Prepublished Copy

NATURE ENVIRONMENT AND POLLUTION TECHNOLOGY

Article In Press

PEER-REVIEWED PREPUBLISHED MANUSCRIPT

This is a peer-reviewed prepublished version of the paper to be published in **Vol. 26, No. 1, March 2027.**

Illuminating India's Diverse Renewable Energy Transition: Evidence from Explainable Artificial Intelligence

Ravi Sharma, Anudeep Arora and Sarvpreet Singh

RECEIVED
5 August 2026

CITE THIS ARTICLE
Sharma, R., Arora, A. and Singh, S., 2027. Illuminating India's Diverse Renewable Energy Transition: Evidence from Explainable Artificial Intelligence. *Nature Environment and Pollution Technology*, 26(1), B4467.

https://doi.org/10.46488/NEPT.2027.v26i01.B4467

EARLY-ACCESS STATEMENT

This manuscript, after peer review, is being released online in its original, unedited form to provide early access to its findings. Prior to final publication, it will undergo additional editing and formatting. Readers may note that, as an unedited version, the manuscript may have linguistic and other minor errors. The DOI will be active after the final publication of the paper.

PUBLICATION STATUS

PEER REVIEW	ONLINE VERSION	FINAL ISSUE
Completed	**Prepublished**	**March 2027**

© The Author(s) 2027. This article is Open Access under the Creative Commons Attribution 4.0 International License, allowing use, sharing, adaptation and reproduction with proper credit. Images or third-party material are covered by this licence unless otherwise noted. For uses beyond the licence, permission must be obtained from the copyright holder. Licence details: creativecommons.org/licenses/by/4.0/.

Type of the Paper: Original Research

Illuminating India's Diverse Renewable Energy Transition: Evidence from Explainable Artificial Intelligence

Ravi Sharma[1]†, Anudeep Arora[2] and Sarvpreet Singh[3]

[1]Symbiosis Institute of International Business, Symbiosis International (Deemed University), Pune, Maharashtra, India; ravi.sharma@siib.ac.in; ORCID ID: 0000-0003-0819-4340

[2]Symbiosis Institute of International Business, Symbiosis International (Deemed University), Pune, Maharashtra, India; anudeep.arora@siib.ac.in; ORCID ID: 0000-0002-4128-9977

[3]Symbiosis Institute of International Business, Symbiosis International (Deemed University), Pune, India; sarvpreet.singh2025@siib.ac.in; ORCID ID: 0009-0003-1417-4685

†Corresponding author: Ravi Sharma; ravi.sharma@siib.ac.in

Abstract

India's shift to renewable energy is not one national trajectory but a fragmented subnational process, with adoption diverging sharply across states. This study combines K-means trajectory clustering, panel machine-learning forecasting and SHAP-based explainable AI to examine renewable-energy (RE) capacity growth across Indian states from 2014 to 2024, using 319 state-year observations for 29 states from open government sources (MNRE, RBI, NITI Aayog). Delhi, unscored in the SDG India Index editions used, is excluded from the predictive analysis (28 states, 280 observations). Four models are evaluated under strictly chronological, expanding-window cross-validation against a benchmark ladder anchored on a naive persistence forecast. Persistence alone reaches R^2 = 0.991; adding lagged capacity, real state GDP, governance quality and a time trend lifts a linear (OLS) specification to R^2 = 0.995, which no ensemble method (Random Forest, XGBoost, LightGBM) improves upon. SHAP attributes most predictive weight to prior installed capacity (mean |SHAP| = 1.795), consistent with strong path dependency; real GSDP is a distant second (0.168) and governance quality smaller still (0.038). These measures model behaviour, not causal effects. K-means clustering (silhouette = 0.478) separates four trajectory groups (Innovators, Early Adopters, Early Majority, Laggards), read through Rogers' diffusion typology rather than as a validation of it. The findings remain stable when excluding COVID-affected years and

when using a GDP-normalised alternative dependent variable. The paper contributes a reproducible, open-data pipeline for subnational energy-transition diagnostics, an explicit prediction–explanation–causation distinction for policy use of SHAP evidence, and cluster-specific findings relevant to India's 2030 renewable-capacity targets.

Keywords

Renewable energy; State-level adoption; Panel machine learning; Explainable AI; K-means clustering; TOE framework; Diffusion of innovations

1. Introduction

Subnational governance has become indispensable to the global clean energy transition (REN21, 2024). India's Nationally Determined Contribution (NDC), updated in 2022, targets 500 gigawatts (GW) of non-fossil electricity generation capacity by 2030 and net-zero carbon emissions by 2070 (Government of India, 2022). By March 2023, India's installed renewable capacity exceeded 175 GW, one of the five largest markets in the world (IRENA, 2023; IEA, 2023). That national figure is not directly comparable with this study's state-wise panel (Section 3.4), whose 29-state total is smaller because it rests on a single, consistently defined MNRE bulletin series. Capacity addition is also uneven across states, and this has not been systematically modelled or explained.

India's federal energy governance framework gives state governments significant powers to plan renewable projects, acquire land, connect renewable sources to the grid and manage the finances of distribution companies (DISCOMs). Rajasthan (26,692.89 MW) and Gujarat (25,471.72 MW) together hold over a third of the 29-state total as of 2024, while 19 of the 29 states have less than 2,000 MW installed. The disparity reflects geographic resource endowments, but equally differences in state fiscal capacity, institutional quality, industrial energy demand and the pace at which states have operationalised national renewable energy policy.

Credible state-level capacity projections underpin plans for infrastructure investment, grid expansion, storage deployment and private developer investment. Yet energy forecasting for India has relied mainly on autoregressive integrated moving average (ARIMA) models or linear extrapolation (Voyant et al., 2017), which assume structural stationarity and fit poorly with the non-linear, policy-shaped adoption dynamics of India's states. Ensemble algorithms such as Random Forest (Breiman, 2001), XGBoost (Chen and Guestrin, 2016) and LightGBM (Ke et al., 2017) have been shown to predict better in high-dimensional, non-stationary regression problems, but whether they add genuine forecasting value

over a simple persistence baseline for subnational renewable capacity in a large developing economy has not been tested.

Interpretability is a second gap. Forecasts of renewable capacity have limited policy relevance unless they can be related to structural drivers: state planners, central agencies and multilateral financiers need to know not only which states will miss the 2030 targets, but which factors are associated with each state's trajectory. SHAP (SHapley Additive Explanations; Lundberg and Lee, 2017), among the more theoretically grounded interpretability methods (Linardatos et al., 2021), attributes a model's predictions to its inputs. It explains which inputs a fitted model relies on, not which factors causally determine the outcome, a distinction developed in Section 3.8.

This study integrates K-means trajectory clustering, panel machine-learning forecasting and SHAP-based XAI in a framework grounded in the Technology-Organisation-Environment (TOE) framework (Tornatzky and Fleischer, 1990) and Diffusion of Innovations theory (Rogers, 2003). Three research questions organise the analysis. **RQ1**: How can the temporal and spatial heterogeneity of state-level renewable energy adoption in India (2014–2024) be characterised, and what clusters of states with similar trajectories emerge? **RQ2**: Under strictly chronological, expanding-window panel cross-validation, how much predictive value do lagged capacity, economic and governance predictors add over a naive persistence forecast, and does any ensemble model improve on a linear benchmark? **RQ3:** Which structural, economic and governance predictors are most strongly associated with the best-performing model's predictions, overall and within each adoption cluster, as revealed by SHAP analysis?

The study makes four contributions. First, it provides an open-data, chronologically validated ML pipeline for subnational energy forecasting that can be replicated without proprietary data. Second, it offers a state-level SHAP decomposition of renewable-adoption heterogeneity in India that explicitly separates predictive attribution from causal claims. Third, it derives an empirically grounded cluster taxonomy of state adoption trajectories for differentiated policy design. Fourth, it shows through a worked illustration (Sections 3.6–3.8) how look-ahead bias can enter panel ML pipelines through derived clustering features, and how to guard against it.

Section 2 reviews the literature; Section 3 sets out the theoretical framework, data, methodology and validation procedure; Section 4 presents the results; Section 5 discusses theoretical, practical and managerial implications; and Section 6 concludes.

2. Review of Literature

2.1. Renewable Energy Adoption: Policy and Economic Drivers

Research on the determinants of RE adoption has grown along several streams. At the country level, the policy instruments most consistently associated with faster deployment are feed-in tariffs, renewable portfolio standards, carbon pricing and technology subsidies (IPCC, 2011). Using panel cointegration, Sadorsky (2009) and later Apergis and Payne (2010) report a positive association between income and renewable energy consumption in emerging economies, usually explained by higher-income states' greater capacity to absorb the upfront capital cost of renewable infrastructure while keeping the grid reliable for industrial consumers.

India-specific studies concentrate on the National Solar Mission (2010), renewable purchase obligations and the financial viability of DISCOMs. Dubash (2011) documents India's shift in energy-supply incentive design from a subsidy-based to a market-based approach. Official reporting confirms wide variation in state renewable-sector performance, associated with DISCOM financial health, land-availability policy and grid-infrastructure quality (NITI Aayog, 2024). Qualitative and single-source state case studies exist, but quantitative modelling of subnational heterogeneity, and rigorous validation of the resulting forecasts, remain underdeveloped.

2.2. Machine Learning in Energy Systems Forecasting

Machine learning is now routine in energy systems research. Voyant et al. (2017) find that ensemble methods outperform statistical time-series models for solar radiation prediction across diverse temporal and spatial conditions, though, as Bergmeir and Benítez (2012) and Roberts et al. (2017) caution, such comparisons are only as credible as the temporal validation design behind them. Random Forest (Breiman, 2001) and XGBoost (Chen and Guestrin, 2016) are standard benchmarks in energy regression for their handling of high dimensionality, missing data and non-linearity; LightGBM (Ke et al., 2017) achieves comparable accuracy at markedly lower computational cost, which suits moderate-sized panels.

