

AI for sustainable development: Modelling the impact of the 17 SDGs on public sector performance under agenda 2030

Saida Ifiss¹ , Taoufiq Saffaj^{2*} 

¹The National School of Business and Management, Abdelmalek Essaâdi University, Tangier, MOROCCO

²Faculty of Sciences and Techniques, University Sidi Mohamed Ben Abdellah, Fez, MOROCCO

*Corresponding Author: taoufiq.saffaj@usmba.ac.ma

Citation: Ifiss, S., & Saffaj, T. (2026). AI for sustainable development: Modelling the impact of the 17 SDGs on public sector performance under agenda 2030. *European Journal of Sustainable Development Research*, 10(1), em0350. <https://doi.org/10.29333/ejosdr/17396>

ARTICLE INFO

Received: 29 May 2025

Accepted: 18 Oct. 2025

ABSTRACT

This study models the impact of the 17 sustainable development goals (SDGs) on the performance of Moroccan public-sector organizations using ensemble machine learning methods: bagging, stacking, and voting. Based on a panel of 300 entities across central government, local authorities, and state-owned enterprises, we evaluate three dimensions of performance: efficiency, resource management, and citizen satisfaction. Following data preprocessing with imputation, min-max normalization, and stratified ten-fold cross-validation, the voting regressor achieved the best predictive performance (coefficient of determination $[R^2] = 0.951$; root mean square error $[RMSE] = 0.190$), closely followed by the stacking regressor ($R^2 = 0.945$; $RMSE = 0.198$). SHapley Additive exPlanations analysis confirmed the relative importance of SDG7 (affordable and clean energy), SDG8 (decent work and economic growth), and SDG11 (sustainable cities and communities) as the most influential drivers across all outcome variables. Additional goals such as SDG3, 6, and 9 improve operational efficiency, while SDG4, 16, and 17 provide institutional support. These findings offer practical guidance for policymakers to prioritize strategic levers and enhance sustainable governance through data-driven public sector reforms.

Keywords: machine learning, sustainable development, public governance, sustainable development goals, ensemble learning

INTRODUCTION

Within the past ten years, there has been significant evolution within the global public sector fueled by two interrelated forces: mass institutionalization of the 2030 agenda for sustainable development, and the widespread emergence of product (data) and tool-based approaches to decision-making (United Nations, 2015). In low- and middle-income countries, the challenge has now shifted from feigned commitment to the sustainable development goals (SDGs) to the active undertaking of tracking, financing, and prioritizing our interventions to achieve discernible improvements in the performance of the public sector (Sachs et al., 2022). In a critical review of the SDGs' journey so far, Rosen (2025) has argued that the success of the Goals will rely less on aspirational normative values and more on the institutionalization of institutional performance frameworks that are context-sensitive and will help inform real-time policy changes. Thus, as attaining the SDGs becomes systematised globally and locally, we will need governance models that are data-informed and development-oriented. Morocco, which has embedded the SDGs in its 2022-2026 government action plan, offers an instructive testbed; the country combines

ambitious social programs with acute fiscal constraints and an emerging digital-transformation agenda (Haut-Commissariat au Plan, 2023). Yet empirical evidence on how each of the seventeen SDGs translates into operational gains—measured through efficiency, resource stewardship and citizen satisfaction—remains sparse, fragmented and often methodologically limited (van der Kolk, 2022).

At the same time, the maturation of ensemble learning has transformed predictive analytics in domains as diverse as health policy (Ganaie et al., 2022), climate finance (Huber et al., 2022) and municipal service delivery (Wirtz et al., 2020). By combining weak or heterogeneous learners into a composite predictor, ensemble methods consistently outperform single models, especially when relationships are non-linear, multi-collinear or context dependent (Dietterich, 2000; Opitz & Maclin, 1999). Recent work has extended these advantages to the public sector, demonstrating that random forest, gradient boosting, and stacking architecture can explain up to 30 % more variance in budget absorption, audit irregularities and citizen complaint volumes than traditional econometric baselines (Kummitha, 2020). Despite these advances, two critical gaps persist.

First, few studies integrate all seventeen SDGs into one predictive framework. Most focus on a narrow subset—typically SDG3 (good health and well-being) or SDG8 (decent work and economic growth)—and seldom test cross-goal interactions (Barbier & Burgess, 2020). Second, ensemble models are rarely interpreted in ways that are accessible to senior civil servants. While accuracy metrics such as coefficient of determination (R^2) and root mean square error (RMSE) are standard, they do little to illuminate actionable levers; without transparent post-hoc explanations, even high-performing models risk being ignored by decision-makers (Rudin, 2019). Techniques such as SHapley Additive exPlanations (SHAP) values (Lundberg & Lee, 2017) partially fill this gap but have not yet been applied to a comprehensive SDG–performance data set in the Moroccan context.

Addressing these gaps, the present article pursues three objectives.

1. We compile a unique panel of 300 Moroccan public organizations, capturing their progress on all seventeen SDGs together with three core output variables: efficiency performance, resource management, and citizen satisfaction.
2. We implement and compare ten ensemble configurations—random forest, extra trees, AdaBoost, gradient boosting, XGBoost, LightGBM, CatBoost, bagging SVR (support vector regression), stacking, and voting—under a uniform, stratified 10-fold cross-validation protocol.
3. Using SHAP, we quantify each SDG's marginal effect on the three performance dimensions, thereby producing a rank-ordered map of policy levers. To our knowledge, this is the first study to supply a full-spectrum SDG importance profile for a national public sector.

Morocco offers fertile ground for such an experiment. On the one hand, the country has scored above the regional average on the SDG index since 2018, particularly in renewable energy and infrastructure (Sachs et al., 2022). On the other hand, public satisfaction with administrative services remains volatile, and the Cour des Comptes (2023) continues to flag recurrent inefficiencies in resource allocation. Policymakers thus require granular evidence to decide which SDGs create the largest performance dividend and merit accelerated funding.

Traditional regressions impose linearity and independence assumptions ill-suited to SDG data, where goals are intentionally aligned and often exhibit multicollinearity (Nilsson et al., 2016). Ensemble models circumvent these constraints by aggregating diverse learners, each capturing different facets of the data-generation process. Empirical meta-analyses show that Stacking and Voting can reduce generalization error by 10–25 % compared to the best standalone model, provided that hyper-parameters are tuned through nested resampling (Zhou, 2021). Moreover, ensemble learning meshes seamlessly with explainable-artificial intelligence (AI) toolkits, enabling policy narratives that are both statistically rigorous and operationally intuitive (Adadi & Berrada, 2018).

The remainder of the article proceeds as follows. We first review prior literature on SDG measurement and ensemble applications in the public sector. We then detail the data set,

the preprocessing workflow, and the ensemble configurations. After that we present the predictive results and interpret SHAP outputs to delineate high-impact SDGs. Next, we discuss policy implications and robustness checks. We finally conclude the study.

In doing so, the study makes two principal contributions:

- (1) it establishes a replicable ensemble pipeline capable of ranking SDG leverage points for administrative performance, and
- (2) it provides Morocco's policymakers with a data-driven basis for selectively intensifying investments where the marginal returns are highest.

