



Munich Personal RePEc Archive

Decarbonizing the Building Sector: The Integrated Role of ESG Indicators

Magaletti, Nicola and Notarnicola, Valeria and di Molfetta, Mauro and Leogrande, Angelo

LUM Enteprise s.r.l., LUM Enteprise s.r.l., LUM Enteprise s.r.l.,
LUM University Giuseppe Degennaro

20 June 2025

Online at <https://mpra.ub.uni-muenchen.de/125068/>
MPRA Paper No. 125068, posted 22 Jun 2025 06:09 UTC

Decarbonizing the Building Sector: The Integrated Role of ESG Indicators

Nicola Magaletti, LUM Enterprise s.r.l., magaletti@lumenterprise.it

Valeria Notarnicola, LUM Enterprise s.r.l., notarnicola@lumenterprise.it

Mauro Di Molfetta, LUM Enterprise s.r.l., dimolfetta@lumenterprise.it

Angelo Leogrande, LUM University Giuseppe Degennaro, leogrande.cultore@lum.it

Abstract

This work tests the relationship of the building sector's carbon dioxide (CO₂) emissions with a set of environmental, social, and governance (ESG) indicators in an international panel of countries. Using machine learning approaches alongside traditional econometric techniques, the work identifies strong predictors of emissions intensity in the nature of scientific productivity, healthcare infrastructure, and good governance. The findings indicate higher scientific productivity and better government governance are associated with reduced CO₂ emissions from building stocks, while the effectiveness of government and R&D expenditures are found to be associated with higher emission rates, possibly due to the broader urban infrastructures of the developed nations. With the help of clustering as well as permutation-based measures of importance, the work establishes the complex dynamics interlinking development, knowledge creation, and environmental efficiency. The result provides practical indications for policymakers who aim to harmonize the national ESG policies with the targets of decarbonization in the built space.

Keywords: Carbon Emissions, ESG Indicators, Building Sector, Machine Learning, Governance Effectiveness

JEL Codes: Q56, O44, C55, H52, O38.

1. Introduction

Climate change has been one of the prominent issues of the 21st century with increased global interest in the reduction of greenhouse emissions. With the building sector ranking amongst the top emitters of carbon dioxide production, the building sector occupies an important but at times neglected position (Adewumi et al., 2024). With the building sector worldwide accounting for almost 40 percent of energy usage and over a third of energy-related CO₂ emissions, the identification of the underlying causes of these emissions holds the secret to the formulation of effective decarbonization strategies in the built environment (Adewumi et al., 2024). At the same time, the increased deployment of Environmental, Social, and Governance (ESG) indicators has added a multi-dimensional element to the approach of sustainability and development performance evaluation (Costantiello et al., 2025). While ESG structures apply extensively in the context of corporate finance as well as in the context of investment analysis, their deployment at the level of macro-policy evaluation remains underexplored (Costantiello et al., 2025). ESG indicators have the ability to act as much more than

the mere tools of investment screens but as proxies for the presence of strong institutional quality as well as innovation systems on the level of societal infrastructure as well as long-term outcomes on the environment (Costantiello et al., 2025). For instance, researchers have discovered innovative healthcare policies as well as good governance on the environment strongly associate with the achievement of carbon footprints at fairly lesser levels, especially in highly industrialized sectors such as the construction sector (Al Amosh & Khatib, 2024).

This work addresses a significant research question: how do national indicators related to ESG—namely governance, education, healthcare, and scientific productivity—interface with carbon dioxide emissions from the building sector? This is done by positing an original empirical model bringing the ESG metrics together with emission datasets from the IPCC-defined residential and building energy use sectors. By connecting emissions to broader socio-institutional and knowledge-based indicators, the work attempts to go beyond the usual economic or first-sonic explanations for carbon production and bring in broader determinants. The work's novelty is in its interdisciplinary combination of the dimensions of ESG with environmental outcomes analysis using the econometric as well as the machine learning approach. While the previous work had focused on energy efficiency measures, the regulation tools, or the scale of economics as the terminal determinants of emissions, the work here acknowledges the role of national scientific capabilities, expenditure on public service sectors, and the efficacy of governance (Zhao et al., 2024). The interplay of clustering methods, variable importance maps, and the specifications of the regressions allows a multi-angular scrutiny of the strong associations and non-funnelled interoperations across countries with equal felicity as the simple associations across countries. The noticeable gap in the literature triggering the work here is stark. Despite an extensive body of work on emissions from the building sector, especially through the lens of technology diffusion as well as urban energy usage, very few efforts engage on how the institutional and societal determinants—particularly those as captured from the metrics of the ESG—define the emissions outcomes (Salarin et al., 2024). Besides, few efforts work out the relative importance of the indicators through model-agnostic methods such as the permutation-based dropout loss function as well as the global heterogeneity of the interoperations across countries by clustering. This work helps bridge the gap by offering an extensive survey of the building-related emissions through a data-driven approach on the social as well as the institutional determinants. It assumes that countries having higher level of activity in the scientific production activity level along with higher expenditure on the education and the healthcare sectors and having stronger governance structure have tended to exhibit the emissions from the building sectors to be lower due to higher presence of the innovation activity as well as the infrastructure plan better due to the efficacy of policies used. Conversely, the article provides an exploration of the dichotomy where intensive R&D spending and sound governance typical of development would be connected to the higher emissions by reasons of broader and energy-intensive infrastructure (Valaskova et al., 2023). By situating building emissions in an ESG-based analytical context, the article provides new insights on the highly complicated determinants of environmental performance as well as evidence-based suggestions to policymakers who aim to balance the goals of sustainability and organizational development.

2. Literature Review

Scholarly and professional literature on the integration of ESG considerations within the built environment has grown considerably over the last few years, witnessing increased concern about the sector's environmental performance and its ability to facilitate sustainable transition. The subsequent critique reviewed a range of recent articles on the application of the ESG principles to building construction, real property stewardship, and operating performance. The concern is the interpretation and application of the ESG system across the varied contexts and application of building management in the application of building sustainability criteria into operational usage. With an examination of literature reviewed across thematic range from certification and governance through tenant activity and innovation in technologies, the resultant synthesis addresses the extent to which design and invest

approaches compatible with the ESG are enforced—or frustrated—by current operational and stewardship practices in the built environment. Agedal (2024) addresses building owners and occupants response to the application of the ESG investments and finds financial responsibility the top priority in the real property market response to sustainability but the dynamic illustrates the social and governance aspects of the ESG in the property stewardship but the application is not translated to resilient environmental performance outcomes. Adewumi et al. (2024) describe the BREEAM system in the UK and find an overwhelming focus on the environment but weak application of the elements of the governance and SOCIAL pillars of the ESG system pertaining to the aspects of social equity or institutional accountability. Banerjee et al. (2024) work on Indian office market and insinuate certified green office stock has detrimental socioeconomic outcomes such as increased rents and so illustrates the possible disconnect between the application of the environmentally-oriented label and inclusive stock building practices. Battisti (2023), through the case of Florence, tries to outline the application of the parameters of the ESG indicators at the level of housing and planning policies where the response to the demand to apply the parameters of sustainability across the varied territorial contexts is the creation of the evaluation tools but the application of the resultant metrics on emissions/performance on the energy/resource usage axis is neglected.

Chungath (2023) suggests a digital infrastructure for monitoring ESG data in commercial real estate with prospects for applying building operations under sustainability metrics. However, the system exists at the theoretical level and is not evaluated for deployability or influence. Cruz et al., (2023) discuss ESG and innovation in building construction and mention increasing interest in the role of technology but fail to describe how innovation at the building level corresponds to enhanced ESG performance. Debrah et al., (2022) name inaccessible finance as the top barrier to the implementation of green building construction projects and obliquely attribute capital access limitations as related to governance weaknesses under the ESG model. Dugo (2021) consider financial products such as ETFs invested in sustainable real assets and observe positive returns but fail to discuss how such investments enhance building-level outcomes in sustainability metrics. Fedchenko et al., (2024) discuss the Russian case and illustrate how regulatory ambiguity obstructs the implementation of ESGs in building construction and reinforces the influence of governance on the creation of sustainability at the building level. Lastly, Gu et al., (2024) discuss ESG reporting as applied in smart cities and illustrate divergent datasets hindering the monitoring of sustainability metrics at the urban infrastructure level including the building level. Ifediora and Igwenagu (n.d.) note the increasing relevance of the ESG factors in real assets by noting environmental certification and social infrastructure as drivers but omit discussion of the building level in favor of the drivers of investment across the building level. Karerat et al., (2023) correlate the healthy building with the ESG's social pillar and promote the well-being of occupants as a performance descriptor. This human-centric approach underscores the ability of building level management to focus on the support of indoor air quality and illumination along with the aspects of the building's influence on the human body these aspects increasingly serve as the focus of the evaluation of the ESGs. Nevertheless, the examination focuses on the level of the building and design as opposed to the ongoing operational measures and tenant morale crucial in sustaining the level of the ESGs after the construction level has been completed. Kempeneer et al., (2021) offer a behavioral model to restore the interconnectivity of the application of the ESGs with the building end-user's level of experience. This work is particularly significant in placing the building occupants and facilities managers as the key actors in the achievement of the level of the ESGs performance.

They argue organizational routines and user practices can significantly mediate organizational efficiency outcomes from the application of ESG policies but base their arguments on thin empirical backup and require practical management practices to validate the model. Kostrikin and Andreeva (2023) describe low-rise residences and detail out the selection of ESG criteria specifically for low-rise residential constructions. Their study suggests the potential application of scaleable schemes of building administration applicable to non-urban cases and resolving environmental and societal problems. But the article remains proscriptive by offering suggestions without analyzing available

management capacities as well as constraints. Kudinova and Shushunova (2023) talk about certification programs in the context of green building constructions and insights about codification of the expectations of ESG through the application of standardized schemes. These schemes for certification usually stipulate managerial necessities about maintenance, monitoring, and leasehold communication and are crucial for the conformance of ESG. But the work does not proceed where such schemes influence building operation on a routine basis. Lam (2024) offers an integrated picture through the case application of the Hong Kong building market as well as exploring the mutual reinforcement of the administration of green building and ESG administration. Lam showcases how the application of the ESG is as much the physical characteristics of the green building as well as facilities shrewd administration such as energy monitoring, awareness campaigns amongst occupants as well as stakeholder reporting. This work strongly places building administration as an administration key to integrate the application of the ESG administration. Li et al. (2022) survey prefabricated infrastructures through the application of clustering techniques in an examination about the associations amongst the progress of the economy as well as the characteristics of the attribute of the ESG. While macro-trends remain the focus of discussion, the work insinuates modular constructions as an enabler of standardized administration regimes conducive to the attainment of the objectives of the ESG such as the minimization of energy as well as the minimization of wastes. Liu et al. (2025) survey how the occupants ascribe monetary values to building stock consistent with the application of the ESG and establish the existence of the rent premiums in support of the certification of the building assets as green. This demand from the occupants places building administration practices under the obligation of retaining certification parameters but to conserve market capitalization as well as meet occupancy expectations. Yet the correlation of occupant satisfaction alongside administration activity has received minimal scrutiny.

Meshcheryakova (2022) addresses the contribution of green standards to the overall ESG revolution of the construction industry. The focus on regulatory change underpins the argument that management systems must evolve to respond to new expectations about compliance but leaves operational detail thin on the ground. Lastly, Miklinska (2024) delves into the case of the green building in the supply chain context, specifically the case of the Polish warehouse market. The report finds building management crucial to embedding ESG in logistics and procurement policies and advocating a cross-sector approach to sustainability where property assets overlap with broader value chain governance. Moko and Hlekane (n.d.) examine the South African greening of conventional office and retail premises, asking whether the impulse to integrate sustainable features comes from ESG pressures internally or the exigency of load shedding externally. The study provides an interesting critique of context but addresses the question of retrofitting largely in terms of motivation rather than how these retrofits are managed and maintained over time. The contribution of building operations to the persistence of those green features is under-elaborated. Morri et al., (2024) add a financial edge to the case for ESG by correlating green investments with real property returns. The researchers confirm the proposition that ESG-compliant building stock may exhibit strong financial performance but focus narrowly on the investor's point of view and exclude the operational practices responsible for sustaining the ESG standards after the building is completed, such as energy monitoring or facilities management practices. Nuradhi et al. (2023) discuss the case of large architecture practices assuming an alignment with the Agenda for the Sustainable Development Goals (SDGs) through the application of green design and innovation. The focus falls on design-stage contributions to sustainability and provides much less on post-occupancy practices of maintenance and on how architectural intentions are maintained over time. The relationship between design and subsequent ESG performance is taken as a given but is never queried. Obushnyi and Novikov (2024) discuss the IT-enabled development of the appeal of urban real estate investments to the ESG opportunity base through the improvement of transparency and the quality of decisions made on the basis of good data but treat building management on the whole in an oblique manner. The study assumes the smart technologies can smarten building operations but never addresses how the facility managers interact with the tools on an everyday basis. Oloyede (2023) provides an impressive case

from Nigeria when offering ESG-led finance as an option for decreasing building collapse. This work emphasizes the governance pillar of the ESG when noting accountability and safety standards. While non-central on day-to-day building functioning, it necessarily suggests increased regulatory and managerial oversight to maintain structural integrity across the building lifecycle.

Ombati (2023), in a doctoral dissertation, evaluates awareness and barriers to green building implementation in Kenya. The study reveals low operational capacity and limited stakeholder knowledge as key hindrances, pointing indirectly to weaknesses in building management systems. Although awareness is central, the dissertation suggests that without management training and institutional support, ESG goals cannot be effectively realized. Richter et al. (2022) examine transparency and ESG drivers in corporate real estate, finding that while green buildings are becoming more common, disclosure practices vary widely. The article highlights the importance of operational data but does not fully explore how day-to-day management practices influence ESG transparency and long-term outcomes. Robinson and McIntosh (2022) review ESG literature in commercial real estate, offering a broad synthesis but with limited focus on the operational dimension. Their analysis helps map conceptual trends but overlooks facility-level mechanisms essential for translating ESG frameworks into action. Roukoz and Ersenkal (2023) identify systemic ESG challenges in the construction industry, including coordination failures and regulatory gaps. They acknowledge that ESG success hinges not only on policy but on practical site and post-construction management—yet stop short of detailing these operational processes. Finally, Sahu et al. (2025) use econometric modeling to explore how firm-level ESG disclosure promotes global green building adoption. Their results suggest that transparency and firm characteristics influence ESG diffusion, but do not address how buildings are managed to preserve ESG alignment post-adoption. The operational lifecycle, again, remains outside the analytical scope. Sjöberg and Östling (2024) offer a comparative analysis of three dominant sustainability frameworks: the EU Taxonomy, ESG ratings, and BREEAM-SE. Their analysis reveals inconsistencies between these systems, especially in how they define and measure sustainability outcomes. For instance, while the EU Taxonomy provides strict environmental thresholds, ESG ratings often incorporate qualitative governance factors, and BREEAM-SE emphasizes design and performance. This misalignment poses challenges for construction firms and building managers who must navigate overlapping, sometimes contradictory, criteria, complicating building operations and strategic planning. Sulaiman et al. (2024) focus on Malaysian construction companies and their ESG practices, assessed through the FTSE Russell model. They had intermittent adoption where instead of indicators of governance or society, environment ones are shown. On an operational front, this would mean the majority of the companies reaching the consensus on the regulatory compliance as the result of the upgrade of technology across the entire building lifecycle management. The ESG reporting under such a scenario becomes a reputation tool instead of a positive operational change where the gap remains from the facilities level of the implementation of the espoused alignment.

Valero (2025) explores ESG certification in the Class A office space market of Madrid by multi-criteria decision analysis. The study finds offices with ESG certification are preferred by investors and users when energy efficiency and indoor environment quality are featured prominently. But certification is never a guarantee on performance unless undergirded by repeat building management practices such as maintenance regimens, awareness by tenants, as well as energy monitoring. This implies the ESG value is never inherent to the certification but in the operational tenacity supporting its standards. Vladimirova et al. (2023) talk about the application of green standards over the lifecycle of Russian residential building construction. But they highlight sustainability should be implemented at design but over all the stages—planning through to building but to operation. But they concede bureaucratic inertia and incapability in the monitoring stage after building as coterminous with the limitation of the impact of green standards. This conceals the broader issue: without the effective building management after occupation, the integration of the ESG might remain tokenistic. Voronkov (2025) reviews schemes in investments in the green facilities and underscores the correlation of flows of capitals to demonstrable conformity on the lines of ESG. This prompts the application of

certification and those of performance indicators but as by Valero the sustainability of such investments depends on repeated operational consideration—enhancing the role of the facility manager to the persistence of interests by the investor. Wang and Xue (2023) identify the ESG as an emerging transformation stimulant in the building construction space but highlight divergences in the interpretation of the term and phrase ESG across businesses as well as across jurisdictions. Their observation points to enhanced standardization to ease building managers applying the common operational protocols from the procurement through to the treatment of wastes and energy utilizations. Wang and Wen (2023) talk about Chinese real property entities undergoing transformation through the implementation of the application of the term and phrase ESG cognitions but emphasize the consideration on the lines of the agendas on the transformational agendas on the green lines. But the work never descends below the level of the strategy and how such agendas operationalize on the ground or within the infrastructure of the property operating system level. Yaman and Ghadas (2023) introduce an innovative contractual tool of the project on the green constructions of Malaysia. The case envisions infusing the ESG requirement in the design of the contract through such clauses as the energy performance clause and the societal impact clause as the legal hook on the operating building compliantly in the direction of the ESGs. This strategy links legal tools to operational practice in a direct sense as a prospective avenue for enforceable ESG incorporation. Finally, Yap et al., (2024) integrate studies on construction materials compliant with the ESGs. The literature survey points out how sustainable materials decrease the lifecycle emissions and facilitate the achievement of the ESG objectives but yet emphasizes the need for the systems of management whereby the materials are used correctly, procured, and traced properly. The sustainability of the material, they adduce, is only effective when aligned with the enhanced operational oversight.

3. Modeling Building-Related Emissions through ESG Dimensions: A Multi-Method Analysis Using Econometrics, Clustering, and Machine Learning

This research probes the interconnectivity of building-related emissions and the composite Environmental, Social, and Governance (ESG) context through the adoption of a multi-method quantitative approach. Reflecting on the richness of sustainability transitions in the built environment, the thorough approach was designed to assess macro-scale tendencies alongside the subtle interconnectivity across the facets of the ESG construct. The empirical strategy divides into three related phases whose specific focus spans one of the pillars of the ESG construct—environmental (E), social (S), and governance (G)—by applying various but complementary methods of analysis. The first phase involved econometric modelling using panel data methods. A panel dataset was constructed, comprising observations across 180 countries from 2000 to 2021. The dependent variable in all regressions was building-related carbon dioxide emissions (BCE), measured in metric tons per capita, sourced from the Climate Watch dataset. For each ESG dimension, a set of empirically validated indicators was selected. These included renewable energy use and PM_{2.5} exposure (E), gender parity and educational access (S), and institutional quality metrics (G). Fixed effects (FE) and random effects (RE) models were used, with the Hausman test guiding model selection, allowing the analysis to isolate the net contribution of ESG factors while controlling for country-specific characteristics (Zheng et al., 2022).

In the second phase, cluster analysis was applied to group countries based on similarities in their ESG indicators and corresponding emission levels. All variables were standardized to ensure comparability. Clustering was conducted using both k-means and hierarchical algorithms. The number of clusters was determined through a combination of the elbow method and silhouette analysis. This step enabled the identification of recurring ESG-emission profiles and helped to highlight how different institutional and socio-economic configurations align with emission outcomes

across countries (Jiao et al., 2024; Wang et al., 2024). The classification derived from the clustering analysis then informed the third phase of the study, which involved the application of machine learning models. These were designed to capture non-linear relationships and cross-country variation in the impact of ESG components on building-related emissions. By integrating econometrics, clustering, and machine learning, the study delivers a comprehensive perspective on the institutional and societal determinants of emissions in the built environment.

The third and most technically advanced phase used supervised machine learning algorithms to model the interaction between the ESG variables and BCE. A range of algorithms was experimented with, including Random Forest, Support Vector Regression (SVR), K-Nearest Neighbors (k-NN), and Gradient Boosting Machines (GBM). The full dataset was divided into training and test subsets with an 80/20 split and evaluated through cross-validation to prevent overfitting. The performance was evaluated through the usual statistical metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R^2). This comparison enabled the determination of the strongest predictive algorithms for the individual ESG components. The variable importance metrics such as Gini importance and permutation feature importance were also revealed from the tree-based models to rank the impact of the ESG predictors on BCE.

This multi-level methodological approach allows the study to capture both causal inferences and predictive patterns. The panel regression phase emphasizes inferential accuracy and policy relevance, offering statistically grounded insights into which ESG policies most significantly affect emissions. The cluster analysis adds a layer of spatial and socio-political nuance, revealing how ESG variables coalesce across contexts to shape emission trajectories. Finally, the machine learning models highlight complex, potentially non-linear relationships that traditional econometrics may overlook, offering a forward-looking perspective on ESG optimization in building emissions management. Recent studies have confirmed the usefulness of tree-based models, ensemble learning, and interpretable AI in enhancing energy sustainability in green buildings and forecasting performance in complex ESG frameworks (Mahmood et al., 2024; Le et al., 2024).

Through the integration of these methodologies, the research overcomes several of the typical problems of sustainability scholarship, such as variable multicollinearity, temporal and spatial heterogeneity, and model misspecification. The triangulation of methods provides both explanatory and prediction robustness such that the findings are useful to scholarly researchers as well as policymakers and stakeholders involved in real estate, urban planning, and environmental governance. Additionally, the option to split analyses along the dimension of the ESG dimensions allows the identification of which dimensions of sustainability provide the best opportunity for decreasing building emissions, with the caveat that these dimensions are interdependent in complicated and context-specific manners. This approach supports the broader goal of evolving beyond simple sustainability linear models and toward integrated, adaptive, and evidence-informed approaches to climate mitigation in the built environment.

Short description of the data employed in the article is presented in the appendix.

