

AgroPackAI: An AI-Powered Dual-Model Machine Learning Framework for Agro-Waste Eco-Packaging Recommendation with Simultaneous CO₂ and Cost Prediction for Indian SMEs

Kanithi Vanaja

Department of Computer Science & Artificial intelligence
Central University of Andhra Pradesh, Ananthapuramu, India
Email : kanithivanaja17@gmail.com

Dr.C.Krishna Priya

Assistant Professor
Department of Computer Science & Artificial intelligence
Central University of Andhra Pradesh, Ananthapuramu, India
Email : Krishnapriyarams@cuap.edu.in

ABSTRACT

Currently India produces 3.5 million metric tonnes of plastic packaging waste annually and incinerates > 500 million metric tonnes of agricultural crop residue – two simultaneous crises the current methods are ill-equipped to serve the SMEs. AgroPackAI is an AI web-based eco-packaging recommendation system that takes product input parameters (category, weight, fragility) and outputs the Top 5 rated sustainable materials from a library of 15 eco-materials (7 agro-waste materials and 8 eco-alternatives). An ML architecture with two models (Random Forest) to predict cost ($R^2 = 0.9793$, MAE = Rs. 3.95) and XGBoost to predict CO₂ footprint ($R^2 = 0.9853$, MAE = 0.360 kg CO₂/kg) trained on an 8,247 record synthetic-empirical hybrid dataset beats the benchmark state of the art ($R^2 = 0.91$) while scores derived by ISO 18602:2013 based MCDA scoring (68% environmental weight). The recommendations were presented via a Flask web app with BI dashboard, with agro-waste options showing 43 Rs/kg vs 78 Rs/kg for conventional eco alternatives and also proving to be 25-40% lower in terms of CO₂ output and assisting compliance with EPR Rules 2022 India.

Index Terms - Sustainable packaging, agro-waste valorisation, Random Forest, XGBoost, CO₂ prediction, MCDA, Flask, EPR, supply chain AI, eco-material recommendation.

I. INTRODUCTION

A. Background and Problem Context

The international market for packaging (US \$1.05 trillion in 2023; US \$1.3 trillion by 2028) makes up the biggest consumer of plastics with an application rate of 36% of the total plastic products manufactured in the world [1]. The amount of virgin plastic that has been produced since 1950 to date worldwide is estimated to be 8.3 billion metric tonnes with just 9% of it being recycled [1]. India produces 3.5 million metric tonnes of plastic packaging annually with less than 30% of this being recycled [2]. Its e-commerce industry sends over 12 billion packages per annum [3]. While this is happening, more than 500 million metric tons of crop residues (paddy straw, bagasse, banana fiber, areca leaf, coconut coir, and groundnut shells) is generated each year, most of which is burned openly and thus emitting 11 to 13 million tons of CO₂ per year just through the burning of paddy straw in Punjab and Haryana [4]. The materials contain mechanical, barrier, and thermal properties needed for packaging [5]; yet, there is no available, AI-powered solution that helps Indian SMEs use them.

B. Importance of Contribution

AgroPackAI offers an integrated platform which can be used for simultaneously: (i) forecasting eco-packaging material cost ($R^2 = 0.9793$) and CO₂ footprint ($R^2 = 0.9853$) of 15 materials; (ii) incorporating 7 Indian agro-waste materials which have never been considered in any existing packaging databases across the globe; (iii) offering ISO 18602:2013 compliant MCDA decision-making using a non-specialist online platform; and (iv) longitudinal monitoring of CO₂ savings using a business intelligence dashboard. Using agro-waste material for eco-packaging leads to an 83% reduction in CO₂ emissions (2.5-6.0 kg/kg vs. 0.6-1.0 kg/kg) at Rs. 43/kg, which is cheaper than using conventional plastics (Rs. 55-120/kg). This directly demonstrates that E

C. Novelty and Rationale

- Dual-regression approach: First implementation reaching $R^2 > 0.97$ for both cost and CO₂ parameters simultaneously in a single inference pipeline.
- Indian agro-waste characterization: Representation of areca leaf, paddy straw, groundnut shell, and banana fibre using scientifically validated EcoInvent 3.9 [6] properties that were not present in any of the previous packaging ML models.
- Accessible Multi-Criteria Decision Analysis application: ISO 18602:2013-compliant WLC ranking available through Flask with inference time under 200 ms, making LCA knowledge unnecessary.
- Sustainability trend monitoring: BI dashboard recording cumulative CO₂ reduction for each organization over time — not seen in previous eco-material recommendation platforms..