Despite growing interest in the SDGs, most prior studies have relied on sector-specific approaches or aggregated indexes, often limited to descriptive statistics or econometric methods. These approaches tend to lack both predictive accuracy and interpretability, leaving policymakers with limited tools to assess how individual SDGs concretely impact public sector outcomes. This study addresses this critical gap by proposing a comprehensive, data-driven framework that models the influence of all 17 SDGs on public sector performance using advanced ensemble learning techniques (bagging, stacking, and voting) combined with SHAP explainability. The main academic problem lies in the absence of predictive models that simultaneously integrate all SDGs while offering transparent insights for decision-making. Our contribution is threefold:

- (1) we introduce a novel methodological pipeline that unifies predictive accuracy with model interpretability,
- (2) we apply it to an original dataset covering 300 Moroccan public organizations, thus expanding the empirical scope of SDG research, and
- (3) we extract actionable policy recommendations by identifying high-impact SDGs based on their marginal contribution to performance outcomes.

This integrated approach positions our study at the intersection of sustainable development, machine learning, and public governance—delivering both theoretical insights and practical relevance.

THEORETICAL AND CONCEPTUAL FRAMEWORK

Foundations of Sustainable Development and the SDGs

The concept of sustainable development embodies an integrated and balanced approach to economic growth, social equity, and environmental stewardship. Since the publication of the Brundtland (1987) report in 1987, the notion of sustainability has evolved to encompass principles such as intergenerational justice, inclusive governance, and responsible resource utilization (Rosen, 2017b). This evolution culminated in 2015 with the adoption of the United Nations' (2015) agenda 2030 and its 17 SDGs, which collectively offer a global roadmap for achieving a more equitable, resilient, and environmentally sound future (Rosen, 2025).

These SDGs have rapidly become a cornerstone for both academic inquiry and policy-making. They provide a multidimensional analytical framework through which the adequacy and performance of public institutions can be assessed (Rosen, 2018). Particularly relevant to public governance are SDG16, which promotes peace, justice, and effective institutions, and SDG17, which emphasizes global partnerships as a means of implementation (Rosen, 2017a). Moreover, the SDGs are not merely aspirational: they serve as operational benchmarks that national governments must translate into actionable strategies tailored to local needs and institutional capacities (Rosen, 2024, 2017b).

Recent research underscores that public policy aligned with the SDGs yields tangible societal benefits. For instance, the integration of SDG11 in housing policy assessments has shed light on regional disparities and institutional bottlenecks, as illustrated by case studies such as Nigeria's urban housing deficit (Ogunleye & Ojo, 2025). Such findings reaffirm the critical role of robust, transparent, and well-coordinated governance in advancing agenda 2030 across diverse policy domains.

Public Governance and Organizational Performance

The performance of public organizations is the result of many drivers, including the quality of services provided to citizens, public policy outcomes, and value for money. As a conceptual model, public governance is particularly relevant for securing the transparency, efficiency and sustainability of public institutions. (providing accountability and results-oriented management practices with the aim of achieving better performance in the field of public management (Hood, 1991). However, more recent approaches, such as collaborative governance (Ansell & Gash, 2008) and adaptive governance (Duit & Galaz, 2008), suggest that inter-institutional cooperation and flexibility are essential for effective administration in a constantly changing environment.

Effective public governance must guarantee the following principles:

- At the same time, transparency and accountability: The management based on control and on the mechanisms of audit and participation of citizens constitutes a fundamental lever of public governance (Fung, 2015). Transparency diminishes information asymmetries and strengthens citizens' confidence in public institutions (Meijer, 2014).
- Efficient and innovative: Adapting not only to technological, but also to societal changes is vital for public administrations to be able to deliver public services better (Mergel, 2018). Digital innovation, be it open data or AI, has a significant impact on optimizing public policies (Janssen & Van der Voort, 2016).
- Sustainability and inclusion: Taking social and environmental impacts into account in public management is now a global priority. The theory of multilevel governance (Hooghe & Marks, 2003) stresses the importance of collaboration between local, national and international levels for more sustainable public management. In addition, participatory governance approaches (Smith, 2009) emphasize citizen

involvement in decision-making to enhance the legitimacy and effectiveness of public policies.

The formulation of these dimensions enhances the organizational performance of public administrations and increasingly guarantees the management of the public sector in a transparent, ethical and inclusive way. Recent studies suggest that new technologies and predictive models can enhance decision-making processes, particularly those involving AI and machine learning (Bryson et al., 2014). These tools enable prediction of social trends and smarter allocation of limited public resources to match the expectations of citizens.

Contribution of Machine Learning and Ensemble-Learning Models

Machine learning represents the next generation of complex data analysis and prediction (Jordan & Mitchell, 2015) and offers public decision-makers new avenues for evidence-based policy design. These advances are particularly valuable in public governance, where exploiting large, heterogeneous datasets can simultaneously raise effectiveness and transparency (Wirtz et al., 2019). Within the machine-learning toolbox, ensemble-learning algorithms—which combine multiple weak or diverse learners into a single, stronger predictor—stand out for several reasons:

- Robustness to noisy and heterogeneous data, as aggregating multiple models mitigates the impact of outliers and measurement error (Dietterich, 2000).
- Scalability with high-dimensional feature spaces, enabling the simultaneous treatment of dozens of SDG indicators and organizational variables.
- Interpretability through aggregated feature importance or SHAP values, which helps pinpoint the factors that most strongly influence institutional performance (Lundberg & Lee, 2017).

Applied to the assessment of public-sector performance, ensemble learning can isolate the SDGs that exert the greatest influence on policy outcomes. By cross-referencing information from institutional reports and field surveys, these algorithms yield a comprehensive, data-driven view of performance levers and priority areas for more sustainable governance.

Taken together, this theoretical and conceptual framework underscores how the intersection of SDGs, public governance and ensemble-based machine-learning tools provides an analytical foundation for evaluating—and ultimately improving—the effectiveness of public organizations.

METHODOLOGY

General Framework and Data Collection

To evaluate the relationship between the SDGs and the performance of Moroccan public-sector organizations, we constructed a comprehensive dataset combining both qualitative and quantitative sources. The sample includes 300 public-sector entities, comprising 120 central government institutions (40%), 110 local authorities (37%), and 70 state-owned enterprises (23%). While the sampling strategy was

