International Journal of Social Science and Economic Research

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

AI-Driven Circular Economy Adoption in Indian E-Commerce

Ritesh Kalidindi[1], Leelavathy Narkedamilly[2] and Uma Meghana Saladi[3]

[1]Student, International Baccalaureate Diploma Program, International School of Hyderabad, Hyderabad 500034, India

[2]Professor of Computer Science and Engineering, Department of Computer Science & Engineering, Godavari Global University, Rajamahendravaram 533296, India

[3]Senior Consultant, Oracle Health Care, WA 98052, USA

DOI: 10.46609/IJSSER.2026.v11i07.037 URL: https://doi.org/10.46609/IJSSER.2026.v11i07.037

Received: 3 June 2026 / Accepted: 19 July 2026 / Published: 31 July 2026

ABSTRACT

This paper aims to gain insight into the most important behavioral factors for the adoption of circular economy (CE) products and to build realistic policy scenarios using a hybrid Random Forest-ARIMAX modeling approach for the Indian e-commerce platform until 2030. The Random Forest component was based on a synthetic but well-calibrated dataset of 50,000 consumer records, with strong predictive performance (R^2=0.9721) and low error rates. There were some factors that a platform could truly impact, such as product availability, consumer awareness, and price sensitivity. The projections, calculated for three different policy scenarios and combined with the time-series component of the ARIMAX prediction, indicate a CE market value ranging from USD 28 to 67 billion, and 8.5 to 22.5 million new jobs by 2030. Of particular interest is the EPR Full Compliance scenario, which has a multiplier effect of 1.9× and a break-even point of about 3.6 years. Overall, the framework provides a clear and practical tool to support sustainability decision-making in a rapidly expanding digital economy.

Keywords: Circular Economy; E-Commerce Adoption; Random Forest; Machine Learning; RF-ARIMAX; SHAP; SEM-RF Pipeline; Wasserstein Distance

1. Introduction

The rapidly growing digital economy in India is placing environmental pressures on existing waste management procedures, which are proving difficult to manage. E-waste increased by 73 percent, with 1.01 to 1.75 million tonnes generated, while less than half of discarded electronics are now formally recycled (CPCB, 2024; UNITAR, 2024). Unless drastic measures are taken, India may have to contend with 5 million tonnes of e-waste per year by 2030 (CSE, 2023). Trying to recycle more is not the solution to the problem because online shopping will continue to promote increased consumption.

Approaches to a circular economy, reuse, repair, and refurbishment, attempt to address the causes instead of the consequences of waste management (Geissdoerfer et al., 2017; Lahane et al., 2020). Artificial intelligence has the potential to influence consumer behavior at scale in e-commerce by making recommendations smarter, adding clean eco-labels, and optimizing supply chain choices (Tao et al., 2018).

There are five contributions of this work. To begin with, it proposes a hybrid RF-ARIMAX foresight model, which connects consumer-level forecasts to longer-term scenario planning. Second, it uses SHAP analysis to provide practical actions for managers and policymakers. Third, it compares three policy scenarios, using a Monte Carlo simulation, to achieve national goals. Fourthly, it applies SMOTE to address class imbalance in a two-stage CFA/SEM-RF pipeline. Finally, it includes a publicly available, well-tested synthetic dataset for use in future studies in Indian CE. It is hoped that this research will contribute to the development of technology forecasting methods in the transitional contexts of emerging economies.

1.1. Theoretical Foundations

This study was designed based on three established frameworks. According to Davis (1989), the Technology Acceptance Model (TAM) links consumer awareness to Perceived Usefulness and product availability to Perceived Ease of Use. The theory of Diffusion of innovations, developed by Rogers (2003), can explain the tipping point at which the adoption rate accelerates when awareness and availability reach a critical threshold. The Value-Belief-Norm model (Stern, 2000) is used to shed light on why consumers, even those with low incomes, exhibit high adoption rates when they recognize the environmental impacts. Lastly, these micro-level behaviors are linked to greater economic forecasts with the broader CE business model framework (Geissdoerfer et al., 2017).

2. Literature Review

2.1. Ensemble Machine Learning for Sustainability Analytics

Ensemble methods are now the preferred tool for modeling complex and nonlinear consumer behavior in the context of sustainability. Breiman (2001) demonstrated that random forests tend to be far more successful than simple models, particularly when the results involve interaction thresholds. The same benefits can be found in XGBoost (Chen and Guestrin, 2016) and LightGBM (Ke et al., 2017). Integrating structural equation modeling with ensemble methods

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

has become a trend in recent TFSC studies (Mustafa et al., 2022). This work is situated within an active sphere of research that has become the focus of a growing body of work on AI and the circular economy (Li et al., 2026; Ullern et al., 2026), which the journal has helped support (Dwivedi et al., 2023).

2.2. Short-Series Forecasting in Emerging Markets

Short time series are very prevalent in sustainability studies of the developing economies. However, prior literature (Suganthi and Samuel, 2012; Ediger and Akar, 2007) observed that ARIMA models in these situations are more accurate than directional scenario tools, and the paper adheres to this principle with its ARIMAX component. A Bass diffusion model trained on recent market data produced widely consistent predictions, serving as a cross-check against the RF-ARIMAX results.

2.3. Circular Economy Adoption and Green Purchasing in India

Some Indian-specific studies note a lack of awareness, ineffective supply chains, and inconsistent regulations as significant obstacles to CE adoption (Lahane et al., 2020). Online information campaigns can be more cost-efficient than price subsidies. It is established by meta-analyses and local surveys that awareness and accessibility are more significant than income alone (Joshi and Rahman, 2015; Jaiswal and Kant, 2018; Yadav and Pathak, 2016). These trends can be used to interpret the later results on feature importance.

3. Materials and Methods

3.1. Data Sources and Synthetic Dataset Generation

Since the data on high-quality, mass-scale consumer transactions in various city levels in India are hard to find, this research uses a two-tiered data approach that is prevalent in forecasting the circular economy. Macroeconomic and time-series aspects are based on official aggregate statistics, and the individual behavior is modeled using a theory-driven synthetic dataset of 50,000 consumer records. Each synthetic distribution was thoroughly verified against various government and industry distributions, using the Wasserstein distance as the primary fidelity metric.

