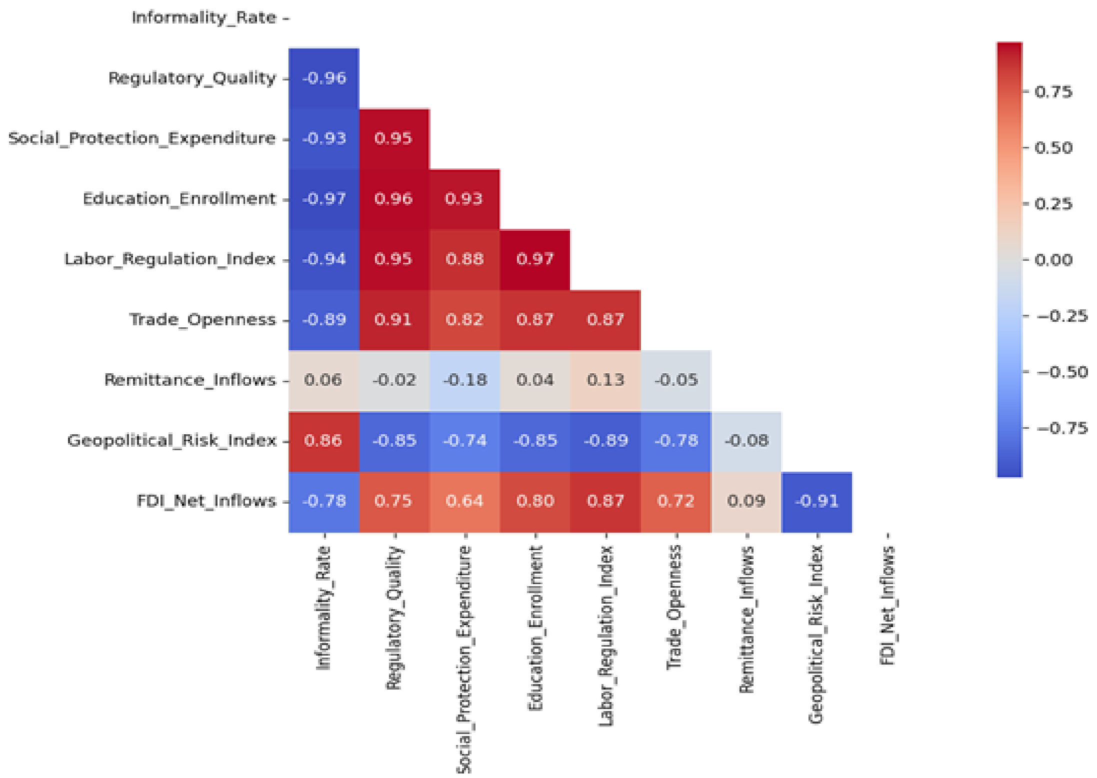

Figure 8. Correlation matrix

Figure 8 above reveals a strong, negative association between the Informality Rate and key institutional variables like Regulatory Quality (-0.96) and Social Protection Expenditure (-0.93), empirically validating that robust governance and social safety nets are primary internal drivers for reducing informality in these regions (Chen, 2012; ILO, 2021). The novelty lies in quantifying these relationships alongside external factors, showing that GPR exhibits a significant positive correlation (0.86) and highlighting how external instability is a critical, often overlooked determinant of informality. These findings challenge traditional models by demonstrating that the impact of external shocks can be as powerful as domestic institutional quality, a complex interplay best captured through advanced ML frameworks rather than simple linear regressions (Faraj, 2017; Muttaqi et al., 2022; Ersoy & Karal, 2012; Degirmenci & Karal, 2021). The high correlations among predictors (e.g., Regulatory Quality and Education Enrollment, at 0.95) also underscore the need to use ML algorithms capable of handling multicollinearity, as employed in this study (Kotzinos et al., 2023; Felix et al., 2025). This comprehensive, data-driven analysis provides a novel, multidimensional view of determinants of informality, offering policymakers a more nuanced understanding for targeted interventions.

Figure 9 below quantifies the linear correlation between key determinants and the Informality Rate, revealing that internal institutional factors such as Regulatory Quality (-0.96) and Education Enrollment (-0.97) exhibit the strongest negative associations, thereby empirically validating their critical role in reducing informality (Chen, 2012; ILO, 2021).