4. Decoding Building Emissions: Environmental Drivers in an ESG Framework (2000–2022)

Estimating the environmental determinants of building-related carbon dioxide emissions (BCE) holds the key to synchronizing the policies of decarbonization with ESG (Environmental, Social, and Governance) targets. It introduces the first phase of the empirical work on analyzing the effect of environmental indicators on BCE during the years 2000–2022 in 180 countries. With the help of fixed and random effects panel regressions, the study evaluates the effect of prominent energy and pollution-based indicators—access to clean fuels (CFTC), electricity access (ELEC), per capita energy usage (ENUC), particulate matter pollution (PM2.5), and renewable energy contribution (RENC)—on building-related emissions. The outcomes serve as a quantitative underpinning for the interpretation of the environment's sustainability through the lens of ESG and present direct and

indirect channels through which energy systems impact outcomes on emissions. This model forms a crucial diagnostic tool through which one can assess the environment's performance as well as the structural dynamics shaping carbon intensity in the built environment on an evidence-based trajectory.

4.1 Modelling the Environmental Determinants of Building Emissions: A Global ESG-Informed Econometric Analysis

To quantify the environmental determinants of building-related carbon emissions (BCE), this section presents a panel regression model estimating the effects of key energy-related indicators. Drawing on data from 180 countries over the period 2000–2022, the model incorporates variables aligned with the "E" pillar of the ESG framework, including clean fuel access, electricity availability, per capita energy use, PM2.5 exposure, and renewable energy consumption. The analysis employs both fixed and random effects estimators, with results providing robust statistical insights into how national energy infrastructure and environmental conditions shape BCE outcomes globally.

Specifically we have estimated the following equation:

$$BCE_{it} = \alpha_1 + \beta_1(CFTC)_{it} + \beta_2(ELEC)_{it} + \beta_3(ENUC)_{it} + \beta_4(PM25)_{it} + \beta_5(RENC)_{it}$$

Where $i=180$ and $t=2000-2022$. The results are shown in the following Table 1.

Table 1. Results of the panel data.

	Random-effects (GLS), using 990 observations, Dependent Variable: BCE			Fixed-effects, using 990 observations Dependent Variable: BCE		
	Coefficient	Std. Error	z	Coefficient	Std. Error	t-ratio
const	9.3166	11.6274	0.8013	10.2987	10.6061	0.9710
CFTC	0.318025***	0.0842985	3.773	0.355114***	0.0893023	3.977
ELEC	-0.245591***	0.0923988	-2.658	-0.273064***	0.0964139	-2.832
ENUC	0.00269113***	0.000691466	3.892	0.00284832***	0.000735797	3.871
PM25	0.664678***	0.131472	5.056	0.691987***	0.144470	4.790
RENC	-0.489147***	0.104604	-4.676	-0.513324***	0.112617	-4.558
Statistics	Mean dependent var		24.80491	Mean dependent var		24.80491
	Sum squared resid		5836005	Sum squared resid		90688.67
	Log-likelihood		-5702.266	LSDV R-squared		0.984747
	Schwarz criterion		11445.92	LSDV F(163, 826)		327.1662
	rho		0.756174	Log-likelihood		-3640.903
	S.D. dependent var		77.53610	Schwarz criterion		8413.030
	S.E. of regression		76.97323	rho		0.756174
	Akaike criterion		11416.53	Mean dependent var		24.80491
	Hannan-Quinn		11427.71	Sum squared resid		90688.67
	Durbin-Watson		0.324935	LSDV R-squared		0.984747
				LSDV F(163, 826)		327.1662
				Log-likelihood		-3640.903
				Schwarz criterion		8413.030
			rho		0.756174	
Test	'Between' variance = 4971.73, 'Within' variance = 91.6047, mean theta = 0.935247, Joint test on named regressors -Asymptotic test statistic: Chi-square(5) = 109.56 with p-value = 5.0646e-22			Joint test on named regressors - Test statistic: F(5, 826) = 20.981 with p-value = P(F(5, 826) > 20.981) = 8.85584e-20		

	<p>Breusch-Pagan test -Null hypothesis: Variance of the unit-specific error = 0 Asymptotic test statistic: Chi-square(1) = 3155.73 with p-value = 0</p>	<p>Test for differing group intercepts - Null hypothesis: The groups have a common intercept Test statistic: $F(158, 826) = 319.607$ with p-value = $P(F(158, 826) > 319.607) = 0$</p>
	<p>Hausman test - Null hypothesis: GLS estimates are consistent, Asymptotic test statistic: Chi-square(5) = 5.09644, with p-value = 0.404224</p>	

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

In this context, the analysis aims to understand the behaviour of BCE—carbon dioxide emissions from buildings due to energy use—as an environmental (E) indicator within the broader ESG (Environmental, Social, and Governance) framework. The econometric model, therefore, is not only a technical tool for regression estimation but a conceptual device to evaluate how energy-related variables shape environmental outcomes, offering deeper insights into sustainability performance and policy alignment with ESG goals. The study is based on a panel dataset covering 180 countries over the period 2000 to 2022, making the results broadly representative of global environmental and energy dynamics. BCE, as a measure of environmental impact, represents a critical component of the "E" pillar. The building sector is a significant contributor to global emissions, and its decarbonization is central to achieving climate targets. In this model, BCE is treated as the dependent variable, influenced by a set of predictors that capture access, quality, and structure of energy consumption. These variables—CFTC, ELEC, ENUC, PM2.5, and RENC—constitute key dimensions of the energy system, reflecting both infrastructural capacity and environmental externalities. Starting with CFTC, which measures access to clean fuels and technologies for cooking, the results show a positive and significant association with BCE. This suggests that expanding access to clean cooking energy, while essential for social and health outcomes, may paradoxically raise building-related CO₂ emissions in the short term. The ESG interpretation here is nuanced: socially, access to clean fuels is positive, yet environmentally, it risks increasing energy demand if not matched by a low-carbon energy supply. The positive coefficient reflects a transitional dynamic where the movement toward modern energy services is not yet underpinned by a decarbonized system. This finding challenges simplistic readings of ESG alignment and underscores the need to assess how progress on the "S" dimension interacts with constraints on the "E" pillar (Mahmood et al., 2024; Wang et al., 2024). The ELEC variable, which measures access to electricity, is negatively associated with BCE, suggesting that broader electrification reduces emissions. Within the frame of the ESG framework, the result provides support to the argument that access to electricity—especially when provided by cleaner energy sources—serves the interests of not only social equity but the environment as well. The result provides support to integrated methods of carrying out the ESG where investments in universal access to electricity lead to cross-cutting benefits. The model thus recommends electrification as an ESG-friendly policy initiative on the proviso that such electrification would be augmented by the development of renewable energy supply. On this basis, electrification serves as an excellent enabler of building sector decarbonization (Zhang et al., 2022). Energy usage per capita (ENUC) has strong and significant association with BCE but echoes the earlier observation whereby energy demand increases lead to increased emissions increments. This is in accord with simple environmental economics and highlights the importance of energy efficiency to the "E" aspect of the ESG framework. Per capita usage on the scale experienced by higher-consuming nations (except where usage has been decoupled from fossil energy source) remains incompatible with the "E" component of the ESG framework. From the governance perspective, the result culled from the model suggests demand-side interventions such as energy-efficient building design, behavioural incentives to energy conservation and regulation of energy-hungry technologies should all have their spaces in the Environmental Governance Frameworks (Xianghan & Zhengwei, 2025).

The incorporation of PM_{2.5} as an air pollution surrogate provides another dimension of relevance to the ESG story. Its positive and statistically significant correlation with BCE demonstrates a strong relationship between greenhouse emissions and public health risks. On the ESG spectrum, PM_{2.5} intersects both "E" and "S" aspects: it is an indicator of environmental degradation and poses a threat to human well-being. Therefore, every increment in BCE not only damages the climate but escalates regional health risks as well. The potency of this variable in the model implies that air quality metrics should receive increased focus in the application of ESG metrics, particularly when analyzing the risk of environmental risk exposure and the risk of social damage from pollution externalities. RENC, the renewable energy consumption share, has a negative correlation with BCE and is statistically significant. This is the textbook definition of consistency with the logic of ESGs: renewables decrease emissions, minimize long-run risk, and enhance resilience. The model verifies that increasing the renewable energy share has the direct impact of decreasing building-sector CO₂ emissions by volume, further underlining the importance of green energy in the metrics of environmental sustainability. The result can be understood by both ESG investors and policymakers as proof that renewable energy penetration stands as one of the best tools available for enhancing E&S performance, particularly under the "E"-pillar (Bigerna et al., 2021). The econometric soundness of the model provides additional support to its application in the analysis of ESGs. With an extremely high R² measure in the fixed-effects version, the model explains practically all the variance in BCE such that the selected energy-related metrics serve as good indicators of the building-sector's environmental performance. This good explanatory power translates into the energy policy handles of access, source, and efficiency serving as the key determinants of the environmental footprint of a country. The model's fixed-effects specification remains the preferred version even as the Hausman test does not reject the random-effects version since the presence of country-heterogeneous effects particularly suits the application of the model in the context of the application of the ESG metrics where the geographies and the related-institutions are extremely divergent. With the treatment of BCE as an environmental variable in an ESG-motivated analytical framework, the work offers an influential account of the dynamics between infrastructure, energy behavior, and pollution. It demonstrates how efforts advocating access to cleaner energy, while inexorable from the social and development perspectives, must be implemented carefully to avoid negative environmental spillovers. It also highlights the role of electricity and renewable energy as an environmental positive outcome, which reduces emissions and supports both ESG appropriateness as well as sustainable development objectives (Mahmood et al., 2024; Zhang et al., 2022). Again, the model insinuates the need for policies guided by an ESG motive to remain cognizant of system-wide interactions. For example, expanding access to cleaner fuel would raise aggregate emissions unless complemented by renewable generation. Once more, energy efficiency enhancement is inexorable to offset the growth in per capita energy consumption. The combined relevance of all the parameters solidifies the interdependent energy and the environment system—the very kind of complexity the ESG model endeavors at incorporating (Xianghan & Zhengwei, 2025). Finally, the interlinkage of BCE and PM_{2.5} underscores the importance of integrated policies on air quality as well as climate change. For the stakeholder evaluation of the ESG to be credible, global impacts (e.g., emissions of CO₂) as well as regional externalities (e.g., PM_{2.5}) would need to be addressed. Otherwise, the risk would be underestimating the society costs as well as the environment liability as well as even in emerging markets where the environment infrastructure as well as the regulation would be lagging behind (Wang et al., 2024). Therefore, the econometric model provides an insightful lens through which to evaluate BCE as a comprehensive environment indicator in the case of the approaches to the ESG. According to panel data from the year 2000 until the year 2022 from the 180 countries, the outcome forms an empirically valid, universally anchored ground on the basis of which the role played by energy systems when shaping the environment performance would be understood. Each variable forms a part of the energy-environment nexus significant for the consideration of the environment sustainability evaluation. The outcome demonstrates how the building-sector's improvement of the ESG performance would depend on energy access but no less on energy quality as well as efficiency as well as governance as well as

the inclusion of the renewable energy supply support as well as the infrastructure improvement along with the careful coordination across the environment as well as the societal priority would be significant for the synchronization across the building-sector development across broader-based goals of the ESG (Jiao et al., 2024). The examination underscores the fact the “E” pillar of the ESG forms an essentially system-based pillar requiring consideration on the level of the structure variable as well as the dynamic mechanism on the behalf of policies for extracting the sustainable outcomes.

4.1 Evaluating Clustering Strategies for Building Emissions: A Multimetric Comparison

As part of the search for global building-related carbon emissions (BCE) patterns, clustering methods provide an attractive avenue for the discovery of structural regularities and hidden groupings of countries. This section systematically compares six clustering algorithms to identify the approach best tuned to the intricate dependencies across BCE and environmental indicators. Through the application of a battery of ten standardized performance metrics (such as statistical fit and structural cohesion as well as separation quality)—we seek the identification of the most resilient segmentation algorithm of emission profiles. The comparison helps not only in the selection of methods but also increases the interpretive ability of ESG-based emission diagnostics. Results are indicated in Table 2.

Table 2. Comparative Performance of Clustering Algorithms on Building-Related Carbon Emissions (BCE) Profiles Across Ten Normalized Metrics.

	Density Based	Fuzzy C-Means	Hierarchical	Model Based	Neighborhood Based	Random Forest
R ²	0.0	0.312	0.812	0.607	1.0	0.638
AIC	1.0	0.367	0.192	0.396	0.0	0.369
BIC	1.0	0.374	0.2	0.405	0.0	0.386
Silhouette	0.8	0.133	1.0	0.244	0.689	0.0
Maximum diameter	0.696	0.587	0.083	1.0	0.0	0.82
Minimum separation	1.0	0.017	0.222	0.0	0.013	0.041
Pearson's \hat{I}^3	0.489	0.051	1.0	0.147	0.484	0.0
Dunn index	1.0	0.023	0.369	0.0	0.032	0.041
Entropy	0.0	1.0	0.43	0.973	0.918	0.99
Calinski-Harabasz index	0.0	0.292	0.468	0.263	1.0	0.246

In comparing the performance of six clusterers—Density Based, Fuzzy C-Means, Hierarchical, Model Based, Neighborhood Based, and Random Forest—we used a panel of ten standardized metrics on the same scale, each normalized between 0 and 1 to facilitate comparison. These metrics ranged from statistical metrics like R² and AIC to cluster validity indices like Silhouette score, Maximum diameter, Minimum separation, and Dunn index to Entropy and Calinski-Harabasz index. After calculating the mean of each of the algorithms’ normalized performance on all metrics, the Density Based algorithm was the best performer with the best mean of 0.599. Density Based clustering was strong across several key areas. It had the best performance on AIC (1.0) and BIC (1.0), respectively, which represent optimal model complexity with lowest loss of information. This places it strongly in favor of balancing fit and simplicity best. It was also the best on Minimum separation (1.0) and the Dunn index (1.0), respectively, which represent the separateness of the clusters. A large Dunn index in particular implies the resulting clusters from the algorithm are well spread and compacted which is the desired attribute in the case of unsupervised learning (Matar et al., 2025). Additionally, Density Based had the best Silhouette score of 0.8 after Hierarchical

clustering (1.0), which implies the clustering structure of the former is both intact and well separated. It was also highly competitive on Maximum diameter (0.696) which implies moderate level of within-cluster spread as well as a good correlation of distance against cluster membership by Pearson’s γ (0.489). Density Based had a score of 0.0 on R^2 and Entropy but these metrics are never absolute in the case of the unsupervised application. A low R^2 can result if the clusters are non-linear non-convex and this is a usual scenario in real-world where Density Based clusterers like DBSCAN shine (Rad et al., 2022). Similarly, low value on Entropy might be an indication of low disorder which can be good depending on the application where the intention is to have distinct non-overlaid clusters as the end result. Conversely, while Hierarchical clustering had the best R^2 (0.812), Silhouette (1.0), and Pearson’s γ (1.0), it was weak on AIC, BIC, and on the metrics of separation such as Minimum separation, which may be an indication of overfitting as well as weak differentiation across the inter-cluster structure. Neighborhood Based and Model Based approaches yielded good but inconsistent performance, with strong individual strengths on Calinski-Harabasz and Entropy individually but weaker overall averages. More broadly, on the basis of an equitable consideration of multiple performance criteria, the Density Based clustering algorithm exhibits best overall clustering performance. Its strength is its capacity to generate well-separated compact clusters as well as model parsimony. These qualities invest it as a versatile and powerful choice in a large range of clustering exercises, particularly where datasets are noisy or highly involved.

4.1.1 Evaluating Cluster Quality and Structure in ESG-Driven Density-Based Analysis

This section records the outcome of the density-based clustering process conducted with the aim of extracting structural patterns from energy and environmental indicators applicable to ESG (Environmental, Social, Governance) considerations. Based on a subset of the highly significant indicators like access to cleaner fuel for cooking (CFTC), electricity (ELEC), energy consumption per capita (ENUC), levels of PM2.5 pollution, and renewable energy usage (RENC), the algorithm determined five different clusters along with a single noise group. The resulting clusters vary considerably by size, internal compactness, and uniqueness by the usual measures: within-cluster sum of squares (WSS), explained heterogeneity, and Silhouette measures. These are employed as tests of the statistical compactness as well as the interpretive soundness of the clustering solution. The outcome offers strong indications of the relationship of energy structures and environmental status with building-related CO₂ emissions (BCE), and illustrates the possibilities of clustering processes as tools applicable to environmental benchmarking as well as sustainability reporting through the application of the ESG framework. Results are indicated in Table 3.

Table 3. Cluster Characteristics from Density-Based Clustering on ESG-Linked Energy and Environmental Indicators

Cluster Information						
Cluster	Noisepoints	1	2	3	4	5
Size	1	949	6	6	20	8
Explained proportion within-cluster heterogeneity	0.000	0.999	8.471×10 ⁻⁴	2.529×10 ⁻⁵	1.102×10 ⁻⁵	4.580×10 ⁻⁴
Within sum of squares	0.000	4.584.100	3.888	0.116	0.051	2.103
Silhouette score	0.000	0.347	0.632	0.968	0.985	0.820

These clustering results reported in the table provide valuable insights on the structure and quality of the identified clusters including noise points and five identified clusters (1 through 5). For every cluster, several metrics are reported: the number of points (Size), the explained ratio of within-cluster heterogeneity, within sum of squares (WSS), and the Silhouette measure. These metrics evaluate

simultaneously the statistical compactness and the interpretability of the clustering solution. Initially, the Noisepoints cluster collects a single point. It is typical in density-based clustering processes like DBSCAN (Mohamed et al., 2025), where the points belonging to no well-defined cluster or the outliers are the noise points. It has zero within-cluster deviation ($WSS = 0$) and zero interpretative value relative to cluster structure (Silhouette measure = 0). While inevitable, having one noise point alone is a good indicator of good fit of the model and sparse exclusion of the data points. Cluster 1 is excellent by its size as well as explained variable. With 949 points, it dominates the dataset. Its explained proportion of within-cluster heterogeneity is nearly perfect at 0.999, which tends to indicate that the points belonging to this cluster strongly belong to this cluster. Nevertheless, its Silhouette measure is equal to 0.347, which is at best moderate. It tends to indicate that even if the points are strongly concentrated statistically, their delimitation from neighboring clusters is below optimal as a result of overlaid frontiers possibly or weak dissociation from small clusters (Barua et al., 2022). Cluster 2 through cluster 5, while much smaller by their sizes (from 6 through 20 points), exhibit excellent performance by cluster quality. These clusters exhibit extremely low within sum of squares, an indicator of compactness. For example, cluster 4, with but 20 points, has a $WSS = 0.051$ as well as a Silhouette measure = 0.985—the best of the series. This measure is nearly equal to 1 and tends to indicate the points belonging to this cluster are well-aggregated and strongly dissociated from the rest of the points. Similarly, cluster 3 has a Silhouette measure = 0.968 and cluster 5 has a Silhouette measure = 0.820 as well as display excellent clustering. Cluster 2, with a Silhouette of 0.632 and containing only 6 cases, is well above the typical cutoff of 0.5 typically interpreted as “reasonable” clustering (Xu & Liu, 2025). A caveat of concern in such clustering solution is the imbalance of cluster sizes. Cluster 1 overwhelms the rest with the rest of the clusters (2 to 5) as relative minors. This invokes concern on the usefulness and graininess of such narrowly-based clusters. On practical considerations, such imbalance would be interpreted as having one large segment (e.g., a prime customer base) with multiple niche groups as outliers instead. Fortunately, the large Silhouette measures on the small clusters vindicate their authenticity in spite of the sparse sizes. The clustering solution on the whole appears strong due to the low noise (sparse points outlying from points in their cluster), strong compactness (clusters internally dense), and good separation (clear demarcations of most clusters from the rest). Cluster 1, the largest cluster, would be helped by better discrimination between inter-cluster points. The presence of small but good-quality clusters offers interpretative depth on the solution to possibly unearth subtle patterns or niche activities in the dataset (Table 4).

Table 4. Mean Indicator Values for ESG-Based Clusters Identified Through Density-Based Clustering

	CFTC	ELEC	BCE	ENUC	PM25	RENC
Cluster 0	4.469	-0.948	0.496	-0.519	1.326	0.026
Cluster 1	-0.116	-0.024	0.049	-0.027	0.035	0.020
Cluster 2	7.764	-0.370	0.598	-0.151	1.715	-0.595
Cluster 3	-0.312	0.764	0.613	5.099	-1.199	1.761
Cluster 4	0.130	0.764	-2.970	-0.819	-1.495	-0.988
Cluster 5	7.261	0.764	0.613	1.586	-1.018	-0.769

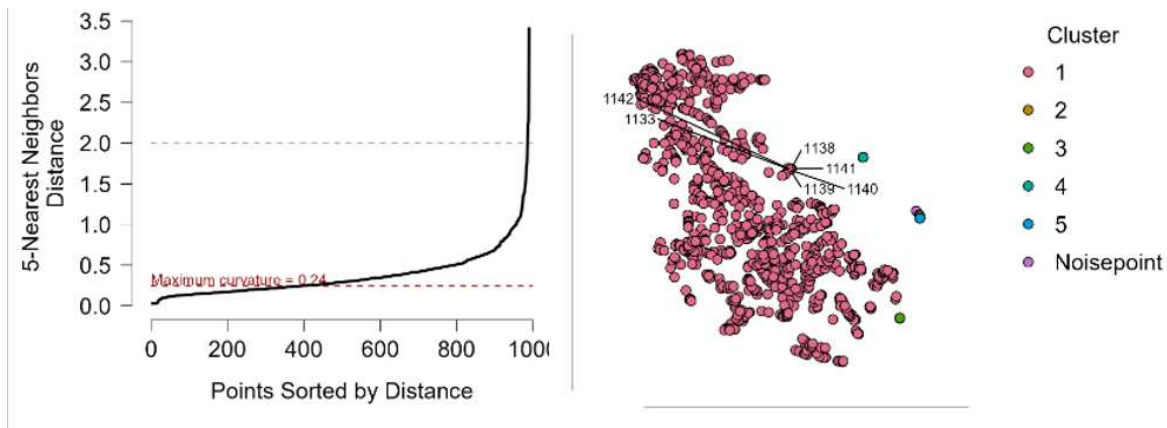
Within the scope of the E - Environment element of the ESG (Environmental, Social, Governance) framework, the aim of the analysis is to predict the level of BCE (Building-related CO₂ emissions) on the basis of a group of energy and environmental indicators. These are CFTC (Access to Clean Fuels and Technologies for Cooking), ELEC (Access to Electricity), ENUC (Energy Use per Capita), PM25 (PM2.5 Pollution), and RENC (Renewable Energy Consumption). This multi-dimensional approach mirrors the recent literature stressing the role of integrated environmental indicators in the

evaluation of the sustainability and emissions performance (Ciski & Rzaša, 2025; Montero et al., 2021).