ML forecasting of renewable capacity at the subnational level, particularly in developing economies, remains limited. Most ML energy-forecasting studies address generation output, load demand or prices at the system or plant level (Hayati et al., 2026; Tang, 2026; Haggag et al., 2026). State-level installed-capacity forecasting, which must integrate economic, policy and geographic predictors across a heterogeneous panel under a validation design that respects its

temporal structure, has received little attention in India, where applications remain largely single-source and single-state, such as solar irradiance modelling in Rajasthan (Tandon et al., 2025).

2.3. Explainable AI in Energy and Sustainability Research

XAI remains an emerging area in energy and sustainability research. Building on Shapley values, Lundberg and Lee (2017) formalised SHAP as an additive feature-attribution method satisfying local accuracy, missingness and consistency. Linardatos et al. (2021), reviewing interpretability methods, find SHAP theoretically sound and broadly implementable, but as an account of the patterns a trained model has learned rather than a test of real-world causal mechanisms. The distinction matters for policy: an attribution that cannot be tied to an actionable lever is of limited use to a planner, and one mistaken for a causal effect can misdirect intervention. Recent applications range from correlates of citizen engagement in renewable energy projects (Taghikhah et al., 2026) to wind-speed prediction for resource assessment (Rahat et al., 2025).

Three gaps motivate this study: (a) whether ensemble ML forecasts of state-level renewable capacity in India add genuine value over a chronologically validated persistence baseline has not been tested; (b) no state-level SHAP decomposition of adoption correlates with an explicit prediction/explanation/causation boundary has been attempted; and (c) the literature lacks a data-driven typology of Indian states by adoption trajectory for differentiated policy design.

3. Materials and Methods

3.1. Technology-Organisation-Environment (TOE) Framework

Tornatzky and Fleischer (1990) divide the context of organisational technology adoption into three dimensions: technological (properties of the technology and available technical infrastructure), organisational (the resources, capacity and governance arrangements of the adopter) and environmental (the external regulatory, policy and market setting). The adopting entity here is the state government, which plans renewable projects, allocates land, negotiates grid connections and bears fiscal responsibility for DISCOM finances (Section 1). Mapped onto it, prior installed capacity and grid-integration experience represent the technological context; the size and real growth of the state economy (GSDP) represent organisational capacity to finance and administer adoption; and the governance and institutional-quality signal in the NITI Aayog SDG score represents the environmental context. The mapping is a simplification (GSDP could equally be read as an environmental, market-size variable), but it is the most defensible one the data support, and it serves only to

classify predictors already justified individually (path dependency, economic capacity, governance quality), not as independent evidence.

3.2. Diffusion of Innovations Theory

Rogers (2003) describes diffusion of innovations as the spread of a new technology through social and institutional channels over time, at a rate shaped by five attributes: relative advantage, compatibility, complexity, trialability and observability. The strong predictive weight on prior installed capacity (Section 4.4) is consistent with, though it does not establish, the observability and trialability mechanisms: states with early, visible deployment plausibly build reputational and infrastructural templates that others can observe and trial. This is an interpretive frame for a strong empirical regularity, not a causal test.

The four K-means trajectory groups (Innovators, Early Adopters, Early Majority and Laggards; Section 3.5) are named after, and interpreted through, Rogers' adopter categories rather than presented as a quantitative validation of them. The groups are built purely from each state's own RE-capacity trajectory (starting level, growth, acceleration) with no reference to the theory; the labels are applied afterwards because the resulting profiles resemble the pattern it describes. Cluster-level differences in predictors (Section 4.4) are consistent with that reading but do not independently confirm it.

3.3. Explainable AI as an Interpretive Principle

SHAP-based XAI is chosen on technical and normative grounds. Lundberg and Lee (2017) formalise Shapley values as locally faithful, globally coherent feature attributions: a feature's SHAP value is its average marginal contribution to a prediction across all possible orderings of the other features. In this policy context, interpretability is a substantive requirement. State governments allocating infrastructure investment toward 2030 targets need to know which model inputs do the predictive work, which, as Section 3.8 sets out, is a narrower claim than knowing which structural factors causally constrain or enable a state's renewable path. SHAP is used at full-panel and cluster level for the former purpose only.

3.4. Data Sources and Collection

All data are public Indian government sources requiring no subscription, data request or licence. State-wise cumulative RE installed capacity (solar, wind, small hydro, bio-power) for 2014–2024 is extracted with tabula-py from 13 annual MNRE 'State-Wise Cumulative Installed Capacity' bulletins (MNRE, 2024) and checked manually against the source PDFs. This check surfaced two issues, both resolved before analysis: (i) 10 of 11 years of Chhattisgarh's series had

been dropped because the PDFs spell the name inconsistently ('Chhatisgarh' in most years, 'Chhattisgarh' from 2024), and standardising the spelling recovers the full series; (ii) West Bengal does not appear under any spelling in the bulletins as extracted and cannot be recovered without re-parsing the original tables, so it is excluded (Section 6). Net state domestic product at current and constant prices is parsed from three worksheets of the RBI Handbook of Statistics on Indian Economy (RBI, 2025); a header-matching rule keyed to state-name length had silently excluded Goa (a three-character name) and was widened to include it. Governance quality is proxied by the NITI Aayog SDG India Index composite score, usable at state level in three of four editions (2018, 2019–20, 2020–21); the 2023–24 edition is not. Delhi receives no composite score in any usable edition, a feature of the source rather than an extraction artefact, so it is excluded from the predictive/SHAP panel (Section 3.6) but kept in the descriptive clustering (Section 3.5), which needs only capacity data. Section 3.4.1 sets out the interpolation. Processing uses Python 3 (pandas, tabula-py, openpyxl, scikit-learn); figures were produced with Seaborn 0.13.2 and Matplotlib 3.10.9 in Python 3.11.

3.4.1. SDG Score Interpolation Procedure

Scores from the three usable editions are assigned to the terminal year of each reference period: 2018 → 2018, 2019-20 → 2020 and 2020-21 → 2021. Within 2014–2024: (a) 2019, the only year strictly between two anchors, is linearly interpolated; (b) 2014–2017 take the 2018 value (constant carry-back), since no earlier edition exists; and (c) 2022–2024 take the 2021 value (constant carry-forward), because the 2023–24 edition could not be incorporated at the state level. The SDG feature for 2022–2024 is therefore fixed at its observed 2021 value and cannot carry information from those years into the cross-validation folds that test on them (Section 3.6). The interpolation choice hardest to justify on data-availability grounds is, for that same reason, the one most immune to look-ahead bias.

Table 1: Variable Description and Sources

Variable	Description	Source	Role
Total RE Installed Capacity (MW)	Cumulative RE capacity excl. large hydro	MNRE (2014-2024 annual bulletins)	Dependent variable
Log RE Capacity	log(total_re_mw + 1)	Computed	Normalised DV
GSDP - Current Prices (Rs. Crore)	Net State Domestic Product at current prices	RBI Handbook of Statistics	Descriptive / source variable only (Section 3.6.2)

GSDP - Constant Prices (Rs. Crore)	NSDP at 2011-12 constant prices	RBI Handbook of Statistics	Predictor: organisational capacity
GSDP Growth Rate	Year-on-year % change in GSDP, untransformed (Section 3.6)	Computed from RBI	Predictor: economic momentum
SDG Composite Score	NITI Aayog SDG India Index overall score (3 anchor editions, interpolated/forward-filled; Section 3.4.1)	NITI Aayog	Predictor: governance quality
Prior RE Capacity (Lag 1)	Log RE capacity at t-1 per state	Computed	Predictor: path dependency
Time Trend	Linear time index, normalised 0-1	Computed	Predictor: secular trend
Cluster Membership	K-means adoption-trajectory group (Innovators / Early Adopters / Early Majority / Laggards)	K-means output (Section 3.5)	Descriptive grouping only - not a model predictor (Section 3.6.1)
RE Intensity (MW/Rs.1000 Cr)	Total RE capacity per unit constant-price GSDP	Computed	Alternate DV (robustness, Section 4.5)

The descriptive panel has 319 state-year observations (29 states and union territories, 2014-2024). The predictive panel used for cross-validation and SHAP analysis (Sections 3.6-3.7) has 280 (28 states, 2015-2024), reflecting the one-year loss to lagging and Delhi's exclusion for missing SDG data (Section 3.4).

3.5. K-Means Trajectory Clustering (Descriptive Analysis)

K-means clustering groups the 29 states by the shape of their renewable-energy adoption curve. It is descriptive and reported separately from the predictive model, which does not use cluster membership (Section 3.6.1). Seven trajectory features are computed per state from the full 2014–2024 series: installed capacity in 2014 and 2024, absolute growth, CAGR, growth acceleration (second-half minus first-half growth), 2024 solar share and 2024 RE intensity. Features are standardised by z-scoring; missing RE-intensity values are imputed with the column median.

Silhouette analysis and the elbow criterion (within-cluster inertia) jointly guide the choice of k over 2–7. Silhouette scores are 0.552 (k=2), 0.516 (k=3), 0.478 (k=4), 0.463 (k=5), 0.485 (k=6) and 0.380 (k=7). k=4 is retained, consistent with the elbow in the inertia curve; its overall silhouette of 0.478 indicates reasonable but not strong separation (Rousseeuw,

1987), weaker still within two of the four clusters (Section 4.2). K-means is run with 50 random initialisations (n_init = 50) to reduce sensitivity to starting values.

3.6. Predictive Model Specification

The dependent variable is log(total_re_mw + 1), which addresses the right skew of installed capacity and captures multiplicative growth. Cluster membership is excluded (Section 3.6.1) and current-price GSDP is dropped as collinear with constant-price GSDP (Section 3.6.2), leaving five predictors: (1) lagged log RE capacity (t−1): path dependency; (2) log GSDP at constant prices: real economic capacity; (3) GSDP growth rate, entered untransformed: economic momentum; (4) SDG composite score: governance quality; and (5) a normalised linear time trend. The growth rate is a year-on-year percentage change that can be zero or negative, for which the logarithm is undefined, so it is never logged. The SDG score stays on its 0–100 scale, being a bounded index rather than a skewed count.