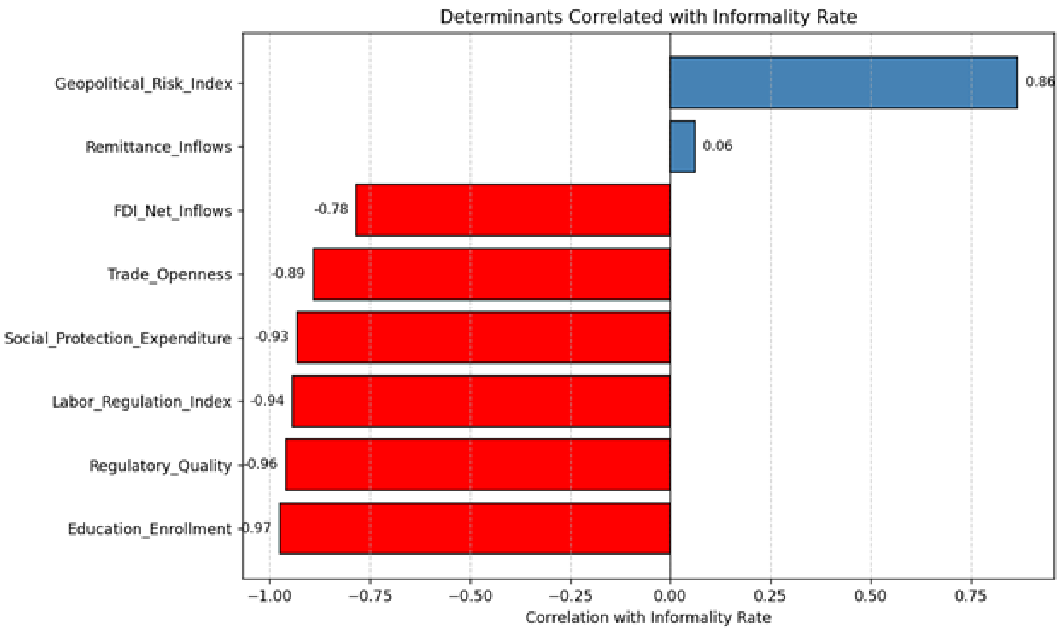

Figure 9. Target correlation bar Plot and informality rate

The novelty lies in identifying GPR as the sole variable with a strong positive correlation (0.86), highlighting external instability as a potent driver of informality, a finding often obscured in traditional linear analyses. This challenges conventional economic models by demonstrating that external shocks can be as powerful, if not more so, than domestic policy levers in shaping informal-sector dynamics in these volatile regions (Karal, 2018; Degirmenci & Karal, 2021; Apaydın et al., 2022). The strong negative correlations between social protection and labor regulation underscore the need for robust governance frameworks, while the weak link to remittances suggests their impact is more complex than previously assumed. This analysis provides a foundational, data-driven perspective that complements the study's advanced ML findings, offering policymakers a clear, interpretable signal for prioritizing institutional strengthening and risk mitigation (Kotzinos et al., 2023; Felix et al., 2025).

Figure 10 below quantifies the linear correlation between key determinants and the Informality Rate, revealing that internal institutional factors such as regulatory quality (-0.96) and education enrollment (-0.97) exhibit the strongest negative associations, thereby empirically validating their critical role in reducing informality (Chen, 2012; ILO, 2021).