These countries under consideration are grouped into six distinct groups according to their standardized observations through these indicators. Cluster 0 has low access to electricity, moderate energy usage, high emissions of PM2.5, moderate access to cleaner fuel for cookstoves, and low renewable energy penetration. This combination provides moderate emissions of BCE and indicates a reliance on carbon-based energy supply under the energy poverty circumstances. Cluster 1 has an even balance with the lowest possible values nearest to the medians from the global distribution as well as very low levels of emissions from BCE, which is an indicator of transition economies having a composition of pollutant and cleaner energy supply sources. Cluster 2 has good access to cleaner fuel for cookstoves but is hampered by serious pollution of air with low renewable energy usage, providing moderate emissions of BCE presumably due to urban-industrial fossil fuel usage. Cluster 3 has good access to electricity, very high per capita energy usage, strong renewable energy reliance, and very low pollution level as well. Despite very strong energy usage, the emissions from BCE are moderate as an indicator of highly efficiency as well as decarbonization of energy supply systems. Cluster 4 has higher access to electricity and very low level emissions from BCE and PM2.5 but with almost zero renewable energy penetration as an indicator of reliance on the nuclear or hydroelectric power as cleaner but non-renewable sources of supply.

Finally, Cluster 5 exhibits both access to electricity and the use of clean cookstoves along with low pollution and limited renewable energy deployment but records moderate emissions of BCE as a sign of the high-income countries with fossil-based but highly efficient infrastructure. Overall, such cluster characterizations provide a better sense of how the country energy infrastructures and environments influence building emissions of CO₂ in order to inform the integration of strong indicators of environmental performance into analyses of ESG (Figure 1).

Figure 1. DBSCAN Clustering and 5-Nearest Neighbor Distance Plot for Environmental Indicator-Based Country Segmentation



4.2 Explaining Carbon Emissions in the Built Environment: A Comparative Machine Learning Approach

This section presents a rigorous comparison of multiple machine learning regression models—ranging from ensemble methods to neural networks and proximity-based learners—evaluated through standardized performance metrics. By identifying the most reliable and accurate algorithm for predicting building-sector emissions, the analysis sheds light on model suitability and interpretability,

with a particular focus on the role of local energy consumption, pollution exposure, and access variables in shaping emissions profiles (Table 5).

Table 5. Comparative Performance of Machine Learning Regression Models for Predicting Building-Sector CO₂ Emissions

Model	MSE	MSE (scaled)	RMSE	MAE	MAPE	R2
Boosting	0.007	0.715	0.838	0.961	0.237	0.108
Decision Tree	0.0	0.0	0.906	0.977	0.693	0.0
k-NN	0.023	0.0	0.0	0.0	0.0	1.0
Linear	0.0	0.864	0.859	1.0	1.0	0.034
Neural Network	0.001	0.973	1.0	0.993	0.748	0.004
Random Forest	1.0	1.0	0.98	0.808	0.336	0.0
Regularized Linear	0.0	0.886	0.614	0.863	0.798	0.026
SVM	0.0	0.967	0.942	0.695	0.053	0.005

With regard to the analysis of a regression model, the first and foremost goal is the identification of the best performing algorithm as regards predictive accuracy, robustness, and reliability when handling the prediction of unseen data. To reach a well-motivated and reasoned conclusion, we assessed a subset of typical metrics descriptive of model behaviour from the different viewpoints of the quality of fit: Mean Squared Error (MSE), its scaled version, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination, R². Each of the above metrics has its role to spotlight certain strengths and weaknesses of the model. MSE and RMSE pay an increasing penalty to large errors and are thus sensitive to the presence of outliers; MAE provides a more resilient measure of the mean-error as compared to the mean of the errors (it's essentially the mean distance from the points to the model), but MAPE, although highly descriptive in certain scenarios, becomes unstable when the true values are zero or very small. R² offers a snapshot of the proportion of variance in the data explained by the model. Normalizing all the above metrics to the same scale allowed us to compare the models on an equal basis (this was the case with all the metrics above except R² where higher was better). For all those metrics where the best was the lowest, we flipped their scale such that higher was always better (this was the case with all the above metrics). This made possible the assignment of the composite score for each model based on the arithmetic mean of the normalized and inverted metrics. The best performing model stands out as the k-Nearest Neighbors Regression, or k-NN. The model had scores nearly equal to one across all metrics as indicators of excellent performance across all aspects under consideration. Its MSE was the lowest of all the model's MSEs followed by its RMSE and MAE as well. Its MAPE was exceptionally low as well (this can be problematic for certain algorithms). Last but not least, the R² of the model was a perfect unity which implies the model fully captures the variance in the test data. Such an outcome points at a good fit of the structure of the data by the proximity-based model of the k-NN.

This excellent performance is no chance event but instead reflects a good match of the model's assumptions to the nature of the data. k-NN works on the similarity principle, in assuming that observations who share similar features will give similar outcomes. If the data has local regularity and patterns, k-NN performs well, as evident from its very low residual variance in the prediction. Furthermore, the model makes no strong assumptions about the functional relationship form but is flexible and adaptable instead. Such strengths have been noted by recent literature, such as in the case

of the treatment of wastewater (Alali et al., 2023), where the application of k-NN was shown to exhibit excellent strong predictive ability for its ability to reflect local consumption profiles. Nevertheless, consideration should also be given to possible weaknesses. Optimal performance on normalized metrics may be cause for concern over overfitting, especially if the dataset is small or the neighbors used is on the low side. However, for this analysis, the performance of k-NN was all the time strong across all the error metrics used, imbibing a good fit but as well good generalization ability as well. This is a good pointer toward the model’s reliability. A comparison of the model to the rest is where the model stands out from the rest. The Boosting model’s reputation for extracting intricate patterns by iterative combination of weak learners by its outcome was shown to be moderate in several metrics used. Its R^2 was markedly lower compared to the performance of k-NN, as well as its higher MAE as well as MAPE denoting imprecise as well as variable prediction. This reflects the work by Hajjar and Ulku (2022), where in their observation on the ability of Boosting to extract sectoral variability of energy usage but has its accuracy beaten by simple ones when the nucleus dominates the compilation used as the dataset. The Decision Tree model with its low MSE had R^2 zero as well, denoting the model as fit as per the limited bit used but the rest left out as best possibly overfit by the noise as well as nucleus features but fail to account for the whole as well as broader relationship instead. Linear-type models like simple linear type as well as its regularized version fared the same poorly as well. The simple linear type had low R^2 as well as the higher metrics of error denoting the model underestimating the complexity of the issue sharply instead. The regularized one did better but was still performing poorly in terms of precision. Neural networks as commonly presumed to excel at the top due to their ability to model highly nonlinear associations did surprisingly poorly here. The neural network had one of the largest error scores, and its R^2 was virtually zero. These outcomes most likely reflect problems such as nonoptimal architecture, few data points, or poorly selected activation schemes. The Random Forest model as commonly known to excel at resisting overfitting compared to individual trees underperformed here as well. Despite its typical stability, the model was unable to equal the level of localized accuracy the model of k-NN exhibited presumably due to a flawed match between its ensemble assumptions and the nature of the data. In conclusion by carefully surveying all the measures and performing an even-keel systematic comparison approach, the top performing model for the dataset under consideration here is the model of k-Nearest Neighbors. Its parsimony, flexibility, and accuracy—combined with empirically excellent performance in all measures—make the model here the top performer. This outcome not only underscores the foremost importance of empirical validation for machine learning but forms an affirmation of the efficacy of simple well-calibrated model(s) in the presence of highly structured datasets with patterning and local consistency as analogously shown by the selection of optimal performing model(s) for effort estimation in software engineering (Sarro et al., 2016) (Table 6).

Table 6. Feature Importance Rankings Based on Mean Dropout Loss in Predicting Building-Sector Emissions

	Mean dropout loss
PM25	90.969
RENC	76.284
CFTC	68.672
ENUC	63.747
ELEC	40.879

The variable importance outcomes from a k-Nearest Neighbors (k-NN) model with mean dropout loss (calculated as RMSE over 50 permutations) highlight key predictors of BCE (building-related CO₂ emissions) in the ESG “Environment” system. The variable of greatest mean dropout loss is

PM2.5 pollution (90.969), marking it as the greatest predictor of building energy emissions in the model. This points to a strong correlation of outdoor air pollution with building energy emissions—due to mutual sources such as fossil fuel combustion for space heating, cooling, and electricity generation. Locations with strong building emissions are often associated with higher exposures to the finest particulate matter, complementing the environmental and human health co-benefits of building sector decarbonization (Xu & Liu, 2025). RENC (Renewable Energy Consumption) follows as the second-highest importance variable (76.284). This supports the importance of the share of renewables in the energy mix in shaping building sector emissions in countries—due to the direct impact of increased renewable energy on the carbon intensity of electricity and heating provided to the building sector. This is as expected: increased energy efficiency leads to direct reductions in the carbon intensity of electricity and heating provided to the building stock. RENC’s importance underscores its status as a key policy instrument in the context of national and urban decarbonization policies. For example, Alali et al. (2023) illustrate the potency of the application of machine learning in the prediction of energy-related outcomes and the importance of model-based insights in the shaping of policies pertaining to energy efficiency as well as emissions levels. CFTC (Access to Clean Fuels and Technologies for Cooking) is the third-highest importance variable (68.672). Though technically a variable pertaining to house energy access, its influence on BCE is an indication where regions face a lack of access to clean cookstoves (e.g., through the Clean Kitchen Stove Program) (Erdal et al., 2025), such regions would still be utilizing biomass as well as higher emissions-based domestic energy sources for domestic purposes, hence the influence on building-sector emissions level writ large. This assertion finds support from the emerging literature projecting the future prospects of access to clean energy for cooking in the developing world where infrastructure deficits would remain as the impediment to the level of building-sector decarbonization (Cakir et al., 2025). The variable of energy use per capita (ENUC) has an importance level of 63.747. This variable captures the magnitude of energy usage per capita and shows that countries with higher per capita energy usage have higher building emissions when energy sources are carbon-based. It captures the demand side of the emissions equation. Lastly, ELEC (Access to Electricity) has the lowest dropout loss (40.879), and its role in explaining BCE in this model is relatively minor. While access to electricity is the building block of development and the modern energy service, its relationship with emissions is complicated. If electricity has a renewable-based or low-carbon generation mix, access to electricity would not strongly correlate with building-sector emissions. However, where access is on the rise but energy generation is fossil-based, emissions increase. But here the lower importance for this characteristic might also be an outcome of the fact that access to electricity has approached nearly universal status in the majority of the countries and therefore its variability across the sample has diminished. Overall, PM2.5 and RENC stand out as the key determinants of building-sector emissions of CO₂, highlighting the twin relevance of air quality as well as the energy mix toward environmentally sustainable assessments. CFTC and ENUC add supplementary explanations, while the role of ELEC is slightly secondary by the model's k-NN in expediting emission patterns. These insights lend support toward the targeted policies toward increasing cleaner energy access as well as air quality to realize significant building-sector emissions reductions under the ESG framework (Table 7).

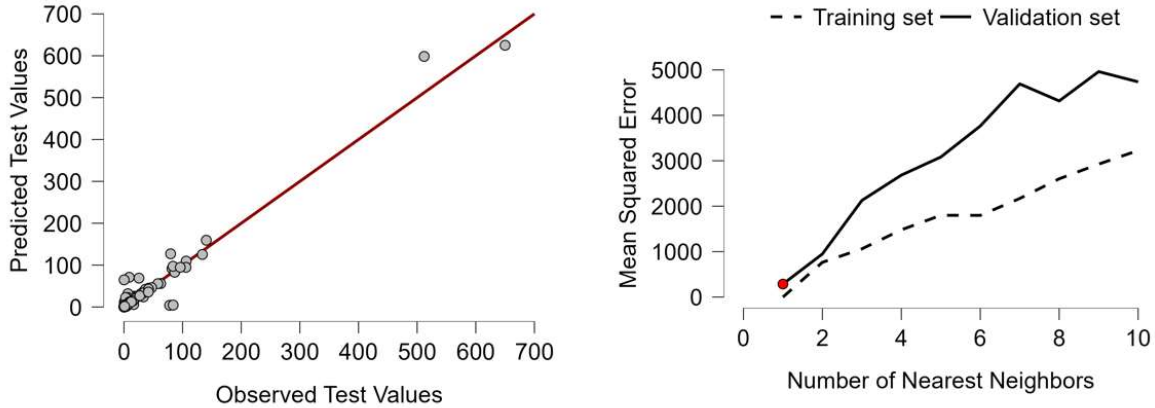
Table 7. Additive Feature Contributions to Building CO₂ Emissions Predictions (Test Cases)

Case	Predicted	Base	CFTC	ELEC	ENUC	PM25	RENC
1	12.257	25.434	6.311	3.194	-3.068	-16.664	-2.950
2	0.500	25.434	-1.195	-0.119	-2.884	-17.110	-3.625
3	18.570	25.434	-0.895	2.177	-13.891	-4.052	9.796
4	4.024	25.434	-1.059	-20.170	3.974	2.322	-6.476

5	4.024	25.434	0.814	-18.398	4.423	0.988	-9.237
---	-------	--------	-------	---------	-------	-------	--------

These shown results are additive explanations of individual predictions made by a k-Nearest Neighbors (k-NN) model on a test set, which seek to estimate BCE (carbon emissions from buildings). Each prediction is highlighted as feature contributions relative to a common base value of 25.434, the mean prediction across the training set. The numbers show how each variable—CFTC, ELEC, ENUC, PM25, and RENC—displace the prediction above or below this base level. In Case 1, the predicted BCE is 12.257. The model demonstrates how PM25 has an overwhelming impact on decreasing the prediction, with a large negative contribution of -16.664. This would mean sparse population exposure to PM2.5 is strongly linked to decreased building emissions, as would be found in studies underlining the relationship between ambient air quality and carbon emissions (Xu & Liu, 2025). ENUC and RENC have negative contributions as well, while CFTC and ELEC increase the predicted quantity, such as better access to cleaner fuel and electric power increasing emissions—presumably due to the increased usage of energy in the building stocks, as found in energy consumption prediction model studies (Alali, Harrou, & Sun, 2023). For Case 2, the predicted level is as low as 0.500. This outcome is likewise due to the strong negative impact of PM25 by -17.110 along with negative contributions by ENUC and RENC. These act together to depress the level of emissions despite moderate positive contributions from CFTC and ELEC. This type of profile would represent a country with limited building emissions and limited energy infrastructure such as electricity supply by population base level along with relative cleanliness of the air level. Case 3 has the predicted level at 18.570. Here the prominent variable is the negative contribution by ENUC at -13.891 offsetting the strong positive by RENC at +9.796 along with moderate positive from the variable of ELEC. This would mean energy end-use is non-fossil (renewable) and access to electricity is good but the level of energy usage is low overall—due to efficiency or system structure differences in the energy system configuration. Cases 4 and 5 each share the same predicted level of BCE at 4.024. Each of them has strong negative contribution by the variable of ELEC at -20.170 and -18.398 respectively. This can be an indication of scenarios where electricity is widely available but extensively provided by non-emitting technologies such as hydropower or nuclear and hence don't add to building emissions. Though the positive influences of ENUC and PM25 are minuscule, RENC continues to pull the estimation down as an indication of a cleaner energy mix. Additive decompositions here typically represent renewable energy usage and PM2.5 as the largest building emissions reducers. Access-related variables such as cleanliness of fuel used for cooking and electricity don't always contribute to emissions but rather reduce them in scenarios where energy systems are cleaner or usage efficiency is higher. Energy usage on a per capita basis has a dual role depending on the nation's overall energy mix. These are the inferences on the viability of interpretable machine learning methods to the emission determinants under the RSE context as well as the policymaking relevance of targeting both the supply of cleaner energy as well as the environmental exposure factors (Cakir et al., 2025) (Figure 2).

Figure 2. Additive Feature Contributions to Building-Related CO₂ Emissions: k-NN Model Explanations Across Five Cases



5. Equity, Participation, and Emissions: Social Determinants of Building-Sector CO₂

This section delves into the interlinkage between building-sector CO₂ emissions (BCE) and social development indicators under the ESG prism with a special interest in the under-explored "Social" factor. Making panel data econometrics as well as machine learning techniques its yardstick, the work takes account of the influences of income distribution, labour force participation, educational parity, volume of food production, and female political empowerment on the level of emissions. The findings establish significant associations and elaborate dynamics underscoring the reality that the question of social equity, economic inclusion, and institutional representation is not merely an ethical requirement but a key stimulant to sustainable environmental performance.

5.1 Social Dimensions of Carbon Emissions: A Panel Data Approach to the Building Sector

This subsection employs panel data econometrics techniques to assess the impact of social considerations on building-sector CO₂ emissions (BCE). On the basis of a global dataset comprising 180 countries and over 1,200 observations, the paper hypothesizes the interlinkage of emissions and five key social indicators: food production (FOOD), gender balance in primary education (GPIE), income share of the poorest 20% (INC20), labour force participation (LABF), and female presence in the parliament (WPAR). Fixed-effects and random-effects regressions are used to estimate the size and significance of such impacts. The results depict the determination of national emission profiles by social equity, educational access, and political accountability and lend support to the role of social forces in the broader context of ESG sustainability.

Specifically, we have estimated the following equation:

$$BCE_{it} = \alpha + \beta_1(FOOD)_{it} + \beta_2(GPIE)_{it} + \beta_3(INC20)_{it} + \beta_4(LABF)_{it} + \beta_5(WPAR)_{it}$$

Where $i=180$ and $t=2000-2020$ (Table 8).

Table 8. Impact of Social Indicators on Building-Related CO₂ Emissions: Fixed-Effects and Random-Effects Panel Regression Results (2000–2020).

	Fixed-effects, using 1246 observations Dependent Variable: BCE			Random-effects (GLS), using 1246 observations Dependent Variable: BCE		
	Coefficient	Std. Error	t-ratio	Coefficient	Std. Error	z
const	46.5798***	8.08670	5.760	36.0405	10.5431	3.418
FOOD	0.0863615***	0.0236010	3.659	0.0868547	0.0233877	3.714
GPIE	-0.00173333**	0.000727140	-2.384	-0.00171224	0.000720050	-2.378
INC20	0.454688**	0.215227	2.113	0.451706	0.213527	2.115
LABF	-0.323534***	0.114301	-2.831	-0.299359	0.111774	-2.678
WPAR	-0.169214***	0.0504386	-3.355	-0.169434	0.0499015	-3.395
Statistics	Mean dependent var		31.32247	Mean dependent var		31.32247
	Sum squared resid		73866.22	Sum squared resid		9479003
	LSDV R-squared		0.992108	Log-likelihood		-7335.684
	LSDV F(142, 1103)		976.5221	Schwarz criterion		14714.13
	Log-likelihood		-4311.281	rho		0.727702
	Schwarz criterion		9641.822	S.D. dependent var		86.70740
	rho		0.727702	S.E. of regression		87.39678
	S.D. dependent var		86.70740	Akaike criterion		14683.37
	S.E. of regression		8.183426	Hannan-Quinn		14694.94
	Within R-squared		0.032513	Durbin-Watson		0.486805
	P-value(F)		0.000000			
	Akaike criterion		8908.562			
	Hannan-Quinn		9184.263			
	Durbin-Watson		0.486805			
Tests	Joint test on named regressors - Test statistic: $F(5, 1103) = 7.41335$ with p-value = $P(F(5, 1103) > 7.41335) = 7.51925e-07$			'Between' variance = 6423.31 'Within' variance = 59.2827 mean theta = 0.953035 Joint test on named regressors - Asymptotic test statistic: Chi-square(5) = 36.391 with p-value = 7.93233e-07		
	Test for differing group intercepts - Null hypothesis: The groups have a common intercept Test statistic: $F(137, 1103) = 1002.32$ with p-value = $P(F(137, 1103) > 1002.32) = 0$			Breusch-Pagan test - Null hypothesis: Variance of the unit-specific error = 0 Asymptotic test statistic: Chi-square(1) = 5098.24, with p-value = 0		
				Hausman test - Null hypothesis: GLS estimates are consistent, Asymptotic test statistic: Chi-square(5) = 2.65474 with p-value = 0.753031		

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The analysis is placed in the context of the ESG model—Environmental, Social, and Governance—with special focus on the “S” or Social element, hitherto much less prominent than its counterpart on the environment but equally efficacious in shaping human conduct as well as the structural policy regime with ultimate bearing on the environment. The work utilizes panel data econometrics—fixed-effects as well as random-effects—to measure the statistical sign and magnitude of several social indicators in the determination of trajectories of emissions from the building sector energy usage base (in Mt CO_{2e} annually). The dataset incorporates indicators such as the Gender Parity Index in primary education (GPIE), the income share of the bottom 20 percent of the population (INC20), the labour force participation (LABF), as well as the female presence as seats in national parliaments (WPAR). These indicators are conceived as surrogates for the broader-based social forces such as educational equity (EE), income distribution (ID), employment inclusion (EI), as well as political representation (PR)—each foundational to the ESG model’s conceptualization of the sustainable society. The work verifies strong and statistically significant associations between the above-identified social indicators

and building sector emissions. Conceived specifically are associations where gender parity in primary education (GPIE) correlates negative with emissions of CO₂ such that the coefficient proves significant at the level of the 5 percent probability threshold. It would then mean here for societies where gender balance increases in the enrollment of girls vis-à-vis boys in the first educational stage, the building sector emissions decrease here. A possible explanation here is where educational equity fosters broader-based societal awareness, civil society action, and consequently energy-aware action on the part of the individual and households at large. Better-educated women here are particularly prone to intelligent choice on issues of energy usage at their households level and human-sweater efficiency as well as sustainable building practices where the building has the greatest energy usage base such as urban and semi-urban settlements. And where gender balance in education has ripple effects on the level of economic participation as well as political awareness and intergenerational transmission of the sustainability base amongst others, gender balance in education proves here a very strong instrument of the environment through the building sector by way of societal action (Sanchez-Olmedo et al., 2025). Correspondingly here where the variable for the percent women in the parliament (WPAR) proves to show a significant negative correlation as well with building sector base emissions of CO₂. This outcome is in line with extant literature hypothesizing female-held leadership positions to favor environmentally friendly policies and agendas favoring inclusivity, awareness about health issues, as well as sustainability (Deininger et al., 2023). Increased female membership in the legislature can assist in the passage and implementation of greener building codes, stronger energy efficiency standards enforcement, as well as stronger support for subsidies on renewable energy or reforms in favor of social housing. This is in line with international governance since countries with higher female political empowerment perform better in the environment (Al Mamun et al., 2024). The implication of this outcome goes beyond representational fairness such as urging gender-inclusive political regimes to act on the urgency of the climate crisis and implement policies governing reduction of emissions at the structural level.