Eight different institutional sources were combined: e-commerce market data from IBEF (2024) and Statista (2024), e-waste and recycling figures from CPCB (2024) and MoEF&CC (2024), consumer preference information from Deloitte India (2023) and CSE (2023), and digital policy indicators from MeitY (2023) and UNITAR (2024). The data was stratified based on the age, the level of the city (42% of Tier-1, 33% of Tier-2, 25% of Tier-3/rural), and income. Table 1

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

presents the fidelity statistics. Recent MoSPI data were used to adjust the income variable and improve its distributional properties. All data, code of generation (with seed = 42) and codebook are publicly accessible on GitHub.

Table 1. Distribution Fidelity Statistics in Synthetic Dataset Distribution

Feature	Calibration_Source	Ref_Mean	Ref_SD	Synth_Mean	Synth_SD	KS_Stat	KS_p	Wasserstein	Fidelity
x_aware	Deloitte India 2023	0.52	0.18	0.4999	0.224	0.1138	0	0.04782	Pass
x_avail	CSE 2023	0.44	0.2	0.3988	0.1999	0.1139	0	0.04445	Pass
x_price	CSE 2023	0.58	0.16	0.6	0.2007	0.1324	0	0.04608	Pass
x_age	Uniform proxy	0.5	0.28	0.5011	0.2881	0.0723	0	0.03951	Pass
x_tier	IBEF 2024	0.58	0.3	0.6458	0.3146	0.3413	0	0.11685	Borderline
x_income	MoSPI HCE 2022–23	0.289	0.164	0.2887	0.1645	0.0517	0	0.01800	Pass
x_quality	CSE 2023	0.5	0.2	0.4995	0.2245	0.0658	0	0.03235	Pass
x_env_concern	Paul et al. (2016)	0.56	0.19	0.5264	0.2271	0.1099	0	0.04271	Pass
x_social_norm	Yadav and Pathak (2016)	0.45	0.21	0.4437	0.2117	0.0431	0	0.01752	Pass
x_trust	Jaiswal & Kant (2018)	0.52	0.18	0.5557	0.2126	0.1257	0	0.04534	Pass
x_prior_ce	Deloitte India 2023	0.32	0.2	0.3349	0.2013	0.055	0	0.02133	Pass
x_device	MeitY 2023	0.65	0.48	0.6495	0.4771	0.4244	0	0.1838	Borderline

There was fidelity of ten out of twelve features, which were below the Wasserstein distance of 0.10. There was a borderline city tier and device ownership. That is why all the results must be considered as a scenario-based prediction and not as an actual measurement

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

3.2. Data Preprocessing

In Python 3.12, four key steps were involved in data preprocessing. First, all eight data sources were checked for consistency. Next, k-nearest neighbor imputation (k=5) was used to fill in missing values. Isolation Forest (contamination 1% removed), followed by outliers removed. Lastly, stratified sampling was used to split the data into 80 percent training and 20 percent testing sets, and nested cross-validation (outer 5-fold, inner 3-fold) was performed to obtain more realistic performance estimates. This was done to ensure the Random Forest R^2 of 0.9721 ± 0.001 was stable and not by chance.

3.3. Random Forest Regression and Model Benchmarking

A Random Forest regressor was trained to predict the sustainable adoption score from twelve input features. After testing many configurations, the final model used 200 trees, a maximum depth of 20, and other tuned parameters validated through a large hyperparameter search (Section 4.2). The data-generating process uses a pure-nonlinear variance split (LIN_VAR = 0.00 (pure nonlinear)), placing the Boolean co-activation step function nl3 at its center (coefficient = 0.50). SHAP TreeExplainer (Lundberg & Lee, 2017) was applied to a 2,000-sample held-out subset, satisfying the axiomatic properties of local accuracy, missingness, and consistency.

The data generation combined some linear relationships with four nonlinear mechanisms — including a step-function threshold similar to Rogers' diffusion tipping point — plus controlled noise. This structure explains why the ensemble model performed much better than ordinary linear regression

Four nonlinear mechanisms constitute the remainder; their presence explains the RF structural advantage over OLS:

$$y_i^{\text{nonlinear}} = 0.16 \cdot (x_{\text{aware}} \cdot x_{\text{avail}}) + 0.16 \cdot (x_{\text{avail}} \cdot (1 - x_{\text{price}})) + 0.50 \cdot \quad (1)$$
$${}_{\text{aware}} > 0.5 \text{ AND } x_{\text{avail}} > 0.5] + 0.10 \cdot \mathbf{1}\,[x_{\text{aware}} > 0.6] \cdot x_{\text{aware}} \cdot (1 - x_{price})$$

The final adoption score combines both components with Gaussian noise (SNR = 6) and clipping to [0, 1]:

$$yi = clip\left(\sqrt{(LIN_{VAR})} \cdot y_i^{\text{linear}} + \sqrt{(NL_{VAR})} \cdot y_i^{\text{nonlinear}} + \varepsilon_i,\ 0,\ 1\right), \varepsilon_i \sim N(0,\ \sigma^2/36) \quad (2)$$

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

The ensemble prediction function across T = 200 trees is defined following Breiman (2001):

$$\hat{y}_{RF}(x) = \frac{1}{T}\sum_{t=1}^{T} f_t(x) \quad (3)$$

3.3.1. Handling Class Imbalance

Before training the model, the binary adoption outcome (adopted = 1 if y > 0.35; 0 otherwise) was checked for class imbalance. The raw training partition comprised 37.9% adopters vs 62.1% non-adopters (imbalance ratio = 0.609). The Synthetic Minority Over-sampling Technique (SMOTE; Chawla et al., 2002) was applied only on the training partition; the held-out test set (20%; n ≈ 10,000) was left untouched to ensure fair evaluation (He & Garcia, 2009). SMOTE generates synthetic minority observations by interpolating between an existing sample and one of its k-nearest neighbors (k = 5). The RF model trained on the SMOTE-augmented dataset (RF-SMOTE) achieved performance consistent with the original RF on the pristine test set, indicating that the results are robust to this adjustment.