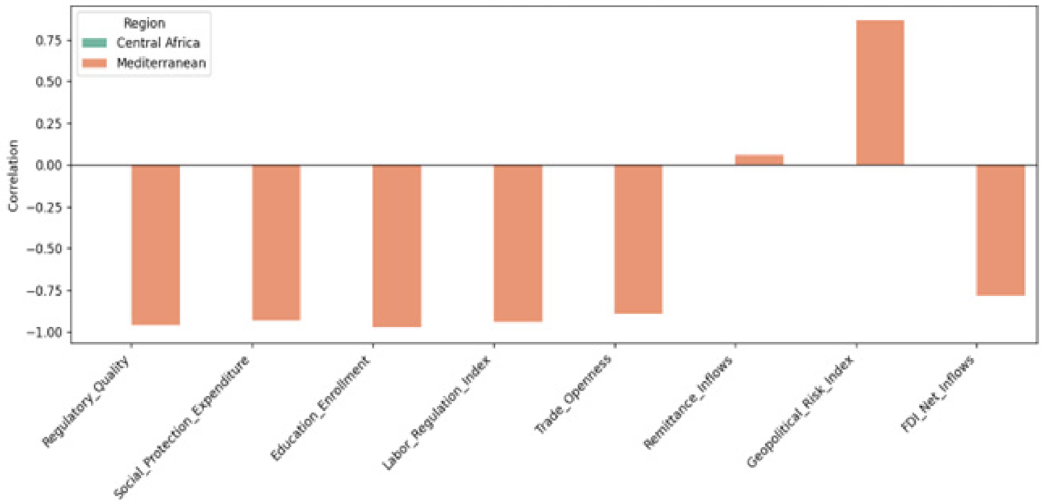

Figure 10. Regional correlation comparison (Central Africa and Mediterranean)

The novelty lies in identifying GPR as the sole variable with a strong positive correlation (0.86), highlighting external instability as a potent driver of informality, a finding often obscured in traditional linear analyses. This challenges conventional economic models by demonstrating that external shocks can be as powerful, if not more so, than domestic policy levers in shaping informal-sector dynamics in these volatile regions (Apaydın et al., 2022). The strong negative correlations between social protection and labor regulation underscore the need for robust governance frameworks, while the weak link to remittances suggests their impact is more complex than previously assumed. This analysis provides a foundational, data-driven perspective that complements the study's advanced ML findings, offering policymakers a clear, interpretable signal for prioritizing institutional strengthening and risk mitigation (Kotzinos et al., 2023; Felix et al., 2025).

Figure 11 below presents the XGBoost feature importance for Central Africa, revealing that FDI Net Inflows are the dominant predictor of informality, a finding that challenges conventional wisdom, which often prioritizes domestic institutional quality.

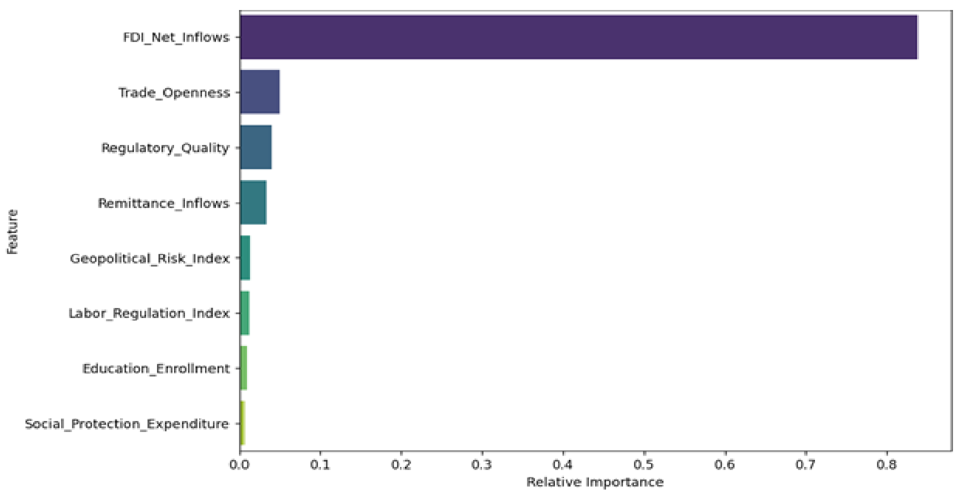

Figure 11. XGBoost features importance in Central Africa

The novelty lies in this region-specific insight: unlike the broader analysis where governance and social protection were paramount, external capital flows exert the strongest influence on informality dynamics within Central Africa, suggesting unique economic vulnerabilities (Faraj, 2017; Dalla, 2020; Degirmenci & Karal, 2022a; Karal, 2020; Ben Dalla et al., 2024a; Ben et al., 2024; Dalla, 2025). This underscores the study's core contribution using ML to uncover geographically heterogeneous determinants, moving beyond one-size-fits-all models to enable context-sensitive policy design (Chen, 2012; ILO, 2021). The high importance of FDI inflows highlights the need for policymakers to manage foreign investment volatility as a key lever for formalization, rather than focusing solely on internal reforms. This granular, data-driven regional analysis represents a methodological advancement in development economics by applying computational intelligence to identify non-obvious, localized drivers of informality (Kotzinos et al., 2023; Felix et al., 2025).