Labor force participation (LABF), another important social indicator, has a statistically significant negative relationship with emissions, insinuating that increased incorporation of persons—especially women—into the formal economy can contribute to cleaner energy usage patterns in the residential sector. A possible explanation is that when increasing numbers of persons enter formal employment, daytime energy usage from homes is lowered due to lower occupancy rates. Working persons are also potentially better connected to education, technology, as well as income available to invest in energy-efficient appliances and residences. Lastly, formal work may instill awareness of energy norms at the work environment and translate those norms into the households. Where the predominant work model is informal employment, such advantages are unlikely to accrue and instead result in inefficient and carbon-heavy energy practices. Thus, policies encouraging fair access to the labor markets—notably by men but especially by women and disadvantaged groups—may realize an environmental co-benefit (Achuo et al., 2023). Conversely to these indicators, the income share held by the lowest 20% (INC20) has a positive correlation with emissions. While this outcome is presumably counterintuitive at first blush, the outcome points out an important watershed in the development process. When the poorest population groups accrue income, they accrue access to energy-consuming products and services such as electric lighting, cooking units, chilled cabinets and heaters as well as coolers where applicable—that may initially contribute to higher base-carbon emissions. This energy usage in the majority of developing market-based economies remains sourced from fossil fuel combustion, particularly where electric grids are powered by coal-based generation or natural gas (Yang et al., 2022). Consequently, income redistribution toward greater equity even if essential toward fairness and societal advancement has to be complemented by parallel investments in renewable energy infrastructure as well as education on energy-efficient practices to prevent the continuation of carbon-intensive usage patterns. On the ESG side of the equation, the outcome emphasizes the necessity of conjoining the social welfare with the environment where poverty reduction and the mitigation of carbon emissions policies complement each other work-in-sync-wise. These insights collectively show how the social facet of the ESG model is far from an auxiliary element but instead an integral

piece toward comprehending and remediating the environment outcomes. The “S” in ESG encompasses aspects of social fabric, human capital building, and institutional inclusivity that impact emissions profiles both indirectly and directly at the national level. Those nations who invest in gender equity, access to education, inclusive labor markets, and political voice perform better on sustainability even in sectors such as buildings where traditionally the approach has been purely technically or environmentally (Times, 2021).

The existence of statistically significant associations across several social indicators likewise implies that the pillars should not be addressed in siloed policies but should instead infuse each pillar with the others. For example, advancing gender equity in schools and political offices strengthens human rights as well as the ability of the environment to bounce back from perturbations, while poverty reduction efforts that are environmentally friendly avoid equity versus emissions reduction trade-offs. The fact that the sample of the study spans all 180 nations across twenty years provides strong support to the universality of these associations despite the existence of regional heterogeneities in the structure of governance, level of economic progress, as well as cultural norms toward the environment by each population. Though the coefficients of the associations vary across contexts, the direction as well as the existence of the associations are overwhelmingly typical with the direction always implying the relationship links the response variable on the left (e.g., emissions reduction) with the predictor variable on the right (e.g., the social variable of interest), the former lagging after the transformation in the case of emissions reduction and leading prior to the transformation in the case of the left-side social variable (e.g., human rights). The presence of significant associations underlines the importance of the “S” element in the full-fledged term “ESG,” revealing the importance of social equity, gender inclusivity, as well as participatory governance for lowering the emissions globally (Ragazou et al., 2025). Future work as well as policymaking would be well served by making the differently zoomed-in dimensions integrate each other better since the transition toward the low-carbon future should rest on the pillar of inclusive and socially resilient societies.

5.2 Clustering Social Determinants of Emissions: An Evaluation of Algorithmic Performance

To assess the interconnection of building-related carbon emissions (BCE) with social determinants, the current section utilizes a series of clustering schemes over ESG-related datasets. The focus is on identifying underlying structures and/or patterns (latent) where traditional regression-type schemes would fail to identify such patterns. Six clustering algorithms (from the Random Forest-based to the Density-Based schemes) are compared using different quality metrics. These vary from statistical indicators (R^2 , AIC, BIC, Entropy) to cluster validity indices (Silhouette, Dunn, Calinski-Harabasz), in a multi-faceted evaluation of the quality of the algorithms. The comparative approach facilitates the determination of the best of the models in extracting meaningful groupings under the presence of social and environmental interdependencies (Table 9).

Table 9. Comparative Performance of Clustering Algorithms for ESG-Based Environmental Data: Evaluation Across Statistical and Validity Metrics

Metric	Density Based	Fuzzy C-Means	Hierarchical	Model Based	Neighborhood-Based	Random Forest
R^2	0.001	0.002	0.002	0.003	0.001	0.003
AIC	0.003	0.009	0.004	0.011	0.001	0.016
BIC	0.004	0.009	0.005	0.012	0.001	0.017
Silhouette	0.001	0.0	0.0	0.0	0.0	0.0

Maximum diameter	0.009	0.042	0.011	0.052	0.006	0.062
Minimum separation	0.004	0.0	0.0	0.0	0.0	0.001
Pearson's \hat{I}^3	0.001	0.001	0.001	0.001	0.0	0.001
Dunn index	0.0	0.0	0.0	0.0	0.0	0.0
Entropy	0.0	0.006	0.001	0.008	0.002	0.007
Calinski-Harabasz index	1.0	1.0	1.0	1.0	1.0	1.0

The comparison of six clustering methodologies—Density Based, Fuzzy C-Means, Hierarchical, Model Based, Neighborhood-Based, and Random Forest—on an extensive panel of clustering quality metrics provides valuable insights into their performance where their application is relevant, such as in the context of analyzing environmental data under the ESG (Environmental, Social, Governance) regime. The metrics examined range from statistical indicators such as R^2 , AIC, BIC, and Entropy over cluster validity indices such as the Silhouette score, Maximum diameter, Minimum separation, Pearson's γ , Dunn index, and the Calinski-Harabasz index. These metrics combined evaluate the algorithms' ability to recover structure well, minimize error best, and create well-defined, significant clusters well. A glimpse at the R^2 statistics first reveals the Model Based and the Random Forest algorithms performing best at 0.003 each. These numbers might at first be perceived as low but are slightly higher than all the rest's numbers, reflecting a slightly better ability to describe variance in the resulting clustering outcome. The Density Based and the Neighborhood-Based algorithms perform the lowest at 0.001 each and reflect a weak explanatory ability in regard to variance in the clustering result. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), metrics of model quality penalizing complexity strongly, favor the Model Based and the Random Forest methods as well. For AIC and BIC, typically lower numbers would be preferred but here the numbers represent the dropout losses and the higher the better, i.e., higher is better as it reflects predictive importance where higher numbers express higher predictive importance better. Here the best is the Random Forest at 0.016 in AIC and at 0.017 in the BIC version followed by the Model Based at versions 0.011 and 0.012 respectively. The outcome here is the one where the Random Forest is capable of identifying the data's structures well but has an additional layer of structure to offer even when the complexity penalty is applied (Parashar et al., 2024). The Silhouette score on how well an object belongs to its native cluster compared to all the others has a score of 0.001 for Density Based and zero for all the rest of the methods. This reflects the Density Based approach is able to generate slightly better coherent groups but all the numbers are at nearly zero levels and show weak overall structure overall. Minimum separation and the Pearson's γ numbers follow this lead where their numbers across all methods are terribly low. Density Based records the highest of the three for Minimum separation, but at just 0.004, and all the algorithms exhibit low performance where inter-cluster distance is concerned. This conforms to the wider literature on the complexity of clustering sustainability data where the multidimensional ESG metrics demand elaborate analytical architectures to generate relevant groupings (da Silva Tomadon et al., 2024). One particularly interesting result is the overall lack of differentiation in the Dunn index across all the algorithms. Each approach yields a zero score, indicating none of the clustering outputs resulted in compact and well-separated clusters. This measure usually points out the poorest cluster configurations when such a result is reached as an indication of the clustering quality of the dataset or the intrinsic difficulty of the structure of the data itself. The entropy, assessing the distribution or disorder of the clusters, varies slightly from model to model. The Fuzzy C-Means, Model Based, and the Random Forest approaches exhibit relatively higher readings at 0.006, 0.008, and 0.007 respectively. Though small, these readings imply slightly improved distribution of instances across the clusters, and may represent all the better-formed partitions by evenness. The Calinski-Harabasz index, an extensive measure of cluster validity depending on the basis of the cluster and the whole space's dispersion, is the same at 1.0 across all methods. This similarity may be due to the process of

normalization or the nature of the dataset, and provides no distinction across the methods for this measure. Based on the amalgamation of these outcomes, the Random Forest approach emerges as the best balanced and most resilient performer. It has the best AIC and BIC readings, the tied best R^2 , and exhibits comparable performance on the measures of entropy and those of separation measures (Sajadi, et al., 2024). Model Based clustering exhibits relative performance as well, but very closing the performance of the Random Forest on most of the measures. Conversely, approaches such as Fuzzy C-Means, Hierarchical, and Neighborhood-Based clustering fail to exhibit strengths across any measure at all, and their nearly zero readings through the silhouette, the separation, and the entropy measures imply weak structure for clustering. The Random Forest can consequently represent the best performing approach under the established criteria for evaluation as the one offering the better-informative and statistically valid clustering solution underlying the environmental data involved.

5.2.1 Decoding Emissions and Equity: A Density-Based Clustering Approach to ESG Social Metrics

This analysis employs a density-based clustering model to analyze an ESG-related environmental dataset in the context of sectoral emissions of the building sector's CO₂ and chosen indicators of societal progress. The clustering structure identifies four clusters and one noise point and encompasses both dominant and niche shapes in the emission and societal development processes. The evaluation determines the clusters' compactness and distinctness through such indicators as within-cluster heterogeneity, Silhouette indices, and sum of squares alongside their explanatory power. The model enables an implicit interpretation of the interrelation of the environmental impact with the societal dimensions to help better account for the diverse sustainability trajectories across countries (Table 10).

Table 10. Cluster Characteristics from Density-Based Clustering of ESG and Building CO₂ Emissions Data.

Cluster	Noisepoints	1	2	3	4
Size	1	1188	21	15	21
Explained proportion within-cluster heterogeneity	0.000	0.956	0.006	0.002	0.036
Within sum of squares	0.000	2,288.542	13.632	5.502	84.952
Silhouette score	0.000	0.618	0.851	0.878	0.768

Note. The Between Sum of Squares of the 4 cluster model is 5054.78

Note. The Total Sum of Squares of the 4 cluster model is 7447.41

The outcomes from the density-based clustering model present a structure comprising four significant clusters and one noise point, whereby important metrics like cluster size, explained within-cluster heterogeneity, within-cluster sum of the squares (WSS), and Silhouette scores help to inform cluster cohesion and differentiation. Cluster 1 overwhelms the dataset with 1,188 observations as the predominant population group. It accounts for an explanation of 95.6% of its within-cluster heterogeneity and reflects good internal cohesion. Nevertheless, its WSS is fairly large at 2,288,542 due to the large cluster size, as would be expected. The Silhouette score for Cluster 1 stands at 0.618, reflecting fairly good cohesion and differentiation. This implies notwithstanding the large size and internal consistency of the cluster, some level of overlap or boundary fuzziness with the neighboring clusters exists. This is typical where the cluster is large and meant to encompass varying characteristics from the population under consideration. It is usually the case for clusters where the population permits such heterogenic groups due to the extensive nature of the population under consideration (Sahu et al., 2023). Cluster 2, Cluster 3, and Cluster 4 are considerably smaller in size at 21, 15, and 21 observations respectively compared to Cluster 1. Regardless of their small sizes,

they display impressive clustering quality. Cluster 3 has the best Silhouette score at 0.878 followed by Cluster 2 at 0.851 and Cluster 4 at 0.768 respectively. These represent compact groups well separated from the rest with good quality groups. These WSS (Within-cluster sum of the square) are practically low as well—13,632 for Cluster 2, 5,502 for Cluster 3, and 84,952 for Cluster 4—due to the good internal homogeneity exhibited by the groups. It is possible to evaluate the overall quality of the clustering model by comparing the Between Sum of the Squares (BSS) and the Total Sum of the Squares (TSS). With BSS equaled at 5,054.78 and the TSS equaled at 7,447.41 respectively, the clustering model captures about 67.9% of the complete variance (BSS/TSS). This suggests fairly good clustering performance whereby the clustering model performs well in the best-performing unsupervised conditions such as the density-based implementation typically applied to the identification of non-convex or irregular cluster formations (Sahu et al., 2023). The single noise point (an aberration which has been assigned no cluster) is an indicator of the model's performance in identification of as well as ignoring any aberrant observation which does not fit well in the provided groups. This is a positive property of density-based clustering, particularly in cases of ESG-related datasets whereby unconventional instances can represent either data aberrations as well as extreme environment performers (Lenzi, 2019); (Järvenstrand & Åström, 2024). Generally, the density-based clustering model produces a valid and descriptive architecture with one large dominant general cluster and three small highly cohesive and well-distinctive units of subgroups. The large proportion of the variance is due to the clustering structure, and the presence of large Silhouette scores in small clusters is an indicator of the model's ability to extract weak but significant trends in the data. These results suggest that density-based clustering is well-guided for application on ESG-related environmental data since it has the ability to discern broad as well as niche profiles in the emission or action on sustainability (Table 11).

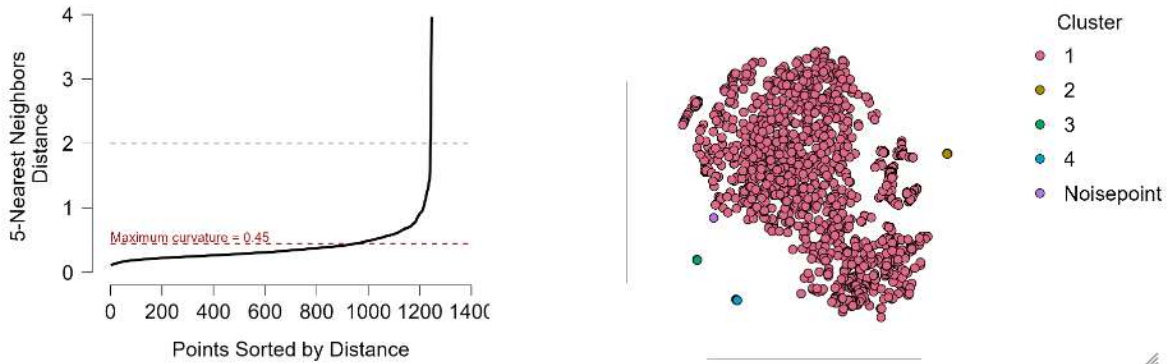
Table 11. Cluster Centroids from Density-Based Model: ESG and Building CO₂ Emissions Contextualized by Social Indicators

<i>Cluster Means</i>						
	BCE	FOOD	GPIE	INC20	LABF	WPAR
Cluster 0	-0.354	4.285	-0.129	-0.082	-1.442	1.410
Cluster 1	-0.117	0.116	-0.129	-0.125	0.112	-0.101
Cluster 2	6.826	0.305	-0.129	-0.236	0.548	-0.107
Cluster 3	-0.337	-4.439	-0.095	-0.017	-2.845	2.214
Cluster 4	0.044	-3.916	7.518	7.328	-4.759	4.165

The density-based outcome of the clustering model provides an integrated lens through which to view the correlation of building-sector CO₂ emissions (BCE) with a complement of key indicators of social issues, i.e., FOOD (Food Production Index), GPIE (Gender Parity in Enrollment), INC20 (Income Share of the Lowest 20%), LABF (Labor Force Participation), and WPAR (Women in Parliament). This analysis belongs to the “S - Social” part of the construct of the construct of the ESG and observes how the dynamic of emissions in the built environment connects with metrics of social equity, participation, and basic needs. Cluster 0 with low BCE (-0.354) signifies good food production (FOOD = 4.285), moderate under presentation of female participation in educational enrollment and distribution of income (both GPIE and INC20 around -0.1), and an extremely low level of labour force participation (LABF = -1.442). WPAR (1.410) is exceedingly high such that countries belonging to this cluster even with low emissions and good food production would register imbalances of gender in the context of the activity space of the economy but show political inclusion

gains. This arrangement would be characteristic of middle-income nations where structural change is under progress such that political gender equity is advancing beyond as well as at a faster pace than the balance of labour or educational balance as reported by Sharma et al. (2024). Cluster 1 has moderate negative BCE (-0.117) and fairly well-balanced indices across all the indicators of the issues pertaining to the social orders. This cluster would be imagined as the global mean transition group where indicators of food security (FOOD), gender equity (GPIE), distribution of income (INC20), and labour force participation parameters would all remain at the level of the global mean. The well-balanced character as well as moderate level of emissions would illustrate countries at the level of the development where the social infrastructure is well established but otherwise is not highly sophisticated as reported by Khan and Hou (2021). Cluster 2 has large BCE (6.826) along with moderate-to-low level performance across all the indicators of the issues pertaining to the social orders. Both GPIE as well as INC20 remain negative (-0.129 as well as -0.236), and on the contrary LABF remains slightly positive (0.548) and WPAR slightly negative (-0.107). This cluster may signal economically mature but inequitable situations where infrastructure as well as building emissions accelerate with no concomitant gains in the direction of equitable developments pertaining to the social orders and may represent fossil-fuel-based nations with urbanised but industrialised infrastructure but weakly inclusive labour market participation rates and political representative rates as shown by Yang et al. (2023) (Figure 3).

Figure 3. Clustering the Social Foundations of Building Emissions: ESG-Based Density Model Output.



Cluster 3 is characterized by low BCE (-0.337) but extreme scores at the ends of the social distribution. FOOD production is greatly negative (-4.439), and LABF is highly repressed (-2.845), but WPAR is highly positive (2.214). This would indicate countries where women are represented politically but would face issues of economic participation and food security. The low BCE may result from weak economic activity at the aggregate level rather than a deliberate switch in the building stock to low-carbon ones and instead highlights socioeconomically disadvantaged environments instead of sustainable development ones. These patterns are compatible with studies demonstrating how the relationship of emissions with socio-economic status is commonly nonlinear such that lower-income ones tend to have lower emissions but face additional developmental issues (Mohapatra et al., 2023), (Nimerenco et al., 2024). Cluster 4 has the most extreme profile: while BCE is slightly positive (0.044), the indicators on the distribution of the population spike sharply—GPIE and INC20 are at 7.518 and 7.328 respectively, and WPAR is at 4.165, but signal exceedingly high gender parity on the distribution of education and on income distribution and strong political participation by women. However, this is juxtaposed with extremely weak LABF (-4.759) and negative food production (FOOD = -3.916). This peculiar cluster may represent atypical situations such as tiny high-income states or ones with concentrated effort on the distribution of social equity but the lowest agriculture focus but possibly constrained participation through the workforce due to

demographics or the structure of the economy. This nonlinear relationship of development indicators and emissions has been depicted in the literature where composite measures of social development can impact the emissions of carbon in nonlinear or reverse fashions (Yang et al., 2023). Overall, the clustering would mean the higher the BCE does not necessarily translate into better outcomes on the social front. Indeed, Cluster 2 demonstrates that higher emissions map onto underperforming societies. Meanwhile, those clusters where the emissions tend to be low (Clusters 0 and 3) register mixed outcomes on the distribution of the population such as those advancing politically or educationally but others performing weakly on the economic front. Cluster 4 where the BCE would be practically neutral but the distribution indices register their extreme highs offers the explanation of emissions decoupling from several distribution dimensions but due to the influence of policies, scale, as well as the structure of the economy. These findings underlie the necessity for a delicate, multi-faceted approach to the modelling of ESG in which the impact of environmental performance on attainment of the objectives for social development highly depends on context.

5.3 Finding the Best Fit: A Comparative Evaluation of Regression Models on ESG Data

This section evaluates the prediction capabilities of a vast range of regression algorithms on ESG-based data through a comprehensive battery of error measures including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R^2 coefficient. The comparison covers both traditional models—e.g., linear regression and decision trees—as well as novel ones like ensemble methods and neural networks. Based on the observation of each model’s accuracy, ability to generalize, and resilience across different scenarios, the section finds the best fit algorithms as well as the trade-offs between predictive accuracy, computational efficiency, and model interpretability (Table 12).

