Machine Learning Project

Approach for Investigating the Impact of Circular Economy on Taiwan's Economy and the Energy efficiency

Yuze Tsai

May 20, 2025

Table of Contents

Overview of Project Structure

- **Goal:** Investigate the impact of circular economy policies on Taiwan's economic development and energy efficiency (2014–2023).
- **Data:**
  - *Dependent Variables:* Income, Electricity consumption per capita, Housing vacancy rate
  - *Independent Variables:* Environmental taxes, recycling rate, green infrastructure, etc.
- **Methods(previously):**
  - Panel data regression (Fixed vs. Random Effects)
  - Machine learning models: XGBoost, Random Forest, SVM

Version updated: Catboost, LightGBM ML models, and take R^2 and SHAP value for interpretation

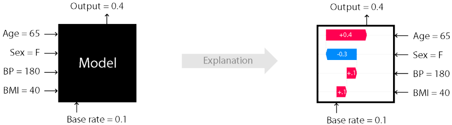
Review the Assumptions

Hypothesis	Statement
1	Environmental budgets and recycling improve electricity efficiency
2	Landfill/incineration harm income, poor recycling lowers income
3	Waste positively linked to income and electricity usage, negatively to vacancy
4	Renewable energy capacity improves income and electricity efficiency

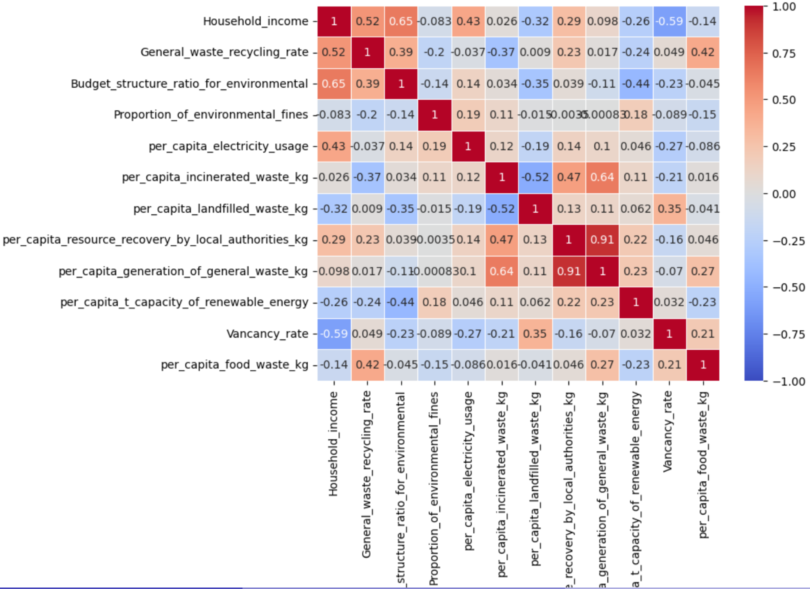
SHAP (SHapley Additive Explanations)

- SHAP assigns each feature an importance value for prediction.
- Based on cooperative game theory—accounts for all possible feature combinations.
- **Advantages:**
  - Global and local interpretability
  - Indicates both direction (+/-) and magnitude of effect
  - Helps identify nonlinear thresholds and interactions
- **Applied to:** XGBoost, CatBoost, LightGBM, RF, and SVM

SHAP



Correlation Overview



Panel Regression: Household Income

- Fixed Effects (FE) vs Random Effects (RE) models.
- Preferred: FE due to focus on within-unit variation.
- **Within R^2 (FE): 0.5755**

Table: Model Comparison of Household Income

	Fixed Effects	**Random Effects**
Dep. Variable	Household_income	Household_income
R-squared	0.5755	0.5269
F-statistic	16.718	15.717
General_waste_recycling_rate	1.656e+04 (0.3543)	1.565e+04 (0.3247)
Budget_structure_ratio_for_environmental	-2.424e+04 (-1.5064)	-2.416e+04 (-1.4588)
Proportion_of_environmental_fines	-1656.4 (-0.3080)	-1520.4 (-0.2741)
per_capita_incinerated_waste_kg	-1.459e+04 (-1.0786)	-1.509e+04 (-1.0815)
per_capita_landfilled_waste_kg	-8152.9 (-0.7027)	-8504.2 (-0.7109)
per_capita_generation_of_general_waste_kg	9.208e+04 (0.9104)	9.113e+04 (0.8736)
per_capita_t_capacity_of_renewable_energy	2.426e+04 (2.6360)***	2.446e+04 (2.5775)***
per_capita_food_waste_kg	-6.258e+04 (-1.0021)	-6.412e+04 (-0.9958)
per_capita_resource_recovery_by_local_authorities_kg	-1.877e+04 (-0.1813)	-1.786e+04 (-0.1673)