3.6.1. Excluding Cluster Membership as a Predictor

Trajectory clusters (Section 3.5) are time-invariant and built from each state's full 2014–2024 series, including its 2024 capacity. Attaching them to every year would give a model predicting, say, 2016 capacity access to a label partly constructed from still-unobserved 2022–2024 outcomes, a look-ahead risk. Cluster membership is therefore excluded from the predictive model, and K-means is retained purely as a descriptive analysis never fed back into prediction. Refitting the clustering within each training fold was considered but not adopted, because fold-specific cluster definitions would be harder to present as a single, stable typology (Section 4.2). The model performs well without cluster information: with no cluster dummies, the five-predictor specification reaches a cross-validated R^2 of 0.9950 and improves on naive persistence (Section 4.3, Table 3).

3.6.2. GSDP Current-Price / Constant-Price Collinearity

Log GSDP at current and at constant prices are correlated at r = 0.995 and produce variance inflation factors above 600 when both are entered alongside the other four predictors (Section 4.5, Table 6), severe multicollinearity. Because the two series track the same economic size and differ only in the deflator, entering both adds no distinguishable information and destabilises the coefficient (and SHAP attribution, Section 3.7) split between them. Only the constant-price (real) series is retained, consistent with the organisational-capacity interpretation in Section 3.1; current-price GSDP remains in Table 1 as a descriptive/source variable.

Four models are trained on the five-predictor specification: OLS, Random Forest (500 trees, max_depth = 8, min_samples_leaf = 3), XGBoost with Optuna Bayesian hyperparameter optimisation (50 trials minimising RMSE) and LightGBM with Optuna optimisation (50 trials); search spaces cover n_estimators (100–600), max_depth (3–8), learning rate (0.01–0.20), subsample (0.6–1.0) and colsample_bytree (0.6–1.0). Cross-validation is strictly chronological, with an expanding window over the last three panel years (2022, 2023, 2024): for each test year τ, only observations with year < τ are used to fit the model evaluated on τ. The panel's one source of look-ahead information, cluster membership (Section 3.6.1), is handled by excluding that feature, not by altering the splitting rule. RMSE, MAE and R^2 are averaged across folds.

3.6.3. Benchmark Ladder

To benchmark the reported R^2 values critically, four nested specifications are evaluated under the same expanding-window cross-validation, all using OLS so that the comparison isolates the value of additional predictors rather than of a more flexible functional form: (i) naive persistence, where the year-τ prediction is simply the state's own lagged log capacity, with no fitting; (ii) lag-only OLS; (iii) lag-plus-economic OLS, adding constant-price GSDP and its growth rate; and (iv) full OLS, adding the SDG score and the time trend. Table 3 reports this ladder alongside the four-model comparison.

3.7. Explainable AI: SHAP Analysis

SHAP values are computed for the best-performing model (Section 3.6). Because cross-validation selects OLS (Section 4.3), they come from the Linear Explainer, which gives the exact, closed-form Shapley decomposition for additive linear models: each feature's contribution is its coefficient times its deviation from the sample mean, adjusted for feature covariance, with no sampling approximation. Where tree-based models (Random Forest, XGBoost, LightGBM) are reported for comparison, SHAP values are computed with TreeExplainer, which is exact for tree ensembles by construction (Lundberg et al., 2020). Outputs are global summary plots, dependence plots for the leading features and cluster-level profiles, the last using the full-panel clustering (Section 3.5) purely as a post-hoc grouping for presentation; cluster membership is never an input to the model being explained (Section 3.6.1).

3.8. Interpretive Scope: Prediction, Explanation and Causal Inference

Prediction, explanation and causal inference are separate exercises, and the paper's claims must stay within what its analysis supports. Three questions arise. The first is predictive: which variables, combined, let the model forecast a state's renewable capacity one year ahead with the lowest out-of-sample error? Section 4.3 answers this for the model as a whole; the lag dominates, as expected of a highly persistent stock, and the benchmark ladder (Section 3.6.3) quantifies what the remaining predictors add. The second is explanatory: how does the fitted model attribute a prediction across its input features? SHAP values (Sections 3.7 and 4.4) answer this, but they describe the model's own accounting, not evidence that a feature causally influences adoption; a feature can carry a large SHAP value because it marks the conditions that drive adoption without itself being a lever that, if moved, would change the outcome. The third question, which variables would cause greater adoption if a policymaker changed them, is causal, and this study is not designed to answer it. The panel is observational, predictors are plausibly endogenous to omitted state-level factors (political commitment, land availability, grid infrastructure, central-scheme allocation rules) correlated with both predictors and outcome, and no identification strategy (instrumental variables, a natural experiment, difference-in-differences) is employed. Sections 4 and 5 therefore describe features as 'associated with', 'a strong signal for' or 'informative of' predicted capacity, and avoid causal verbs ('drives', 'leads to') unless explicitly flagged as a hypothesis for future testing. This scoping does not diminish the contribution: knowing what the model uses to predict adoption, attributed transparently, is directly useful for forecasting and monitoring while the causal structure remains open for future research.

4. Results

4.1. Descriptive Analysis: Trends and Heterogeneity

Across the 29-state panel, installed renewable capacity grew at a compound annual rate of 15.0 per cent, from 35.22 GW in 2014 to 142.70 GW in 2024. This is a state-bulletin panel total, not India's nationally reported capacity, which uses wider MNRE and Central Electricity Authority (CEA) aggregates and a different large-hydro treatment (Section 3.4). Solar overtook wind in 2022 (53.6 GW versus 43.4 GW) and reached 81.4 GW by 2024 against 48.9 GW for wind. Small hydro, bio-power and other residual categories form a single 'Other RE' component in Fig. 1, because the MNRE data do not report these sub-categories consistently across all eleven years.

Heterogeneity is pronounced. In 2024, Rajasthan (26,693 MW) and Gujarat (25,472 MW) together account for 36.6 per cent of panel capacity, with Tamil Nadu (19,983 MW) and Karnataka (17,753 MW) a clear second tier and Maharashtra (14,483 MW) and Andhra Pradesh (9,419 MW) a third tier of steady but modest growth. Nineteen of the 29 states remain

below 2,000 MW in 2024, and the three smallest, Manipur (18.5 MW), Tripura (34.5 MW) and Nagaland (35.8 MW), are still close to negligible after eleven years.

Fig. 1 shows the national shift from wind to solar dominance, including the 2022 crossover. Fig. 2, a state-by-year heatmap of log capacity ranked by 2024 level, makes the concentration of capacity in western and southern India immediately visible; the northeastern and several eastern states remain in the lightest shading band across all eleven years.

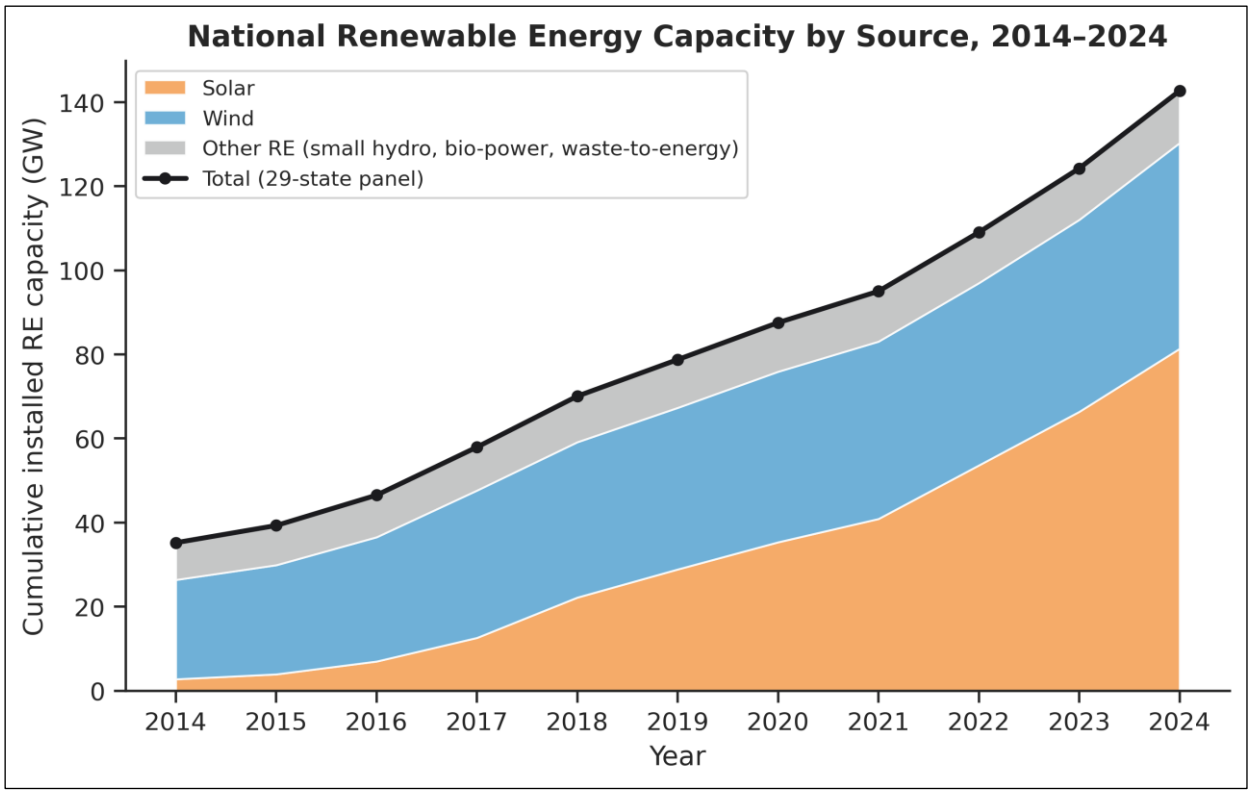


Fig. 1: National Renewable Energy Capacity by Source, 2014-2024 (GW). Stacked area shows solar, wind and the residual 'Other RE' category (small hydro, bio-power, waste-to-energy); the black line marks total panel capacity.