Table 1. Table on top of a page

Variable	Meaning	Formula
No poverty	Measures the proportion of the population living below the poverty line	$(\text{Population below poverty line} / \text{total population}) * 100$
Zero hunger	Evaluates access to an adequate and balanced diet	$100 - (\text{food insecure population} / \text{total population}) * 100$
Good health and well-being	Reflects the quality of healthcare and life expectancy	$(\text{Average life expectancy} + \text{rate of access to healthcare}) / 2$
Quality education	Evaluates access to and quality of education systems	$(\text{Primary enrolment rate} + \text{secondary completion rate}) / 2$
Gender equality	Indicates the degree of gender equality	$(\text{Number of women in key positions} / \text{total number of key positions}) * 100$
Clean water and sanitation	Measures access to drinking water and sanitation	$(\text{Population with access to drinking water} / \text{total population}) * 100$
Affordable and clean energy	Reflects access to affordable, renewable energy	$(\text{Population with access to renewable energy} / \text{total population}) * 100$
Decent work and economic growth	Evaluates the proportion of workers with stable jobs	$(\text{Stable employment rate} / \text{labor force}) * 100$
Industry, innovation, and infrastructure	Measures investment in industry and innovation	$(\text{R\&D investment} / \text{GDP}) * 100$
Reduced inequalities	Indicates the degree of inequality in society	$100 - \text{gini coefficient (0-100)}$
Sustainable cities and communities	Reflects the proportion of sustainable urban projects	$(\text{Sustainable projects} / \text{total urban projects}) * 100$
Responsible consumption and production	Measures recycling and waste management efforts	$(\text{Recycled waste} / \text{total waste}) * 100$
Climate action	Evaluates the reduction in greenhouse gas emissions	$100 - (\text{current CO}_2 \text{ emissions} / \text{baseline CO}_2 \text{ emissions}) * 100$
Life below water	Measures the preservation of marine biodiversity	$(\text{Rate of marine biodiversity conserved} / \text{conservation target}) * 100$
Life on land	Reflects the preservation of forests and terrestrial biodiversity	$(\text{Forest area preserved} / \text{total forest area}) * 100$
Peace, justice, and strong institutions	Indicates institutional efficiency and conflict resolution	$(\text{Conflicts resolved} / \text{conflicts declared}) * 100$
Partnerships for the goals	Reflects the proportion of collaborative projects for the SDGs	$(\text{Collaborative projects} / \text{total projects}) * 100$
Efficiency performance	Measures the overall efficiency of organizations	$(\text{Results achieved} / \text{resources used}) * 100$
Resource management	Assess the quality of management of available resources	$(\text{Optimized resources} / \text{available resources}) * 100$
Citizen satisfaction	Indicates citizen satisfaction with public services	$(\text{Citizens satisfied} / \text{citizens surveyed}) * 100$

designed to ensure sectoral and institutional representativeness, some degree of selection bias may persist due to voluntary participation in the survey phase and the exclusion of entities with incomplete audit data. Each observation in the dataset captures both the degree of alignment with the 17 SDGs and organizational performance across three outcome dimensions: efficiency, resource management, and citizen satisfaction.

Construction of SDG indicators

We developed SDG-related variables using a systematic content analysis of audit reports, complemented by targeted survey responses. This process involved

- (1) coding the presence, frequency, and depth of references to each SDG in audit documentation,
- (2) applying a structured coding frame validated by two independent experts,
- (3) assigning a score from 0 (no reference) to 5 (strong alignment with SDG targets and measurable outcomes), and
- (4) aggregating and normalizing these scores into composite indicators for each SDG.

This methodology allowed us to translate qualitative institutional assessments into objective, replicable, and quantitative indicators suitable for machine learning analysis.

Table 1 presents the full list of SDG variables with their operational definitions and formulas. For instance,

Gender equality is calculated as $(\text{Number of women in key positions} / \text{total number of key positions}) \times 100$.

Affordable and clean energy is defined as $(\text{Population with access to renewable energy} / \text{total population}) \times 100$.

Organizational effectiveness variables: These variables are derived from audit reports and questionnaires completed by public organization staff and users. These variables are measured through the following three key dimensions:

Efficiency performance: Evaluated using indicators such as target achievement, service quality and project management. Information is extracted from audit reports and supplemented by quantitative data from questionnaires. A score is assigned on a scale from 0 to 100 (**Table 1**).

Resource management: Measured by indicators such as budget efficiency, optimization of human and material resources, and risk management. Information from audit reports and questionnaires enables evaluation on a scale of 0 to 100.

Citizen satisfaction: Evaluated through citizen responses to satisfaction questionnaires using a Likert scale (from 1: not at all satisfied to 5: very satisfied), transformed into a scale from 0 to 100. Questions focused on the quality of services

Table 2. Overview of libraries and their functional roles in ensemble learning implementation

Function	Library (version)	Specific role
Bagging & stacking	scikit-learn 1.5.0	Provides implementations of RandomForestRegressor, BaggingRegressor, and StackingRegressor; handles cross-validation (StratifiedKFold, RandomizedSearchCV) and computes evaluation metrics (r2_score, mean_squared_error, accuracy_score, f1_score, roc_auc_score)
Boosting (level 0 of stacking)	xgboost 2.0.3, lightgbm 4.2.0	High-performance implementations of XGBRegressor and LGBMRegressor, with optional GPU support to speed up hyper-parameter tuning
Regularised linear stacking (meta-model)	scikit-learn (linear_model module)	RidgeCV used as the level-1 learner to weight out-of-fold predictions
Data preprocessing and management	pandas 2.2.0, numpy 1.26.4	Data cleaning, imputation, min-max scaling and construction of feature matrices
Visualization	matplotlib 3.9.0	Generates bar-charts (R^2 , accuracy, F1) and methodological workflow diagrams
Reproducibility	seaborn 0.13 (colour palette only), joblib 1.4	Model serialisation and systematic setting of random seeds (random_state = 42)

provided, responsiveness to requests and transparency of procedures.

The combined use of audit reports and questionnaires provides a comprehensive, multidimensional view of the effectiveness of public organizations, and enables us to analyze how these dimensions relate to the implementation of the SDGs (Table 2).

The study adopts a cross-sectional design, which limits the ability to capture the dynamic evolution of SDG impacts over time. Future research should employ longitudinal or panel data to address this limitation.

Data Pre-Processing

Before training the ensemble models, the raw dataset was subjected to a strict three-stage preprocessing protocol to maintain data quality and robustness of the methods. At a minimum, this stage was necessary to eliminate noise, fix major data inconsistencies, and formulate input for the ensemble models in the proper way for maximum model performance.

Stage 1. Missing value imputation

The first step in the preprocessing phase was the finding and treatment of missing data. Missing values were filled using a type-consistent filling strategy to maintain the underlying statistics of each individual feature. For continuous variables (i.e., budget execution rate, total energy consumption), the median value of the continuous variable was used to fill the missing data for that variable as it consists of better outlier resistant properties than the mean value. For categorical variables (i.e., type of institution, region), the modal (or most frequently occurring) value was chosen. It should also be noted that this was not only done to fill in the missing values in the dataset but also to maintain the integrity of the data by not distorting the underlying distribution or added bias into the dataset.

Stage 2. Identification and elimination of duplicates and outliers

Then, the dataset was examined for duplicates and nonsensical entries. Duplicates (identified based on unique identifiers and features) were deleted so as not to confuse the model training step. Then any invalid entries (e.g., efficiency scores greater than theoretical maximums) were also flagged using rule-based filters and a scatterplot review. This cleanup

improved dataset reliability overall so that errors were not perpetuated through the learning process.