Panel Regression: Electricity Usage

- Dependent variable: Per Capita Electricity Usage.
- **HOrriBIE explanatory power: Within $R^2 = 0.0538$.**

Table: Model Comparison of Per Capita Electricity Usage

	Fixed Effects	Random Effects
Dep. Variable	p-c_electricity_usage	p-c_electricity_usage
R-squared	0.0538	0.0430
F-statistic	0.7588	0.6891
General_waste_recycling_rate	-1748.9 (-1.0245)	-1630.1 (-0.9591)
Budget_structure_ratio_for_environmental	1469.0 (1.5002)	1385.3 (1.4779)
Proportion_of_environmental_fines	-56.664 (-0.1633)	-9.4697 (-0.0272)
per_capita_incinerated_waste_kg	-446.12 (-0.5155)	-573.76 (-0.6610)
per_capita_landfilled_waste_kg	-246.51 (-0.3294)	-371.66 (-0.4985)
per_capita_generation_of_general_waste_kg	-3261.6 (-0.7386)	-2803.4 (-0.6335)
per_capita_t_capacity_of_renewable_energy	374.06 (0.6450)	415.61 (0.7202)
per_capita_food_waste_kg	5519.8 (1.7388)*	4844.5 (1.5677)
per_capita_resource_recovery_by_local_authorities_kg	4079.9 (0.9360)	3577.6 (0.8217)

Panel Regression: Vacancy Rate

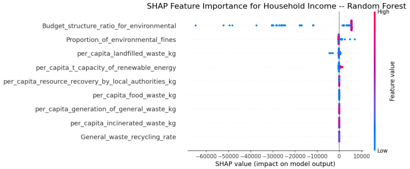
- Dependent variable: Vacancy Rate.
- Within $R^2 = 0.3796$.

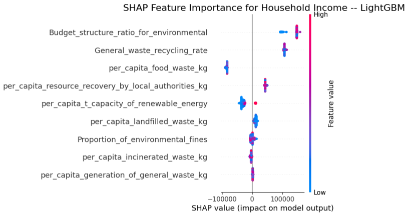
Table: Model Comparison of Vacancy rate

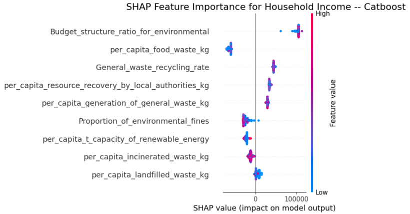
	Fixed Effects	**Random Effects**
Dep. Variable	Vacancy_rate	Vacancy_rate
R-squared	0.3796	0.3371
F-statistic	8.1573	7.7976
General_waste_recycling_rate	1.1738 (4.8452)***	1.1819 (4.7618)***
Budget_structure_ratio_for_environmental	0.1066 (0.7670)	0.1001 (0.7037)
Proportion_of_environmental_fines	0.0098 (0.1999)	0.0108 (0.2138)
per_capita_incinerated_waste_kg	-0.1464 (-1.1922)	-0.1513 (-1.2023)
per_capita_landfilled_waste_kg	-0.3178 (-2.9923)***	-0.3194 (-2.9346)***
per_capita_generation_of_general_waste_kg	2.6392 (4.2111)***	2.6646 (4.1490)***
per_capita_t_capacity_of_renewable_energy	0.1055 (1.2815)	0.1081 (1.2824)
per_capita_food_waste_kg	-0.7543 (-1.6744)*	-0.7786 (-1.6881)*
per_capita_resource_recovery_by_local_authorities_kg	-2.7926 (-4.5143)***	-2.8214 (-4.4511)***

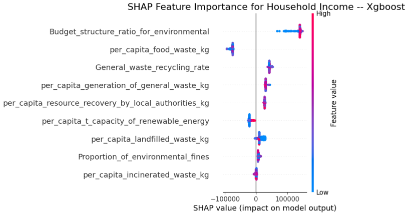
ML Model Comparison: Household Income

- Applied ensemble methods: RF, LightGBM, CatBoost, XGBoost.
- Compared feature importance via SHAP values.









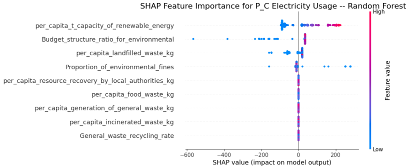
Where is SVM?