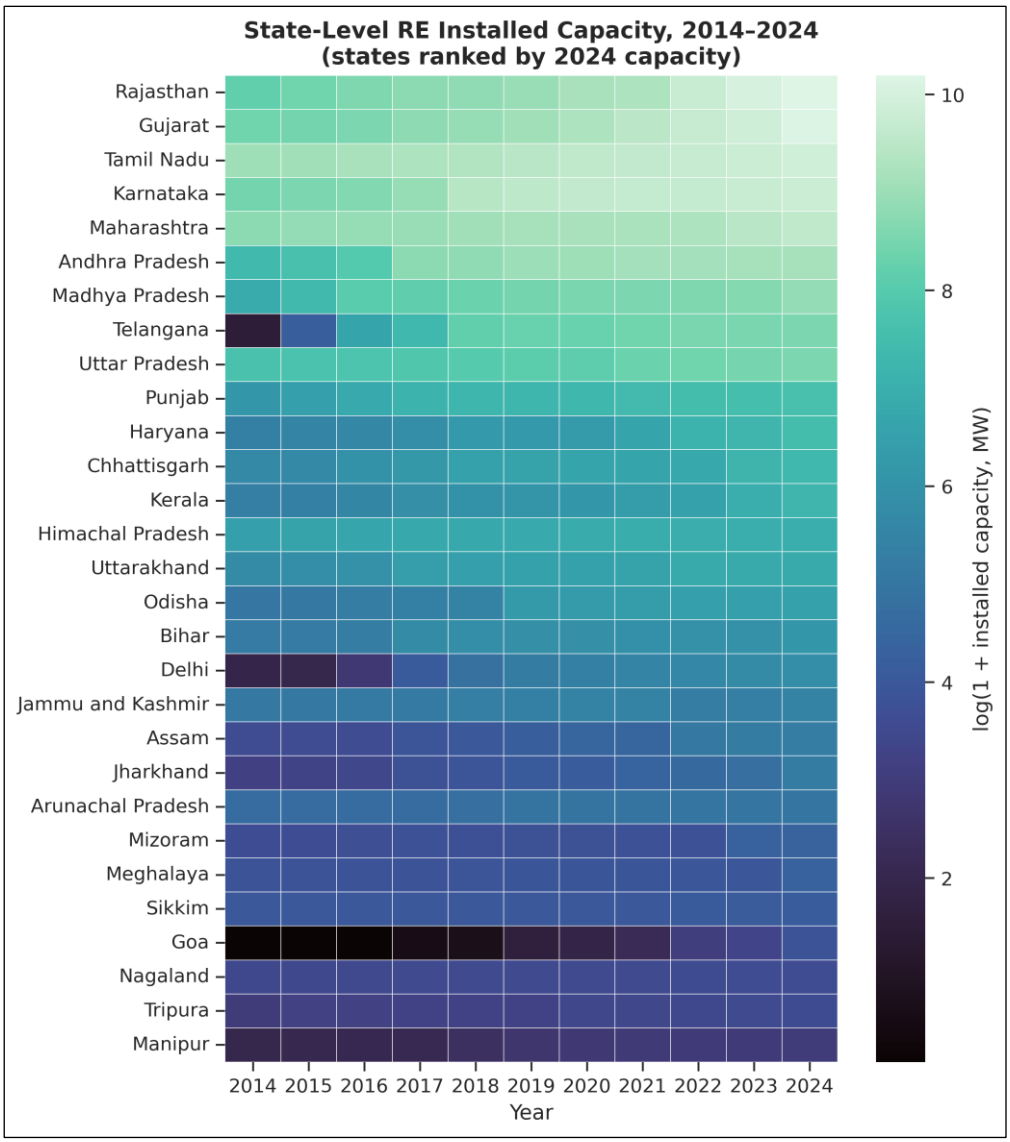


Fig. 2: State-Level Installed RE Capacity, 2014-2024, log(1+MW) scale, states ranked by 2024 capacity.

4.2. K-Means Clustering Results

K-means clustering (k = 4, overall silhouette = 0.478) partitions the 29-state descriptive panel into four empirically distinguishable trajectory groups (Section 3.5), profiled in Table 2.

Table 2: K-Means Cluster Profiles (Mean Values per Group), 29-State Panel

Cluster	n	Member States	Cap. 2014 (MW)	Cap. 2024 (MW)	Abs. Growth (MW)	CAGR (%)	Silhouette
Innovators	2	Gujarat, Rajasthan	4,031	26,082	22,051	20.6	0.547

Early Adopters	3	Karnataka, Maharashtra, Tamil Nadu	6,489	17,406	10,918	10.6	0.443
Early Majority	3	Delhi, Goa, Telangana	3	1,862	1,859	79.5	0.339
Laggards	21	Remaining 21 states	366	1,559	1,193	12.2	0.496

The Early Majority cluster's high CAGR (79.5%) is a small-base artefact: the cluster starts from a combined 2014 capacity of only 3 MW and is reported descriptively, not as evidence of a distinct adoption mechanism. Only Goa and Telangana enter the predictive panel, as Delhi lacks SDG data (Section 3.4).

Fig. 3 presents the elbow and silhouette diagnostics for k = 2 to 7 (Section 3.5) and the mean RE capacity trajectory per cluster, 2014–2024. Innovators and Early Adopters accelerate from around 2021, consistent with utility-scale solar scaling under central schemes such as PM-KUSUM; Laggards stay nearly flat, consistent with the structural constraints discussed in Section 5. Fig. 4 adds the state-year distribution of key trajectory variables by cluster: means are well separated, but the Early Majority and Laggard distributions overlap appreciably, mirroring their lower silhouette scores in Table 2.

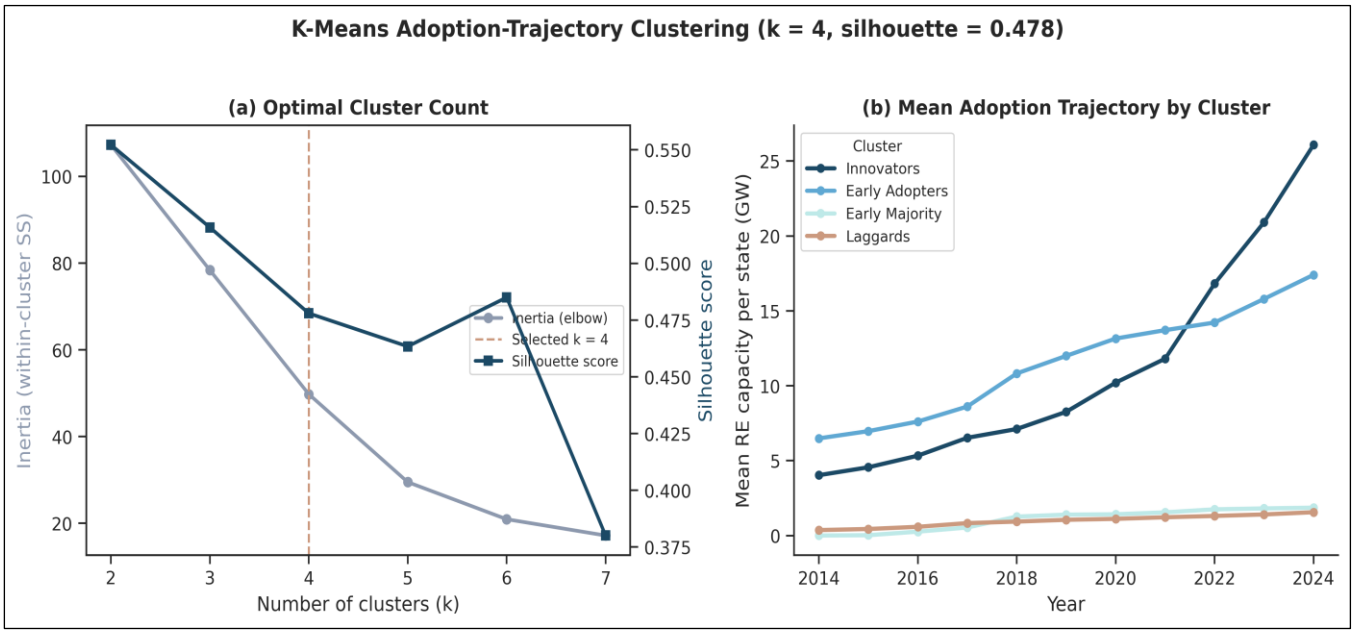


Fig. 3: (a) Optimal cluster count by elbow (WCSS) and silhouette criteria, k = 2-7; (b) mean adoption trajectory by cluster, 2014-2024.