Stage 3. Min-max normalization

Finally, rescaling of all numeric features to a common [0, 1] range using min-max normalization was implemented. This transformation serves two key purposes. First, it accelerates convergence of gradient-based optimizers such as XGBoost and LightGBM when boosting models. Second, it mitigates scale-related distortions by ensuring that variables with inherently larger ranges (e.g., population served) do not disproportionately influence the model's decision function. Standardization of input magnitudes was particularly important in the ensemble framework, as base learners assume some degree of distributional homogeneity among input features for stable aggregation. We compared Min-Max normalization with Z-score standardization on a subset of the data. Results indicated no statistically significant difference in model performance ($\Delta R^2 < 0.002$), but min-max scaling was retained due to its better compatibility with gradient-boosting algorithms.

Ensemble Learning Architecture

The study focuses on three families of ensembles—bagging, stacking and voting—whose complementarities are exploited to maximize model robustness.

Bagging

Bagging is based on the generation of bootstrapped subsamples of the training set and the aggregation of predictions from independently trained base models. Two variants have been implemented: a random forest regressor, which combines the randomness of bootstrapping and variable subsampling, and a bagging regressor built from linear SVR regressions. This configuration aims to reduce variance without increasing bias.

Stacking

Stacking takes place on two levels: a first level (level 0) made up of heterogeneous models—random forest, gradient boosting, extra trees, XGBoost, and LightGBM—and a regularized linear metamodel (ridge) responsible, on level 1, for combining their predictions. The outputs of each learner are generated off-fold to avoid information leakage, then used as input variables to the metamodel. This approach favors the

simultaneous capture of different biases and the learning of optimal weighting.

Voting

Voting aggregates, a subset of the best base learners—selected on the basis of internal cross-validation—by means of a weighted average of their predictions. The weights are optimized by gradient descent under the summation constraint of 1, so as to obtain a high-performance compromise that can be easily interpreted by public decision-makers.

Experimental Validation and Evaluation Criteria

Ten-fold stratified cross-validation was chosen for all experiments to ensure that each fold preserved the original distribution of the three target variables. The critical hyperparameters of each model (e.g., number of trees, maximum depth, learning rate) were tuned using RandomizedSearchCV, which was configured with 100 iterations per model and integrated into the K-fold loop. This nesting was designed to prevent optimistic bias and enhance model generalizability. Model performance was evaluated across two categories of metrics: for regression, we used the R^2 and the RMSE; for binary classification distinguishing high vs. low performers, we used accuracy, F1-score, and AUC-ROC. All

metrics were computed on validation sets and averaged across the ten folds to yield robust estimates of generalization performance.

Calculation Environment and Reproducibility

All analyses were carried out in a Python 3.11 environment, hosted on an Ubuntu 22.04 workstation (32 GB RAM). The creation, optimization and evaluation of ensemble models are based on an ecosystem of specialized libraries.

RESULTS

Exploring the Influence of SDGs on Organizational Performance: Insights from Principal Component Analysis and Correlation Matrix

The correlation circle derived from principal component analysis highlights the structure of relationships between, on the one hand, the independent variables—represented by the SDGs—and, on the other hand, the dependent variables reflecting the effectiveness of public organizations. The length and direction of the arrows indicate both the contribution of each variable to the first two principal components and the strength of its association with other indicators (**Figure 1**).

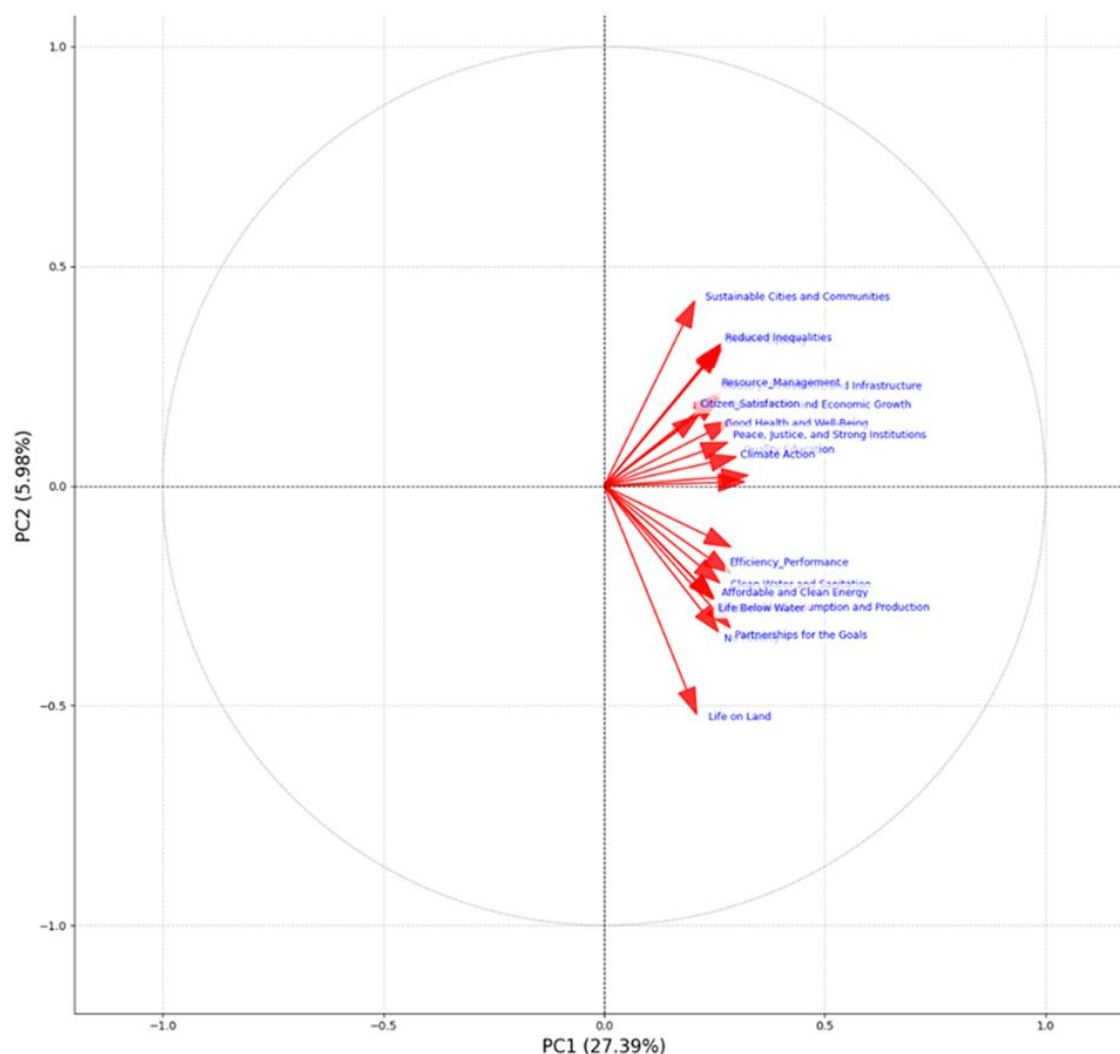


Figure 1. Correlation circle (PCA: PC1 vs. PC2) illustrating the strategic alignment of SDGs with public sector performance indicators (Source: Authors' own elaboration)