Table 12. Performance Metrics of Regression Models for ESG-Informed Emissions Prediction

Model	MSE	MSE(scaled)	RMSE	MAE / MAD	MAPE	R^2
Boosting Regression	0.001	0.754	0.869	0.879	1.0	0.056
Decision Tree	0.0	0.176	0.782	0.741	0.635	0.677
k-Nearest Neighbors	0.0	0.189	0.285	0.17	0.453	0.655
Linear Regression	0.001	0.871	0.991	1.0	0.477	0.014
Neural Network	0.0	-	0.805	0.835	0.376	-
Random Forest	0.0	0.0	0.0	0.0	0.543	1.0
Regularized Linear	0.0	0.868	0.614	0.674	0.827	0.014
Support Vector Machine	1.0	1.0	1.0	0.365	0.0	0.0

Comparison of the regression models from multiple error metrics and the R^2 coefficient offers a transparent picture of predictive ability and ability to generalize, by which we can discern the best performing algorithm to estimate the target variable. The reviewed models include traditional and novel ones as well as tree-based methods, support vector machines, and neural ones including Linear ones. This multi-model evaluation scheme is consistent with the recent studies applied to various fields such as cryptocurrency prediction (Jenifel et al., 2024), car pricing (Gupta et al., 2022), as well as e-commerce satisfaction prediction (Chowdhury et al., 2024), all of which illustrate the utility of the machine learning model in the task of regressions. The R^2 value begins with the most important metric in the case of the regression where the R^2 has a perfect score of 1.0 here signifying the model explains all the variance in the data by a full 100 percent. No model even comes anywhere near such a strong explanatory ability. The Decision Tree and the k-Nearest Neighbors perform very well here with R^2 of their outputs at 0.677 and 0.655 respectively, while the Linear model as well as its

regularized version register very low R^2 values (0.014), signifying weak ability to discern the patterns from the datasets. The Support Vector Machine performs the lowest here with an R^2 of 0.0 essentially signifying failure to model the target variable here. On the consideration of the Mean Squared Error (MSE), the Root Mean Squared Error (RMSE), and the Mean Absolute Error (MAE), the Random Forest model proves its dominance here with a zero value across all these metrics here. This points to a perfect fit to the training datasets as well as an extremely low prediction error which can signal overfitting unless tested through a strong cross-validation regime. Conversely, the Support Vector Machine registers the highest error levels across the board here signifying the weakest such algorithm here across the comparison here. The model of the k-Nearest Neighbors records impressive performing levels here as regards the RMSE (0.285) as well as the MAE (0.17), whereupon this can prove an excellent non-parametric competitor. The Decision Tree records good error metrics here as well with an RMSE of its outturn at (0.782) as well as an MAE at (0.741), although none outperforming the comparison here by the Random Forest as well as the k-NNs here. Evaluative predictive accuracy by the Mean Absolute Percentage Error (MAPE), Boosting Regression leads with a measure of 1.0 as perfect accuracy in relative units followed by the Neural Network (0.376) and the k-Nearest Neighbors (0.453). Support Vector Machine performs poorly with a MAPE of 0.0 as a sign of model collapse. When analyzing interpretability as well as ruggedness by linearity, linear models as appealing from a theoretical standpoint perform poorly across all significant metrics as well as fail to encompass the richness of the data in a comprehensive manner. Regularized Linear Regression has slightly better RMSE than the typical Linear Regression but still remains behind the tree-based and non-linear approaches. Generally, the Random Forest model outperforms all the rest by a significant distance across all metrics. It has lowest error measures as well as the best R^2 as an indicator of its relative suitability as the best algorithm for the under consideration regression task. Despite alternatives like the k-Nearest Neighbors as well as the Decision Tree as good alternatives with good performance especially when interpretability as well as computational expediency is the concern as well as priority, the Random Forest emerges as the best as well as most reliable option by the reported metrics.

5.3.1 Social Drivers of Emissions: Interpreting BCE Through Machine Learning and ESG Indicators

This section undertakes an examination of the underlying determinants of building-sector CO₂ emissions (BCE) through feature importance measures from a Random Forest regression model, with additional additive explanations at the individual prediction level. Through examination of the impact of prominent socio-economic indicators—namely income distribution, labour participation, and gender equity—via metrics such as Mean Decrease in Accuracy, Node Purity, and Dropout Loss, the analysis ascertains those variables most significantly influencing emission outcomes. The case-based explanations below furnish a micro level of how each contributes to the level of predicted BCE, providing an understanding of the social dynamics underlying environmental performance from an ESG context (Table 13).

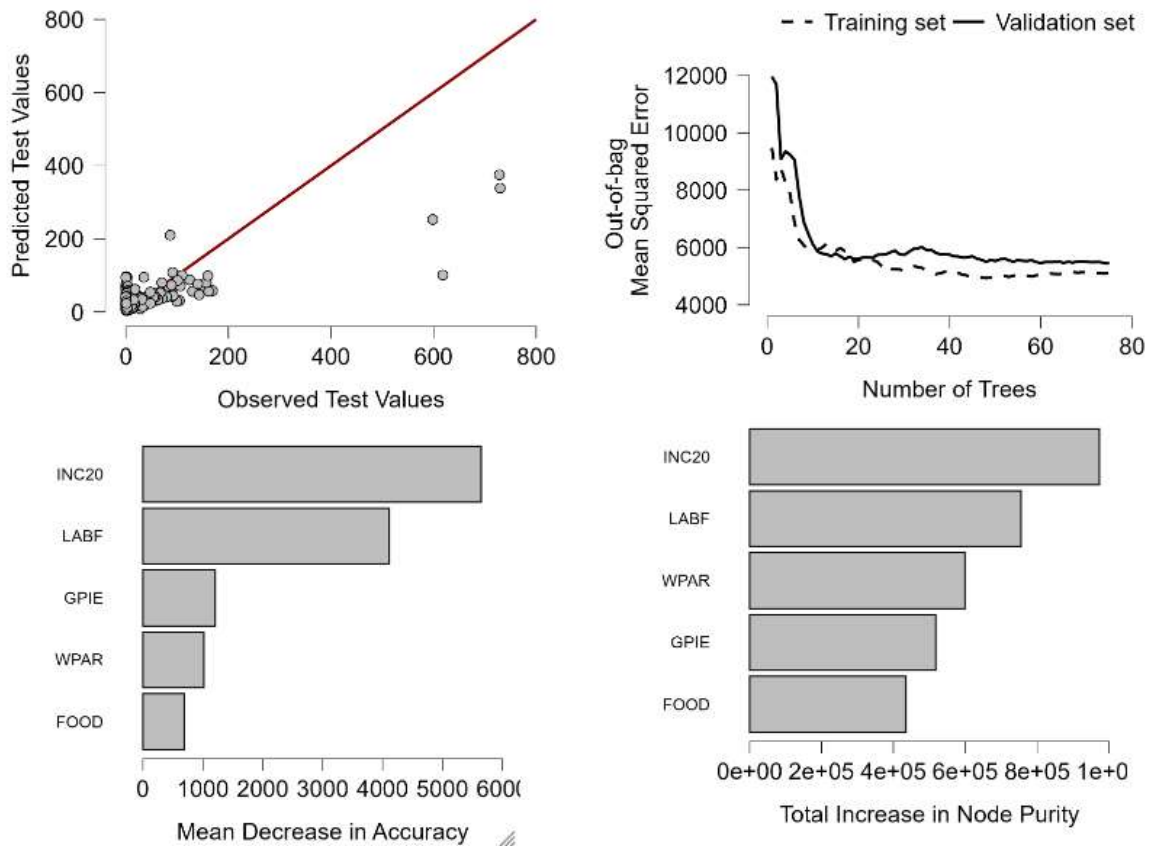
Table 13. Feature Importance Analysis of Socio-Economic Determinants in Building CO₂ Emissions via Random Forest Model

	Mean decrease in accuracy	Total increase in node purity	Mean dropout loss
INC20	56.464	973.192	83.505

LABF	4.113	755.169	81.929
WPAR	1.018	599.087	64.896
GPIE	1.209	517.854	61.618
FOOD	688.200	434.772	58.385

Importance measures from the Random Forest model present the relative contribution of six socio-economic and political parameters to BCE (Carbon dioxide emissions from buildings). These measures are exemplified by three crucial indicators: Mean Decrease in Accuracy (MDA), Total Increase in Node Purity (TINP), and Mean Dropout Loss (MDL, RMSE-based). Taken individually, they present a multi-aspect observation of each variable's impact on prediction under the ensemble learning model. Among the parameters under consideration, INC20—the income share of the bottom 20% of the population—emerges as the largest predictor in all three indicators. With a Mean Decrease in Accuracy of 56.464, a Total Increase in Node Purity of 973.192, and a Mean Dropout Loss of 83.505, it strongly dominates the internal structure of the model. This is an indication income distribution, and income distribution in the poorest segments of the population in particular, has an overwhelming impact on building-related emissions. A possible explanation is the increase in income inequality matches with increased access gaps in energy-efficient residences, renewable energy sources, and infrastructure ensuing on aggregate energy demand and emissions from residential sectors. The second largest variable after INC20 is LABF, the labor force participation. Although its Mean Decrease in Accuracy is significantly lower at 4.113, the variable has a highly significant Total Increase in Node Purity (755.169) and a Mean Dropout Loss comparable to INC20 (81.929). This reflects its highly strong prediction but low direct impact on model accuracy level. LABF is likely a variable mimicking socio-economic development and energy demand. Where the participation is very significant in the population, the likelihood is higher for higher economic activity and energy usage due to urban infrastructure and building usage with the resultant higher emissions. Where the population's participation is weak but due to structural vulnerabilities such as energy market imperfections and energy access gaps in the past, the building emissions would represent slower energy demand but inefficient energy usage.

Figure 4. Relative Importance of Socio-Economic Indicators in Predicting Building CO₂ Emissions Using Random Forest.



WPAR, women's presence in national parliaments, is also highly significant with a Mean Decrease in Accuracy of 1.018, Node Purity of 599.087, and Dropout Loss of 64.896. Its impact would be even more indirect but significant as higher female presence correlates with more inclusive and environmentally progressive policies affecting urban planning, building codes, and energy efficiency laws. GPIE, the Gender Parity Index in education, is next with moderate but consistent scores of predictive power. With a Mean Decrease in Accuracy of 1.209, Total Increase in Node Purity of 517.854, and Dropout Loss of 61.618, the variable reflects long-term structural determinants. Better-balanced societies in educational opportunities might be better equipped to adopt and sustain the low-emission building practices with added access to knowledge and energy and sustainability skills. FOOD, the food production index, has the role to play in the model as somewhat less direct but highly significant nonetheless. Its Node Purity of 434.772 and Dropout Loss of 58.385 would mean interlinkages by agricultural productivity and land-use practices with infrastructure type and energy systems connected to the built environment as relevant particularly within the rural or agro-industrial context. These analyses are in accord with the broader literature on emission forecasting where the direct and indirect forces time and again show the relevance of the socioeconomic and structural indicators as determinants of the trajectory of the CO₂ on the trajectory of emissions (for e.g., application of advanced time-series model approaches of the machine learning as validating the role of the direct and indirect forces (Costantini et al., 2024). Wen and Yuan (2020) used a hybrid neural network model for the prediction of emissions by China's commercial sector with the inclusion of the social and the economics indicators (Wen & Yuan, 2020). Zhou et al. (2024) investigated emissions in the China's "Four Provinces of Mountains and Rivers," applying time-series machine

learning techniques where the regional development gaps would be bridged (Zhou et al., 2024). As a whole, the variable importance statistics from the Random Forest provide strong support for the presence of strong links between the indicators of social equity, specifically income distribution and labour participation, and the building sector's environmental outcomes. INC20 is the most significant and consistent predictor of BCE and highlights the importance of social inclusion and equity in the building sector's environmental performance evaluation through the system of the ESG. These analyses provide support for the application of integrated ESG models where the social context of development is viewed as an outcome as well as an input of its impact on the environment (Table 14).

Table 14. Additive Feature Contributions to Predicted Building CO₂ Emissions: Social Indicators and Random Forest Interpretability

Case	Predicted	Base	FOOD	GPIE	INC20	LABF	WPAR
1	35.045	35.711	-5.862	6.325	0.438	-12.268	10.702
2	15.396	35.711	-0.881	2.596	-15.411	-1.366	-5.253
3	21.101	35.711	0.201	-0.242	2.898	-15.392	-2.075
4	18.029	35.711	-1.575	-6.516	7.341	-14.523	-2.408
5	45.117	35.711	-10.631	2.439	-15.323	12.822	20.099

Note. Displayed values represent feature contributions to the predicted value without features (column 'Base') for the test set.

The additive explanation results from the model predictions give an insightful picture of how the social indicators contribute to predicted building-related CO₂ emissions (BCE) across five test cases. The model maintains a base value of 35.711 Mt CO₂e for prediction of BCE as the mean predicted level in the absence of any explanation variable. The contribution from each feature—FOOD, GPIE, INC20, LABF, and WPAR—are depicted as increasing or decreasing the end predicted value by each observation so as to assess the direction as well as magnitude of influence. The end predicted value from Case 1 is 35.045 below the base level. Hereby, LABF contributes the largest negative influence by a strong downward pull of -12.268 as an indication where lower the labour force participation is associated with decreasing emissions but where higher the women's representation in the parliament contributes extensively in the opposite direction (+10.702), whereby higher women's parliamentary representations maybe associated with higher emissions as an indirect relation with the emerging market index from the otherwise governing the governance presence for an otherwise internalised conception on the applied CBA impact from the cases. GPIE contributes additively as well in the positive direction by (+6.325), while FOOD contributes largely negative by (-5.862), whereby stronger food production contributes to the decreasing building emissions maybe due to agriculture-based economies with lower infrastructure intensity. Case 2 with prediction at BCE of 15.396 reflects an abrupt drop from the base level. Hereby, the prime driver as the cut down level is INC20 contributes at -15.411 as an indication where increasing the income distribution as the lowest allotment of twenty percent strongly facilitates the lowest building emissions level. This here insinuates where higher the income distribution level by equitability basis maybe contributes to decreasing the residential energy usage wherein the correlated emissions where maybe concentrated but where equitably used instead by the residences. WPAR and LABF depict negative contribution as well sustaining the low emission idiosyncrasy while GPIE provides modest upward revision by (+2.596). Thus case 3 prediction equates to 21.101 where LABF provides extensive level by -15.392 as an indication where weak level of labour participation keeps on correlating the BCE by reducing

level always. INC20 provides positive by (+2.898) as an indication where in such context an equitability level on the income share distribution maybe contribute by increased energy access otherwise infrastructure level maybe due to the societal level of investments. On the whole, the positive as well as negative impacts nearly cancel out each other and result in moderate reduction from the base level.

Case 4 predicts 18.029, also below the baseline. LABF continues to show strong negative influence (-14.523), and GPIE significantly reduces the score by -6.516, suggesting that gender parity in education in this scenario correlates with lower building emissions, potentially due to broader access to knowledge and efficient practices. On the positive side, INC20 contributes +7.341, indicating some potential trade-off where equitable income can coincide with increased energy consumption or building activity. Case 5 stands out with a much higher predicted BCE of 45.117, well above the base. This sharp increase is driven predominantly by WPAR (+20.099) and LABF (+12.822), implying that in this scenario, high female political representation and strong labor participation are linked to increased emissions. This could reflect a setting where institutional inclusiveness aligns with advanced infrastructure and energy use. Notably, INC20 contributes negatively (-15.323), while FOOD has the most suppressive effect (-10.631), consistent with the trend where higher food production appears to correlate with lower building emissions. Overall, the results show that LABF is the most consistent downward driver of building-related emissions across nearly all cases, while WPAR and GPIE show context-dependent effects. INC20 is influential but varies in direction depending on the broader socio-economic setting. FOOD consistently suppresses emissions, reflecting perhaps a rural or agriculture-based development model with lower building-sector intensity.

6. Governance and Carbon: Unpacking the Institutional Drivers of Building-Sector Emissions

This section describes how the governance factors influence the building sector emissions of carbon dioxide (BCE), which maps to the "G" (Governance) pillar of the ESG approach. The empirics apply the fixed-effects as well as the random-effects versions of the regression model to the panel data of 982 observations to account for the influence on BCE from the level of political stability, legal infrastructure, public spending on infrastructure, and scientific capacity. The variables of interest are government effectiveness (GOVT), expenditure on education (EDUE), political stability (STAB), rule of law (LAWR), expenditure on R&D (RNDG), expenditure on hospitals (HOSP), and production of scientific publications (SCIE). The results display complicated and at instances surprising associations where higher governance capacity is linked to higher emissions reflecting the development-linked growth of infrastructure. These remarks are complemented by clustering analysis by Section 6.2 where countries are clustered by typical governance as well as emission profiles and display additional insights on the heterogenous trajectories through which governance has an influence on the built environment's environmental performance.

6.1 Governance and the Carbon Cost of Development: A Panel Analysis of Building Emissions

This section examines the impact of governance-related variables on the building sector emissions (BCE), an important "G" variable of the ESG framework. With a panel of 982 observations, fixed- and random-effects regressions are applied to assess how institutionally strong countries, government spending, legal fairness, and scientific abilities influence BCE levels. The prime predictors include government efficiency, political stability, rule of law, education and R&D spending, infrastructure by hospitals, and publications on the sciences. The findings exhibit subtle and sometimes counterintuitive dynamics—e.g., increasing emissions with higher governance quality—tracing out the intricate interdependence of the state's ability, development, and environmental performance. The

study provides an empirical basis for linking governance quality and emissions outcomes in the built environment.

We have estimated the following equation:

$$BP = \alpha + \beta_1(GOVT)_{it} + \beta_2(EDUE)_{it} + \beta_3(STAB)_{it} + \beta_4(RNDG)_{it} + \beta_5(LAWR)_{it} + \beta_6(HOSP)_{it} + \beta_7(SCIE)_{it}$$

Where $i=180$ and $t=2000-2020$ (Table 15).

Table 15. Panel Regression Results: Governance Indicators and Building-Sector CO₂ Emissions (BCE).

	Fixed-effects, using 982 observations Dependent variable: BCE			Random-effects (GLS), using 982 observations Dependent variable: BCE		
	Coefficient	Std. Error	t-ratio	Coefficient	Std. Error	z
const	14.7556**	6.12013	2.411	4.37823	9.08037	0.4822
GOVT	12.7921***	2.42217	5.281	11.2943***	2.15647	5.237
EDUE	0.450092*	0.245066	1.837	0.404457*	0.237995	1.699
STAB	-3.19087***	1.16516	-2.739	-2.29649**	0.964108	-2.382
RNDG	-4.11812**	1.71293	-2.404	-4.25068**	1.65044	-2.575
LAWR	-4.29609**	2.00928	-2.138	-1.36151*	0.755280	-1.803
HOSP	3.80015***	0.612354	6.206	3.75312***	0.589134	6.371
SCIE	0.000433491***	2.56973e-05	16.87	0.000462643***	2.50122e-05	18.50
Statistics	Mean dependent var		44.83643	Mean dependent var		44.83643
	Sum squared resid		88339.25	Sum squared resid		6648511
	LSDV R-squared		0.992464	Log-likelihood		-5724.171
	LSDV F(108, 873)		1064.614	Schwarz criterion		11503.46
	Log-likelihood		-3602.578	rho		0.766045
	Schwarz criterion		7956.121	S.D. dependent var		109.3165
	rho		0.766045	S.E. of regression		82.57715
	S.D. dependent var		109.3165	Akaike criterion		11464.34
	S.E. of regression		10.05935	Hannan-Quinn		11479.22
	Within R-squared		0.299880	Durbin-Watson		0.464047
	P-value(F)		0.000000			
	Akaike criterion		7423.156			
	Hannan-Quinn		7625.898			
Durbin-Watson		0.464047				
Tests	Joint test on named regressors - Test statistic: $F(7, 873) = 53.4184$ with p-value = $P(F(7, 873) > 53.4184) = 1.58042e-63$			'Between' variance = 5929.08 'Within' variance = 89.9585 mean theta = 0.942699 Joint test on named regressors - Asymptotic test statistic: Chi-square(7) = 437.883 with p-value = 1.77433e-90		
	Test for differing group intercepts - Null hypothesis: The groups have a common intercept Test statistic: $F(101, 873) = 78.0485$ with p-value = $P(F(101, 873) > 78.0485) = 0$			Breusch-Pagan test - Null hypothesis: Variance of the unit-specific error = 0 Asymptotic test statistic: Chi-square(1) = 2207.29 with p-value = 0		
				Hausman test - Null hypothesis: GLS estimates are consistent Asymptotic test statistic: Chi-square(7) = 73.1884 with p-value = 3.34305e-13		

--	--	--

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

This empirical work focuses on analyzing the influence of governance-related aspects—standing for the "G" factor in the ESG (Environmental, Social, Governance) approach—on building sector carbon dioxide emissions (BCE). The study addresses specifically how the institutional quality, the ability of the government, and investments in scientific-technological sectors, as these are reflected by certain governance indicators, impact yearly energy-use-based CO₂-equivalent emissions.

This empirical work focuses on analyzing the influence of governance-related aspects—standing for the "G" factor in the ESG (Environmental, Social, Governance) approach—on building sector carbon dioxide emissions (BCE). The study addresses specifically how the institutional quality, the ability of the government, and investments in scientific-technological sectors, as these are reflected by certain governance indicators, impact yearly energy-use-based CO₂-equivalent emissions.

The variable BCE covers emissions from residential, commercial, and unspecified building activities, in the classification of the IPCC 2006 and the AR5 global warming potential approach. The model is estimated as a panel using a dataset containing 982 observations. Two methods are used: a fixed-effects (LSDV) model and a random-effects generalized least squares (GLS) model. Both methods rely on BCE as the dependent variable and the independent variables on several facets of governance: government effectiveness (GOVT), political stability (STAB), rule of law (LAWR), government spending on education (EDUE), gross spending on research and development (RNDG), hospital infrastructure (HOSP), and the volume of scientific publications (SCIE). The case for including these variables is their conceptual correspondence with the governance pillar in the framework of ESG. Governance here encompasses something broader than corporate governance alone but includes broader public governance arrangements such as the strength of institutions, transparency, spending on research, ability to implement policies, and social infrastructure by which a country's environmental performance might be determined. A few recent papers support this approach: in its findings on how good governance makes countries cleaner through better governance structures (Siriteanu, 2023); how governance indicators represent the key determinants of the environment across the countries of the BRICS group when combined with trade and population flows (Sachan et al., 2024); and how governance quality and productive public spending show significant impact on the emissions of CO₂ in the building and energy sectors (Hassan et al., 2024).

Thus, the estimation of the relationship between these governance indicators and BCE sheds light on the role of effective and stable governance in promoting or dampening the environmental externalities of the built environment. The fixed-effects specification adjusts for time-invariant heterogeneity at the unit level—e.g., countries or regions—whereas the random-effects specification posits no correlation of unit-specific effects with regressors. The statistical results from both specifications are qualitatively similar but the formal test strongly supports the fixed-effects specification. Government effectiveness (GOVT) exhibits a positive and strongly significant correlation with BCE in both specifications.