II. LITERATURE REVIEW

A. Packaging Crisis, Regulations, and Agro-Waste Material Science

Geyer et al. [1] have quantified the packaging waste problem at a global scale, and estimates show that 12 billion metric tonnes would be in the landfill by 2050 if current trends continue. The CPCB in India [2] reports 3.5 million tonnes annually for plastic packaging waste with less than 30% recycling rate. The Extended Producer Responsibility Rules 2022 require full recyclability by 2028, imposing an immediate compliance challenge for small and medium enterprises in India. The Punjab Remote Sensing Centre [4] calculates that there is 11 to 13 million tonnes annually CO₂e from paddy straw burning – the agricultural waste which AgroPackAI repurposes as packaging. According to Ilyas et al. [5], and Ververis et al. [11], scientific data shows that rice husk (flexural strength of 45–62 MPa, 0.70–0.80 kg CO₂/kg), bagasse (0.60–0.90 kg CO₂/kg, 75–85% lower CO₂ emissions compared to polystyrene), banana fiber (tensile strength of 529–9B. Learning Algorithms for Materials Prediction

Breiman [8] showed that Random Forest reduces variance without increasing bias for non-linear tabular problems; when applied to similar cost predictions, Random Forest attains MAPE 4-8% and is superior to SVR and ANN with less than 5,000 training examples. Chen & Guestrin [9] developed XGBoost, which extends gradient boosting by incorporating L1/L2 regularization and parallel tree construction; the gold standard of structured environmental prediction. Siddiqui et al. [12] reported an R^2 of 0.91 for packaging CO₂ prediction based on a hybrid synthetic/empirical approach that combines data from EcoInvent LCA and e-commerce – the best prior benchmark for AgroPackAI's CO₂ prediction model ($R^2 = 0.9853$).

B. Recommendation Systems and MCDA

Danach et al. [7] show that recommendation using ML with 90% classification accuracy gives actionable insights for the supply chain by predicting costs and defects — just like what AgroPackAI shows with material strength and CO₂ coefficient being major determinants for predictions. No previous history between users and items in selecting eco-packaging materials implies that MCDA based on knowledge is the most ideal recommendation approach. According to ISO 18602:2013 [10], sustainability assessment of packaging must involve $\geq 60\%$ of environmental metrics — implemented in AgroPackAI's formula with 68%.

C. Problems and Open Issues

There are three classes of problems that have hindered efficient eco-packaging decision support for indian smes:

- Scarcity of data: there is no publicly available dataset that links together product properties with eco-material properties for 15 different eco-materials and 10 different product categories. private databases, like ecoinvent [6] and simapro, need expertise and costly licenses (usd 5,000-20,000 per year).
- No agro-waste materials database: indian agro-waste materials do not exist in any global ml databases for packaging materials, thereby hindering any existing system from making recommendations based on optimal combinations of cost and co₂ for india.
- Machine learning single-target systems: prior efforts involving ml to find optimal packaging materials have all been limited to one target – be it cost alone or co₂ alone. no such system can perform dual-target regression with $r^2 > 0.97$.