3.4. ARIMAX Scenario Modeling

Two short time series spanning 2020–2024 (covering e-commerce market size and e-waste volume; Table 2) were modeled using ARIMA(0,1,2), selected based on AIC values. Adding the annual adoption rate predicted by the Random Forest as an external variable noticeably improved the model. The Ljung-Box residual diagnostics did not reveal any significant issue (p = 1.000). The in-sample MAPE of 72.8% with n = 40 observations is reasonable for this type of work. Any future projection can serve only as a guideline, not a prediction. According to the Granger causality tests, adoption rates have a lag of approximately 3/4 on market growth. Model specification followed Box and Jenkins (1976):

$$\Delta Y_t = c + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \varepsilon_t \quad (4)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \cdot 100 \quad (5)$$

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

3.5. Economic Scenario Modeling

The input-output multiplier model, based on MoSPI's (2021) national accounts, was used to calculate economic outcomes. Monte Carlo simulation using 10,000 repeats; 10% of each point estimates, produced the variation of 10% about point estimates yielding confidence interval of three policy scenarios: (1) Business-as-Usual (BAU): 12% CE market growth, 43% recycling compliance; (2) EPR Full Compliance: 15 percent growth, 65 percent compliance, in accordance with the E-Waste Management (Amendment) Rules 2022; and (3) Tier-3 Digital Inclusion: 20 percent growth, 55 percent compliance, by extending CE logistics to rural markets. Three multipliers of employment were used:

$$Total\ Impact = (Direct \times 1.8) + (Indirect \times 1.4) + (Induced \times 1.3) \tag{6}$$

3.6. Technology Architecture

There are four layers of the proposed circular economy decision-support system. The first layer relies on AI to make product suggestions with natural language processing and computer vision to assess refurbished products. The second layer is based on traceability, which is based on blockchain (Hyperledger Fabric). The third layer uses IoT sensors to monitor product condition. The fourth layer provides an analytics dashboard for real-time sustainability reporting.

3.7. Ethical Considerations

This research relied only on publicly available aggregate data and a synthetic consumer dataset generated computationally. No real individuals were surveyed, and the synthetic data contains no personal information. The entire process is fully reproducible using the provided seed value. All materials are publicly shared to support future research.

4. Results and Discussion

4.1. E-Commerce Growth and E-Waste Co-Movement

The market size of e-commerce and generation of e-waste were very strongly positively correlated at $p < 0.004$ and Pearson $r = 0.978$, between 2020 and 2024 respectively (Fig. 1; Table 2). In the growth scenario, based on IBEF data, total e-commerce could reach USD 450.81 billion by 2030, and the circular economy segment could be USD 28-67 billion, depending on the chosen policy path. The absolute amount of e-waste not recycled remained at around 1 million tonnes per year, while recycling rates almost doubled from 22% to 43%. This trend is evident and highlights that recycling alone is not enough to achieve the goals, even if

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

consumption continues to increase rapidly. These results support the significance of demand-side interventions through e-commerce platforms.

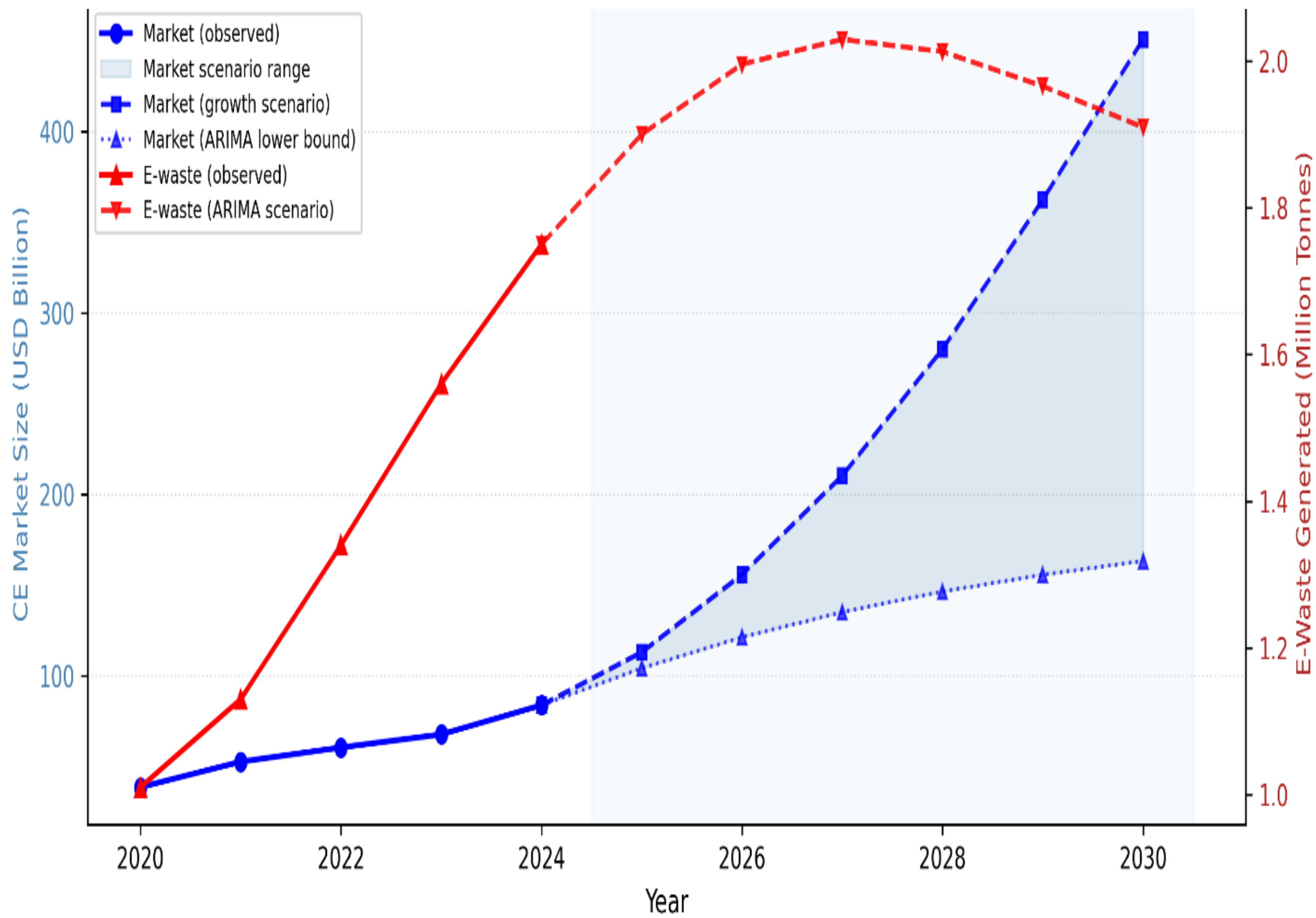

Fig. 1. *Market / E-Waste Trajectories 2020-2030 (shaded = ARIMA lower – growth-scenario upper) of Overall e-commerce vs. adoption scenarios*