The coefficient is 12.79 in the fixed-effects version and 11.29 in the random-effects specification, significant at the 1% level. This result suggests that, counterintuitively, good governments are associated with higher building-sector emissions. This may be taken as an expression of the presence of stronger state capacities in the more advanced countries, whereby good governments governally manage complex urban systems and undergird economically active processes implicating higher energy usage. Under such circumstances, stronger governance does not necessarily mean lower emissions but rather can support growth processes non-decoupled from fossil energy and increasing the pressure load of the environment. This process corresponds to recent work by Siriteanu (2023), who asserts good governance attracts international capital and supports the process of economic growth but may overestimate carbon emissions unless specifically aligned with achieving sustainable development goals (Adelina-Andreea, 2023). Education expenditure (EDUE), which can be considered the long-term expression of the public expenditure on human capital and the process of

building strong institutions, has a positive and weakly significant relationship with the BCE as well. Despite the fact its coefficient is small and significant but at the 10% level, the result suggests public expenditures on education are not necessarily reducing the influence on the environment by the term used here over the short term period. Rather, they may be correlated with the processes encompassed by broader-based process-oriented economic modernization processes increasing the demand for energy usage, particularly in urban and infrastructure-intensive sectors such as construction work and building activities. Political stability (STAB) is significant but negative for BCE. The coefficient has an estimated -3.19 in the fixed-effects version and -2.30 in the random-effects version but significant at conventional levels. This would suggest the politically stable countries tend toward the creation of the lowest possible emissions of the CO₂ type in the building sectors. Political stability may represent fertile grounds on which the consistent implementation of the environment policies and rules and long-term climatic policies are followed through. Conversely, politically unstable situations may lack the consistency and the strong-institutional base on the grounds of which the sustainability standards can be enforced on the building environment. This corresponds to the work by Sadaoui et al. (2024), who show in the MENA countries governance effectiveness significantly moderates the environment outcome when combined specifically with the impact of natural resources (Sadaoui et al., 2024). The case of research and development expenditure (RNDG) becomes particularly significant here. The R&D coefficients are negative and significant across both specifications at around -4.12 and -4.25, respectively. These results show that increased national spending on R&D correlates with decreased building emissions, presumably through the creation and deployment of cleaner technologies, energy-efficient building materials, and innovative construction techniques. From a governance point of view, this provides support for the assertion that prospective public spending and coordination of policies in the fields of science and technology can pay real environmental dividends. To support this case, Hassan et al. (2024) report the significant combined impact of R&D expenditure and governance quality on the reduction of air pollution and emissions of CO₂ in the case of the BRIC nations (Hassan et al., 2024). The rule of law (LAWR) is another key governance indicator found to exert a negative influence on BCE. The fixed-effects version of the model places the coefficient at -4.30 (5% significant level), but the random-effects version provides a smaller but negative coefficient at -1.36 (10% significant level). This outcome would mean that nations whose legal institutions are strong, whose contracts are enforced, and whose property is safeguarded are better placed to regulate emissions from the building sector. Legal certainty and adequate regulatory enforcement represent key elements of ESG-compatible governance toward fostering the building as well as real estate sectors toward higher degrees of environmental compliance. The availability of hospital beds (HOSP), an indicator of the ability of the healthcare system, exhibits a strong positive correlation with BCE. Both versions of the model give the coefficient as about 3.80 highly significant statistically. This outcome highlights the carbon-intensive nature of the healthcare infrastructure base. While never conceived as a governance variable per se, the volume and quality of public service provision—incorporating the cases of healthcare—are commonly referenced in the evaluation of the relationship with Environmentally Sustainable Governance (ESG). These would be the consequences of the development on the infrastructural side on the social front, even as this is worthwhile from the human development point of view, would promote an energy as well as an environmental expense if unrestrained without the complementariness of an energy-efficient course of action. This is corroborated by Al Fannah et al. (2023), who exhibit how hospitals contribute a lot to the production of carbon emissions and must adopt greener practices so as to prevent inflicting an environmental impact (Al Fannah et al., 2023), and by Kadrić et al. (2025), who exhibit how certain renovation initiatives can have a strong influence on reducing hospitals' energy consumption as well as production of emissions of the CO₂ (Kadrić et al., 2025). Scientific publication output (SCIE) is strongly and positively correlated with BCE but with the coefficient's magnitude low due to the variable's scale. This would mean that countries publishing increasing scientific work also have higher building-sector emissions. This result would most likely reflect the underlying building-sector-economic and institutional development of such countries

where universities and research facilities contribute to innovation as well as to energy-intensive activities. While long-term innovation is spurred by scientific publication, the same is also connected with increased energy usage now. The fixed-effects model in the end exhibits stronger statistical performance with an immensely large R-squared of 0.992, denoting fully explaining nearly all the BCE's variance. The fixed-effects model exhibits better performance even on the information criteria such as AIC and BIC over the random-effects specification. Finally, the Hausman test strongly rejects the null hypothesis of the consistency of the random-effects estimator such that the fixed-effects model is more reliable for such work. The test statistic is 73.19 with p-value well below the level of 0.01, paving the way for the confirmation from the test that the governance indicators are associated with the unobserved individual effects and consequently favor fixed-effects estimation. Overall, the study establishes the crucial role played by governance in the determination of the eventual impact on the environment in the building sector. It points out how crucial aspects of the ESG governance dimension such as the strength of the institution, the Stability, legal integrity as well as innovative spending influence the level of emissions. Importantly, even while better governance capabilities are usually assumed to lead to the reduction of the emissions rates, the relation here establishes much deeper than the relation where better governance capabilities associate with increased emissions rates due to the developmental as well as infrastructural realities on the ground. This subtlety gets support in Baldi and Lambertides (2024), who demand the consideration of performance-risk tradeoffs as well as context-based sustainability dynamics inadequately captured by the traditional belief in the role of the role of ESG integration in infrastructure investing (Baldi & Lambertides, 2024). Accordingly, the governance ought to be equally understood in the context of the orientation of the policies as the implementation of the regulatory system as well as the bolstering of the policies on the environment so as to realize the ability of the efficiency of the establishments translates into sustainable outcomes on the ground.

6.2 Governance and Emissions: Clustering Insights from Neighborhood-Based Algorithms

As a complement to the regression analysis, this section performs multiple clustering methods to search for latent group structures across countries on the basis of their governance features and building-sector CO₂ emissions (BCE). By comparing across different statistical and structural criteria such as R², AIC/BIC, silhouette statistics, and entropy, the analysis finds that Neighborhood-Based Clustering performs best. This approach does not only attain perfect explanatory power (R² = 1.0) but has superior cluster cohesion and separation as well and provides a solid ground for the interpretation of how varied governance profiles are connected with emission profiles (Table 16).

Table 16. Comparison of Clustering Algorithms on Governance Indicators and Building CO₂ Emissions (BCE)

	Density Based	Fuzzy c-Means	Hierarchical	Model Based	Neighborhood Based	Random Forest
R ²	0.45	0.0	0.762	0.845	1.0	0.144
AIC	1.0	0.094	0.609	0.0	0.308	0.093
BIC	0.538	1.0	0.234	0.165	0.0	0.981
Silhouette	1.0	0.087	0.362	0.217	0.449	0.0
Maximum diameter	0.294	1.0	0.142	0.271	0.0	0.717
Minimum separation	1.0	0.0	0.088	0.006	0.018	0.019
Pearson's \hat{I}^3	1.0	0.007	0.499	0.172	0.191	0.0
Dunn index	1.0	0.0	0.122	0.015	0.048	0.015

Entropy	0.0	1.0	0.367	0.832	0.962	0.969
Calinski-Harabasz index	0.872	0.0	0.598	0.574	1.0	0.015

Best performing algorithm by the criteria given is the Neighborhood Based Clustering. This is rendered by its remarkable and stable performance across the entire spectrum of clustering quality indices given. The best R^2 (1.0) by Neighborhood Based signifies the model accounts for 100% of the variance in the data, an essential criterion when assessing model fit. It is likewise best on the Calinski-Harabasz index (1.0), the ratio of the inter-cluster dispersion to the intra-cluster dispersion—a higher figure is an indication of well-defined and compact clusters. While its BIC and AIC are not the lowest (higher complexity), its best metrics are on practical clustering validity indices such as Silhouette score (0.449), where the points are well-clustered with regard to each other but distant from points from different groups. It has excellent metrics on the Entropy (0.962), an indication of an overall extensive level of order in the clustering structure. Additionally, while its Minimum Separation and its Pearson's γ are good but not the best, the balance across the structural and the statistical measures are an indication of a clustering solution well-fitted but interpretable as well as reliable to implement. The best and all-around the most effective solution from the comparison here is the Neighborhood Based Clustering, especially where an extensive explanatory power and strong cluster quality are called for.

6.2.1 Mapping Governance-Emission Profiles: Insights from Neighborhood-Based Clustering

To better illuminate the interaction of governance traits and building-related CO₂ emissions (BCE), the section below undertakes an examination of the outcome of the Neighborhood-Based Clustering algorithm found to be the best approach from the previous model comparisons. With the clustering of nations along eight governance and infrastructure indicators (including BCE, government effectiveness, education, R&D, and institutional quality), the examination derives country profiles different from each other distinctly. The clusters establish varied governance-emissions dynamics and provide an in-depth sense of the alignment of institutional qualities to environmental objectives in varied developmental contexts (Table 17).

Table 17. Country Clusters by Governance Profiles and Building CO₂ Emissions: Results from Neighborhood-Based Clustering

Cluster	1	2	3	4	5	6	7	8	9	10
Size	42	204	15	161	108	30	144	35	142	101
Explained proportion within-cluster heterogeneity	0.054	0.220	0.013	0.082	0.132	0.073	0.121	0.048	0.140	0.117
Within sum of squares	56.725	231.768	13.300	86.907	139.552	76.412	127.810	50.303	147.958	122.906
Silhouette score	0.314	0.165	0.911	0.416	0.283	0.706	0.223	0.507	0.273	0.326
Center BCE	0.363	-0.071	-0.096	-0.333	-0.352	5.163	-0.341	0.842	-0.197	-0.148
Center GOVT	-0.930	0.366	6.765	-0.236	0.547	0.079	-0.287	0.422	-0.591	-0.493
Center EDUE	-0.243	-0.379	-3.211	-0.731	0.105	0.373	0.603	-0.855	-0.090	1.848
Center STAB	-0.452	0.136	7.360	0.008	0.098	-0.245	-0.075	0.141	-0.549	-0.395
Center RNDG	-0.326	0.420	0.552	-0.498	1.860	1.044	-0.684	1.827	-0.812	-0.816
Center LAWR	-0.236	-0.058	8.003	-0.139	-0.039	-0.114	-0.143	-0.052	-0.191	-0.185
Center HOSP	1.860	-0.071	1.506	0.609	-0.165	-0.661	0.324	2.239	-1.097	-1.146

Center SCIE	-0.120	0.072	-0.411	-0.312	-0.226	5.061	-0.377	1.033	-0.230	-0.295
-------------	--------	-------	--------	--------	--------	-------	--------	-------	--------	--------

Dataset offers clustering of countries for eight most important indications: CO2 emissions because of buildings, government effectiveness, spending for education, political stability, spending for R&D, rule of law, hospital bed density, and scientific publication production. Each cluster is a group of countries with same profile defined by mean (center) for each indication. Cluster 3 is remarkable with extremely high silhouette measure (0.911), being compact and disconnected group. The cluster offers extremely high values for government effectiveness, political stability, and rule of law, suggesting sound institutions and stable governments. But it offers unexpectedly low spending for education and relatively low scientific production, possibly suggesting centralized or authoritarian regimes with strong institutions and governments over educational and scientific development. Cluster 6 is also remarkable with highest building emissions and extremely high level of scientific production. But it offers strong spending for R&D and relatively poor infrastructure for healthcare. The cluster likely offers technologically advanced and highly industrialized countries with high environmental impact and strong innovations systems. These findings are similar to broader clustering and decarbonization literature, e.g., Jiménez-Preciado et al. (2024), who used K-means with PCA and t-SNE to reveal global patterns of emissions based on structural indicators. Correspondingly, Chovancová et al. (2025) identify that EU countries adhere to distinct decarbonization paths based on their energy mix and root of their economy. Cluster 2, with highest country number (204), offers average values close to zero for each indication and relatively low silhouette measure. The implication is that it offers a basic or generic group with countries that do not indicate strong or strong tendencies for some area. Cluster 7 includes many countries spending most spending for education but behind others spending for R&D. The latter may be emerging economies or developing countries with emphasis placed over expanding educational access while building innovations and research capacity. Cluster 10 offers extremely high spending in education, however, with spending in R&D, and infrastructure of healthcare being low. This shortfall can represent a hard time for educational spending to be translated to more inclusive development outcomes. Cluster 3 and 6 represent by far the most specified information and patterns. The former includes politically stable, institutionally strong countries with narrow educational concentration, and the latter most likely includes industrially developed societies with sustainability and infrastructure trade-off outcomes. These specified patterns represent country-oriented strategies for CO₂ mitigation, as Matviychuk et al. (2024) argue and indicate such implication for modeling country-oriented carbon mitigation trajectories based on country patterns.

6.3 Predicting Emissions with Precision: Machine Learning Models for Governance and BCE

In supplement to the econometric analysis, in this subsection we utilize a set of machine-learning regression models to predict building-sector CO₂ emissions (BCE) based on governance-related features. Experimenting with models—including are K-Nearest Neighbors, Decision Trees, Random Forests, and Boosting—using traditional metrics of performance such as MSE, RMSE, MAE, MAPE, and R², analysis reveals best-performing algorithm(s) for capturing nonlinear, complex relationships. The results establish a data-driven baseline for tracking forecasting capacity of environmental outcomes defined by governance indicators (Table 18).

Table 18. Regression Model Performance on Governance-Based Prediction of Building CO₂ Emissions (BCE)

Model	MSE	MSE (scaled)	RMSE	MAE / MAD	MAPE	R ²
-------	-----	--------------	------	-----------	------	----------------

Boosting	1.0	0.136	0.373	0.248	0.135	0.843
Decision Tree	0.147	0.037	0.045	0.048	0.0	0.957
K-Nearest Neighbors	0.09	0.0	0.0	0.0	0.054	1.0
Linear Regression	0.647	0.11	0.269	0.219	0.092	0.873
Neural Network	0.003	1.0	1.0	1.0	1.0	0.0
Random Forest	0.0	0.125	0.599	0.198	0.043	0.857
Regularized Linear	0.797	0.148	0.316	0.25	0.124	0.83
Support Vector Machine	0.607	0.126	0.256	0.152	0.033	0.855

Based on the metrics of evaluation shown, the KNN algorithm is unequivocally the best-performing model of those attempted. The model demonstrates perfection or almost perfection across each of the most important performance indicators. Namely, both Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) are zero for the model, with Mean Absolute Percentage Error (MAPE) just 0.054. The measure of R-squared for variance explained by the model is 1.0. This demonstrates that there is complete prediction of the target variable within the scope of data employed for assessment. These figures put the KNN algorithm far ahead of others in terms of detail and predictive consistency. To contrast, although there are strong performances by the Decision Tree model with 0.957 R² and extremely low errors, there is not quite the same level of perfection. The Boosting and Random Forest models are both strong with 0.843 and 0.857 respectively for R² and fairly low error values. But they remain behind KNN across each significant measure. The Linear Regression and Regularized Linear remain sensible but exhibit higher errors and marginally lower R² values, a sign of less accurate prediction. The Neural Network model, a notoriously strong algorithm in so many difficult prediction tasks, is poor in this assessment. Admittedly possessing a highly low MSE value of 0.003, it fares poorest in each of the others with RMSE being RMSE; MAE being MAE; MAPE being MAPE; and R² being 0.0. The implication is that there was failure to generalize by the model or that perhaps it was not adequately trained and, while not typical, is certainly a consequence of inadequate data preparation and/or poor adjustment. The seeming superiority of the model by KNN prompts some important considerations. An important area to overcome is that of overfitting. KNN is a non-parametric, instance-based machine that makes prediction by finding most similar data points within a training set. The consequence is that it is highly accurate and extremely adaptable to data that are similar or highly structured. Such flexibility, however, can also cause overfitting when used with evaluation data that is defined by a training set. If such metrics were computed over the train set or with inadequate validation methods such as through cross-validation or test over a distinct data set, then the performance of such a model would not generalize adequately to new, unseen data. Recent literature supports such findings and subtleties. Srisuradetchai and Suksrikran (2024) proposed a random kernel KNN method to improve flexibility and decrease overfitting in such regression problems (Srisuradetchai & Suksrikran, 2024). Ihzaniah et al. (2023) contrasted KNN with multi-linear regression for predicting house prices and found that consistently, KNN had higher accuracy but with some word of caution regarding its ability to scale (Ihzaniah et al., 2023). Additionally, in such complex, high-variance environments as that of forecasting finance, for example, Jenifel et al. (2024) experimented with machine learning for predicting Bitcoin prices, affirming that non-parametric such as KNN can be highly accurate but with necessary caution to prevent overestimating such performance scores during prediction valuations (Jenifel et al., 2024). Another factor to consider is that of scalability in such instances by using KNN. While it excels dramatically in such instances, computation cost increases dramatically with datasets because during prediction, distances to be computed are those to each point within that train set by that query point and thus grows with data and complexity. That makes it prohibitive for compute over large or more complex data pieces. That being said, strictly by the metrics of such evaluation available, however,

KNN proves to be the most accurate and most reliable of the available models. Its ability to reduce errors by maximizing explanatory data power renders such a model quite desirable for such prediction jobs within similar data structures. Assuming that further such validation affirms such and that such data structure accommodates such assumptions by that of KNN and that such should be best within such data structures, such proves to be that best choice within such evaluated model choices.

6.3.1 What Drives Emissions? Feature Importance of Governance and Knowledge Indicators

This section explores the predictive influence of governance-related variables on building-sector CO₂ emissions (BCE) using feature importance metrics and additive model explanations. Feature importance, measured through mean dropout loss across 50 permutations, highlights the central role of scientific output (SCIE), healthcare infrastructure (HOSP), and education and R&D investment in driving model accuracy (Table 19).

Table 19. Governance Feature Importance in Predicting Building CO₂ Emissions: Dropout Loss-Based Ranking

	Mean dropout loss
SCIE	170.266
HOSP	33.765
EDUE	31.846
RNDG	29.399
GOVT	21.112
STAB	19.785
LAWR	3.374

Information provided elucidates relative significance of different governance, scientific, environmental, and socio-economic predictors to forecast or influence a particular outcome, by virtue of measure of mean dropout loss with computation of RMSE over 50 permutations. The technique helps to check for each’s significance to a prediction model by each feature with higher dropout losses being likely to be more significant impacts to model performance with computation of variable drop out. For each predictor, count of scientific articles (SCIE), appears to be most significant by far with 170.266 dropout loss. The enormously huge number points to volume of Peer-reviewed scientific articles being a principal causal driver within prediction paradigm. What possibly exists is that those are those nations that are strong in their outputs in sciences perform better across various development fronts such as technological progress, health care innovations, and educational attainment. The variable would equally be used to represent overall research infrastructure, human resources, and knowledge-intensive expenditures. The outcome is similar to previous research that found that there is strong predictivity by virtue of scientific indicator; e.g., that by Liu et al. that revealed that some learner behavior features—involving temporal patterns within academics outputs—would be used to predict such outcomes such as drop outs and found to be strong predictors of such outcomes with strong dependence of such measures in prediction models (Liu et al., 2020). SCIE is followed by hospital bed availability with 33.765 dropout loss in importance. That would make physical capacity of health systems be significant to underlying phenomenon that prediction model seeks to describe. Availability of staffed beds is basic but significant measure of country’s capacity to deliver health services and thus likely to be critically related to outcomes such as those to do with public health, wellbeing, and possibly even preparedness to respond to crisis. Education spending (EDUE), with 31.846 in loss value, and research and development spending (RNDG), with 29.399, are also important. These findings underscore the strategic importance of expenditures on education and

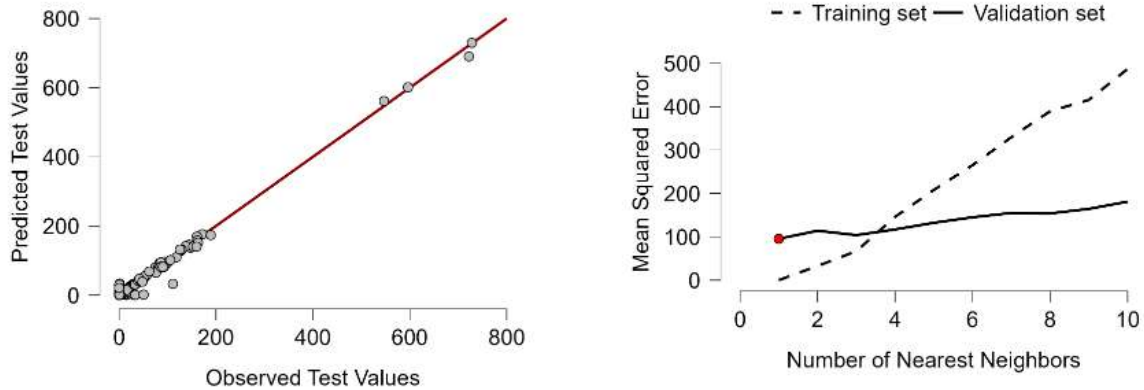
innovation. High spending on education tends to be reflective of a regular longitudinal policy bias towards building human capital, and spending on R&D tends to reflect a country’s drift towards technological and scientific competitiveness. The similar weighting used for both of those measures suggests that both foundation (education) and advanced (research) investment both play important contributions to national performance measures. These are in line with comparative modelling literature, such as Ihzaniah et al. (2023), who found that data-driven methodologies such as K-Nearest Neighbors regression better duplicated performance patterns with more structured data—quite frequently beating linear methodologies where interdependence amongst data’s constituent variables is significant (Ihzaniah et al., 2023). Governing effectiveness (GOVT), with 21.112, and political stability (STAB), with 19.785, are also important to performance by the model. These assist in underscoring institutional and government facets of development. Effective, credible, and stable governments allow for sustained and sound performance within economics, strong public provisioning within servicing to peoples, and social faith. Their significance, while somewhat lagging that of the measures of science and education, continues to reflect them as important enablers to national capacity. Interestingly, rule of law (LAWR) demonstrates lowest dropout loss amongst considered variables, with only 3.374. Whilst still significant, it appears relatively indirectly effective within that modelling context. That might reflect that while integrity of laws and faith within institutions are important for medium-to-long-term societal operation, they play more indirect impacts compared to more concrete measures such as measures of scientific output or health infrastructure. On balance, data reveals strong focus towards knowledge creation, healthcare provisioning, and spending for educational and research investment as having highest influence. Governance facets such as effectiveness and stability are important, with rule of law, however institutionally important, having more modest impact over shaped modelling results (Table 20).