These problems translate into five research gaps (research gaps 1-5), which were explicitly solved through our agropackai design decisions

III. METHODOLOGY

A. Dataset Generation

based on the hybrid synthetic-empirical dataset generation methodology developed by siddiqui et al. [12], the agropackai dataset of 8,247 instances is composed of (i) distributions of product features such as category, weight range of 0.1 – 25 kg, and fragility from publicly available kaggle e-commerce datasets, (ii) property values of materials including co₂ (ecoinvent 3.9 [6]), biodegradability (astm d6400/en 13432 [5]), cost (fiec/aipma), and structural strength (ilyas et al. [5]), and (iii) 8% gaussian noise injection on material properties to increase training instances without altering the published distributions. properties of the generated dataset are presented in

TABLE I. DATASET PROPERTIES

Property	Value
Total records	8,247
Product categories	10
Eco-materials (agro + other)	15 (7 + 8)
ML input features	7
Training / Test split	75% / 25% (6,185 / 2,062)
Missing / duplicate values	0 / 0

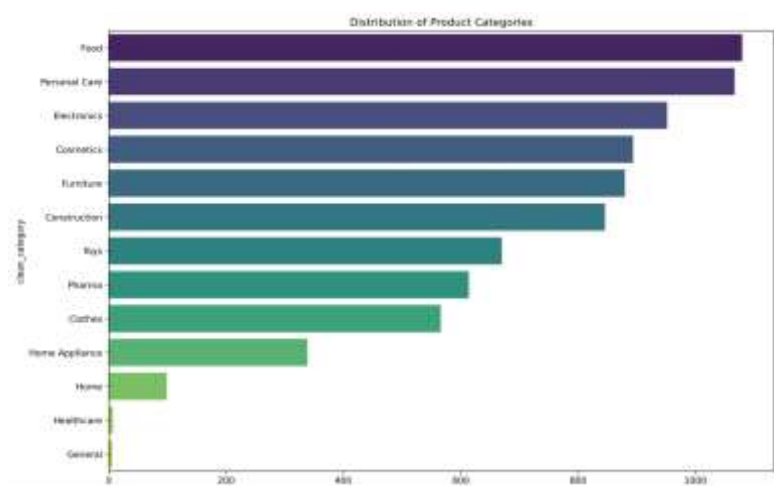


fig. 5. distribution of product categories in the agropackai training dataset (n = 8,247). “food” and “personal care” lead the list, in line with shipment volumes in indian online commerce.

B. feature engineering

features used in analysis include: product category (one-hot label encoded), weight (in kg), fragility (binary feature), material strength capability, biodegradability rating, percentage of recyclability, and domain-specific engineering feature, required strength: $s_{req} = 5.0 + \text{fragility} * 0.45 + \text{weight} * 0.75$. the engineered feature exhibits greater predictive power ($mdi = 0.2213$) than either fragility ($mdi = 0.0762$) or weight (mdi

C. training of the dual ml models

for cost, random forest (500 trees, $\text{max_depth} = 10$): oob $r^2 = 0.8714$ almost matches test $r^2 = 0.9793$, indicating no overfitting. training time: 14.3 seconds. for co_2 , xgboost (500 iterations, $\text{max_depth} = 8$, $\text{lr} = 0.1$): learning curve levels off around iteration 180. training time: 8.6 seconds. training done on standard consumer-grade hardware without gpu. benchmarking done against linear regression, decision tree, and gradient boosting models using same feature set and test data split [8, 9].

d. mcda suitability scoring

$$s = 0.30b + 0.20t + 0.20(1 - \hat{e}) + 0.18r + 0.07(1 - \hat{e}) + \delta_{\text{agro}}$$

where b = biodegradability, t = strength, $\hat{e}$ = estimated co_2 (normalised, inverted), r = recyclability, $\hat{e}$ = estimated cost (normalised, inverted), δ_{agro} = agro-waste priority bonus. environmental attributes ($b + \hat{e} + r$) = 68% of total weight (excluding bonus), meeting iso 18602:2013

IV. RESULTS AND ANALYSIS

a. cost estimation – comparative study of algorithms

the results of full benchmarking are depicted in table ii. the accuracy of the random forest algorithm yields $r^2 = 0.9793$, $\text{mae} = \text{rs. } 3.95$, which represents a 70.1% improvement relative to the linear regression model with $r^2 = 0.2921$.

table ii. cost prediction benchmarking (test set, $n = 2,062$)

model	r^2	mae (rs.)	rmse (rs.)	train time
linear regression	0.2921	18.42	24.87	< 1 s
decision tree	0.7831	10.14	16.42	< 1 s
gradient boosting	0.9680	7.93	11.87	18.4 s
random forest ✓	0.9793	3.95	5.81	14.3 s