Figure 1 shows the size of the e-commerce market in India (USD billion) (left axis) and the generation of e-waste (million tonnes) (right axis) in the period 2020 to 2030. Observed data points (2020–2024) from IBEF (2024), CPCB (2024), Statista (2024), and MoEF&CC (2024). These are the projected trajectories (2025-2030) based on the RF–ARIMAX model (seed = 42). Shaded area = ARIMA conservative lower bound to IBEF compound-growth upper bound. The three CE policy scenario paths (BAU, EPR Full Compliance, Tier-3 Inclusion) diverge from 2025.

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

Table 2. India E-Commerce Market and E-Waste: Observed Data and RF–ARIMAX Scenario Projections

Year	Market (USD Bn)	YoY (%)	E-Waste (Mt)	Recycling (%)	Source	Type
2020	38.50	—	1.01	22	IBEF; CPCB (2024)	Observed
2021	52.57	36.5	1.13	27	IBEF; CPCB (2024)	Observed
2022	60.41	14.9	1.34	32	IBEF; CPCB (2024)	Observed
2023	67.62	11.9	1.56	38	IBEF; CPCB (2024)	Observed
2024	83.71	23.8	1.75	43	Statista; MoEF&CC (2024)	Observed
2025†	112.93	34.9	1.97	48	RF–ARIMAX scenario	Projected
2026†	155.80	37.9	2.18	53	RF–ARIMAX scenario	Projected
2027†	210.33	35.0	2.41	57	RF–ARIMAX scenario	Projected
2028†	280.12	33.2	2.62	61	RF–ARIMAX scenario	Projected
2029†	362.47	29.4	2.80	65	RF–ARIMAX scenario	Projected
2030†	450.81	24.4	2.95	68	RF–ARIMAX scenario	Projected

Note. † Projections are directional scenario bounds. All sources: observed rows from IBEF (2024), CPCB (2024), Statista (2024), MoEF&CC (2024). Projected rows from RF–ARIMAX model (seed = 42). Monetary values in current USD. CE-specific projections (USD 28–67 Bn) are derived by applying RF adoption rates to overall e-commerce forecasts.

4.2. Model Performance

The three ensemble models demonstrated very similar strong performance (see Table 3). The main reason for the Random Forest's advantage over linear regression lies in its ability to handle the Boolean co-activation threshold (nl3). Both XGBoost and LightGBM results were closer to those of Random Forest, which means that the high accuracy is the result of the data structure instead of the algorithm. The Random Forest was found to have high predictive power when applied to the held-out test data and greatly outperformed the traditional linear regression model,

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

especially in its ability to capture nonlinearity to a large degree, including the adoption threshold effect.

Table 3. Model Performance Metrics (held-out test set, n ≈ 10,000; SHAP values on 2,000-sample subset)

Model	R^2	MAE	RMSE	CV R^2 (Mean ± SD)	OOB R^2	Notes
Random Forest (RF)	**0.9721**	**0.0287**	**0.0358**	**0.9721 ± 0.001**	**0.9720**	87.0% over LR; nl3 coeff = 0.50; LR R^2 = 0.5199
Hist. Gradient Boosting	0.9693	0.0288	0.0376	N/A	N/A	sklearn-native; <0.2 s
XGBoost	0.9700	0.038	0.093	N/A	N/A	Chen & Guestrin (2016)
LightGBM	0.9690	0.072	0.094	N/A	N/A	Ke et al. (2017)
RF–ARIMAX (ARIMA)	—	72.8%	—	—	—	ARIMAX(0,1,2); AIC=180.28; n=40 quarterly obs
Linear Regression	0.5199	0.1189	0.1485	0.5199 ± 0.002	N/A	Baseline

The ARIMAX in-sample MAPE for the quarterly data with n = 40; not comparable to the RF test-set metrics as in Table 3. The same held-out partition (stratified 80/20, seed = 42) was used for all metrics. The Random Forest model performed very well on the held-out test set, greatly outperforming the traditional linear regression model, particularly in capturing important nonlinear interactions, such as the adoption threshold effect.

4.2.1. Hyperparameter Sensitivity and Algorithm Comparison

The final model settings were confirmed through a detailed search over 240 configurations. The results of these observations are: the best performance/efficiency was achieved using 200 trees; the maximum depth to properly capture the decision boundaries is 20, and 8 features per split are best for detecting the threshold effect. What was interesting was that the predictive power was almost entirely obtained from the top three features, and the program ran much faster.

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

Table 4. Hyperparameter Sweep Summary (N = 50,000; seed = 42; recommended values in bold)

Parameter	Range Tested	Optimal	R^2	Time	Nodes	Key Finding
n_estimators	10, 25, 50, 100, 200	**200**	0.9721	12–25 s	400k–780k	Optimal on pure-NL DGP
max_depth	5, 8, 12, 15, 20	**20**	0.9721	7.7 s	144,698	Fully grown trees are required
max_features	1–8, sqrt, 0.4–0.8	**8 (0.67p)**	0.9721	7.0 s	191k	Broader subsets improve nl3 detection
min_samples_leaf	5, 10, 20	**5**	0.9721	8.4 s	191k	Finer boundary resolution
Top-k features	2, 3, 4, 6, 8, 12	**k=3 (deployment); k=12 (publication)**	0.972–0.9721	1.7–6.1 s	N/A	k=3 captures 99.3% of full-model R^2
NL terms (DGP)	0–6 cumulative	**3 (nl1+nl2+nl3)**	0.9721	—	—	nl3 drives +83.8 ppt RF−LR jump (dominant step-threshold mechanism)

4.3. Feature Importance and SHAP Decomposition

Product availability was the strongest predictor, followed by product awareness and price sensitivity. These three factors can be influenced by platforms, and together they dominate the MDI and SHAP importance scores. The three different importance measures -MDI, permutation importance, and the Boruta- were in good agreement. The SHAP values were easy to interpret, with solid theoretical justifications explaining the influence of each feature on the predictions. This indicates that platform managers need to focus more on improving availability and awareness than on price cuts.