Table 20. Additive Explanations of Governance Feature Contributions to BCE Predictions

Case	Predicted	Base	GOVT	EDUE	STAB	RNDG	LAWR	HOSP	SCIE
1	43.573	50.263	4.841	5.392	9.240	14.110	0.691	-7.606	-33.359
2	42.221	50.263	6.761	4.669	4.266	16.392	0.670	-7.442	-33.359
3	18.043	50.263	-11.495	-6.511	0.024	10.360	0.000	3.709	-28.306
4	21.695	50.263	-9.247	-6.022	-5.643	13.223	0.000	1.710	-22.589
5	26.966	50.263	-5.707	-7.277	-2.725	-3.935	0.000	1.141	-4.794

Table 20 demonstrates additive feature contributions for five test set cases, demonstrating how each predictor contributes to offsets based on a common base prediction of 50.263. For each case, the most negative predictor driver of predictions is the predictor variable SCIE, for number of scientific articles. For each case, SCIE contributes a large negative offset relative to that base, ranging by -33.359 for Cases 1 and 2 to -4.794 for Case 5. The suggestion is that there is a systematic negative correlation for each number of scientific papers with that outcome variable predicted—almost certainly carbon emissions from buildings (BCE). The trend may also indicate that countries with broader scientific outputs damp down CO₂ emissions that come from buildings, by virtue of superior innovations for building energy efficiency, sustainable building styles, or better policy choices based on scientific inputs (Figure 5).

Figure 5. Additive Effects of Governance Features on Predicted Building CO₂ Emissions: Case-Level Decomposition



Government effectiveness (GOVT) and R&D expenditures (RNDG) are generally positive contributors. For Case 1, for instance, both GOVT and RNDG add close to 24 units to predictions, and also such types of patterns are observable in Case 2 and Case 4. As such, this is in line with the explanation that effective government and spending on research are linked with higher emissions in the universe of the model, perhaps because more advanced or industrialized societies—typically with more effective government and spending to conduct research—also tend to have more advanced building structures, adding more to emissions unless checked by others. Education spending (EDUE) and political stability (STAB) are mixed contributors. For Case 1 and Case 2, both are positive contributors and are taking predictions upwards, and for Case 3 to Case 5, both are negative contributors. The latter could be because of non-linear relationships or relationships that are context-specific. For example, in societies with underdeveloped infrastructure, spending more on education would not yet spell less environmental impact, whereas in others, spending more on education is linked to a lesser environmental impact because there is more awareness and better regulatory capacity. Rule of law (LAWR) has a generally positive but unobtrusive impact, with its highest impact seeming to be in Case 2 (+16.392). The hospital bed supply (HOSP) is relatively nonvariable and adds little to predictions, so there seems to be limited explanatory power for this variable in this specific case. These additive explanations together show that SCIE is the dominant force impelling predictions down—something that reflects its strong negative association with emissions—while GOVT and RNDG are consistently linked with higher predicted values. That suggests a fine balance in that science is linked with mitigation but that development proxies such as government and research spending are linked with structural forces that would otherwise increase emissions but are checked by knowledge-driven interventions.

7. Harnessing ESG Dimensions for Effective Building Sector Climate Action

This paper's results offer significant policy lessons to national governments, multilateral agencies, and sustainable development and climate action stakeholders. Revealing that building sector carbon dioxide emissions are highly correlated with technological and economic factors but equally so with more inclusive ESG metrics such as governance, education, scientific output, and healthcare infrastructure, there are calls for recalculation of decarbonization trajectories with more institutional and social embedding. Firstly, significant negative connectivity between scientific output and

emissions intensity in the building sector indicates that building national research systems is a strategic climate policy lever. Policymakers should not deem scientific research as a fringe attribute of environmental planning but regard it as a mainstay tool for decreasing energy usage and emissions. Countries with more volumes of scientific literature, particularly technical disciplines such as engineering, materials science, and environmental sciences, could realize more advanced green building technologies, better simulation of energy flows, and better urban design. Public investment in higher educational institutions and research labs should thus be viewed as a long-term environmental resilience investment. The paper by Junsiri et al. (2024) highlights how combining advanced forecasting tools with sustainability policies in industrial structures assists in having a positive impact upon planning for cutting emissions by a huge margin. The governments should consider targeting expanded spending for transdisciplinary research projects with goals such as optimized building energy efficiency, low-carbon building materials, and sustainable building design and consider brokering university, research center, and building sector partnerships. Secondly, discovery that government effectiveness assists in explaining building sector emissions variability indicates that institutional quality is a necessary condition for successful environmental regulation and enforcement of climate policy. Effective government capacity at a high level supports building code enforcement, enforcement of energy standards, and compliance for decentralized urban environments. The implication is to conduct capacity building initiatives within public administration, particularly with regard to urban governance, environmental regulation, and policy enforcement. Samah et al. (2020) found that institutional response and spending for healthcare systems, particularly during periods of crisis such as with COVID-19, play a crucial role in influencing national patterns of emissions and associate social spending with environmental impacts. Furthermore, Andreoni (2021) showed that firm institutional reaction to the restrictions caused by COVID-19 generated significant drops in emissions in European countries, further confirming that good governance counts immensely for climate-related results.

International development agencies and multilateral institutions can help introduce programs building civil service capacity and regulatory convergence, especially in emerging marketplaces where gaps in governance can delay sustainability policy execution. Third, the discovered importance of expenditures on health and education as predictors of emissions performance underscores climate mitigation's broader social foundations. Countries that spend more of their GDP on spending for health and Education are developing social infrastructure to support sustainable conduct and adaptive capacity because they are developing the population's human capital. A population that is better-educated better values energy-efficient building advantages, assumes sustainable consumption patterns, and supports environmental initiatives. That is in line with Dong et al.'s (2024) research, who discovered that educational spending by China is space-wise heterogeneous over CO₂ emissions with regionalside educational policy sustainability results animating CO₂ emissions dynamics. Correspondingly, higher healthcare systems enhance public health and that is commensurably linked to environmental quality. As Samah et al. (2020) discovered, expenditures and sector growth for health sectors can play a significant role in CO₂ emissions, especially during systemic disruptions such as during the COVID-19 pandemic, where expansion of health sectors can also induce energy-intensive infrastructure buildup. Such discoveries call for integrated policy frameworks inscribing social spending towards environmental targets rather than as distinct or competing agendas. Furthermore, mixed impacts for R&D spending by itself to emissions reflect nuanced ties with innovation and environmental endpoints. R&D, despite being universally taken to lead greener technologies, data reflects that it can be linked to higher emissions in some contexts because of its concomitant association with expansion of infrastructure and expansion of industry. As Petrović and Lobanov (2020) discovered, impacts over CO₂ emissions for R&D are mixed over OECD countries contingent upon technological development and balance and balance of industrial intensity respectively (Petrović & Lobanov, 2020). Policymakers would thus need to establish innovations-based policies that prioritize sustainability targets directly. That would necessitate development of green R. & D. tax incentives, environmental conditionality for grants for innovations, and monitoring

systems that account for environmental implication of product of research. Development of sector R.& D. for sustainable building and building automation could be combined with public expenditures for basic research.

Also, analysis supports the thesis that rule of law and political stability, however indirectly significant they proved to be compared to others, play a crucial enabling role. Stable polities are more amenable to infrastructure investment, policy stability and to long-term planning, and these are conditions for sustainable decarbonization of building sector structures. And similarly, a strong rule of law ensures that property rights are protected, that judicial conflicts are not brought to bear to threaten green building projects and that contracts are respected. And so there is a case for considering governance reform and institutional development to be commensal planks of environmental policy, not least in those nations where poor legal systems and political instability weaken implementation capacity. Coming to international cooperation, there is reason to hope that ESG indicators can be deployed as diagnostics to identify nation-specific strengths and weaknesses in building towards decarbonization of building sector structures. Far from applying blanket policy prescriptions, climate finance institutions and development partners should each country's country-specific conditions of social and institutional infrastructure adapt their interventions to. Thus, for example, in those nations that have strong scientific basis with poor structures of governance, interventions could be crafted to turn research into enforceable policy. And in those situations where relatively strong structures of governance are combined with poor scientific output, capacity building within research and innovative sectors would be given strong emphasis. And indeed, cluster analysis conducted in this study reveals that there are recognizable groups that by virtue of similarities in ESG profile and patterns of emissions that nations fall into. And that such a circumstance would invite regionally coordinated responses to policy and platforms for peer-learning. Nations within similar cluster would be placed to learn lessons from each other, to mutually devise technical standards and converge regulatory methodologies. International institutions could stimulate such dialogue by convening transnational networks for research, by hosting policy forums and by financing pilot initiatives that reflect cluster members' similarities. Such strategies are commensal with emerging directions in ESG scholarship, where there was citation of cluster methodology being applied to map sustainability technologies and stakeholder environments across sectors (Alqudah et al., 2024). And with significant consequence for those sectors that are urban and local government sectors. For there, governments are normally responsible for implementing building codes, for running infrastructure and with public services. Strengthening ESG considerations at the municipal level—in research centres, public health, and schools—can facilitate co-benefits for emissions mitigation. Decentralized policies that enable cities and communities to innovate, experiment, and spread solutions need to be encouraged, particularly in federated or devolved administrative systems. Thirdly, the broader relevance of this research is that decarbonization of the building sector cannot be attained by technological fixes or market solutions alone. Whilst energy standards, carbon pricing, and clean technologies are important, they must be placed within a broader template that includes social equity, institutional integrity, and knowledge production. ESG metrics are a strong toolkit for designing such integrated responses, and facilitate policymakers to look towards broader conceptualizations of sustainability that are more inclusive and systemic conceptualizations of environmental progress drivers. The integrative perspective that underlies it is similarly shared by Lagodiyenko et al. (2023), who argue that ESG must transition to respond to new global challenges through a more inclusive conceptual breadth that encompasses socio-institutional dynamics (Lagodiyenko et al., 2023). Similarly, Paolone and Bitbol-Saba (2025) underscore that there is a need to identify thematic groups within ESG research in order to orchestrate multidisciplinary strategies towards sector-specific sustainability responses. In conclusion, this research underscores that there is a multi-dimensional policy response that is necessary to respond to the puzzle of reducing carbon emissions from buildings. Governments must make high-priority investments in sciences, schools, and health, and must reinforce systems of governance and legal institutions. International agencies must deploy ESG-informed diagnostics to

underpin support and cooperation strategies. Sustainable longevity within built environments will be realizable only by bringing environmental objectives within social and institutional development.

8. Conclusions

This paper has explored relationships between building sector CO₂ emissions and a large collection of Environment, Social, and Governance (ESG) indicators to offer a multi-angle environmental performance drivers analysis at the country level. Through combined econometric modeling, machine learning, and clustering analysis, the paper has deciphered significant and multi-dimensional interactions between knowledge production, institution building, and CO₂ emissions. These results oppose traditionally narrow focus to solely technological or cost-based explanations of environmental downturn and support more inclusive, systemic view. Amongst most correlated results is that of strong negative association between building sector CO₂ emissions and scientific output. The implication is that those countries with more research and more research outputs are likely to have better strategies for emissions mitigation and possibly due to better knowledge diffusion, technological innovations towards sustainable development, and better capacity to design environmental policies. The finding accentuates that far from solely being a technological vehicle, sciences play a foundational role as a strategic asset towards environmental governance and sustainability policy. The results further demonstrate that government effectiveness and spending towards research and development, despite being in most contexts associated with higher capacity and higher degree of economic development, are associated with higher degree of emissions in some contexts. The apparent contradiction most likely owes to more developed countries having more built and more energy-intensively built and urban infrastructure stocks. Thus, strong institutions and innovation remain necessary drivers for environmental performance but are evidently inadequate when they are not joined by focused sustainability agendas and environmental regulation. The inclusion of social and health and educational spending and bed provision more intensively inserts deeper analysis to reveal interconnections amongst social development and environmental achievement. Public health and educational spending are not just desirable endpoints in and by themselves but also play some role to increment society's capacity to adopt sustainability practices, to engage in forward-thinking planning, and to usher in environmental reforms. These results make a case towards climate mitigation's strategy to value social investment and environmental influence synergy. The paper's most significant contribution is most likely to be methodological innovations. By combining traditional regression analysis with machine-learning-based feature importance metrics and unsupervised clustering, analysis can disclose not only what matters most but also to what extent different country profiles respond to what matters most. The latter provides more nuanced policy conclusions and better understanding of different national settings as driving forces behind building sector decarbonization challenges and opportunities. Overall, analysis indicates achieving building sector decarbonization to be more challenging than technological modernization or economic incentives. It requires sound governance, active scientific community, socially inclusive investment policy, and comprehensive ESG framework capable of capturing multi-faceted sustainability. Future policies should be capable of taking due account of environmental objectives being rooted in and confluent with institutional, social, and knowledge-based development aspects. Only then can notable progress be made towards emissions reduction and achieving sustainable, climate-resilient development.

9. References

- Aagedal, M. T. (2024). Who's paying for green?-Property owners' and tenants' perspectives on ESG investments (Master's thesis, NTNU).
- Achuo, E. D., Nchofoung, T. N., Zanfack, L. J. T., & Epoge, C. E. (2023). The nexus between labour force participation and environmental sustainability: Global comparative evidence. *Heliyon*, 9(11).
- Adelina-Andreea, S. I. R. I. T. E. A. N. U. (2023). THE EFFECT OF GOOD GOVERNANCE ON CO2 EMISSIONS AND FDI INFLOWS. *Journal of Public Administration, Finance & Law*.
- Adewumi, A. S., Opoku, A., & Dangana, Z. (2024). Sustainability assessment frameworks for delivering environmental, social, and governance (ESG) targets: a case of building research establishment environmental assessment method (BREEAM) UK new construction. *Corporate Social Responsibility and Environmental Management*, 31(5), 3779-3791.
- Al Amosh, H., & Khatib, S. F. (2024). Toward sustainable healthcare: linking environmental governance, business innovation and carbon emission in Europe. *Management of Environmental Quality: An International Journal*, 35(7), 1461-1483.
- Al Fannah, J., Al Sabahi, S., Al Harthi, H., Al Bahrani, M., & Al Salmi, Q. (2023). Towards a green hospital approach in Oman: A case study of quantifying an environmental impact. *The International Journal of Health Planning and Management*, 38(5), 1583-1589.
- Al Mamun, M., Baghdadi, G., Hossain, M. Z., & Rokonzaman, M. Towards Net-Zero: Legislative Gender Diversity and Cleaner Energy Transition. Available at SSRN 5187015.
- Alali, Y., Harrou, F., & Sun, Y. (2023). Unlocking the potential of wastewater treatment: Machine learning based energy consumption prediction. *Water*, 15(13), 2349.
- Alqudah, M. Z., Sierra-garcia, L., & Garcia-benau, M. A. (2024). ESG sustainable technologies literature review using cluster-based methods. *PressAcademia Procedia*, 18(1), 88-89.
- Andreoni, V. (2021). Estimating the European CO2 emissions change due to COVID-19 restrictions. *Science of the Total Environment*, 769, 145115.
- Baldi, F., & Lambertides, N. (2024). Exploring the role of ESG for the performance and risks of infrastructure investing: evidence from the international funds' market. *Managerial Finance*, 50(1), 92-117.
- Banerjee, A., Das, P., & Fuerst, F. (2024). Are green and healthy building labels counterproductive in emerging markets? An examination of office rental contracts in India. *Journal of Cleaner Production*, 455, 141838.
- Barua, S., Adeleye, B. N., Akam, D., Ogunrinola, I., & Shafiq, M. M. (2022). Modeling mortality rates and environmental degradation in Asia and the Pacific: does income group matter?. *Environmental Science and Pollution Research*, 1-20.

Battisti, F. (2023). SDGs and ESG Criteria in Housing: Defining Local Evaluation Criteria and Indicators for Verifying Project Sustainability Using Florence Metropolitan Area as a Case Study. *Sustainability*, 15(12), 9372.

Bigerna, S., Bollino, C. A., & Polinori, P. (2021). Convergence in renewable energy sources diffusion worldwide. *Journal of Environmental Management*, 292, 112784.

Cakir, M. A., Unlu, R., Cakir, S. C., & Xanthopoulos, P. (2025, March). Future of Clean Cooking Energy Access in Emerging Economies by 2030. In *Operations Research Forum* (Vol. 6, No. 1, pp. 1-30). Springer International Publishing.

Chovancová, J., Petruška, I., Pata, U. K., & Adamišin, P. (2025). Diverse Pathways to Decarbonization: Cluster-Specific Impacts of Energy Sources on CO₂ Emissions in the European Union. *Energy Nexus*, 100367.

Chowdhury, M. S., Shak, M. S., Devi, S., Miah, M. R., Al Mamun, A., Ahmed, E., ... & Mozumder, M. S. A. (2024). Optimizing E-Commerce Pricing Strategies: A Comparative Analysis of Machine Learning Models for Predicting Customer Satisfaction. *The American Journal of Engineering and Technology*, 6(09), 6-17.

Chungath, J. (2023). Modeling Architecture to manage Energy, Social, and Governance (ESG) data of commercial buildings.

Ciski, M., & Rzaša, K. (2025). The Environmental Dimension of Sustainable Development in Relation to the Transition from Brown to Green Energy—A Case Study of Poland from 2005 to 2023. *Energies*, 18(11), 2993.

Costantiello, A., Laureti, L., Quarto, A., & Leogrande, A. (2025). Methane Emissions in the ESG Framework at the World Level. *Methane*, 4(1), 3.

Costantini, L., Laio, F., Mariani, M. S., Ridolfi, L., & Sciarra, C. (2024). Forecasting national CO₂ emissions worldwide. *Scientific Reports*, 14(1), 22438.

Cruz, P. L., Ávila, L. V., Dinis, M. A. P., & Baggio, D. K. (2023). Environmental, social and governance (ESG) and innovation in the construction sector: Systematic Literature Review. *Revista de Administração da UFSM*, 16(4), e1.

da Silva Tomadon, L., do Couto, E. V., de Vries, W. T., & Moretto, Y. (2024). Smart city and sustainability indicators: a bibliometric literature review. *Discover Sustainability*, 5(1), 143.

Debrah, C., Chan, A. P. C., & Darko, A. (2022). Green finance gap in green buildings: A scoping review and future research needs. *Building and Environment*, 207, 108443.

Deiningner, F., Woodhouse, A., Kuriakose, A. T., Gren, A., & Liaqat, S. (2023). Placing gender equality at the center of climate action. *World Bank Group Gender Thematic Policy Notes Series*, 179911.

Dong, Y., Gao, J., Qiu, J., Cui, Y., & Guo, M. (2024). The dynamic spatial effects of education investment on carbon emissions: heterogeneous analysis based on north-south differences in China. *Frontiers in Environmental Science*, 12, 1432457.

- Dugo, I. (2021). Finance embraces sustainability: An empirical analysis of the financial performances of ETFs investing in sustainable real estate and green building.
- Fedchenko, E. A., Gusarova, L. V., & Uskenbayeva, A. R. (2024). Green building in the ESG agenda for sustainable development of Russia: Conditions and trends. *Constr. Mater. Prod*, 7(9).
- Galloppo, G., & Nexus, P. (2025). *A Journey into ESG Investments*. Palgrave Studies in Impact Finance.
- Gu, Y., Katz, S., Wang, X., Vasarhelyi, M., & Dai, J. (2024). Government ESG reporting in smart cities. *International Journal of Accounting Information Systems*, 54, 100701.
- Gupta, R., Sharma, A., Anand, V., & Gupta, S. (2022, July). Automobile Price Prediction using Regression Models. In *2022 International Conference on Inventive Computation Technologies (ICICT)* (pp. 410-416). IEEE.
- Hajjar, M. K., & Ulku, I. (2022, June). Forecasting of Turkey's Total Electricity Consumption in Sectoral Bases Using Machine Learning Algorithms. In *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)* (pp. 1-7). IEEE.
- Hassan, S. U., Basumatary, J., & Goyari, P. (2024). Impact of governance and effectiveness of expenditure on CO₂ emission (air pollution): lessons from four BRIC countries. *Management of Environmental Quality: An International Journal*, 35(8), 1721-1743.
- Ifediora, C. O., & Igwenagu, C. L. Identifying Environmental, Social and Governance (ESG) Factors as Key Factors in Residential and Commercial Properties/Real Estate Investment Decision.
- Ihzaniah, L. S., Setiawan, A., & Wijaya, R. W. N. (2023). Perbandingan Kinerja Metode Regresi K-Nearest Neighbor dan Metode Regresi Linear Berganda pada Data Boston Housing. *Jambura Journal of Probability and Statistics*, 4(1), 17-29.
- Järvenstrand, R., & Åström, L. (2024). Evaluating thematic exposures using natural language processing: an unsupervised topic modeling approach to improve classification and mitigate risk.
- Jenifel, M. G., Jasmine, R. A., & Umanandhini, D. (2024, June). Bitcoin Price Predictive Dynamics Using Machine Learning Models. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.
- Jiao, J., Shuai, Y., & Li, J. (2024). Identifying ESG types of Chinese solid waste disposal companies based on machine learning methods. *Journal of Environmental Management*, 360, 121235.
- Jiménez-Preciado, A. L., Cruz-Aké, S., & Venegas-Martínez, F. (2024). Identification of Patterns in CO₂ Emissions among 208 Countries: K-Means Clustering Combined with PCA and Non-Linear t-SNE Visualization. *Mathematics*, 12(16), 2591.
- Junsiri, C., Sutthichaimethee, P., & Phong-a-ran, N. (2024). Modeling CO₂ Emission Forecasting in Energy Consumption of the Industrial Building Sector under Sustainability Policy in Thailand: Enhancing the LISREL-LGM Model. *Forecasting*, 6(3), 485-501.