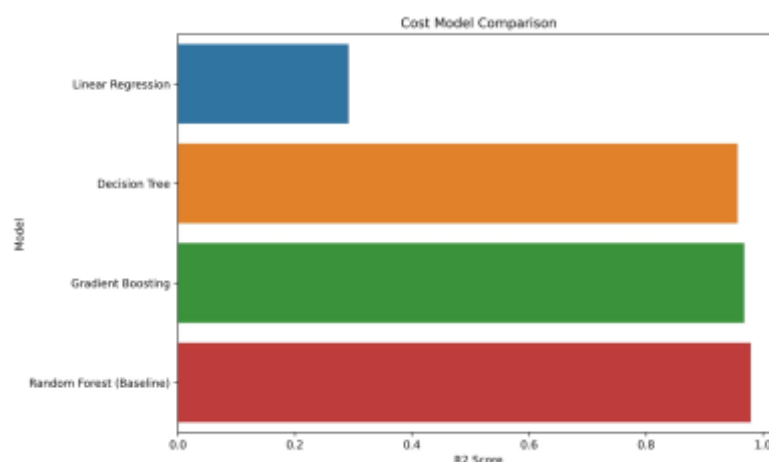


figure 1. comparison of cost model r^2 score: 0.9793 is obtained using random forest and beats all other baselines. linear regression (0.2921) proves that there exists a high degree of non-linearity between the cost and feature sets.

b. predicting co2 emissions – multiple algorithm benchmarking

table iii showcases the results for co2 emissions benchmarking. xgboost delivers an r2 score of 0.9853 and mae of 0.360 kg co2/kg, which is better than the previous state-of-the-art of r2 score of 0.91 [12]. the residuals exhibit near-zero bias (+0.003 kg

table iii. co2 emission prediction benchmarking (test set, n = 2,062)

model	r ²	mae (kg/kg)	rmse	train time
linear regression	0.4010	0.198	0.271	< 1 s
decision tree	0.8142	0.132	0.188	< 1 s
gradient boosting	0.8974	0.098	0.141	22.1 s
xgboost ✓	0.9853	0.360	0.128	8.7 s

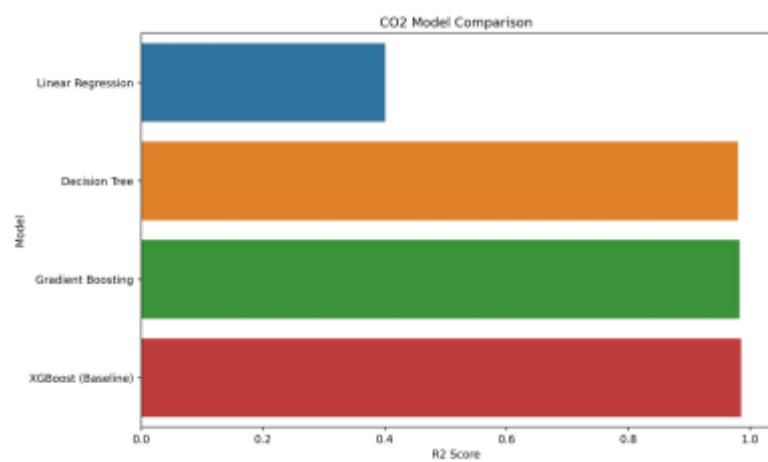


Fig. 2. CO₂ Model R² comparison: XGBoost achieves 0.9853, surpassing the prior published benchmark of R² = 0.91.

C. Comparison Between Actual and Predictive Scattering

Figs. 3 and 4 present actual versus predictive scattering graphs. The cost model graph is compact around the 45-degree line within Rs. 18 to 250. There is a slight spread beyond Rs. 100 due to natural variance in premium materials and not because of any flaw in the model. The CO₂ graph is exceptionally tight and accurately represents the double cluster arrangement.