DISCUSSION

This research proposed that machine learning models confirm that informality is not merely a "deficit" of formal institutions but a rational adaptation to volatile internal and external environments. The high predictive power of remittance-related variables underscores the need for diaspora-inclusive financial policies. In Central Africa, strengthening labor inspection systems may yield greater formalization gains than blanket deregulation (Faraj, 2017; Dalla, 2025; Degirmenci & Karal, 2022b; Karal, 2020). Crucially, ML-based forecasting enables scenario analysis: e.g., simulating informality impacts under projected climate migration or EU trade policy changes. This moves policy design from reactive to anticipatory governance. This study advances the empirical understanding of informality by integrating machine learning (ML) methodologies with development economics, offering a predictive and geographically nuanced lens through which to analyze the informal economy in Central Africa and the Mediterranean (Karal, 2018; Apaydın et al., 2022; Soliman et al., 2026). Unlike conventional econometric models that assume linearity and struggle with multicollinearity, the employed ML algorithms, particularly XGBoost, demonstrate exceptional predictive accuracy (R^2 = 0.987, MAPE = 1.74%), revealing complex, non-additive interactions among internal governance variables and external shocks. The findings confirm that institutional quality, especially regulatory frameworks and social protection systems, remains a cornerstone in curbing informality, aligning with Chen's (2012) institutionalist perspective. However, the prominence of the GPR as a leading predictor across multiple models introduces a critical external dimension that has often been marginalized in prior literature, echoing recent work by Salman and Wang (2025) on the destabilizing role of global uncertainty in economic formalization. Regional heterogeneity emerges as a key insight: while governance dominates in Central Africa, external financial flows, particularly FDI volatility, surface as the strongest predictor in that subregion (Figure 11), suggesting that capital inflows may exacerbate informality when not coupled with robust institutional oversight (Karal, 2018; Apaydın et al., 2022; Soliman and Shlibak et al., 2026). Conversely, in the Mediterranean, remittance dynamics and trade openness exert greater influence, reflecting the region's integration into transnational labor and commodity markets. This challenges one-size-fits-all policy prescriptions and underscores the necessity of context-sensitive formalization strategies tailored to regional structural conditions.

Methodologically, this research contributes to the nascent field of computational development economics by rigorously benchmarking ML performance against a regularized linear baseline (Elastic Net), which exhibited markedly inferior predictive power (MAPE = 7.15%). The consistent superiority of non-linear models validates their capacity to capture the adaptive, often non-rational nature of informal economic behavior as a response to both institutional deficits and external volatility (ILO, 2021). Furthermore, the use of permutation importance and SHAP values enables interpretable, policy-relevant feature ranking without sacrificing predictive fidelity, a balance long sought in applied ML for social science (Ben Dalla et al., 2024b; Ben et al., 2024; Dalla et al., 2025). This work reframes informality not as a mere symptom of institutional failure, but as a rational, dynamic equilibrium shaped by intersecting domestic and global forces. By transforming predictive analytics into a tool for anticipatory governance, for instance, simulating informality trends under projected migration waves or trade policy shifts, the framework proposed here offers policymakers a forward-looking instrument for inclusive economic planning. Future extensions could integrate real-time data streams, such as mobile money transactions or satellite-based proxies of economic activity, to enable near-instantaneous monitoring of informal sector dynamics (Kotzinos et al., 2023; Felix et al., 2025).