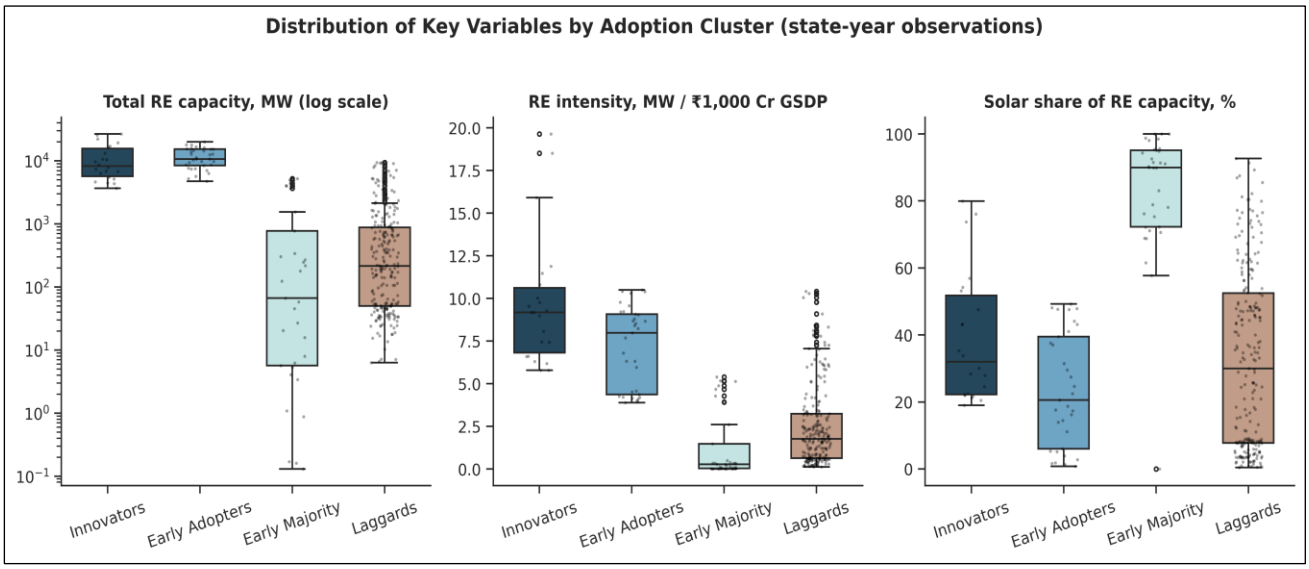


Fig. 4: Distribution of key trajectory variables by adoption cluster (state-year observations).

The taxonomy is read through Rogers' (2003) typology, not as a validation of it (Section 3.2). Innovators (Gujarat, Rajasthan) established large-scale solar parks early, consistent with strong irradiance, available land and proactive state solar policy. Karnataka's wind resource and Tamil Nadu's early wind deployment are consistent with Early Adopter leadership. The Early Majority's moderate capacity from a very small 2014 base plausibly reflects land-availability and grid-configuration constraints in Goa and Telangana rather than a lack of policy intent, though the cross-sectional design cannot separate these channels. Laggard states show patterns consistent with structural constraints (DISCOM financial health, industrial electricity demand, topography in the northeast) that this study proxies only partially and cannot attribute to any single cause.

4.3. Model Performance Comparison

Table 3 reports the benchmark ladder (Section 3.6.3) together with the four-model comparison, all under the same chronological expanding-window cross-validation (test years 2022, 2023, 2024), so each layer's added value can be read from one table.

Table 3: Benchmark Ladder and Model Performance, Chronological Expanding-Window Cross-Validation (test years 2022-2024)

Specification / Model	R2	+-SD	RMSE	+-SD	MAE	+-SD
Naive persistence (y_t = y_t-1)	0.9910	0.0027	0.2112	0.0313	0.1364	0.0137

Lag-only OLS	0.9945	0.0023	0.1641	0.0327	0.1274	0.0176
Lag + economic OLS	0.9949	0.0016	0.1590	0.0250	0.1224	0.0177
Full OLS (lag+econ+governance+time) - best	0.9950	0.0011	0.1585	0.0177	0.1115	0.0108
Random Forest	0.9929	0.0021	0.1868	0.0267	0.1261	0.0169
XGBoost (Optuna-tuned)	0.9913	0.0024	0.2078	0.0296	0.1411	0.0172
LightGBM (Optuna-tuned)	0.9782	0.0082	0.3252	0.0650	0.2412	0.0345

RMSE and MAE in Table 3 are on the log scale; lower is better. Two findings follow. First, naive persistence alone already achieves R^2 = 0.991: nearly all apparent predictive power comes from installed capacity being a slow-moving stock, not from the structural predictors. The full specification lifts R^2 by only 0.004 but, more informatively, cuts RMSE from 0.211 to 0.159 (25 per cent) and MAE from 0.136 to 0.112 (18 per cent). The economic and governance predictors earn their place as error-reducers on a persistence baseline that R^2 alone understates, not as the source of overall fit. Second, full OLS (R^2 = 0.9950) is the best of the four models, marginally ahead of Random Forest (R^2 = 0.9929), XGBoost (R^2 = 0.9913) and LightGBM (R^2 = 0.9782). This fits the ladder finding: once the lagged-stock relationship is specified, predictor-outcome relationships in this panel are close to linear and additive, and an ensemble's added flexibility adds variance without reducing bias, so held-out performance does not improve.

Fig. 5 shows the benchmark ladder and the four-model comparison with standard-deviation error bars. Fig. 6 plots actual against predicted values for full OLS on the full sample; points cluster tightly around the 45-degree line from low-capacity Laggards to high-capacity Innovators, so fit is not concentrated in any one part of the capacity distribution.

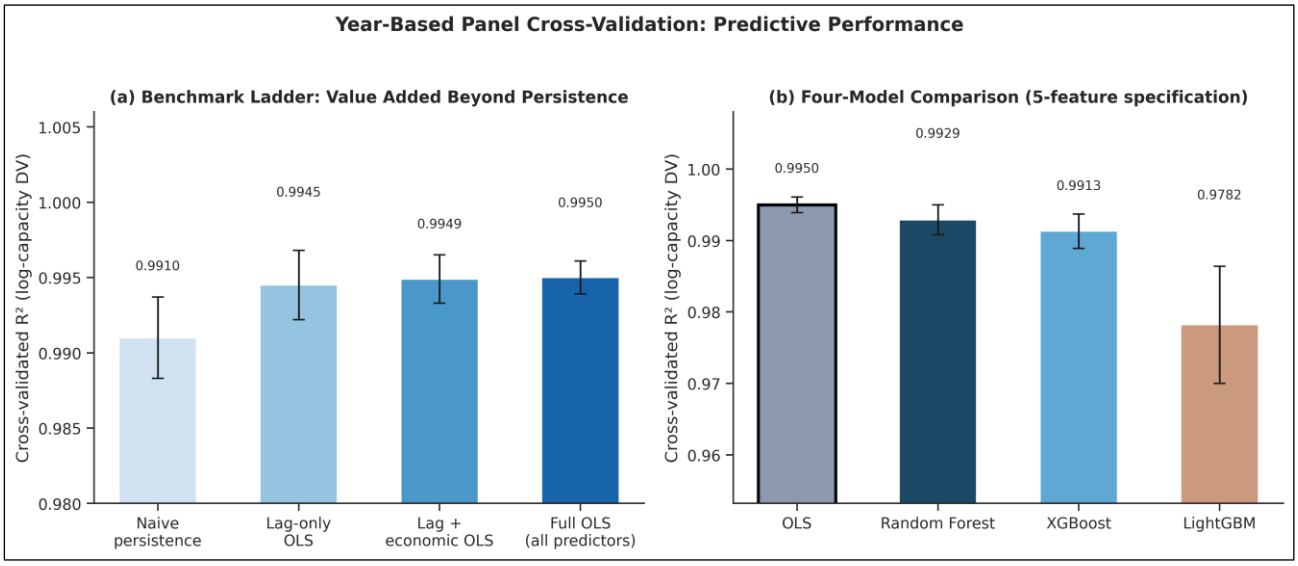


Fig. 5: (a) Benchmark ladder - value added beyond persistence; (b) four-model comparison on the 5-feature specification. Year-based expanding-window cross-validation, test years 2022-2024.

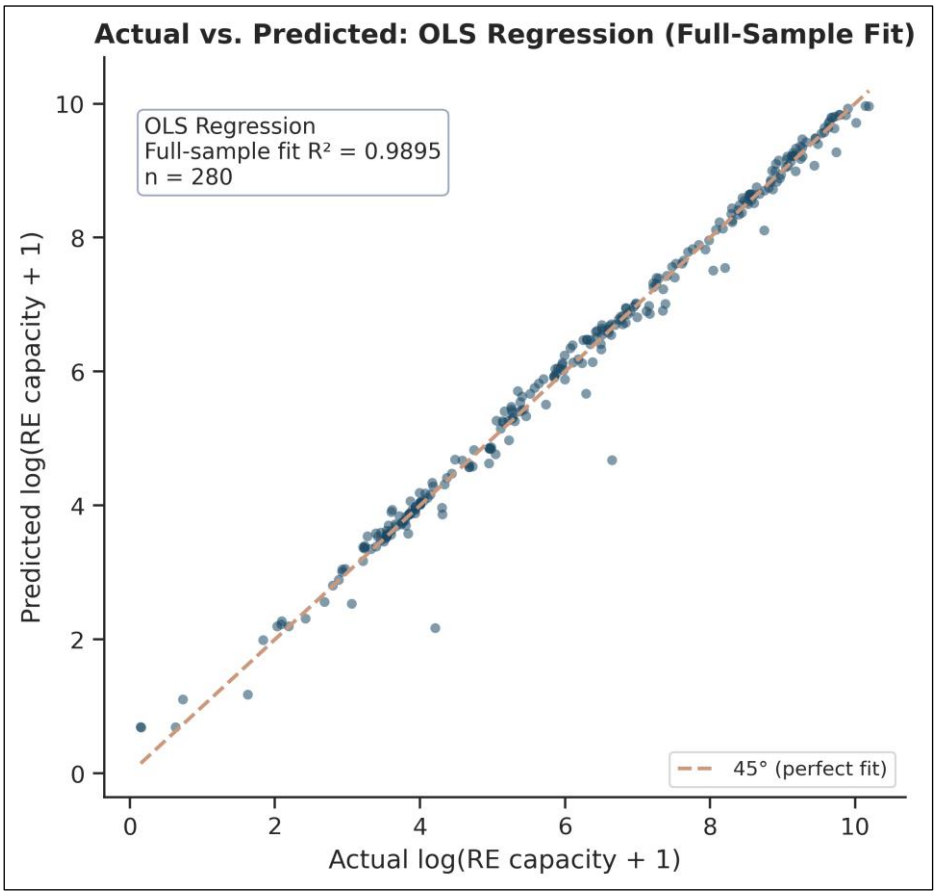


Fig. 6: Actual vs. predicted log RE capacity, full OLS (best model), full-sample fit.