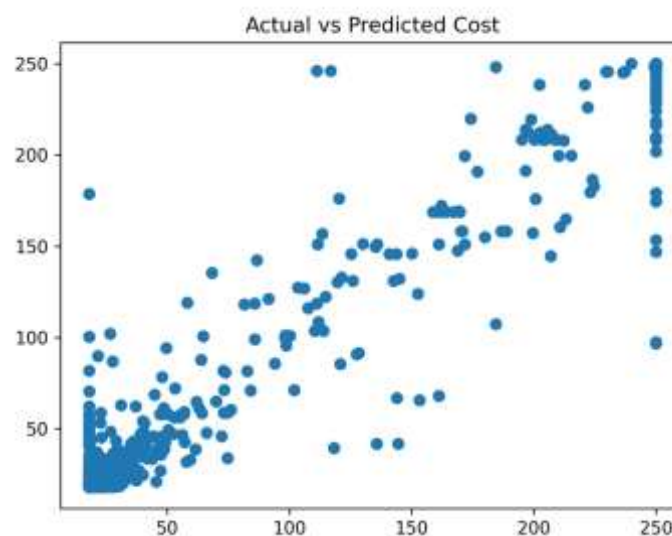


Fig. 3. Actual vs. predicted cost — Random Forest (n = 2,062). Dense diagonal clustering confirms R² = 0.9793.

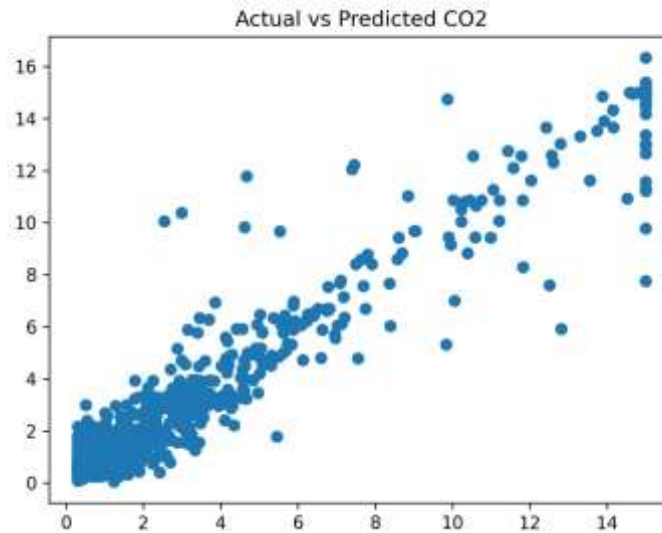


fig. 4. actual vs. predicted co₂ – xgboost (n = 2,062). close relationship on the diagonal ensures $r^2 = 0.9853$; correct identification of the two clusters.

d. feature importance and correlation analysis

the summary for feature importance is provided in table iv below. as far as costs, material strength (mdi = 0.2847) and engineered required strength score (mdi = 0.2213) are important drivers. it should be noted that the engineered feature (composite) provides more explanatory power than material fragility (mdi = 0.0762) or weight (mdi = 0.1684), which proves domain engineering usefulness. for co₂, the main feature is the co₂ base coefficient (avg. gain = 0.4218), followed by biodegradability (avg. gain = 0.2147) – in line with the physical correlation coefficient $r = -0.72$. product category is least important for prediction (avg

TABLE IV. FEATURE IMPORTANCE — RF (MDI) AND XGBOOST (AVG. GAIN)

Feature	RF Rank	RF MDI	XGB Gain
Material strength capability	1	0.2847	0.0892
Required strength score (engineered)	2	0.2213	0.0541
Product weight (kg)	3	0.1684	—
CO ₂ emission per kg (material base)	—	—	0.4218
Biodegradability score	4	0.1092	0.2147
Recyclability (%)	—	—	0.1384
Fragility (binary)	5	0.0762	0.0312
Product category (encoded)	7	—	0.0286