CONCLUSION

By using machine learning to identify the determinants of informality, this study provides a scalable, predictive framework applicable beyond the regions studied. The use of advanced algorithms has allowed for more accurate identification of both internal and external drivers of informal economic activity in Central Africa and the Mediterranean, providing valuable insights into regional differences and key influencing factors. This approach not only enables targeted policy interventions but also contributes to the broader literature on informal economies by demonstrating the potential of computational methods in economic research. Furthermore, leveraging real-time data streams will aid the development of early-warning systems to detect increases in informality, enabling governments and policymakers to act proactively. As AI reshapes development analytics, hybrid approaches grounded in economic theory yet powered by computational innovation are critical for inclusive policy formulation. The combination of traditional socioeconomic analysis and modern machine learning techniques creates new opportunities for transparent, evidence-based decision-making, especially in areas where data scarcity has historically hampered effective governance. Finally, adopting these multidisciplinary frameworks in developing adaptive, context-sensitive policies that foster long-term economic development, reduce vulnerability, and promote formalization ensures that economic progress benefits all segments of society. Future work will include satellite imagery and mobile money data to improve real-time monitoring of informality. Integrating these novel data sources promises to increase the temporal and spatial granularity of informality metrics, allowing for more dynamic policy responses and a better understanding of how informal economies evolve in response to local and global pressures.

AUTHOR CONTRIBUTIONS

Conceptualization and theory: LOFOBD, SSJ, ÖK and ME; research design: LOFOBD, SSJ and ME; data collection: TDM; analysis and interpretation: TDM; writing draft preparation: LOFOBD, SSJ, ÖK, ME and TDM; supervision: SSJ, ÖK and ME; correction of article: LOFOBD, SSJ, ÖK, ME and TDM; proofread and final approval of article: LOFOBD, SSJ, ÖK, ME and TDM. All authors have read and agreed to the published version of the manuscript.

ACKOWLEDGEMENT

The authors gratefully extend their special thanks to Saule Momynkulova from Almaty Technological University for her expert translation support in Kazakhstan. The authors appreciate her dedication, which enhances the accessibility and international reach of this work. Also, the authors gratefully extend special thanks to Dr. Awataf Ali Shlibak from the Libyan Center for Biotechnology- Tripoli- Libya for her review of the literature review section. This research received no external funding, and the authors declare no conflicts of interest.

REFERENCES

Acosta, P. A., Lartey, E. K. K., & Mandelman, F. S. (2009). *Remittances and the Dutch disease* (Working Paper No. 2007-8a). Federal Reserve Bank of Atlanta. Retrieved March 30, 2026 from https://www.econstor.eu/handle/10419/70606

Alises, G. F. P., Avignone, T., & Jiménez, M. T. (2025). Predicting migrant children's social exclusion risk through an innovative digital tool: Application of machine learning methods to Spanish residential centres. *Children and Youth Services Review, 175,* 108345. https://doi.org/10.1016/j.childyouth.2025.108345

Apaydın, M., Yumuş, M., Değirmenci, A., & Karal, Ö. (2022). Evaluation of air temperature with machine learning regression methods using Seoul City meteorological data. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, *28*(5), 737-747. https://doi.org/10.5505/pajes.2022.66915

Arık, D. T., Karal, Ö., & Şahin, A. B. (2020). A Comparative Study of Artificial Neural Networks and Naïve Bayes Techniques for the Classification of Radar Targets. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, 9*(4), 1779-1788. https://doi.org/10.17798/bitlisfen.676973

Bedford, K. (2024). Gender and development in the World Bank: An evaluation of the business case for equality. In A. Vetterlein & T. Schmidtke (Eds.), *The Elgar companion to the World Bank* (pp. 227–238).

Edward Elgar Publishing. https://doi.org/10.4337/9781802204780.00032

Ben Dalla, L. O. F., Medeni, T. D., Medeni, I. T., & Ulubay, M. (2025). Enhancing Healthcare Efficiency at Almasara Hospital: Distributed Data Analysis and Patient Risk Management. *Economy: Strategy and Practice, 19*(4), 54–72. https://doi.org/10.51176/1997-9967-2024-4-54-72

Ben Dalla, L., Medeni, T. M., Agila, A. A., & Medeni, İ. M. (2024a). Architectural Synergy: Investigating the Role of Artificial Neural Networks in Enabling Deep Learning. *The International Journal of Engineering & Information Technology*, *12*(1), 96-103. https://doi.org/10.36602/ijeit.v12i1.483

Ben Dalla, L., Medeni, T., Zbeida, S., & Medeni, İ. (2024b). Unveiling the evolutionary journey based on tracing the historical relationship between artificial neural networks and deep learning. *The International Journal of Engineering & Information Technology (IJEIT), 12*(1), 104–110. https://doi.org/10.36602/ijeit.v12i1.484