- Kadrić, D., Blažević, R., Bajrić, H., Peco, A., & Kadrić, E. (2025). Renovation Measures for Reduction of Primary Energy Consumption and CO₂ Emissions of Hospital Building. *Tehnički glasnik*, 19(1), 26-33.
- Karerat, S., Houghton, A., & Dickinson, G. (2023). Leveraging healthy buildings as a tool to apply a people-centric approach to ESG. *Corporate Real Estate Journal*, 13(1), 32-43.
- Kempeneer, S., Peeters, M., & Compernelle, T. (2021). Bringing the user back in the building: an analysis of ESG in real estate and a behavioral framework to guide future research. *Sustainability*, 13(6), 3239.
- Khan, I., & Hou, F. (2021). The impact of socio-economic and environmental sustainability on CO₂ emissions: a novel framework for thirty IEA countries. *Social Indicators Research*, 155(3), 1045-1076.
- Kostrikin, P., & Andreeva, A. (2023, May). ESG-criteria for low-rise housing construction projects. In *AIP Conference Proceedings* (Vol. 2560, No. 1). AIP Publishing.
- Kudinova, Y., & Shushunova, N. (2023, August). Modern certification systems for green construction. In *AIP Conference Proceedings* (Vol. 2791, No. 1). AIP Publishing.
- Kumar, A., & Vasu, V. (2024). Tool wear prediction using multi-sensor data fusion and attention-based deep learning. *Production Engineering*, 1-17.
- Lagodiyenko, O., Lagodiienko, V., Ivanchenkova, L., Romanashenko, I., & Laskaiev, O. (2023). Prospects for the development of the esg concept in the face of new challenges. *Collection of papers new economy*, 1(1), 179-195.
- Lam, O. T. N. (2024). Exploring the mutual reinforcement of ESG and green building management for long-term sustainability: a case study of Hong Kong.
- Le, T. T., Priya, J. C., Le, H. C., Le, N. V. L., Duong, M. T., & Cao, D. N. (2024). Harnessing artificial intelligence for data-driven energy predictive analytics: A systematic survey towards enhancing sustainability. *International Journal of Renewable Energy Development*, 13(2), 270-293.
- LENZI, E. (2019). SIMBA: systematic clustering-based methodology to support built environment analysis.
- Li, R. Y. M., Li, B., Zhu, X., Zhao, J., Pu, R., & Song, L. (2022). Modularity clustering of economic development and ESG attributes in prefabricated building research. *Frontiers in Environmental Science*, 10, 977887.
- Liu, K., Tatinati, S., & Khong, A. W. (2020, December). A weighted feature extraction technique based on temporal accumulation of learner behavior features for early prediction of dropouts. In *2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)* (pp. 295-302). IEEE.
- Liu, N., Zhao, Y., Yan, B., & Hutchison, N. (2025). Tenants' ESG: influence on preferences and rent premiums for green buildings in commercial real estate. *Journal of Property Research*, 1-35.

- Mahmood, S., Sun, H., Ali Alhussan, A., Iqbal, A., & El-Kenawy, E. S. M. (2024). Active learning-based machine learning approach for enhancing environmental sustainability in green building energy consumption. *Scientific Reports*, 14(1), 19894.
- Matar, J., El Khoury, H., Charr, J. C., Guyeux, C., & Chrétien, S. (2025). Biological sequence clustering: novel approaches and a comparative study. *SN Computer Science*, 6(4), 369.
- Matviychuk, A., Zhytkevych, O., & Osadcha, N. (2024). Modeling carbon dioxide emissions reduction. *Energy Reports*, 12, 1876-1887.
- Meshcheryakova, T. (2022, March). Development of “Green” Building Standards in the Context of Global ESG-transformation. In *International Scientific and Practical Conference "Young Engineers of the Fuel and Energy Complex: Developing the Energy Agenda of the Future"(EAF 2021)* (pp. 188-194). Atlantis Press.
- Miklinska, J. (2024). Green buildings for greening supply chains: selected organisational and legal solutions on the Polish commercial warehouse space market.
- Mohamed, A. M. O., Paleologos, E. K., Mohamed, D., Fayad, A., & Al Nahyan, M. T. (2025). Critical Minerals Mining: A Path Toward Sustainable Resource Extraction and Aquatic Conservation.
- Mohapatra, B. B., Kumari, A., Mohapatra, S., & Sahoo, B. K. (2023). Dynamic Relationship Between Carbon Dioxide Emissions and Gross Domestic Product for Low, Middle-and High-Income Countries. *Journal of Quantitative Economics*, 21(4), 873-898.
- Moko, A., & Hlekane, N. Traditional Commercial Buildings Converted to Green Features: A Result of ESG or Load Shedding? South African Context.
- Montero, J. M., Fernández-Avilés, G., & Laureti, T. (2021). A Local Spatial STIRPAT Model for Outdoor NOx Concentrations in the Community of Madrid, Spain. *Mathematics*, MDPI.
- Morri, G., Yang, F., & Colantoni, F. (2024). Green investments, green returns: exploring the link between ESG factors and financial performance in real estate. *Journal of Property Investment & Finance*, 42(5), 435-452.
- Nimerenco, I., SOARE, B. E. L., & ZANESCU, D. (2024). Assessment of the Link between CO2 \$ emissions and Socio-Economic Indicators. In *Proceedings of the International Conference on Business Excellence (Vol. 18, No. 1, pp. 2594-2608)*. Sciendo.
- Nuradhi, M., Wibowo, H., Paulus, & Fatimah, R. A. (2023, November). Beyond Green Building, How Giant Architecture Firm Contribute to Sustainable Development Goals. In *International Conference on Sustainability in Creative Industries* (pp. 249-260). Cham: Springer Nature Switzerland.
- Obushnyi, S., & Novikov, A. (2024). INFLUENCE OF IT SOLUTIONS ON ESG REAL ESTATE DEVELOPMENT AND INVESTMENT ATTRACTIVENESS OF THE URBAN PROJECTS. *Європейський науковий журнал Економічних та Фінансових інновацій*, 1(13), 96-106.
- Oloyede, B. (2023). ESG-LED Sustainable Finance: A Blueprint For Ending The Menace Of Frequent Building Collapse In Nigeria. Available at SSRN 5043885.

Ombati, P. M. (2023). Evaluating Awareness Levels and Barriers to Green Building Implementation (Doctoral dissertation, University of Nairobi).

Paolone, F., & Bitbol-Saba, N. (2025). Identifying thematic clusters of ESG in contemporary accounting studies through systematic literature review. *Measuring Business Excellence*.

Parashar, M., Jaiswal, R., & Sharma, M. (2024). A quantitative analysis of ESG disclosure and financial performance in renewable energy companies: a two-step approach using unsupervised machine learning. *International Journal of Energy Sector Management*.

Petrović, P., & Lobanov, M. M. (2020). The impact of R&D expenditures on CO2 emissions: evidence from sixteen OECD countries. *Journal of Cleaner Production*, 248, 119187.

Rad, D., Cuc, L. D., Lile, R., Balas, V. E., Barna, C., Pantea, M. F., ... & Rad, G. (2022). A cognitive systems engineering approach using unsupervised fuzzy C-means technique, exploratory factor analysis and network analysis—A preliminary statistical investigation of the bean counter profiling scale robustness. *International Journal of Environmental Research and Public Health*, 19(19), 12821.

Ragazou, K., Garefalakis, A., & Samara, A. (2025). In the fight against climate change: the importance of environmental, social, and governance initiatives. In *Environmental Sustainability and Global Change* (pp. 313-321). Elsevier.

Richter, T. J., Soliva, E., Haase, M., & Wrase, I. (2022). Corporate real estate and green building: prevalence, transparency and drivers. *Journal of Corporate Real Estate*, 24(4), 241-255.

Robinson, S., & McIntosh, M. G. (2022). A literature review of environmental, social, and governance (ESG) in commercial real estate. *Journal of Real Estate Literature*, 30(1-2), 54-67.

Roukoz, K., & Ersenkal, D. (2023). Environmental, Social and Governance-related Challenges in the Construction Industry (CIC 2023).

Ruan, M., & Zhang, H. (2025). A Study of ESG. Available at SSRN 5239259.

Sachan, A., Pradhan, A. K., & Mohindra, V. (2024). How do governance indicators, trade openness, industrialization, and population growth affect environmental degradation in BRICS Nations?. *Discover Sustainability*, 5(1), 193.

Sadaoui, N., Zabat, L., Sekrafi, H., & Abid, M. (2024). The moderating role of natural resources between governance and CO2 emissions: Evidence from MENA countries. *Energy & Environment*, 35(3), 1597-1615.

Sahu, A., Pahi, D., Dwibedi, P., Mishra, A. P., & Mishra, B. (2025). Examining the role of ESG disclosure and firm characteristics in promoting global green building adoption: a panel probit approach. *Socio-Ecological Practice Research*, 7(1), 77-91.

Sahu, R. T., Verma, M. K., & Ahmad, I. (2023). Density-based spatial clustering of application with noise approach for regionalisation and its effect on hierarchical clustering. *International Journal of Hydrology Science and Technology*, 16(3), 240-269.

Sajadi, S. M. A., Barak, S., & Fereydooni, A. (2024). Online Portfolio Selection Using Macroeconomic Pattern Matching. Available at SSRN.

Salarin, T., Staiano, J., Bazzana, F., & Paterlini, S. (2024). Examining Intraday Stock Market Response to ESG News Classified by AI: An Event Study Approach. Available at SSRN 5040946.

Samah, I. H. A., Abd Rashid, I. M., Husain, W. A. F. W., Ibrahim, S., Hamzah, H., & Amlus, M. H. (2020). The impact of healthcare expenditure and healthcare sector growth on CO2 emission using dynamic panel data system GMM estimation model during COVID 19 crisis. *International Journal of Energy Economics and Policy*, 10(6), 235-241.

Samah, I. H. A., Abd Rashid, I. M., Husain, W. A. F. W., Ibrahim, S., Hamzah, H., & Amlus, M. H. (2020). The impact of healthcare expenditure and healthcare sector growth on CO2 emission using dynamic panel data system GMM estimation model during COVID 19 crisis. *International Journal of Energy Economics and Policy*, 10(6), 235-241.

Sanchez-Olmedo, D., Ortiz-Yepe, P., & Faytong-Haro, M. (2025). Bridging the Gap: The Impact of Gender Equality on CO2 Emissions Across Countries. *World*, 6(1), 26.

Sarro, F., Petrozziello, A., & Harman, M. (2016, May). Multi-objective software effort estimation. In *Proceedings of the 38th International Conference on Software Engineering* (pp. 619-630).

Sendecki, A. (2025). Analiza genotypowo-fenotypowa pacjentów ze zwyrodnieniem plamki związanym z wiekiem oparta na wynikach badań obrazowych: OCT, OCTA i kolorowych zdjęć dna oka (Doctoral dissertation, Katedra i Oddział Kliniczny Okulistyki).

Sharma, L., Poddar, P. N., & Singhal, P. B. (2024). INVESTING IN WOMEN, INVESTING IN THE PLANET: QUANTIFYING THE IMPACT OF WOMEN'S EMPOWERMENT ON ENVIRONMENTAL SUSTAINABILITY. *Revista de Gestão Social e Ambiental*, 18(6), 1-16.

Siriteanu, A. A. (2023). THE EFFECT OF GOOD GOVERNANCE ON CO2 EMISSIONS AND FDI INFLOWS. *Journal of Public Administration, Finance and Law*, (30 Special), 225-233.

Sjöberg, K., & Östling, M. (2024). Frameworks for Sustainable Construction: Comparing the EU Taxonomy, ESG Ratings and BREEAM-SE.

Srisuradetchai, P., & Suksrikran, K. (2024). Random kernel k-nearest neighbors regression. *Frontiers in big Data*, 7, 1402384.

Sulaiman, N. N., Omar, A. J., Pakir, F., & Mohamad, M. (2024). Evaluating environmental, social and governance (ESG) practice among Malaysian public listed construction companies using FTSE Russell rating model. *International Journal of Sustainable Construction Engineering and Technology*, 15(1), 25-40.

Tawosi, V., Sarro, F., Petrozziello, A., & Harman, M. (2021). Multi-objective software effort estimation: A replication study. *IEEE Transactions on Software Engineering*, 48(8), 3185-3205.

Times, U. (2021). *Unsettled Lives: Shaping our Future in a Transforming World*. Human development report, 2022.

Valaskova, K., Gajdosikova, D., & Lazaroiu, G. (2023). Has the COVID-19 pandemic affected the corporate financial performance? A case study of Slovak enterprises. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 18(4), 1133-1178.

Valero, A. (2025). The Impact of ESG Certifications on Class A Office Buildings in Madrid: A Multi-Criteria Decision Analysis.

Vladimirova, I., Dmitriev, A., Kallaur, G., Oleynikova, N., & Tsygankova, A. (2023). Trends in the application of “green” standards on the life cycle of capital housing construction objects. In *E3S Web of Conferences* (Vol. 403, p. 02018). EDP Sciences.

Voronkov, A. (2025). Mechanisms for Attracting Investment into Green Building Projects. *The American Journal of Management and Economics Innovations*, 7(05), 83-89.

Wang, J., & Xue, F. (2023, August). Emerging trends of ESG in the construction sector: A promising pathway to sustainable and responsible development. In *International Symposium on Advancement of Construction Management and Real Estate* (pp. 2033-2044). Singapore: Springer Nature Singapore.

Wang, W., Li, D., Zhou, S., Wang, Y., & Yu, L. (2024). Exploring the key influencing factors of low-carbon innovation from urban characteristics in China using interpretable machine learning. *Environmental Impact Assessment Review*, 107, 107573.

Wang, X., & Wen, H. Research on Green Transformation of Real Estate Enterprises under ESG Concept. In *ICCREM 2023* (pp. 500-505).

Wen, L., & Yuan, X. (2020). Forecasting CO2 emissions in Chinas commercial department, through BP neural network based on random forest and PSO. *Science of The Total Environment*, 718, 137194.

Xianghan, Z., & Zhengwei, L. (2025). Research on the Impact of Corporate Green Transformation on Organizational Performance under ESG Strategy Orientation. *Risk and Financial Management*, 6(1), p24-p24.

Xu, X. P., & Liu, L. (2025). EFFECTS OF CLIMATE POLICY ON URBAN CARBON EMISSIONS--NEW INSIGHTS FROM CHINA'S LOW-CARBON CITY PILOT POLICY. *Applied Ecology & Environmental Research*, 23(2).

Yaman, K. M., & Ghadas, Z. A. A. (2023). The Role of Contract in Advancing ESG in Construction: A Proposed Framework of ‘Green’ Standard Form of Contract for Use in Green Building Projects in Malaysia. In *Artificial Intelligence and Transforming Digital Marketing* (pp. 269-279). Cham: Springer Nature Switzerland.

Yang, B., Ali, M., Hashmi, S. H., & Jahanger, A. (2022). Do income inequality and institutional quality affect CO2 emissions in developing economies?. *Environmental Science and Pollution Research*, 29(28), 42720-42741.

Yang, Q., Wang, J., Liu, X., & Liu, Y. (2023). Complexity influence of societal development comprehensive indicators on building carbon emission: empirical evidence from China. *Environmental Science and Pollution Research*, 30(55), 117179-117200.

Yap, C. K., Leow, C. S., & Goh, B. (2024). Sustainable construction materials under ESG: A literature review and synthesis. *MOJ Biology and Medicine*, 9(1), 1-6.

Zhang, H., Li, R., Cai, X., Zheng, C., Liu, L., Liu, M., ... & Wang, X. (2022). Do electricity flows hamper regional economic–environmental equity?. *Applied Energy*, 326, 120001.

Zhao, X., Jiang, S., & Gao, J. Spatial Impacts of Green Finance Reform Pilot Zones on Renewable Energy Technology Innovation: Pathways for Accelerating Low-Altitude Economic Development. Available at SSRN 5122716.

Zheng, J., Khurram, M. U., & Chen, L. (2022). Can green innovation affect ESG ratings and financial performance? Evidence from Chinese GEM listed companies. *Sustainability*, 14(14), 8677.

Zhou, X., Liu, Z., Wu, L., & Wang, Y. (2024). Study on CO2 Emission Forecast of “Four Provinces of Mountains and Rivers” Based on Time-Series Machine Learning. *Atmosphere*, 15(8), 949.

Appendix

Table A1. We used the following variables acquired from the World Bank website.

Acronym	Variables	Definition
BCE	Carbon dioxide (CO2) emissions from Building (Energy) (Mt CO2e)	Total annual carbon dioxide equivalent (CO2e) emissions from energy use in the buildings sector, covering IPCC 2006 categories 1.A.4 (Residential and other) and 1.A.5 (Unspecified), converted to CO2e using global warming potentials from the IPCC Fifth Assessment Report (AR5). Unit: Mt CO2e per year.
CFTC	Access to Clean Fuels and Technologies for Cooking	Access to clean cooking fuel and technology estimates come from the WHO Global Household Energy Database with national representative household surveys as the sole data source (e.g., DHS, MICS, LSMS, WHS, national censuses). A multivariate hierarchical model—split by urban and rural—estimates fuel-type trends by grouping them as 'clean' (e.g., gas, electricity, alcohol) and 'polluting' (e.g., biomass, charcoal, coal, kerosene). There are estimates for 191 countries by Stoner et al. (2020). High incomes countries (by World Bank 2022 classification) have universal clean fuel access assumed.
ELEC	Access to Electricity	Reliable and secure electricity is essential for economic growth, poverty reduction, and human development. As countries decarbonize, dependence on clean, efficient power will grow. Electricity access enables basic services (lighting, refrigeration, appliances) and is a key indicator of energy poverty. Especially in developing countries, governments are prioritizing electrification through rural programs and national agencies. While vital for raising living standards, electricity generation can harm the environment—its impact depends on the energy sources used, with fossil fuels like coal being especially carbon-intensive.
ENUC	Energy Use per Capita	Total energy consumption gauges final energy use after conversion into end-use fuels (e.g., electricity, processed oil). It encompasses energy from combustible renewables and waste—like biomass, biogas, and municipal waste. Biomass describes plant materials used as such or converted into fuel, heat, or power. Figures, as gathered by the IEA, use per capita estimates from World Bank population. National non-OECD data are converted to IEA equivalence. Figures are imprecise and not completely comparable for countries because of limited data quality, particularly for waste and renewables. Energy values have been computed in terms of oil equivalents on the basis of 33% thermal conversion for nuclear and 100% for hydropower.
PM25	PM2.5 Pollution	Population-weighted exposure to ambient PM2.5 refers to the average level of fine particulate matter (PM2.5) pollution that a country's population is exposed to. PM2.5 particles, with a diameter smaller than 2.5 microns, can penetrate deep into the lungs and pose serious health risks. This measure is calculated by weighting the annual average PM2.5 concentrations by the population distribution across urban and rural areas.
RENC	Renewable Energy Consumption	The share of total final energy consumption derived from renewable sources, based on data from IEA, IRENA, UNSD, WHO, and the World Bank (Tracking SDG 7, 2023).
FOOD	Food Production Index	The Food Production Index reflects the output of edible crops that offer nutritional value. It excludes items like coffee and tea, which, despite being consumable, do not contribute meaningfully to nutrition. This metric emphasizes food sources that support dietary needs, aligning production data with human nutritional requirements rather than general edibility alone.

GPIE	Gender Parity in Enrollment	The Gender Parity Index (GPI) in primary education is calculated by dividing female gross enrollment by male gross enrollment. Data are collected by UNESCO from national education surveys and aligned with ISCED standards to ensure international comparability. The current methodology was adopted in 2011. Reference years reflect when the school year ends. A GPI below 1 indicates girls are disadvantaged; above 1 indicates boys are. Achieving gender parity enhances women's opportunities and contributes to broader social and economic development.
INC20	Income Share Lowest 20%	The percentage share of income or consumption reflects the portion received by population subgroups, typically divided into deciles or quintiles. Due to rounding, quintile shares may not total exactly 100%. Data come from household surveys via national statistics agencies and World Bank departments, with high-income country data largely from the Luxembourg Income Study. These measures support the World Bank's goal of shared prosperity—focusing on income growth among the bottom 40%—and help assess inequality within and across countries.
LABF	Labor Force Participation	The labor force participation rate represents the share of the population aged 15 and older that is economically active, including all individuals engaged in the production of goods and services during a specific period. Data, sourced from the ILO's modelled estimates, highlight persistent gender disparities: women's labor force participation is generally lower than men's due to social, legal, and cultural norms. In low-income countries, women often work unpaid in family enterprises, while in high-income nations, higher education has expanded their access to better employment opportunities, though inequalities persist.
WPAR	Women in Parliament	Women in parliaments refers to the percentage of seats held by women in a single or lower house of national parliaments. Although progress has been made, women remain significantly underrepresented in decision-making roles, especially in developing countries. Gender inequality in political participation limits women's influence on policy and national priorities. Equal representation is essential for inclusive governance and sustainable development. True democracy requires full participation of women, whose perspectives and leadership are vital for shaping equitable and effective public policies.
GOVT	Government Effectiveness	Government Effectiveness: Estimate measures perceptions of public service quality, civil service independence, policy formulation and implementation, and government credibility. Scores range from -2.5 to 2.5, based on a standard normal distribution.
EDUE	Gov. Expenditure on Education	General government expenditure on education, including current spending, capital outlays, and transfers, is measured as a percentage of GDP. It accounts for education funding from all government levels—local, regional, and central—and includes international transfers to the government. This indicator reflects the government's financial commitment to the education sector relative to the country's economic output.
STAB	Political Stability	Political Stability and Absence of Violence/Terrorism reflects perceptions of the risk of political unrest, government instability, and politically-motivated violence or terrorism. Countries are ranked by percentile, from 0 (least stable) to 100 (most stable), allowing global comparison. Percentile ranks are adjusted over time to ensure consistency despite changes in the number of countries included in the Worldwide Governance Indicators (WGI).
RNDG	R&D Expenditure	Gross domestic expenditures on research and development (R&D), measured as a percentage of GDP, represent a country's financial commitment to innovation and technological progress. This includes both capital and current spending across four key sectors: business enterprises, government institutions, higher education, and private non-profits. It encompasses all R&D activities—basic research, applied research, and experimental development—supporting economic and scientific advancement.
LAWR	Rule of Law	Rule of Law reflects perceptions of how much confidence individuals and institutions have in societal rules, particularly regarding contract enforcement, property rights, police effectiveness, and judicial independence. It also considers the likelihood of crime and violence. Countries receive a score ranging from approximately -2.5 (weak rule of law) to 2.5 (strong), based on a standard normal distribution.
HOSP	Hospital Beds	Hospital beds refer to the total number of beds that are maintained, staffed, and immediately available for the admission of patients. These include inpatient beds in public and private hospitals, general and specialized institutions, and rehabilitation centers. The count typically covers beds used for both acute and chronic care, reflecting the overall healthcare system's capacity for treatment and recovery.
SCIE	Scientific Articles	Scientific and technical journal articles represent the total number of peer-reviewed publications in key research areas, including physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences. These articles reflect ongoing advancements, innovation, and collaboration within the global scientific community, contributing to knowledge expansion and technological development across multiple disciplines and industries.