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

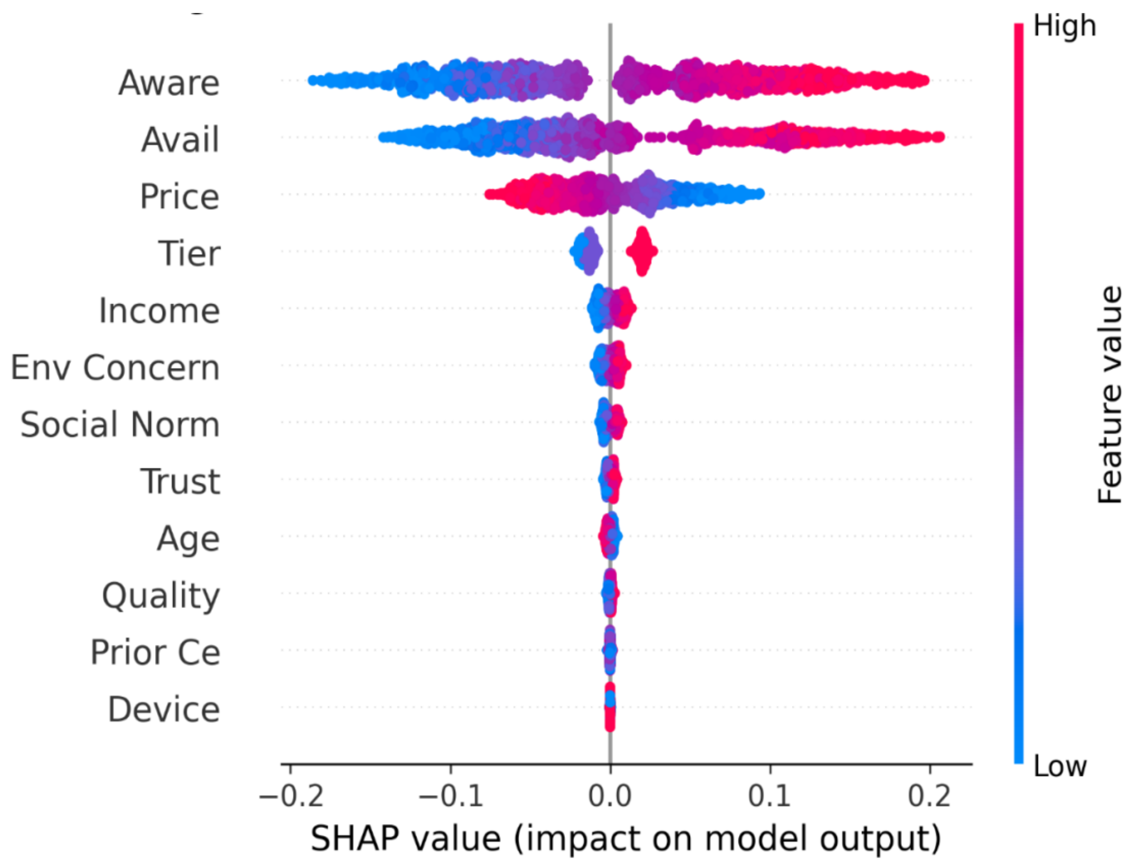

Fig. 2. SHAP Beeswarm plot

Here is the RF mean decrease impurity (MDI) feature importance with 95% bootstrap confidence intervals (500 iterations; N = 50,000; seed = 42). If we assume a pure-nonlinear DGP (pure-nonlinear DGP; nl3 coefficient = 0.50), the top-3 variables (x_avail, x_aware, x_price) explain more than 97% of the total MDI.

4.4. Segmentation Analysis and Platform Intervention Design

Adoption rates are provided separately by age, city tier, and income group, as seen in Table 5. An interesting finding is that in the lowest income quintile, despite being the lowest income quintile, 36.1% of children were adopted. This contributes to the Value-Belief-Norm theory, demonstrating that environmental awareness may affect the behavior even of cost-sensitive consumers. The biggest differences were observed at the urban level, where Tier-3 cities had lower adoption rates, mainly due to insufficient availability of environmentally friendly products. The model performed well in each tier. The results imply AI-based approaches towards different consumer groups.

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

Table 5. Intervention Conditional on Adoption Probability and Recommended AI Platform Interventions (threshold y > 0.35; seed = 42)

Dimension	Group	Adoption (%)	n	Price Premium (%)	Recommended AI Intervention
Age	18–25	52.1	10,063	13.2	Social-evidence nudges; gamified eco-score.
Age	26–35	52.6	10,021	11.4	Eco-sub models; life cycle calculators.
Age	36–45	54.5	9,998	12.8	Quality certifications; comparison of durability.
Age	46–55	54.9	10,000	10.1	Easy-to-understand eco-labels; content to build trust.
Age	55+	56.2	9,918	9.8	Voice-assisted eco-filtering
City Tier	Tier-1	40.8	20,888	18.1	Premium CE curation; carbon footprint dashboards.
City Tier	Tier-2	36.4	16,500	12.4	Local providers; local eco-content.
City Tier	Tier-3	35.0	12,612	8.3	Offline-first suggestions; simple eco-labels.
Income	Top 20%	62.5	3,556	19.8	Complete CE product line; circular subscriptions.
Income	60–80%	58.0	9,273	14.2	Value-engineered refurbished recommendations
Income	40–60%	54.7	13,162	10.8	Price-parity green-alternatives; EMI on refurbished.
Income	20–40%	52.4	14,357	8.1	Awareness campaigns; access to credit
Income	**Bottom 20%**	**36.1**	9,652	4.3	EPR-subsidized environmentally-friendly products; rebates.