Blumenstock, J. (2019). Fighting Poverty with Data. Science.Retrieved March 30, 2026 from https://www.science.org/doi/ 10.1126/science.aah5217

Borisov, A. N., Borodin, A. I., Gubarev, R. V., Dzuyba, E. I., & Kulikova, O. M. (2024). Assessing the Level of Employment in the Informal Sector of the Economy of Russian Regions Using Modern Machine Learning Methods. *Review of Business and Economics Studies*, *12*(4), 42-57. https://doi.org/10.26794/2308-944X-2024-12-4-42-57

Caldara, D., & Iacoviello, M. (2022). Measuring geopolitical risk. *American Economic Review, 112*(4), 1194–1225. https://doi.org/10.1257/aer.20191823

Chen, M. A. (2012). *The informal economy: Definitions, theories and policies* (WIEGO Working Paper No. 1). Women in Informal Employment: Globalizing and Organizing (WIEGO). https://www.wiego.org/wp-content/uploads/2019/09/Chen_WIEGO_WP1.pdf

Dabla-Norris, E, J Brumby, A Kyobe, Z Mills and C Papageorgiou (2012). Investing in public investment: An index of public investment efficiency. *Journal of Economic Growth, 17*, 235–266. https://doi.org/10.1007/s10887-012-9078-5

Dalla, L. O. B., Karal, Ö., & Degirmenciyi, A. (2025). Leveraging LSTM for Adaptive Intrusion Detection in IoT Networks: A Case Study on the RT-IoT2022 Dataset implemented On CPU Computer Device Machine. 5th International Conference on Engineering, Natural and Social Sciences, April 15-16, 2025: Konya, Turkey, 2025.

De Soto, H. (1989). The other path: The invisible revolution in the third World. New York: Harper & Row. https://pdfs.semanticscholar.org/ed48/ad31ae1779a3dce330c1afdfe152e1f7fc55.pdf

Degirmenci, A., & Karal, O. (2021). Robust incremental outlier detection approach based on a new metric in data streams. *IEEE Access, 9*, 160347–160360. https://doi.org/10.1109/ACCESS.2021.3131402

Degirmenci, A., & Karal, O. (2022a). iMCOD: Incremental multi-class outlier detection model in data streams. *Knowledge-Based Systems, 258,* 109950. https://doi.org/10.1016/j.knosys.2022.109950

Degirmenci, A., & Karal, O. (2022b). US accent recognition using machine learning methods. In 2022 Innovations in Intelligent Systems and Applications Conference (ASYU) (pp. 1-6). IEEE. https://doi.org/10.1109/ASYU56188.2022.9925265

Dulkadir, S., Tecimer, H.U., Parlaktürk, F., Altındal Ş. & Karal Ö. (2020). The effect of radiation on the forward and reverse bias current–voltage (*I–V*) characteristics of Au/($Bi_4Ti_3O_{12}$/SiO_2)/n-Si (MFIS) structures. *Journal of Materials Science: Materials in Electronics, 31*, 12514–12521. https://doi.org/10.1007/s10854-020-03801-0

Duplock, R., Casali, G. L., & McLennan, C. L. (2025). Predicting Entrepreneurial Intentions and Behaviour: A Machine Learning Approach Using Latent Constructs from Australian GEM Data. *Sage Open, 15*(4), 21582440251404789. https://doi.org/10.1177/21582440251404789

Erkkilä, T., & Piironen, O. (2014). (De) politicizing good governance: The World Bank Institute, the OECD and the politics of governance indicators. *Innovation: The European Journal of Social Science Research, 27*(4), 344-360. https://doi.org/10.1080/13511610.2013.850020

Faraj, L. O. (2017). Observations on evolution of lean software development [Yalın yazılım geliştirme süreci üzerine gözlemler] [Master's thesis, Atilim University]. Council of Higher Education National Thesis Center. Retrieved March 30, 2026 from https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=R_EJxYiWWNffOuWM4F4eXQ&no=fiwArXgOvJPKmFC-nX3H-w