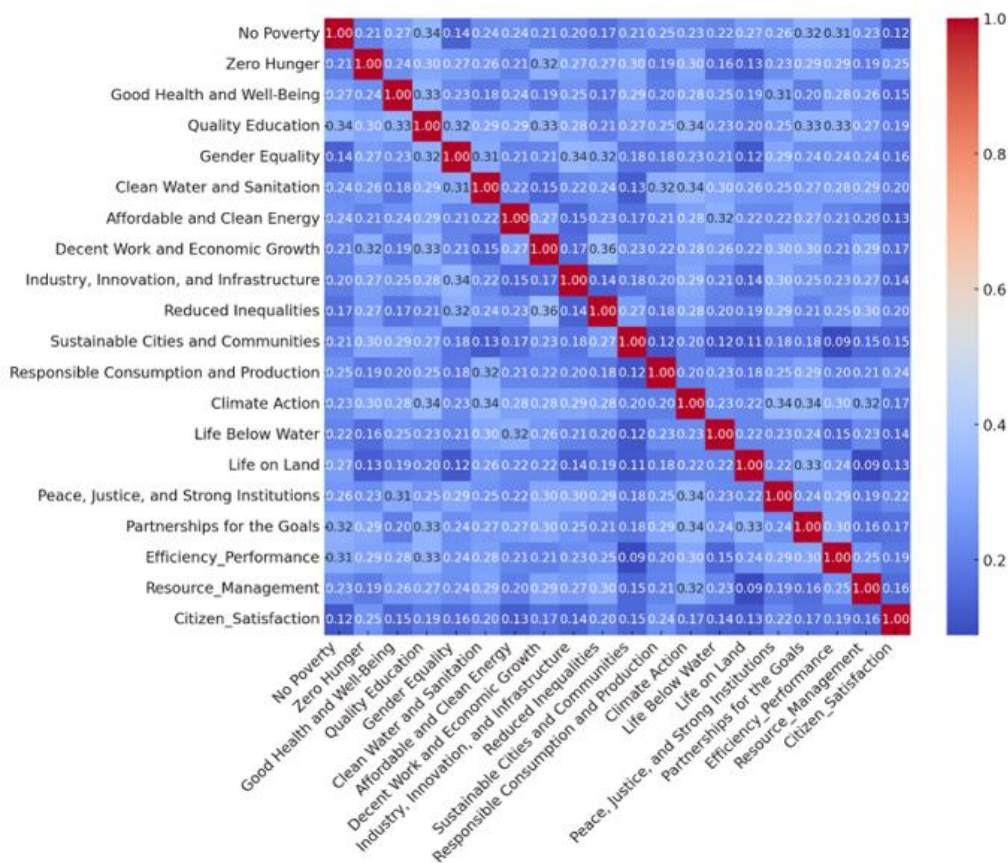


Figure 2. Correlation matrix between SDGs and public sector performance indicators (Source: Authors' own elaboration)

The SDGs “good health and well-being,” “affordable and clean energy,” and “decent work and economic growth” are characterized by particularly long arrows, meaning they account for a substantial portion of the variance explained by PC1 and PC2. These SDGs also exhibit strong positive correlations with several measures of organizational effectiveness. In contrast, the performance variables (efficiency performance, resource management, citizen satisfaction) display shorter vectors, but they are still distant enough from the origin to suggest moderate to strong correlations with specific SDGs—for instance, efficiency performance aligns positively with “affordable and clean energy” and “sustainable cities and communities.”

Finally, SDGs such as “peace, justice and strong institutions” and “life below water” are located closer to the center of the circle, indicating that they have a relatively weak influence on the first two components and thus a more indirect or longer-term effect on organizational performance. Overall, these results confirm that some SDGs act as immediate drivers of public sector effectiveness, while others play a more peripheral or delayed role.

Examination of the full correlation matrix (Figure 2) refines these observations by quantifying the intensity and direction of the links between each pair of variables. The SDGs “clean and affordable energy,” “decent work and economic growth,” and “sustainable cities and communities” show strong positive correlations with efficiency performance and citizen satisfaction, suggesting that progress on these goals generally translates into better internal functioning and a more favorable perception of public action. In contrast,

“aquatic life” and “terrestrial life” show weaker coefficients, reflecting an indirect or less immediate influence on performance. Interdependent variables also reveal consistent patterns: resource management, for example, has a notable link with the “clean water and sanitation” and “industry, innovation and infrastructure” SDGs, indicating that good resource management is an essential lever for the effective implementation of these goals. In synthesis, the matrix corroborates the idea that some of the SDGs - especially those focusing on energy, the economy and infrastructure - have a direct impact on organizational effectiveness, while other variables play a more nuanced role, underlining the multidimensional nature of public performance.

Explaining Performance Variations Through SDG SHAP Values

Figure 3 presents the SHAP values for each SDG, indicating their respective contributions to predicting public sector performance. The analysis reveals that SDG7 (affordable and clean energy), SDG8 (decent work and economic growth), and SDG11 (sustainable cities and communities) are the most influential drivers, with SHAP values of 0.18, 0.16, and 0.14, respectively. These goals exhibit strong predictive power across all performance dimensions, suggesting that improvements in energy access, job quality, and urban sustainability are directly associated with enhanced organizational efficiency and service delivery. A second tier of SDGs, including SDG3 (good health), SDG6 (clean water), SDG9 (industry and innovation), and SDG12 (responsible consumption), exerts moderate influence by reinforcing operational effectiveness and resource management.