4.4. SHAP Analysis: Global and Cluster-Level Feature Attribution

The Linear Explainer applied to the best-performing model (full OLS, Section 4.3) gives exact attributions for all 280 predictive-panel observations (Section 3.7). Table 4 presents the global SHAP ranking for the 5-feature specification.

Table 4: Global SHAP Feature Importance Ranking, Full OLS (Linear Explainer), 5-Feature Specification

Rank	Feature	Mean |SHAP|	Mean SHAP	Std SHAP	TOE Dimension
1	Prior RE Capacity (Lag)	1.7953	-0.1212	2.1097	Technology context
2	GSDP (Constant Prices)	0.1683	-0.0063	0.1965	Organisation
3	SDG Composite Score	0.0375	+0.0006	0.0457	Environment
4	Time Trend	0.0286	-0.0006	0.0328	Secular trend
5	GSDP Growth Rate	0.0104	+0.0019	0.0140	Organisation

SHAP values in Table 4 are on the log scale. Prior RE capacity is by far the strongest predictive contributor (mean |SHAP| = 1.795), more than an order of magnitude above every other feature. In line with Section 3.8, this is a predictive and explanatory finding, not a causal one. The fitted model's reliance on prior capacity is unsurprising for a slow-moving stock and consistent with, but not proof of, the reinforcing mechanisms proposed in the diffusion-of-innovations literature (Section 3.2): falling incremental costs, demonstrated policy stability and accumulated grid-management experience. Distinguishing among them would require data this study lacks (state investment-cost series, investor surveys, or a natural experiment isolating one channel). Fig. 7 adds the per-observation SHAP distribution; the wide spread of positive and negative attributions for prior capacity is consistent with it acting as an accelerant for states already scaling and a weaker signal for states with little installed base, again an association within the fitted model.

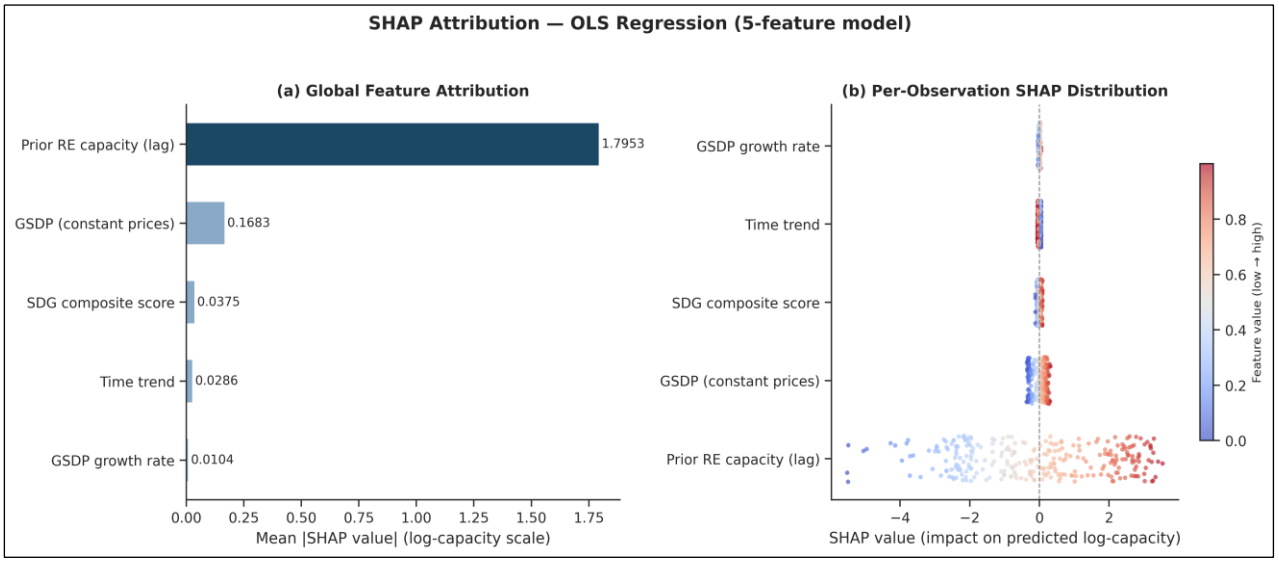


Fig. 7: (a) Global feature attribution, mean |SHAP value| (full OLS, 5-feature model); (b) per-observation SHAP value distribution.

The economic-capacity channel is represented by constant-price GSDP alone (mean |SHAP| = 0.168), current-price GSDP having been dropped for collinearity (Section 3.6.2). Higher real GSDP is associated with higher model-predicted RE capacity, consistent with, though not evidence for, the TOE organisational channel of greater capital-expenditure headroom and lower relative DISCOM financial risk. The dependence plots in Fig. 8 (prior capacity, constant-price GSDP, SDG score) show a positive, approximately monotonic GSDP association across the observed range, with no visible diminishing returns (non-linearity is not formally tested).

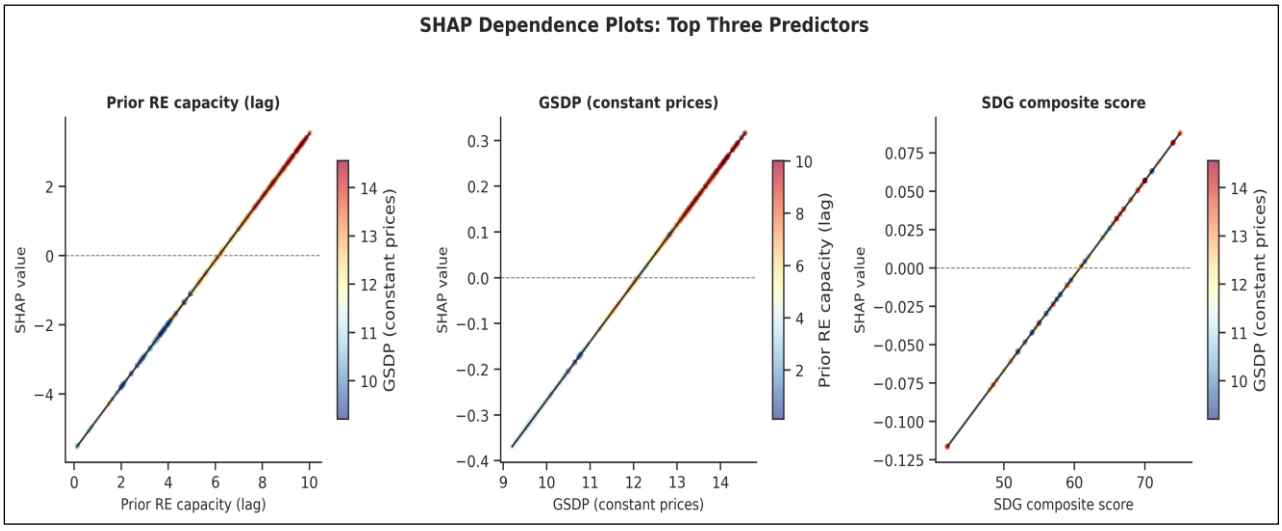


Fig. 8: SHAP dependence plots for the top three predictors: prior RE capacity, GSDP (constant prices), SDG composite score.

The SDG composite score, the governance-quality proxy, ranks third globally (mean |SHAP| = 0.038), ahead of the time trend and GSDP growth rate: modest in absolute terms, but not negligible relative to the economic-capacity channel. In the cluster-level SHAP profiles (Fig. 9), governance attribution is broadly similar across the four clusters (Innovators 0.021; Early Adopters 0.043; Early Majority 0.034; Laggards 0.039), a roughly two-fold range with Early Adopters highest. The evidence does not support a claim that governance quality matters more for states earlier or later in the adoption sequence (Section 5.2).

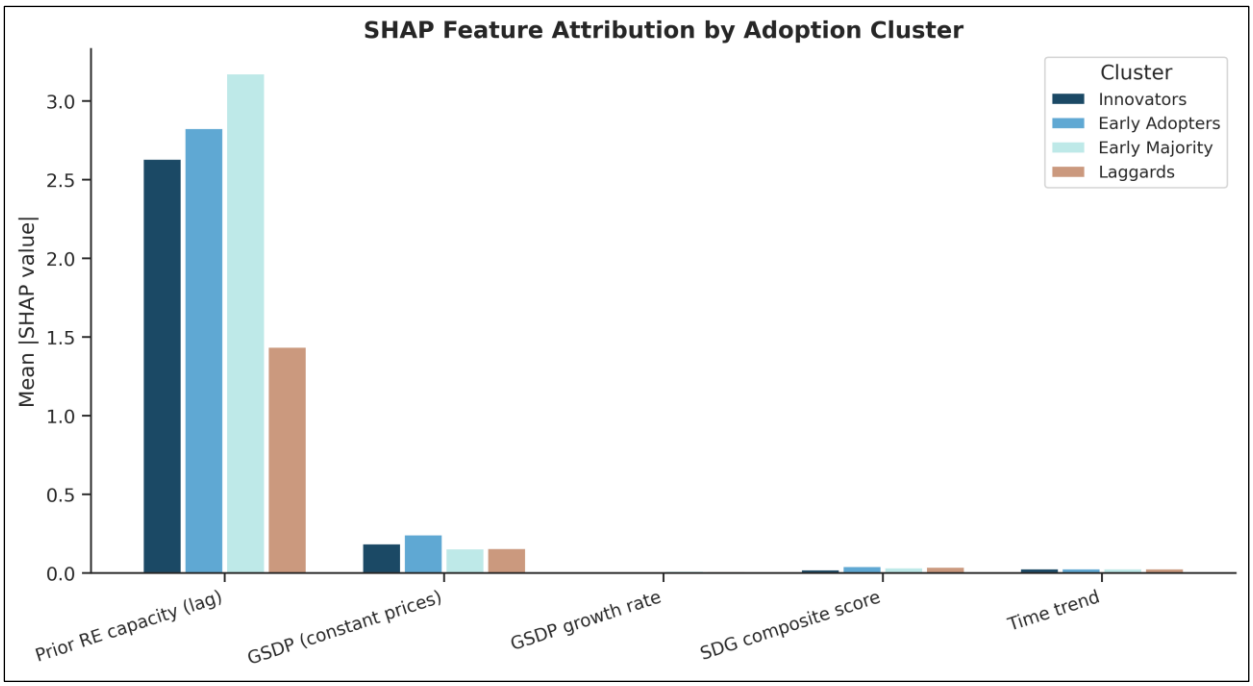


Fig. 9: SHAP feature attribution by adoption cluster, mean |SHAP value| per cluster per feature (5-feature specification).