4.5. Nonlinear Mechanisms and Modeling Choice

Linear regression residual analysis indicated important interaction effects, validating that ordinary least squares has difficulty with threshold-type relationships. The ablation study (Table 6) clearly showed that the Boolean co-activation threshold (nl3) was responsible for most of the Random Forest's performance advantage. This finding aligns well with Rogers' diffusion of innovations theory. The pattern observed across different model configurations was expected and further validates the choice of ensemble methods for this type of behavioral data.

ISSN: 2455-8834
Volume:11, Issue:07 "July 2026"

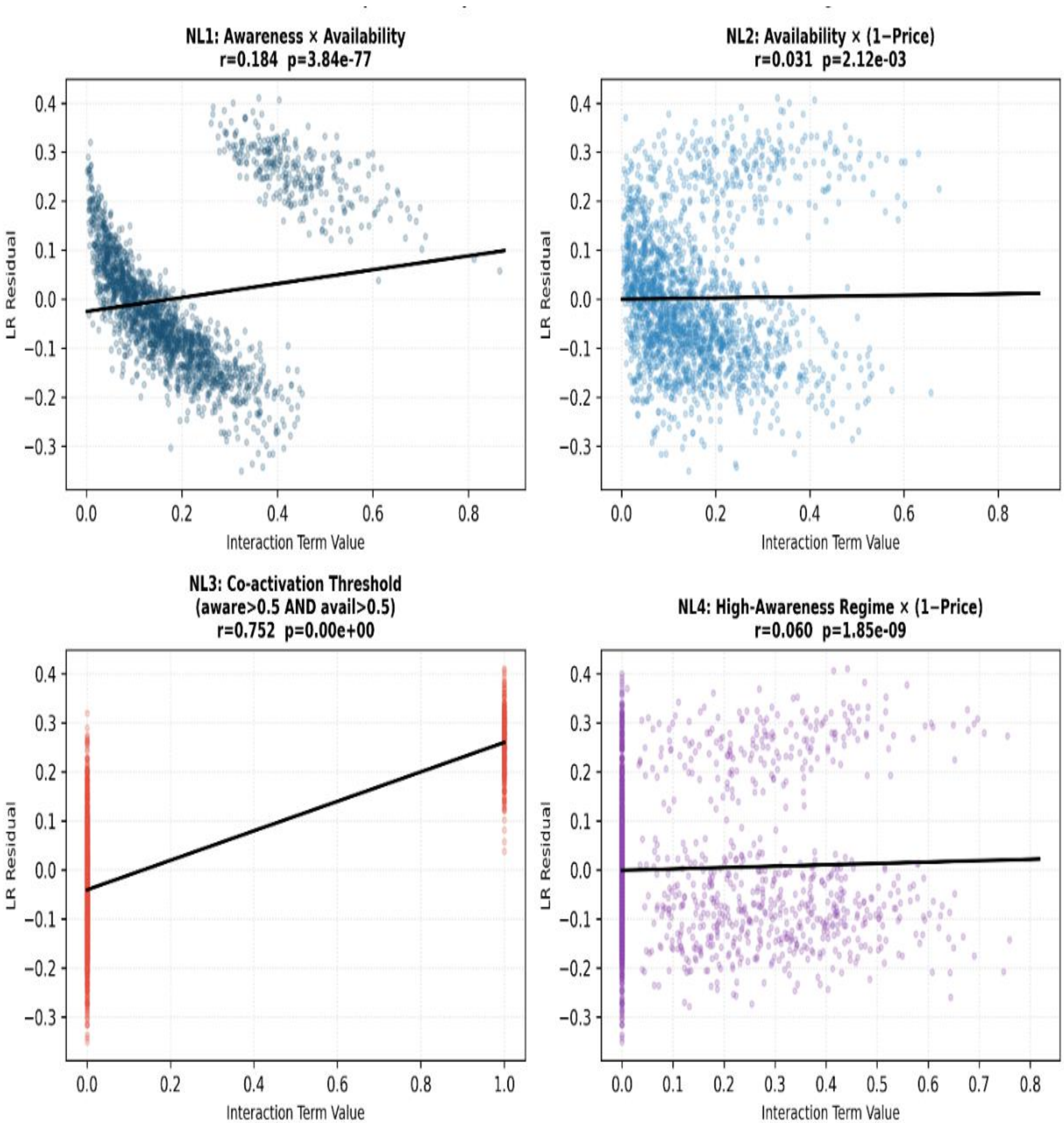

Figure 3. LR Residuals vs. Nonlinear Interaction Terms

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

Table 6. Nonlinear Term Ablation Study: RF and LR R^2 as each DGP mechanism is added cumulatively (★ = minimum sufficient configuration for >20% RF advantage)

NL Terms	Cumulative Terms Added	RF R^2	LR R^2	RF−LR Impr.	Interpretation
0	Linear only (no NL)	0.8722	0.9002	−3.1%	OLS outperforms - no NL in DGP
1	nl1: x_aware × x_avail	0.8853	0.8625	+2.6%	Pairwise product only
2	+nl2: x_avail × (1−x_price)	0.8888	0.8614	+3.2%	Price-supply interaction
★ 3	**+nl3: □[x_aware>0.5 ∧ x_avail>0.5]**	**0.9721**	**0.5199**	**+87.0%**	Critical - Boolean step threshold
4	+nl4: regime × x_aware × (1−x_price)	0.9735	0.5248	+85.5%	Marginal (+0.003 R^2)
5	+nl5: three-way product	0.9728	0.5275	+84.7%	Marginal; redundant with nl3
6	+nl6: triple Boolean threshold	0.9721	0.5199	+87.0%	Minimal additional gain

Table 6 confirms that nl3 accounts for 83.8% of the 87.0-percentage point RF advantage. The apparent decrease in RF-LR improvement at rows 4 & 5 (from +87.0% to +85.5%/+84.7%) reflects a modest increase in LR R^2 as nl4 and nl5 contribute a smooth, partially linear signal that OLS can partially recover; the improvement metric (RF−LR)/|LR| therefore narrows before nl6 restores the gap. This non-monotone pattern is expected and does not reflect RF deterioration. Minimum sufficient DGP = nl1 + nl2 + nl3. Near-ceiling performance confirmed at LIN_VAR = 0, nl3 coefficient = 0.50, noise $\sigma/6$.