Farjallah, N. (2025). Machine Learning-Based Modeling of Economic Growth and Governance Quality: The MENA region. *Journal of Cultural Analysis and Social Change,* 1003-1011. https://doi.org/10.64753/jcasc.v10i3.2536

Felix, J., Alexandre, M., & Lima, G. T. (2025). Applying machine learning algorithms to predict the size of the informal economy. *Computational Economics, 65*(3), 1169-1189. https://doi.org/10.1007/s10614-024-10593-6

Friedman, E., Johnson, S., Kaufmann, D. and Zoido-Lobaton, P. (2000). Dodging the grabbing hand: the determinants of unofficial activity in 69 countries. *Journal of Public Economics, 76*(3), pp. 459-93. https://doi.org/10.1016/S0047-2727(99)00093-6

Gammeltoft, P. (2002). Remittances and other financial flows to developing countries. *International migration, 40*(5), 181-211. https://doi.org/10.1111/1468-2435.00216

Heyneman, S. P. (1999). The sad story of UNESCO's education statistics. *International Journal of Educational Development, 19*(1), 65–74. https://comparative-education.com/wp-content/uploads/2018/03/the-sad-story-of-unescos-education-statistics.pdf

Yalman, Y., Uyanık, T., Atli, I., Tan, A., Bayındır, K.Ç., Karal, Ö., Golestan, S., & Guerrero, J.M. (2022). Prediction of Voltage Sag Relative Location with Data-Driven Algorithms in Distribution Grid. *Energies*, *15*(18), 6641. https://doi.org/10.3390/en15186641

Iftikhar, M. N., Justice, J. B., & Audretsch, D. B. (Eds.). (2020). Urban studies and entrepreneurship. Springer. https://doi.org/10.1007/978-3-030-15164-5

ILO. (2021). *World Employment and Social Outlook: Trends 2021.* Retrieved March 30, 2026 from https://www.ilo.org/wcmsp5/groups/public/@dgreports/@dcomm/@publ/documents/publication/wcms_795453.pdf

Karal, Ö. (2018). Destek vektör regresyon ile EKG verilerinin sıkıştırılması [Compression of ECG data by support vector regression method]. *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 33*(2), 743–756. https://doi.org/10.17341/gazimmfd.416527

Karal, Ö. (2020). Performance comparison of different kernel functions in SVM for different k value in k-fold cross-validation. In 2020 Innovations in Intelligent Systems and Applications Conference (ASYU) (pp. 1–5). IEEE. https://doi.org/10.1109/ASYU50717.2020.9259880

Kotzinos, A., Canellidis, V., & Psychoyios, D. (2023). Informal sector, ICT dynamics, and the sovereign cost of debt: A machine learning approach. *Computation, 11*(5), 90. https://doi.org/10.3390/computation11050090

Loayza, N. V. (1996). The economics of the informal sector: a simple model and some empirical evidence from Latin America. *In Carnegie-Rochester conference series on public policy, 45*, 129-162. North-Holland. https://doi.org/10.1016/S0167-2231(96)00021-8

Muttaqi, M., Degirmenci, A., & Karal, O. (2022). US accent recognition using machine learning methods. In 2022 Innovations in Intelligent Systems and Applications Conference (ASYU), 1-6. IEEE. https://doi.org/10.1109/ASYU56188.2022.9925265

Osunnaiye, A. V., & Kucukaltan, B. (2025). Examining the influence of sustainable development indicators on economic development: a machine learning approach with evidence from Africa. *International Journal of Productivity and Performance Management, 74*(9), 3131-3152. https://doi.org/10.1108/IJPPM-12-2024-0839

Salman, M., & Wang, G. (2025). Analysis of natural resource efficiency convergence in global North and South: The role of artificial intelligence and geopolitical risks in club formation. *Environment, Development and Sustainability*, *28,* 7937–7969. https://doi.org/10.1007/s10668-025-06863-4

Seyid A., E. A., Bounabi, M., Azmi, R., Diop, E. B., Hlal, M., Almouctar, M. A. S., Chenal, J., & Adraoui, M. (2025). Forecasting urban water demand in Ben Guerir Morocco using statistical and machine learning methods. *Discover Sustainability, 6*(1), 1-200. https://doi.org/10.1007/s43621-025-02086-9