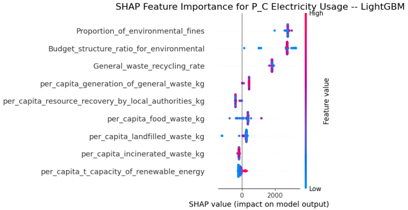
SHAP Summary: Top Features for Household Income

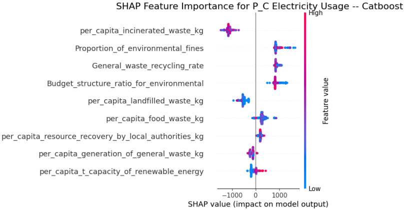
- SHAP values quantify each feature's impact on prediction.
- **Strongest positive influence: Budget structure ratio and Recycling rate**
- **Negative: Food waste per capita and renewable energy capacity.**

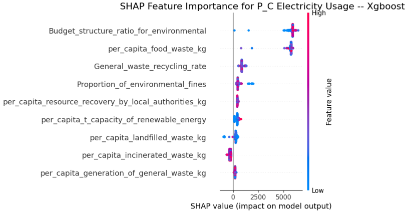
Feature	XGBoost	RF	CatBoost	LightGBM
Budget ratio (env.)	135392	184066	103778	140778
Recycling rate	43613	56781	44104	106733
Waste generation	30847	11152	29307	1692
Food waste per capita	-76820	-72989	-63354	-84387
Renewable capacity	-20539	-16840	-22892	-28026

Results for Per Capita Electricity Usage: Ensemble Models







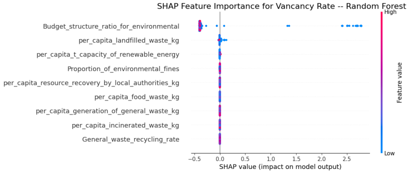


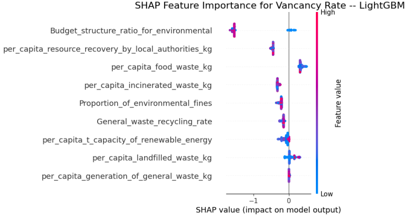
SHAP Summary: Feature Importance for Electricity Usage

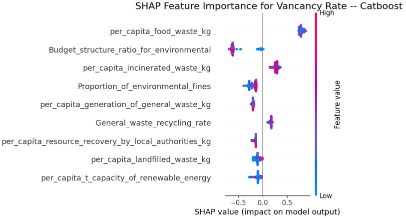
- SHAP values indicate feature impact on model output.
- **Key features with highest influence: Budget structure ratio, Food waste per capita, Recycling rate.**
- **Lower influence or negative impact: Incinerated waste, landfilled waste.**

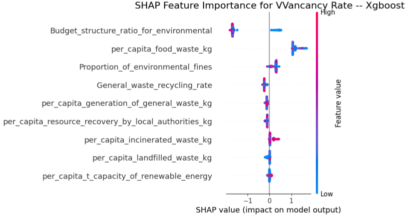
Feature	XGBoost	RF	CatBoost	LightGBM
Budget ratio (env.)	5818	3	872	2683
Food waste per capita	5680	0	292	307
General waste recycling rate	903	0	880	1825
Environmental fines prop	552	-1	902	2782
Incinerated waste per capita	-283	0	-1143	-195

Results for Vacancy Rate: Ensemble Models









SHAP Summary: Feature Importance for Vacancy Rate

- SHAP values indicate feature impact on model output.
- **Key features with highest influence: Budget structure ratio, Food waste per capita, Recycling rate.**
- **Inconsistent results: Environmental fines prop.**

Feature	XGBoost	RF	CatBoost	LightGBM
Food waste per capita	1.146	0.000	0.786	0.334
Environmental fines prop	0.273	0.000	-0.201	-0.229
Incinerated waste per capita	0.093	0.000	0.268	-0.320
Resource recovery by local gov	-0.113	0.000	-0.149	-0.457
Budget ratio (env.)	-1.418	0.017	-0.547	-1.327

R-squared Comparison Across Models

Feature	XGBoost	RF	LightGBM	CatBoost	FE Panel
Household Income	0.720	0.688	0.699	**0.752**	0.578
Per Capita Electricity	0.427	**0.479**	0.474	0.420	0.054
Vacancy Rate	0.736	0.659	0.600	**0.743**	0.380

- **Household Income:** CatBoost leads (0.752), showing strong prediction capability (overall $R^2 > 0.5$).
- **Electricity Usage:** All models weak ($R^2 < 0.5$), Random Forest best (0.479); FE is **terrible**.
- **Vacancy Rate:** CatBoost and XGBoost both Strongest

Variable Importance: Overview

- Significance based on SHAP for ML and statistical significance for FE.
- Summary scores computed: $+1$ (positive effect), -1 (negative), 0 (non-significant).