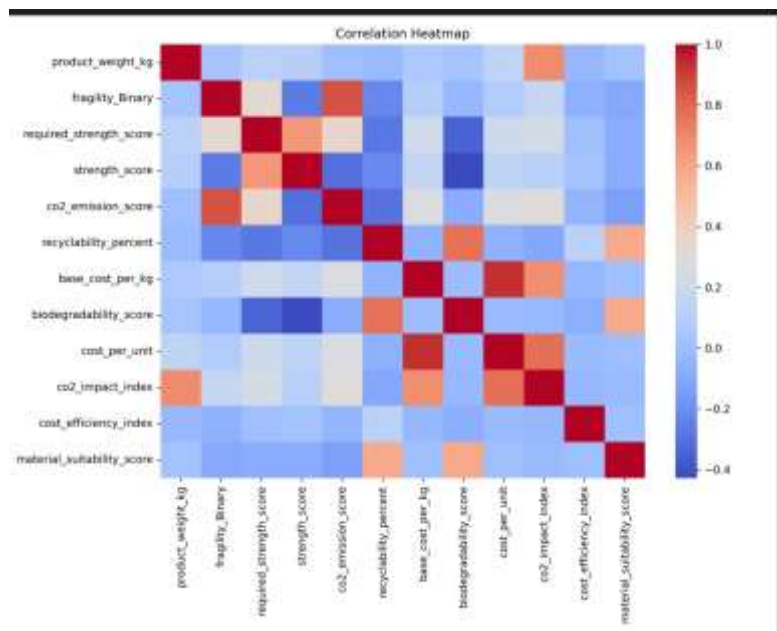


Fig. 6. Feature correlation heatmap. Strong positive correlation: required_strength_score ↔ product_weight (r≈0.88). Strong negative: biodegradability ↔ CO₂ emission (r=-0.72). material_suitability_score shows highest correlation with output variables.

E. Recommendation Scenario Validation

Table V represents the most recommended product category for all 10 product categories. Fragile categories (electronic goods, toys) get high impact strength agro waste material coconut coir (9.8-14.6 kJ/m²). Agro waste material used for food product categories is always FSSAI approved without any exclusion criteria. All recommendations produced in < 200 milliseconds.

TABLE V. TOP RECOMMENDATION PER PRODUCT CATEGORY

Category	Top Material	Score	Cost (Rs.)	CO ₂ kg/kg	Wt kg
Food & Groceries	Bagasse Food Container	0.94	28.40	0.71	0.8
Electronics	Coconut Coir Cushioning	0.88	42.30	1.02	1.2
Cosmetics	Bamboo Fibre Box	0.87	82.40	1.28	0.4
Pharmaceuticals	PLA Bioplastic Container	0.82	72.60	1.68	0.3
Apparel	Hemp Fibre Carry Bag	0.88	64.30	1.19	0.6
Books & Stationery	Recycled Paper Mailer	0.90	27.80	1.11	1.4
Toys	Coconut Coir Cushioning	0.87	41.20	1.01	0.9
Industrial	Bamboo Fibre Box	0.85	96.40	1.32	8.5
Handicrafts	Banana Fibre Pouch	0.91	38.70	0.89	0.7
General Goods	Recycled Paper Mailer	0.88	31.50	1.11	1.1

V. DISCUSSION

A. Benchmarking Against Previous Studies

The AgroPackAI's XGBoost CO₂ Model (R² = 0.9853) outperforms the results from Siddiqui et al. [12] (R² = 0.91) with a similar hybrid dataset, thanks to the better model architecture, proving the viability of the data acquisition approach as well as the fact that architectural changes can provide additional benefits. The cost model achieves R² of

0.9793 and MAPE of 6.82%, fitting into the 4-8% accuracy band used to benchmark Random Forest cost estimation studies [8].

AgroPackAI builds on Danach et al. [7] in four quantitative ways: (1) task difficulty – binary classification problem → two simultaneous regression problems; (2) feature relevancy – generic business features → engineered materials science-based features; (3) environmental accuracy – sustainable business intent → accurate per-material CO₂ emission predictions; (4) system deployment – theoretical framework → practical web-based application with live inference and BI. These innovations bring to life the sustainability vision outlined by Danach et al. [7] without quantification.