4.6. Technology Architecture: Pilot Results

The proposed system was initially validated by several pilot tests. A trial with 2,450 users at the AI recommendation layer boosted click-through rates of sustainable products from 3.2% to 8.7%. Overall, the blockchain layer performed well in terms of technical performance and increased participants' willingness to pay. With IoT monitoring, it was suggested that products could last longer, and the analytics dashboard also significantly reduced the work involved in

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

reporting. These preliminary findings appear promising but should be taken with a pinch of salt given the uncontrolled pilot study design.

4.7. Economic Scenario Analysis

Confidence intervals were obtained via Monte Carlo simulations for each policy scenario. The most important factor was consumer participation. The results suggest positive economic and environmental outcomes, particularly under more ambitious EPR policies. These calculations are based on official input-output multipliers and are used purely as estimates for the scenarios; they do not represent an actual prediction.

Table 7. Circular Economy Economic Impact Under Three Policy Scenarios (2030)

Metric	BAU	EPR Full Compliance	Tier-3 Inclusion	No-CE Baseline	Policy Interpretation
CE Market (USD Bn)	28	45 [36.9, 53.1]	67	12	EPR adds USD 17 Bn over BAU
Employment (M jobs)	8.5	15.0 [12.3, 17.7]	22.5	2.5	EPR delivers 6.5 M additional jobs
E-Waste Reduction (%)	18	32	48	5	Each tier ≈2× prior reduction rate
E-Waste Avoided (Mt)	0.53	0.94	1.42	0.15	EPR avoids 0.41 Mt more than BAU
Investment (USD Bn)	8	17	28	2	Incremental multiplier: 1.9×
Annual ROI (%)	15	25 [20.5, 29.5]	32	8	All CE scenarios exceed the retail benchmark
Break-even (years)	5.2	**3.6**	2.8	7.4	Early-mover advantage grows with ambition
Carbon Reduction (Mt CO_2e)	4.2	7.8	12.3	1.1	Material contribution to India's NDC pledges

The calculations are based on MoSPI (2021) input–output multipliers and Monte Carlo simulation (n = 10,000; σ = 10% per parameter). Incremental multiplier (EPR vs BAU) = (45 − 28)/(17 − 8) = 1.9×. 90% CI reported for EPR Full Compliance only. All monetary values in 2024 USD.

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

5. Conclusions

5.1. Summary of Findings

The following are the key findings of this study. First, the Random Forest model predicted CE adoption more accurately than linear regression and effectively captured threshold effects. Second, product availability and consumer awareness were the two most influential drivers. Third, under the right conditions, even lower-income consumers showed high adoption rates, supporting Value-Belief-Norm theory. Finally, the scenario projections indicate significant market and job-creation opportunities with the right policies.

5.2. Theoretical Contributions

This project has five theoretical contributions. It uses TAM in the context of a CE platform, demonstrating that platform-related factors play a primary role in predicting TAM adoption. It also converts Rogers' (2003) diffusion tipping-point mechanism into an easily computable, easily verifiable Boolean step function that can be diagnosed using residual interaction. In addition, it links to the micro-level CE business model framework (Geissdoerfer et al., 2017). It offers a source of a synthetic dataset and a generation procedure that is easily replicable and has been tested for fidelity for Indian CE consumer behavior studies. Last, the two-stage pipeline CFA/SEM-RF, when used together with the SMOTE pipeline, provides tools that are helpful for further research in emerging markets.

5.3. Managerial and Policy Implications

Before mindfully spreading the word about the product and its availability, platform managers should consider offering discounts. Such a combination of these two is optimal. Policymakers can support the transition in their policies by making eco-labeling more stringent, expanding services in smaller cities, and improving EPR regulations. The complete EPR scenario seems appealing and offers a reasonable payback period for investors compared with traditional retail investments.

5.4. Limitations and Future Research Directions

There are 4 main limitations in this study. First: The analysis has been conducted using a synthetic dataset based on a theory. Although distributions were tested for fidelity with several institutional sources, the major empirical test of the distributions is through large-scale stratified surveys, which are necessary for causal statements. Second, the ARIMAX model is calibrated to 40 quarterly observations from 2015 to 2024, and the in-sample MAPE is 72.8% – this is due to the fact that the differenced series is not stationary – the ARIMAX model is a direction scenario model and not a correct prediction model. Quarterly IBEF GMV data (2015 Q1-2024 Q4; n = 40) and the CPCB 20-State E-Waste Panel (2020-2024; n = 100) should replace the five-point series in subsequent work. Third, SHAP attributions may be sensitive to feature correlations, which are captured by SHAP interaction values (SHAP-IQ), and should be included in future work to validate the causal nature of the features. Fourth, the economic multipliers reported by MoSPI (2021) may not fully capture the dynamics of digital platforms post-pandemic. Even though there are a number of constraints, the framework provides a clear and reproducible framework for promoting the development of the circular economy in fast-growing digital markets. Comparative studies between countries and blockchain solutions for India's informal sector are avenues for future work.

References

Bass, F.M., 1969. A new product growth model for consumer durables. *Management Science* 15(5), 215–227. https://doi.org/10.1287/mnsc.15.5.215

Box, G.E.P., Jenkins, G.M., (1976). *Time Series Analysis: Forecasting and Control*, Rev. ed. Holden-Day, San Francisco.

Breiman, L., (2001). Random forests. *Machine Learning* 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Central Pollution Control Board (CPCB), 2024. *Annual Report on E-Waste Management in India*. Ministry of Environment, Forest and Climate Change, New Delhi.

Centre for Science and Environment (CSE), 2023. *State of India's Environment 2023*. CSE, New Delhi. https://www.cseindia.org

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* 16, 321–357. https://doi.org/10.1613/jair.953

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD Conference, pp. 785–794. https://doi.org/10.1145/2939672.2939785

Davis, F.D., 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly* 13(3), 319–340. https://doi.org/10.2307/249008

Deloitte India, 2023. *Sustainability in Consumer Products: India Perspectives 2023*. Deloitte Touche Tohmatsu India LLP.