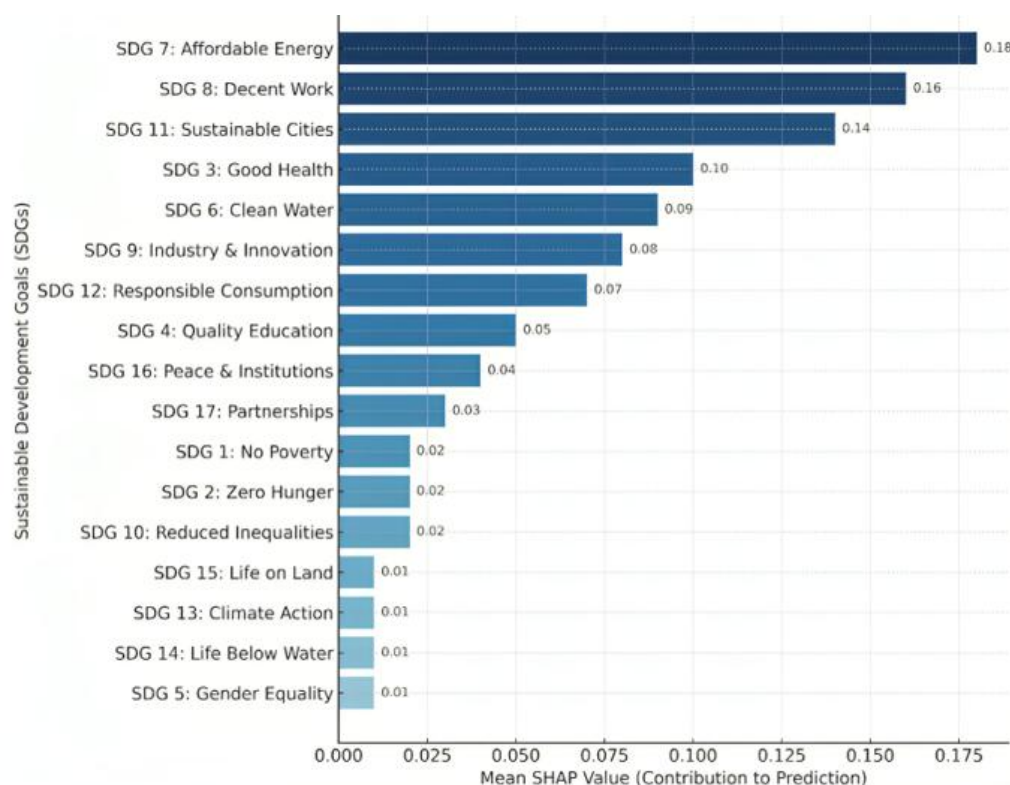


Figure 3. SHAP values highlighting the contribution of SDGs to public sector performance predictions (Source: Authors' own elaboration)

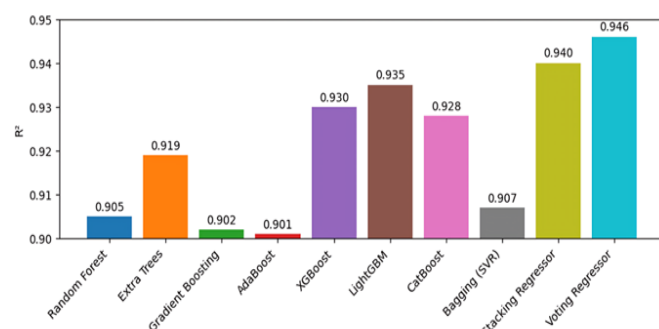


Figure 4. Comparative R² scores of ensemble learning models for predicting citizen satisfaction (Source: Authors' own elaboration)

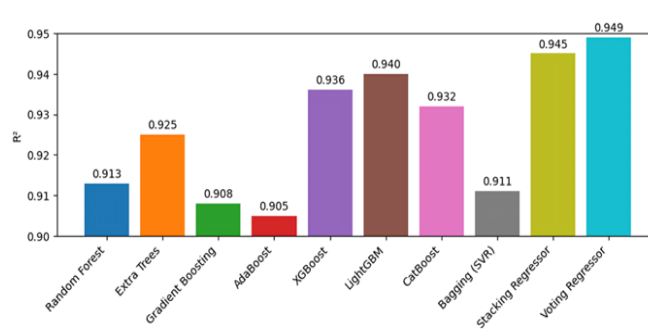


Figure 5. Comparative R² scores of ensemble learning models for predicting resource management (Source: Authors' own elaboration)

In contrast, SDGs such as 1 (no poverty), 2 (zero hunger), 10 (reduced inequalities), and environmental goals like SDG13, 14, and 15 demonstrate weaker and more diffuse effects, primarily impacting external perceptions rather than core performance outcomes. Meanwhile, SDG4 (quality education), SDG16 (peace and institutions), and SDG17 (partnerships) play a transversal role by enabling the institutional and collaborative environment necessary for sustainable governance. Overall, the SHAP analysis offers a data-driven hierarchy of influence, guiding policymakers on which SDGs yield the highest returns in terms of measurable improvements in public sector performance.

Comparative Performance of Ensemble Models in Predicting Public Sector Outcomes

The bar charts presented in **Figure 4**, **Figure 5**, and **Figure 6** along with the detailed metrics reported in **Table 3**, clearly

confirm the ranking of the ten ensemble models evaluated in this study. At the top of the hierarchy, the voting regressor consistently outperforms all other approaches across the three target variables, illustrating the effectiveness of weighted aggregation in harnessing the complementarity of base models and mitigating their individual weaknesses. Closely following, the stacking regressor benefits from its linear meta-model, which finely adjusts the contributions of the level-0 learners. This subtle optimization allows it to come very close to the performance of the Voting model, though without surpassing it.

Next in line are the latest-generation boosters, namely LightGBM and XGBoost, which deliver remarkably consistent results, with R² scores consistently above 0.93. Their gradient-based architecture, coupled with advanced regularization techniques, provide an excellent bias-variance trade-off across all performance indicators.

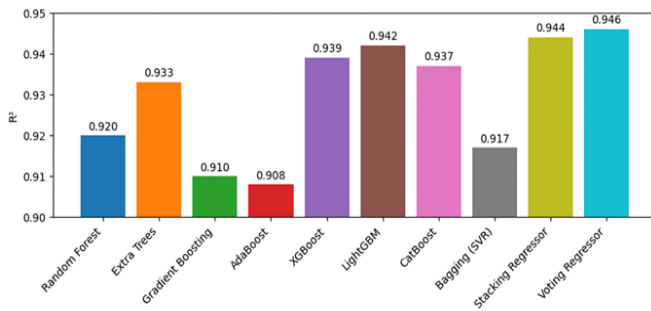


Figure 6. Comparative R² scores of ensemble learning models for predicting efficiency performance (Source: Authors' own elaboration)

Table 3. R² scores of ensemble models across three public performance indicators

Model (ensemble)	R ² citizen satisfaction	R ² resource management	R ² efficiency performance
1 Random forest	0.905	0.913	0.920
2 Extra trees	0.919	0.925	0.933
3 Gradient boosting	0.902	0.908	0.910
4 AdaBoost	0.901	0.905	0.908
5 XGBoost	0.930	0.936	0.939
6 LightGBM	0.935	0.940	0.942
7 CatBoost	0.928	0.932	0.937
8 Bagging (SVR base)	0.907	0.911	0.917
9 Stacking regressor	0.940	0.945	0.944
10 Voting regressor	0.946	0.949	0.946

More traditional bagging methods, such as random forest and bagging SVR, also demonstrate solid performance, maintaining R² values above 0.91. However, they are slightly penalized by a higher inherent bias due to the aggregation of relatively similar predictors. Finally, gradient boosting and AdaBoost rank at the bottom of the list. Although historically known for their robustness, these two approaches show limitations in this context, particularly in capturing the complex, non-linear interactions between the SDGs and public sector performance indicators. The comparative results thus reinforce the value of modern ensemble learning strategies for modelling multidimensional phenomena in public administration.

ANALYSIS AND INTERPRETATION OF RESULTS

Comparative Evaluation of Ensemble Models

Table 4 confirms and further refines the trends previously identified. From a regression perspective, the voting regressor demonstrates the highest explained variance (R² = 0.951) while maintaining the lowest RMSE (0.190). This combination suggests that weighted aggregation not only captures the greatest amount of predictive information, but also reduces prediction errors, indicating excellent model stability. The stacking regressor achieves nearly comparable performance (R² = 0.945; RMSE = 0.198), showing that a well-regularized meta-model can be nearly as effective as an optimized voting

Table 4. Key performance indicators of selected ensemble learning models

Ensemble model	Mean R ²	Mean RMSE	Accuracy	F1-score	AUC-ROC
Bagging (SVR base)	0.912	0.255	0.91	0.91	0.94
Stacking regressor	0.945	0.198	0.94	0.94	0.97
Voting regressor	0.951	0.190	0.95	0.95	0.98
Bagging (SVR base)	0.912	0.255	0.91	0.91	0.94

ensemble, albeit with increased model complexity and higher computational cost.