Variable	Household Income	Pc Electricity Usage	Vacancy Rate
Budget structure ratio for environmental	$+4$	$+4$	-2
General waste recycling rate	$+3$	$+2$	$+1$
Per capita generation of general waste (kg)	0	0	$+1$
Per capita resource recovery by local authorities	0	0	-2
Per capita landfilled waste (kg)	0	-1	0
Proportion of environmental fines	-1	$+2$	0
Per capita incinerated waste (kg)	0	-1	0
Per capita capacity of renewable energy	$+1$	$+1$	-1
Per capita food waste (kg)	-4	$+2$	$+1$

Table: Summary of Variable Scores across Models

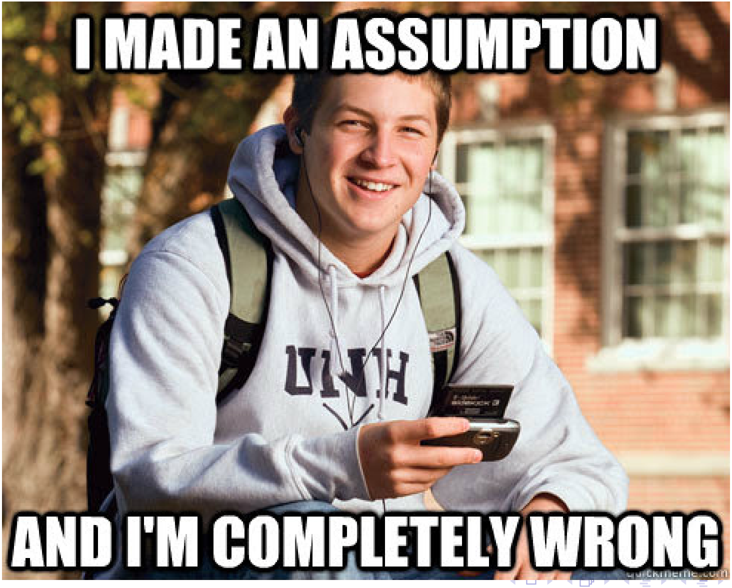
Selected Variable Interpretations

- **Environmental Budget:** Positively linked to income and electricity use, lowers vacancy.
- **Recycling Rate:** Generally positive across all $\Rightarrow$ boosts income, slight energy trade-off.
- **Food Waste:** Negatively affects income, increases usage and vacancy $\Rightarrow$ inefficiency signal.
- **Renewable Capacity:** Some income gains, inconsistent electricity effects.

Evaluation of Assumptions

Hypothesis	Statement	Conclusion
1	Environmental budgets and recycling improve electricity efficiency	Inversely supported
2	Landfill/incineration harm income, poor recycling lowers income	Partially supported
3	Waste positively linked to income and electricity usage, negatively to vacancy	Partially supported
4	Renewable energy capacity improves income and electricity efficiency	Partially supported

Reconsideration



Future Directions I: Data and Comparative Scope

Richer Circular Economy Indicators:

- Extend circular economy metrics:
  - Material Use Rate, Secondary Raw Materials Share
  - Proportion of Resource-Circulation Enterprises
- Improve measurement of policy implementation intensity

Comparative Scope:

- Cross-country or cross-regional comparisons
- Robustness checks in different policy contexts

Additional Outcomes:

- Forecast CO_2 emissions, AQI, or house prices using ML
- Explore long-run dynamics of sustainability indicators

Future Directions II: Methods Insight

Causal Inference:

- Explore Regression Discontinuity on environmental budget shares
- Investigate feasibility of quasi-experimental designs

Structural Analysis:

- Integrate circular economy into the Environmental Kuznets Curve (EKC) framework
- Combine theory-driven models with data-driven tools

Conclusion: Summary of Findings

Study Goal: Assess how circular economy policies affect income, electricity use, and vacancy rates across Taiwan's 22 cities (2014–2023).

Methods: Panel regression vs. Machine Learning (CatBoost, XGBoost, etc.)

Key Findings:

- Environmental budgets, renewables $\Rightarrow \uparrow$ income
- Food waste $\Rightarrow \uparrow$ electricity use and $\uparrow$ vacancy
- **ML models outperform regression**, especially CatBoost (SHAP used for interpretation)
- SVM underperformed due to computational inefficiency

Conclusion: Policy and Methodological Takeaways

Model Implications:

- **ML excels at prediction —** best for income and vacancy
- **Electricity usage hard to model —** needs richer features
- **Method should match goal:** prediction vs. explanation

Policy Relevance:

- Recycling and environmental investment show economic benefits
- Mixed/uncertain effects call for **re-examining assumptions**
- Circular economy indicators can inform **evidence-based, local-level policy**

Thanks !