As cluster membership is not part of the predictive model (Section 3.6.1), no cluster-indicator SHAP values are reported; clustering remains a descriptive framework (Section 4.2), not a predictive channel.

4.5. Robustness Checks

Table 5 presents the robustness specifications, run on the 5-feature, no-cluster model (Section 3.6); Table 6 presents the corresponding VIF diagnostics.

Table 5: Robustness Check Summary, 5-Feature Specification

Check	Specification	R^2	Delta R^2 vs Baseline	Purpose
Baseline	Full model (Random Forest, all 28 states w/ complete predictors)	0.9929	-	Reference specification
RC1	Exclude COVID years (2020-2021)	0.9859	-0.0070	Structural break test
RC2	Log RE intensity as alternate DV	0.8410	N/A (different DV)	Alternate dependent variable
RC3	VIF analysis - all predictors < 5	N/A	N/A	Multicollinearity diagnostic (Table 6)
SA1	Drop path dependency (lag)	0.8896	-0.1033	Largest single-group drop
SA2	Drop GSDP level (constant prices)	0.9913	-0.0017	Economic channel test
SA3	Drop GSDP growth rate	0.9927	-0.0003	Economic-momentum channel test
SA4	Drop SDG score	0.9936	+0.0007	Governance channel test
SA5	Drop time trend	0.9929	-0.0001	Secular-trend channel test

Baseline R^2 (0.9929) in Table 5 is the Random Forest cross-validation figure used for the drop-one comparisons; the best-model result (full OLS, $R^2 = 0.9950$) is in Table 3. Excluding the COVID-affected years (2020-2021) lowers R^2 only to 0.9859, so the predictor-outcome relationship is not a pandemic artefact. With log RE intensity (RE capacity per unit constant-price GSDP) as an alternate dependent variable, $R^2 = 0.841$: lower, as expected for a harder normalised target, but the predictors retain substantial explanatory power once the direct effect of state economic size is removed. All five predictors show VIF below 5 (Table 6), so dropping current-price GSDP fully resolves the multicollinearity identified in Section 3.6.2. Dropping prior capacity costs 0.103 R^2, an order of magnitude more than any other predictor; dropping constant-price GSDP costs a further 0.002, and dropping GSDP growth rate, the SDG score or the time trend individually costs less than 0.001 each. This matches the SHAP ranking in Table 4; no remaining predictor does outsized, undisclosed work.

Table 6: Variance Inflation Factor (VIF) Results, 5-Feature Specification

Predictor	VIF	Threshold Status	Concern

Prior RE Capacity (Lag 1)	3.433	< 5	None
GSDP (Constant Prices)	3.215	< 5	None
SDG Composite Score	1.358	< 5	None
Time Trend (Normalised)	1.254	< 5	None
GSDP Growth Rate	1.004	< 5	None

As noted in Section 3.6.2, entering both GSDP series yields VIF values above 600 (r = 0.995), which is why current-price GSDP is kept only as a descriptive variable (Table 1).

Fig. 10 brings the three robustness dimensions together (specification checks against the baseline R2, the VIF diagnostics of Table 6 and the drop-one-group sensitivity of Table 5) so the stability of the specification can be assessed at a glance.

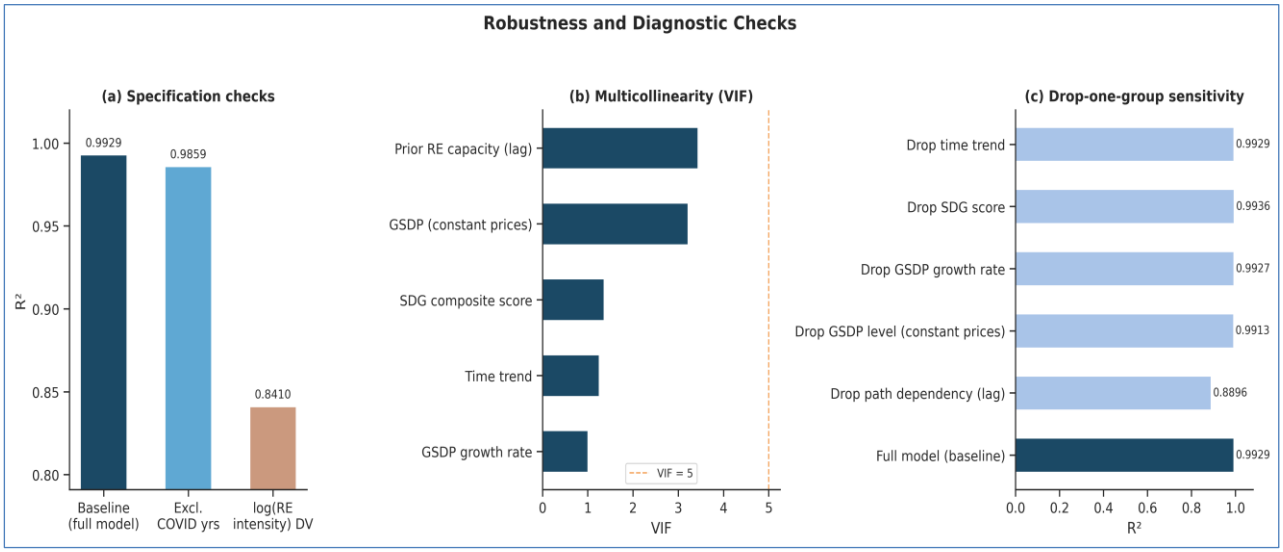


Fig. 10: Robustness and diagnostic checks - (a) specification checks (COVID exclusion, alternate DV); (b) multicollinearity (VIF); (c) drop-one-group sensitivity.

5. Discussion

5.1. Interpretation of Findings in Relation to Prior Evidence

The findings speak to the literature on renewable energy adoption and subnational energy transitions in three ways, within the limits set in Section 3.8. First, prior RE capacity is by far the strongest predictive contributor (Section 4.4), a finding consistent with, though not a test of, the path-dependency mechanisms discussed in the Diffusion of Innovations

literature (Rogers, 2003), applied here to subnational government units in a federal developing economy. Read predictively rather than causally, states with an established RE base are those the model most confidently expects to keep scaling.

Second, mapping the five predictors onto the three TOE dimensions (Section 3.1) shows that each dimension carries measurable SHAP attribution: prior capacity for technology, constant-price GSDP and its growth rate for organisation, and the SDG score for environment. All three contribute; none renders the others redundant, so the TOE categories correspond to empirically separable predictor groups in this setting. This is not a causal test of TOE mechanisms, but, to the authors' knowledge, it is among the first attempts to examine TOE dimensional structure through SHAP-based attribution in the renewable energy adoption literature.

Third, the cluster-level SHAP analysis (Section 4.4, Fig. 9) does not support a claim that governance-quality attribution varies sharply by adoption stage: the range across the four clusters (0.021-0.043) is modest, with Early Adopters highest. This is a null result on that specific question, not evidence for a differential-governance narrative. Nor does it make governance quality unimportant, since it ranks third globally in Table 4; the analysis simply finds no evidence that its predictive contribution is markedly stronger at one stage than another.

5.2. Practical Implications

For central efforts to even out DISCOM finances and grid readiness, the strong path-dependency signal (Section 4.4) is consistent with the view that Laggard states are unlikely to close the gap without deliberate intervention, since the model's strongest predictor of future capacity is the capacity a state already has. Central schemes such as PM-KUSUM and the Revamped Distribution Sector Scheme (RDSS) are plausibly well suited to reinforcing capacity where infrastructure momentum already exists. Whether they can offset the structural constraints of the Laggard cluster is a causal question this associative analysis (Section 3.8) cannot answer; that requires an evaluation design isolating the schemes' effect.

Removing constant-price GSDP costs the model 0.002 R^2 beyond path dependency (Table 5), a small but consistently positive contribution across robustness specifications, suggesting that economic capacity is an associated predictor worth policy attention independent of a state's existing RE base. Where it appears to bind, instruments that reduce upfront capital costs (viability-gap funding, green bonds or other blended-finance mechanisms for distributed

renewables) are a plausible complement to policy or tariff interventions, though this study identifies which predictors the model associates with capacity outcomes, not which interventions would work.

Governance-score attribution is broadly similar across clusters (Sections 4.4, 5.1). Because the SDG proxy is nonetheless the third-ranked predictor overall (Table 4), governance-quality considerations (regulatory streamlining, land-acquisition processes, DISCOM financial health) warrant attention as one lever among several across the adoption spectrum, not preferential targeting at any one cluster. Differentiated fiscal-transfer targeting by adoption stage may still be warranted on other grounds, but this analysis offers no SHAP-based support for it.