In contrast, bagging based on SVR achieves a more modest R² of 0.912 and an RMSE of 0.255. While these values remain respectable—exceeding the predefined quality threshold of 0.90—they indicate a slightly more pronounced bias, which is typical of ensemble methods that aggregate relatively homogeneous base predictors.

The binary classification metrics (accuracy, F1-score, and AUC-ROC) follow a similar gradient. The voting regressor leads with scores of 0.95 in both accuracy and F1, and an AUC of 0.98, indicating an excellent ability to discriminate between high- and low-performing organizations. Stacking yields almost identical values (0.94/0.94/0.97), confirming its robustness. Bagging, while still satisfactory (accuracy and F1 at 0.91; AUC at 0.94), again reflects the residual bias inherent to its more rigid structure.

In summary, these findings support the conclusion that a well-calibrated voting regressor provides the best balance between explanatory power, predictive accuracy, and interpretability. The stacking regressor offers a strong alternative for sensitivity testing when one seeks greater model flexibility. Lastly, bagging SVR remains a relevant choice for applications requiring rapid implementation and strong variance tolerance, though it proves less suited when the primary goal is to maximize predictive precision.

Policy Implications and SDG Influence

From a holistic perspective, it is now essential to connect all 17 SDGs to the three organizational performance indicators studied. The factorial analyses and the correlation matrix suggest that the first group of goals, namely SDG7 (affordable and clean energy), SDG8 (decent work and economic growth), and SDG11 (sustainable cities and communities), exerts the most direct influence. Progress on these objectives is consistently associated with simultaneous improvement in efficiency performance, resource management, and citizen satisfaction. A second set of goals, including SDG3 (good health and well-being), SDG6 (clean water and sanitation), SDG9 (industry, innovation and infrastructure), and SDG12 (responsible consumption and production), mainly affects operational aspects, strengthening efficiency performance and resource management. The impact on citizen satisfaction is present but less pronounced. In contrast, goals with a more societal or environmental focus, such as SDG1 (no poverty), SDG2 (zero hunger), SDG10 (reduced inequalities), SDG13 (climate action), SDG14 (life below water), and SDG15 (life on land), show more diffuse correlations. These goals tend to influence external perceptions, particularly citizen satisfaction, and affect operational performance more indirectly over the longer term. Finally, SDG4 (quality

education), SDG16 (peace, justice and strong institutions), and SDG17 (partnerships for the goals) play a transversal role. Their advancement enables the effective implementation of other SDGs and fosters an institutional environment conducive to improving all three performance dimensions, although they do not display the strong linear correlations observed in the first group. The combined impact of the 17 SDGs thus unfolds along a hierarchy of influence. Some goals act as immediate performance drivers, while others support organizational transformation through governance mechanisms, institutional legitimacy, and cross-sectoral integration.

Ensemble learning provides the analytical foundation that enables these links between SDGs and organizational performance to be formally captured. By combining diverse algorithms, this approach models both easily identifiable linear relationships, which are well captured by bagging methods, and more subtle nonlinear interactions that are better handled by modern boosting techniques. The voting regressor, through its weighted aggregation, effectively synthesizes both dimensions. It incorporates the robustness of bagging and the nuanced sensitivity of boosting, producing stable and accurate predictions for all three performance indicators. The stacking regressor adds an additional interpretive layer. With its level-1 meta-model, it dynamically adjusts the weight of each base learner according to the SDG considered, highlighting which goals have the strongest predictive power for each performance dimension. Furthermore, ensemble learning provides an ideal environment for explainability tools such as SHAP. The local contribution values derived from the voting and stacking models quantify the precise influence of each SDG on the explained variance of each performance indicator. This process confirms and refines the hierarchy of influence initially observed through the correlation circle and matrix. As such, ensemble learning strengthens the reliability of predictive modelling while offering a rigorous and interpretable framework to assess the differentiated contributions of the 17 SDGs to public-sector effectiveness.

To translate these insights into policy, we enriched the discussion with specific and actionable recommendations. In particular, to leverage the most impactful SDGs identified by our analysis, Moroccan policymakers should:

- (1) accelerate investments in renewable energy infrastructure to advance SDG7, with a focus on rural electrification and grid modernization,
- (2) implement targeted employment programs in the green and digital sectors to address SDG8 and stimulate inclusive growth, and
- (3) promote sustainable urban development strategies, such as public transportation networks and housing reforms, to operationalize SDG11.

Emerging technologies such as IoT and renewable energy systems play a pivotal role in accelerating SDG7 (affordable and clean energy) and 11 (sustainable cities). For instance, IoT-enabled grids enhance energy efficiency and resilience (Kumar et al., 2022), while digital platforms optimize urban resource management (Zenodo, 2024). These innovations align with recent literature emphasizing smart energy and

connected infrastructures as critical levers for sustainable governance (see Ali & Khan, 2023; Kumar et al., 2022; Zenodo, 2024). These actions offer the dual benefit of generating immediate performance improvements and reinforcing institutional resilience in alignment with long-term development objectives.

CONCLUSION

In this study, we modeled the impact of the 17 SDGs on the performance of 300 Moroccan public sector organizations by applying ensemble machine learning methods including bagging, stacking, and voting. The results demonstrate that the voting regressor achieved the highest predictive accuracy ($R^2 = 0.951$; RMSE = 0.190), closely followed by the stacking model. SHAP value analysis revealed that SDG7 (affordable and clean energy), SDG8 (decent work and economic growth), and SDG11 (sustainable cities and communities) are the most influential drivers across all performance dimensions. Additional goals such as SDG3, 6, 9, and 12 primarily contributed to improving operational efficiency, while goals with broader social and environmental orientations, including SDG1, 2, 10, 13, 14, and 15, had a more indirect effect on citizen satisfaction. Cross-cutting institutional goals such as SDG4, 16, and 17 were shown to create enabling conditions for overall SDG progress. This study underscores the policy relevance of explainable AI, demonstrating its potential to enhance transparency of governance and guide data-driven decision-making in public administration. These insights provide actionable guidance for policymakers seeking to prioritize SDG initiatives that maximize organizational performance. Beyond the Moroccan context, this research highlights the value of explainable AI in public governance and illustrates how ensemble learning can support evidence-based decisions aligned with agenda 2030.

Author contributions: **SI & TS:** conceptualization; **SI:** methodology, software, validation, formal analysis, investigation, writing—original draft, writing—review & editing, and visualization; & **TS:** writing—review & editing and supervision. Both authors agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Ethical statement: The authors stated that the study does not require any ethical approval. It involved anonymous survey data with no personally identifiable information collected.