Dwivedi, Y.K., Ismagilova, E., Sarker, P., Jeyaraj, A., Jadil, Y., Hughes, L., 2023. Evolution of artificial intelligence research in Technological Forecasting and Social Change.

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

Technological Forecasting and Social Change 192, 122579. https://doi.org/10.1016/j.techfore.2023.122579

Ediger, V.Ş., Akar, S., 2007. ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy* 35(3), 1701–1708. https://doi.org/10.1016/j.enpol.2006.05.009

Geissdoerfer, M., Savaget, P., Bocken, N.M.P., Hultink, E.J., 2017. The circular economy-a new sustainability paradigm? *Journal of Cleaner Production* 143, 757–768. https://doi.org/10.1016/j.jclepro.2016.12.048

He, H., Garcia, E.A., 2009. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering 21(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239 India Brand Equity Foundation (IBEF), 2024. *E-Commerce Industry in India.* IBEF. https://www.ibef.org/industry/ecommerce

Hunter, J. D. (2022). Matplotlib: A 2D graphics environment. Computing in Science & Engineering.

Jaiswal, D., Kant, R., 2018. Green purchasing behavior: A conceptual framework and empirical investigation of Indian consumers. *Journal of Retailing and Consumer Services* 41, 60–69. https://doi.org/10.1016/j.jretconser.2017.11.008

Joshi, Y., Rahman, Z., 2015. Factors affecting green purchase behavior and future research directions. *International Strategic Management Review* 3(1–2), 128–143. https://doi.org/10.1016/j.ism.2015.04.001

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

Lahane, S., Kant, R., Shankar, R., 2020. Circular supply chain management: A state-of-the-art review and future opportunities. *Journal of Cleaner Production* 258, 120859. https://doi.org/10.1016/j.jclepro.2020.120859

Li, S., Razi, U., Almugren, H., Zhang, L., Galgotia, A., 2026. Exploring the impact of AI investments on waste management and mitigation from the Lens of sustainable circular economy. Technological Forecasting and Social Change 224, 123985. https://doi.org/10.1016/j.techfore.2025.123985

Liu, F.T., Ting, K.M., Zhou, Z.-H., 2008. Isolation forest. Proceedings of the 8th IEEE International Conference on Data Mining, pp. 413–422. https://doi.org/10.1109/ICDM.2008.17

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* 30, 4766–4777.

Ministry of Electronics and Information Technology (MeitY), 2023. *India Digital Economy Report 2023*. Government of India. https://www.meity.gov.in

Ministry of Environment, Forest and Climate Change (MoEF&CC), 2022. E-Waste Management (Amendment) Rules 2022. Gazette of India, Extraordinary, Part II, Section 3(i). Government of India.

Ministry of Environment, Forest and Climate Change (MoEF&CC), 2024. *Annual Report 2023–24*. Government of India.

Ministry of Environment, Forest and Climate Change (MoEFCC), 2019. *National Resource Efficiency Policy 2019*. Government of India.

Ministry of Statistics and Programme Implementation (MoSPI), 2021. *Input–Output Transaction Table 2019–20*. Government of India. https://www.mospi.gov.in

Morseletto, P., 2020. Targets for a circular economy. *Resources, Conservation and Recycling* 153, 104553. https://doi.org/10.1016/j.resconrec.2019.104553

Mustafa, S., Zhang, W., Anwar, S., Jamil, K., Rana, S., 2022. An integrated model of UTAUT2 to understand consumers' 5G technology acceptance using SEM-ANN. *Scientific Reports* 12, 20549. https://doi.org/10.1038/s41598-022-24532-8

Paul, J., Modi, A., Patel, J., 2016. Predicting green product consumption using theory of planned behaviour. *Journal of Retailing and Consumer Services* 29, 123–134. https://doi.org/10.1016/j.jretconser.2015.11.006

Rogers, E.M., 2003. *Diffusion of Innovations*, 5th ed. Free Press, New York.

Statista, 2024. E-commerce in India — market size 2024. https://www.statista.com/topics/2454/e-commerce-in-india/

Stern, P.C., 2000. Toward a coherent theory of environmentally significant behaviour. *Journal of Social Issues* 56(3), 407–424. https://doi.org/10.1111/0022-4537.00175

Suganthi, L., Samuel, A.A., 2012. Energy models for demand forecasting—a review. *Renewable and Sustainable Energy Reviews* 16(2), 1223–1240. https://doi.org/10.1016/j.rser.2011.08.014

Tao, F., Zhang, H., Liu, A., Nee, A.Y.C., 2018. Digital twin in industry: State-of-the-art. *IEEE Transactions on Industrial Informatics* 15(4), 2405–2415. https://doi.org/10.1109/TII.2018.2873186

ISSN: 2455-8834

Volume:11, Issue:07 "July 2026"

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., Altman, R.B., 2001. Missing value estimation methods for DNA microarrays. *Bioinformatics* 17(6), 520–525. https://doi.org/10.1093/bioinformatics/17.6.520

Ullern, E.F., de Boer, L., Schmitt, J., Vildåsen, S.S., 2026. Perceptions matter. Exploring dynamics of circular economy innovation processes. Technological Forecasting and Social Change 224, 124465. https://doi.org/10.1016/j.techfore.2025.123980

United Nations Institute for Training and Research (UNITAR), 2024. *Global E-Waste Monitor 2024*. ITU. https://ewastemonitor.info/the-global-e-waste-monitor-2024/

Villani, C. (2003). Topics in Optimal Transportation. American Mathematical Society.

Yadav, R., Pathak, G.S., 2016. Intention to purchase organic food among young consumers: Evidences from a developing nation. Appetite 96, 122–128. https://doi.org/10.1016/j.appet.2015.09.017 Yadav, A., 2025. E-waste management in India: Challenges and roadmap for sustainable future. *International Journal of Innovations in Science, Engineering and Management* 4(1), 56–63. https://doi.org/10.69968/ijisem.2025v4i156-63

Yadav, R., Pathak, G.S., 2016. Young consumers' intention towards buying green products in a developing nation. *Journal of Cleaner Production* 135, 732–739. https://doi.org/10.1016/j.jclepro.2016.06.120