5.3. Management Implications

The cluster taxonomy (Section 4.2) gives renewable energy developers and private investors a data-informed, descriptive starting point for comparing states, not a validated risk model. Innovator and Early Adopter states show established regulatory pathways, prior grid-integration experience and demonstrated offtake, consistent with lower incremental project-development friction; Laggard states carry more structural uncertainty but may offer first-mover advantages to developers willing to invest in early-stage stakeholder engagement. The SHAP evidence (Section 4.4) can flag factors meriting closer due diligence in a given grouping, as one input among many, not a substitute for state-specific credit or regulatory assessment.

The path-dependency finding is relevant for multilateral development finance institutions, such as the Asian Development Bank and the Green Climate Fund, that allocate catalytic capital to early-stage state RE adoption. If the association between early infrastructure and continued scaling reflects a genuine reinforcing mechanism, capital that establishes initial infrastructure templates and demonstrable examples of success could have a larger effect than the same capital deployed later. The evidence here is consistent with that possibility without establishing it, so funders should treat it as a hypothesis for targeted piloting and evaluation.

6. Conclusions

This paper shows that renewable energy adoption at the Indian state level is structurally heterogeneous and is not adequately described at the national level alone. Using a leakage-free framework on 29 states and union territories over eleven years (2014-2024), it finds prior installed capacity to be the strongest predictive contributor to model-predicted RE capacity, with real economic capacity and governance quality secondary but non-negligible (Section 4.4). With

cluster membership excluded as a leakage source and collinear current-price GSDP dropped (Sections 3.6.1-3.6.2), strictly chronological expanding-window cross-validation shows a five-predictor OLS specification (R^2 = 0.9950) modestly outperforming Random Forest, XGBoost and LightGBM. As naive persistence alone achieves R^2 = 0.991, the structural predictors' incremental value is better read from RMSE and MAE reductions than from R^2. Four empirically distinguishable adoption-trajectory clusters emerge, with broadly similar governance attribution across them. Findings hold without COVID-affected years, under an alternate dependent variable and under multicollinearity diagnostics (Section 4.5).

Methodologically, the study contributes an open-data ML pipeline for subnational energy forecasting that separates prediction, explanation and causal inference (Section 3.8), a leakage-audited SHAP analysis of state-level adoption predictors in India, and a descriptive cluster taxonomy for differentiated policy discussion. The findings are relevant to India's 2030 renewable planning and state-level implementation of its updated NDC, though the observational design means they should inform, not substitute for, policy-specific evaluation.

Several limitations apply. The SDG composite covers many non-energy policy areas and is therefore only an indirect proxy for energy-specific governance. Only three editions (2018, 2019-20, 2020-21) were usable, with 2022-2024 carried forward rather than interpolated between two anchors (Section 3.4.1), which understates any genuine governance change in the panel's final three years. West Bengal could not be recovered from the MNRE PDFs and is excluded entirely; Delhi, lacking an SDG score, is excluded from the predictive and SHAP analysis but retained in the clustering. State-level tariff, DISCOM-finance and grid-capacity data, which would sharpen the model, are not incorporated. The window begins in 2014, and an earlier start would better capture adoption-initiation dynamics. As Section 3.8 establishes, the design is predictive and explanatory, not causal; no SHAP-based finding here shows that changing a predictor would change RE adoption. Future research should pursue quasi-experimental or instrumental-variable designs to identify causal effects, apply spatial econometrics to test for adoption spillovers between neighbouring states, disaggregate solar and wind to test whether they follow distinct structural logics, and incorporate DISCOM and grid-capacity data as consistent state-level series become available.

Meeting India's 2030 renewable energy targets and 2070 net-zero commitment plausibly requires substantially narrowing the gap between Innovator and Laggard states. The path-dependency signal is consistent with the view that this is unlikely without deliberate, structural intervention aimed at each cluster's constraints, though predictive evidence alone

cannot establish which interventions would work. The framework developed here, transparent, open-data and explicit about the boundary between prediction, explanation and causal claims, is one empirical input into that targeting, alongside the causal and implementation evidence that state policy design requires.

7. Acknowledgments

Funding acknowledgement: This research received no external funding.

8. References

Apergis, N. and Payne, J.E., 2010. Renewable energy consumption and growth in Eurasia. Energy Economics, 32(6), pp.1392-1397.

Bergmeir, C. and Benitez, J.M., 2012. On the use of cross-validation for time series predictor evaluation. Information Sciences, 191, pp.192-213.

Breiman, L. 2001. Random forests. Machine Learning, 45(1), pp.5-32.

Chen, T. and Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.785-794.

Dubash, N.K., 2011. From norm taker to norm maker? Indian energy governance in global context. Global Policy, 2(S1), pp.66-79.

Government of India, 2022. India's updated nationally determined contribution. Ministry of Environment, Forest and Climate Change. Available online: https://unfccc.int/sites/default/files/NDC/2022/08/India%20Updated%20First%20Nationally%20Determined%20Contribution%202022.pdf (accessed on 12 August 2026).

Haggag, M., Abdelhady, K., Guirguis, M., Saudy, M. and El Dakhakhni, W., 2026. Machine learning long-term electricity demand forecasting system for strategic energy investments. Scientific Reports, 16, p.12471.

Hayati, A., Gharehveran, S.S. and Shirini, K., 2026. Electricity price forecasting with ensemble meta-models and SHAP explainers: A PCA-driven approach. Scientific Reports, 16, p.6466.

IEA, 2023. World energy outlook 2023. International Energy Agency, IEA. Available online: https://www.iea.org/reports/world-energy-outlook-2023 (accessed on 5 August 2026).

IRENA, 2023. Renewable power generation costs in 2022. International Renewable Energy Agency (IRENA). Available online: https://www.irena.org/Publications/2023/Aug/Renewable-Power-Generation-Costs-in-2022 (accessed on 8 August 2026).

IPCC, 2011. Special report on renewable energy sources and climate change mitigation (SRREN). Prepared by Working Group III of the Intergovernmental Panel on Climate Change. Edenhofer, O., Pichs-Madruga, R., Sokona, Y., et al. (eds.). Cambridge University Press, Cambridge, UK and New York, NY, USA.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.Y., 2017. LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, pp.3146-3154.

Linardatos, P., Papastefanopoulos, V. and Kotsiantis, S., 2021. Explainable AI: A review of machine learning interpretability methods. Entropy, 23(1), p.18.

Lundberg, S.M. and Lee, S.I., 2017. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, pp.4765-4774.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N. and Lee, S.I., 2020. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), pp.56-67.

MNRE, 2024. Year-wise achievements: State-wise cumulative installed capacity 2014-2026. Ministry of New and Renewable Energy, Government of India. Available online: https://mnre.gov.in/year-wise-achievement/ (accessed on 16 August 2026).

NITI Aayog, 2024. SDG India index and dashboard 2023-24. Government of India. Available online: https://sdgindiaindex.niti.gov.in (accessed on 2 September 2026).

Rahat, M.M.A., Khan, J.U., Shahnan, I., Islam, M.I. and Al Mansur, A., 2025. Explainable machine learning framework for wind speed prediction: A comprehensive analysis using long-term meteorological data. In: 2025 IEEE 7th International Conference on Sustainable Technologies for Industry 5.0 (STI 2025).

REN21, 2024. Renewables 2024 global status report. REN21 Secretariat. Available online: https://www.ren21.net/gsr-2024/ (accessed on 26 July 2026).

Reserve Bank of India, 2025. Handbook of statistics on the Indian economy 2024-25. RBI. Available online: https://rbi.org.in/Scripts/AnnualPublications.aspx?head=Handbook%20of%20Statistics%20on%20Indian%20Economy (accessed on 18 August 2026).

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schroder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F. and Dormann, C.F., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40(8), pp.913-929.

Rogers, E.M. 2003. Diffusion of innovations, 5th ed. Free Press.

Rousseeuw, P.J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, pp.53-65.

Sadorsky, P. 2009. Renewable energy consumption and income in emerging economies. Energy Policy, 37(10), pp.4021-4028.

Taghikhah, F.R., Malik, A., Voinov, A., Govindan, K. and Mohamadpour Tosarkani, B., 2026. Exploring drivers and policy enablers of citizen engagement in renewable energy projects through explainable AI. Energy Policy, 212, p.115141.

Tandon, A., Awasthi, A., Pattnayak, K.C., Tandon, A., Choudhury, T. and Kotecha, K., 2025. Machine learning-driven solar irradiance prediction: Advancing renewable energy in Rajasthan. Discover Applied Sciences, 7, p.107.

Tang, T. 2026. Predicting energy prices and renewable energy adoption through an optimized tree-based learning framework with explainable artificial intelligence. Scientific Reports, 16, p.6771.

Tornatzky, L.G. and Fleischer, M., 1990. The processes of technological innovation. Lexington Books: Lexington, MA, USA, pp.1-298.

Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F. and Fouilloy, A., 2017. Machine learning methods for solar radiation forecasting: A review. Renewable Energy, 105, pp.569-582.

Author Contributions: "Conceptualization, R.S., A.A. and S.S.; methodology, R.S.; software, R.S.; validation, R.S. and A.A.; formal analysis, R.S. and S.S.; investigation, R.S. and A.A.; resources, A.A.; data curation, R.S., S.S.; writing-

original draft preparation, A.A. and S.S.; writing-review and editing, R.S. and A.A.; visualization, S.S.; supervision, R.S.; project administration, R.S. and A.A. All authors have read and agreed to the published version of the manuscript."

Institutional Review Board Statement: "Not applicable."

Informed Consent Statement: "Not applicable."

Conflicts of Interest: "The authors declare no conflicts of interest."

Data and Code Availability: All source data are public (MNRE state-wise cumulative capacity bulletins; RBI Handbook of Statistics on Indian Economy; NITI Aayog SDG India Index, 2018/2019-20/2020-21 editions). The data-extraction, modelling and figure-generation pipeline described in Sections 3.4-3.7 is available from the corresponding author on reasonable request.