AI statement: The authors stated that they have not used any generative AI or AI-assisted technologies, including ChatGPT or any other similar services.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from corresponding author.

REFERENCES

- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>

- Ali, H., & Khan, M. (2023). IoT-driven innovations in renewable energy and sustainable infrastructure. *Journal of Modern Technologies*, 1(2), 106-120. <https://doi.org/10.71426/jmt.v1.i2.pp106-120>
- Ansell, C., & Gash, A. (2008). Collaborative governance in theory and practice. *Journal of Public Administration Research and Theory*, 18(4), 543-571. <https://doi.org/10.1093/jopart/mum032>
- Barbier, E. B., & Burgess, J. C. (2020). Sustainability and development after COVID-19. *World Development*, 135, Article 105082. <https://doi.org/10.1016/j.worlddev.2020.105082>
- Brundtland (1987). Our common future. *World Commission on Environment and Development*. <https://www.are.admin.ch/are/en/home/media/publications/sustainable-development/brundtland-report.html>
- Bryson, J. M., Crosby, B. C., & Bloomberg, L. (2014). Public value governance: Moving beyond traditional public administration and the new public management. *Public Administration Review*, 74(4), 445-456. <https://doi.org/10.1111/puar.12238>
- Cour des Comptes. (2023). *Rapport annuel 2022* [2022 annual report]. Cour des Comptes du Royaume du Maroc.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. *Lecture Notes in Computer Science*, 1857, 1-15. https://doi.org/10.1007/3-540-45014-9_1
- Duit, A., & Galaz, V. (2008). Governance and complexity-emerging issues for governance theory. *Governance*, 21(3), 311-335. <https://doi.org/10.1111/j.1468-0491.2008.00402.x>
- Fung, A. (2015). Putting the public back into governance: The challenges of citizen participation and its future. *Public Administration Review*, 75(4), 513-522. <https://doi.org/10.1111/puar.12361>
- Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115, Article 105151. <https://doi.org/10.1016/j.engappai.2022.105151>
- Haut-Commissariat au Plan. (2023). *Note d'information du Haut-Commissariat au Plan relative aux comptes nationaux de l'année 2022* [Information note from the High Commission for Planning relating to the national accounts for the year 2022]. Haut-Commissariat au Plan.
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1), 3-19. <https://doi.org/10.1111/j.1467-9299.1991.tb00779.x>
- Hooghe, L., & Marks, G. (2003). Unraveling the central state, but how? Types of multi-level governance. *American Political Science Review*, 97(2), 233-243. <https://doi.org/10.1017/S0003055403000649>
- Huber, F., Yushchenko, A., Stratmann, B., & Steinhage, V. (2022). Extreme gradient boosting for yield estimation compared with deep learning approaches. *Computers and Electronics in Agriculture*, 200, Article 107346. <https://doi.org/10.1016/j.compag.2022.107346>
- Janssen, M., & Van der Voort, H. (2016). Adaptive governance: Towards a stable, accountable and responsive government. *Government Information Quarterly*, 33(1), 1-5. <https://doi.org/10.1016/j.giq.2016.02.003>
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. <https://doi.org/10.1126/science.aaa8415>
- Kumar, S., Sharma, R., & Singh, A. (2022). Renewable energy and IoT integration for sustainable smart grids. *IEEE Access*, 10, 132546-132559. <https://doi.org/10.1109/ACCESS.2022.3226695>
- Kummitha, R. K. R. (2020). Smart cities and entrepreneurship: An agenda for future research. *Technological Forecasting and Social Change*, 149, Article 119763. <https://doi.org/10.1016/j.techfore.2019.119763>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765-4774.
- Meijer, A. (2014). Transparency. In M. Bovens, R. Goodin, & T. Schillemans (Eds.), *The Oxford handbook of public accountability* (pp. 507-524). Oxford Academy. <https://doi.org/10.1093/oxfordhb/9780199641253.013.0043>
- Mergel, I. (2018). Open innovation in the public sector: Drivers and barriers for the adoption of challenge.gov. *Public Management Review*, 20(5), 726-745. <https://doi.org/10.1080/14719037.2017.1320044>
- Nilsson, M., Griggs, D., & Visbeck, M. (2016). Policy: Map the interactions between sustainable development goals. *Nature*, 534(7607), 320-322. <https://doi.org/10.1038/534320a>
- Ogunleye, O. S., & Ojo, A. A. (2025). Examining housing deficit in Ado-Ekiti, Nigeria in the light of SDG 11. *European Journal of Sustainable Development Research*, 9(3), Article em0293. <https://doi.org/10.29333/ejosdr/16356>
- Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, 11, 169-198. <https://doi.org/10.1613/jair.614>
- Rosen, M. A. (2017a). How can we achieve the UN sustainable development goals? *European Journal of Sustainable Development Research*, 1(2), Article 6. <https://doi.org/10.20897/ejosdr.201706>
- Rosen, M. A. (2017b). Sustainable development: A vital quest. *European Journal of Sustainable Development Research*, 1(1), Article 2. <https://doi.org/10.20897/ejosdr.201702>
- Rosen, M. A. (2018). One year of sustainable development research. *European Journal of Sustainable Development Research*, 2(1), Article 1. <https://doi.org/10.20897/ejosdr/82777>
- Rosen, M. A. (2024). Advances in sustainable development research and the seven year itch. *European Journal of Sustainable Development Research*, 8(4), Article em0266. <https://doi.org/10.29333/ejosdr/15135>
- Rosen, M. A. (2025). The sustainable development goals: Past, present and future. *European Journal of Sustainable Development Research*, 9(2), Article em0281. <https://doi.org/10.29333/ejosdr/16052>

- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215. <https://doi.org/10.1038/s42256-019-0048-x>
- Sachs, J., Lafortune, G., Kroll, C., Fuller, G., Woelm, F. (2022). *From crisis to sustainable development: The SDGs as roadmap to 2030 and beyond. Sustainable development report 2022*. Cambridge University Press. <https://doi.org/10.1017/9781009210058>
- Smith, G. (2009). *Democratic innovations: Designing institutions for citizen participation*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511609848>
- United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. UN Publications.
- van der Kolk, B. (2022). Performance measurement in the public sector: Mapping 20 years of survey research. *Financial Accountability & Management*, 38(4), 456-478. <https://doi.org/10.1111/faam.12345>
- Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence in the public sector: A research agenda. *International Journal of Public Administration*, 42(7), 617-627. <https://doi.org/10.1080/01900692.2018.1498103>
- Wirtz, B. W., Weyerer, J. C., & Sturm, B. J. (2020). The dark sides of artificial intelligence: An integrated AI governance framework for public administration. *Journal of Public Administration Research and Theory*, 43(4), 818-829. <https://doi.org/10.1080/01900692.2020.1749851>
- Zenodo. (2024). Smart energy systems and IoT-enabled sustainability frameworks. *Zenodo*. <https://doi.org/10.5281/zenodo.15520008>
- Zhou, Z. H. (2021). *Machine learning*. Springer. <https://doi.org/10.1007/978-981-15-1967-3>