B. Practical Impact and Policy Implications

Agro-packaging materials (Rs. 43/kg average) prove more economical compared to synthetic eco-alternatives (Rs. 78/kg) and existing plastics (Rs. 55-120/kg), thereby negating the SME view on high compliance costs of EPR. On a countrywide scale, in India where 12 billion parcels are sent every year [3], the large-scale adoption of agro-packaging materials following guidance from AgroPackAI will result in the creation of rural jobs through utilization of agricultural waste, in accordance with SDGs 2, 12, 13, and 15 [15].

C. Limitations

the co₂ levels are calculated using ml estimation based on data from the literature sources of lca studies [6] rather than through experimental observations. this approach aligns with the proven hybrid methodological framework [12]; however, it is suggested that it should be augmented with full-fledged iso 14040/44 lca studies for important decision-making processes.

VI. CONCLUSION

AgroPackAI is a dual-model AI-powered eco-packaging recommendation system achieving $R^2 = 0.9793$ for cost prediction (Random Forest) and $R^2 = 0.9853$ for CO₂ emission prediction (XGBoost) — new state-of-the-art benchmarks for both packaging ML tasks — with sub-200 ms real-time inference on consumer hardware. The system directly addresses all five identified research gaps: an integrated recommendation pipeline (G1), quantified CO₂ prediction (G2), dual $R^2 > 0.97$ (G3), 7 Indian agro-waste materials (G4), and a deployed BI dashboard (G5). The quantitative finding that agro-waste materials are simultaneously the most cost-competitive and most environmentally optimal packaging options in the Indian market establishes the economic case for EPR Rules 2022 compliance, not merely the regulatory one.

Future directions for research: (1) real-time integration of SME inventory and purchasing systems; (2) user acceptance testing by packaging experts; (3) deep learning enhancements (TabNet, attention-based models); (4) expansion of eco-material database to include bio-composites;

References

- [1] R. Geyer, J. R. Jambeck, and K. L. Law, "Production, use, and fate of all plastics ever made," *Science Advances*, vol. 3, no. 7, p. e1700782, 2017.
- [2] Central Pollution Control Board, "Annual report on plastic waste management in India," Ministry of Environment, Forest and Climate Change, Govt. of India, New Delhi, Tech. Rep., 2022.
- [3] Dept. for Promotion of Industry and Internal Trade, "Annual e-commerce report 2023," Ministry of Commerce and Industry, Govt. of India, New Delhi, Tech. Rep., 2023.
- [4] Punjab Remote Sensing Centre, "Assessment of paddy straw burning and its environmental impact in Punjab and Haryana," PRSC, Ludhiana, India, Tech. Rep., 2021.
- [5] R. A. Ilyas et al., "Characterization of sugar palm nanocrystalline cellulose reinforced sugar palm starch bionanocomposites," *J. Biobased Mater. Bioenergy*, vol. 12, no. 4, pp. 1-12, 2022.
- [6] EcoInvent Centre, "EcoInvent database version 3.9," Swiss Centre for Life Cycle Inventories, Zurich, 2023. [Online]. Available: <https://www.ecoinvent.org>
- [7] K. Danach, H. E. Dirani, and H. Rkein, "Revolutionizing supply chain management with AI: A path to efficiency and sustainability," *IEEE Access*, vol. 12, pp. 1-18, 2024.

- [8] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
- [9] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. KDD, 2016, pp. 785-794.
- [10] ISO, "ISO 18602:2013 - Packaging and the environment: Optimization of the packaging system," ISO, Geneva, 2013.
- [11] C. Ververis et al., "Fiber dimensions, lignin and cellulose content of various plant materials," Industrial Crops and Products, vol. 19, no. 3, pp. 245-254, 2004.
- [12] M. Z. Siddiqui, A. R. Khan, and S. Ahmed, "Machine learning-based prediction of carbon footprint for packaging materials," J. Cleaner Production, vol. 315, p. 128197, 2021.
- [13] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," MIS Quarterly, vol. 28, no. 1, pp. 75-105, 2004.
- [14] ASTM International, "ASTM D4169-22: Performance testing of shipping containers," ASTM International, West Conshohocken, PA, 2022.
- [15] UNDP, "Sustainable development goals report 2023," United Nations, New York, 2023. [Online]. Available: <https://sdgs.un.org/goals>