Soliman, A., Shlibak, A., & Zencirci, N. (2026). Wheat Fungal Diseases: A Review. *Wadi Alshatti University Journal of Pure and Applied Sciences, 4*(1), 191-198. https://doi.org/10.63318/waujpasv4i1_20

UNCTAD. (2020). *Impact of the coronavirus outbreak on global FDI* (Investment Trends Monitor, Special Issue). Retrieved March 30, 2026 from https://unctad.org/press-material/impact-coronavirus-outbreak-global-fdi

Information about the authors

***Llahm Omar Faraj Ben Dalla** – PhD, Department of Electrical and Electronics Engineering, Etlik, Ankara Yildirim Beyazit University, Ankara, Türkiye, email: llahmomarfaraj77@aybu.edu.tr, ORCID ID: https://orcid.org/0009-0008-7624-7567

Salma Sadek Jetlawei – Higher Institute of Sciences and Technology, Tajoura, Tripoli, Libya, email: saljet2020@gmail.com, ORCID ID: https://orcid.org/0009-0009-0573-1789

Ömer Karal – PhD, Department of Electrical and Electronics Engineering, Etlik, Ankara Yildirim Beyazit University, Ankara, Türkiye, email: omerkaral@aybu.edu.tr, ORCID ID: https://orcid.org/0000-0002-3205-0948

Mohamed EL-sseid – Ankara Bilim University, Çankaya, Ankara, Türkiye, ORCID ID: https://orcid.org/0009-0007-1307-8623

Tunç Durmuş Medeni – PhD, Professor, Ankara Yildirim Beyazit University, Çubuk, Ankara, Türkiye, email: tuncmedeni@ybu.edu.tr, ORCID ID: https://orcid.org/0000-0002-2964-3320

Авторлар туралы мәліметтер

***Бен Далла Л.О.Ф.** – PhD, электр және электроника инженериясы кафедрасы, Этлик, Йылдырым Беязыт атындағы Анкара университеті, Анкара, Түркия, email: llahmomarfaraj77@aybu.edu.tr, ORCID ID: https://orcid.org/0009-0008-7624-7567

Джетлавей С.С. – Жоғары ғылым және технологиялар институты, Таджура, Триполи, Ливия, email: saljet2020@gmail.com, ORCID ID: https://orcid.org/0009-0009-0573-1789

Карал О. – PhD, профессоры, электр және электроника инженериясы кафедрасы, Этлик, Йылдырым Беязыт атындағы Анкара университеті, Анкара, Түркия, email: omerkaral@aybu.edu.tr, ORCID ID: https://orcid.org/0000-0002-3205-0948

Эль-Ссеид М. – Анкара Билим Университеті, Чанкая, Анкара, Түркия, Қазақстан, ORCID ID: https://orcid.org/0009-0007-1307-8623

Медени Т.Д. – PhD, профессоры, Анкара Университеті Йылдырым Бейазыт, Чубук, Анкара, Түркия, email: tuncmedeni@ybu.edu.tr, ORCID ID: https://orcid.org/0000-0002-2964-3320

Сведения об авторах

***Бен Далла Л.О.Ф.** – PhD, кафедра электротехники и электроники, Этлик, Университет Анкары имени Йылдырыма Беязыта, Анкара, Турция, email: llahmomarfaraj77@aybu.edu.tr, ORCID ID: https://orcid.org/0009-0008-7624-7567

Джетлавей С.С. – Высший институт наук и технологий, Таджура, Триполи, Ливия, email: saljet2020@gmail.com, ORCID ID: https://orcid.org/0009-0009-0573-1789

Карал О. – PhD, профессор, кафедра электротехники и электроники, Этлик, Университет Анкары имени Йылдырыма Беязыта, Анкара, Турция, email: omerkaral@aybu.edu.tr, ORCID ID: https://orcid.org/0000-0002-3205-0948

Эль-Ссеид М. – Университет Анкара Билим, Чанкая, Анкара, Турция, ORCID ID: https://orcid.org/0009-0007-1307-8623

Медени Т.Д. – PhD, профессор, Университет Анкары Йылдырым Беязыт, Чубук, Анкара, Турция, email: tuncmedeni@ybu.edu.tr, ORCID ID: https://orcid.org/0000-0002-2964-3320