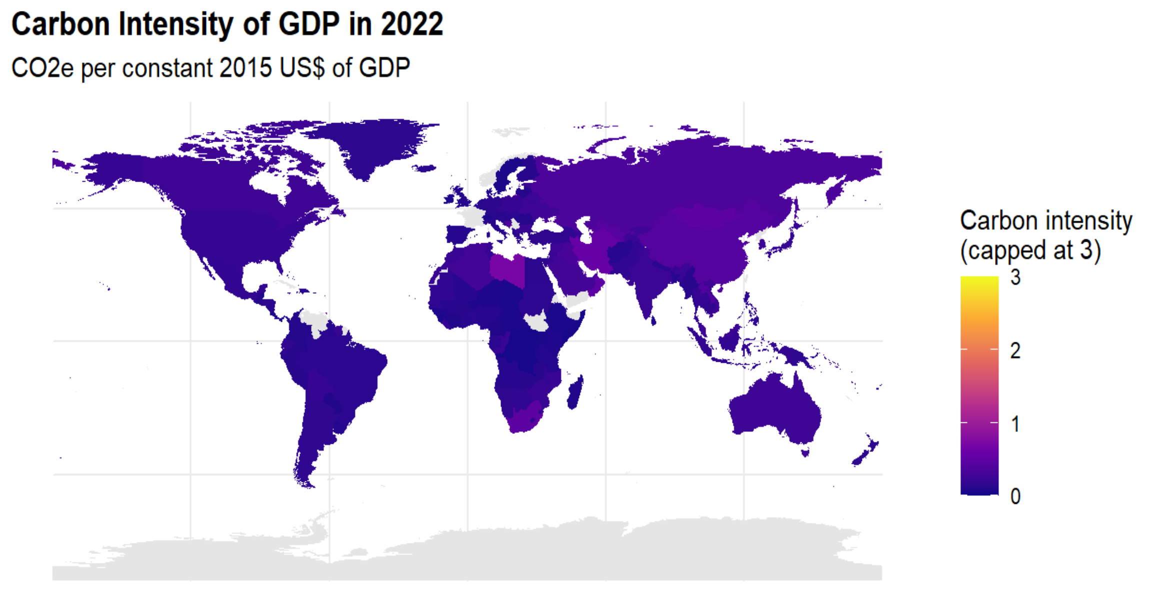
Analyzing Carbon Intensity Drivers and Forecasting Future Trajectories

I. Problem Motivation



- Carbon intensity has become a central indicator for tracking countries' progress toward decarbonization
- As global climate goals tighten, understanding what factors drive carbon intensity over time is increasingly important for policymakers

I. Problem Motivation

Existing Methods and its Restraints

- Relies on traditional econometric approaches that often assume linear, homogeneous relationships across countries
- Carbon intensity is shaped by complex interactions among economic development, industrial structure, innovation capacity, and natural-resource endowments

Machine Learning Methods

- Apply machine learning methods to explore nonlinear patterns in carbon intensity that vary significantly across regions and income groups
- The abundant data from the World Bank provide a unique opportunity to examine how national characteristics shape carbon-intensity trajectories

II. Model Selection

- **Data:** 2000-2023 World Bank World Development Indicators by country
- **Response Variables**: Carbon intensity of GDP (kg CO_2e per constant 2021 PPP GDP dollar)
- **Predictor Variables**:

Category	Variable Names	Variable Codes	Unit
Energy	Energy use	energy_use_per_capita	kg of oil equivalent per capita
	Energy imports	net_energy_imports	% use
	Fossil fuel share	fossil_fuel_share	% of total
	Electricity production from renewable sources, excluding hydroelectric	renewable_electricity_share	% of total
GDP	GDP per capita	gdp_per_capita	constant 2015 US$
	GDP growth	gdp_growth_rate	annual %
Industry	Industry (including construction), value added	industry_share_gdp	% of GDP
	Manufacturing, value added	manufacturing_share_gdp	% of GDP
Population	Urban population	urban_pop_growth	% of total population
	Urban population growth	urbanization_rate	annual %
Research	R&D expenditure	rd_expenditure_gdp	% of GDP
	Researchers per million	researchers_per_million	per million people
Energy rents	Oil rents	oil_rents_gdp	% of GDP
	Natural gas rents	gas_rents_gdp	% of GDP

II. Model Selection

- **Model Performance**

Model	Train RMSE	Train MAE	Test RMSE	Test MAE
OLS	0.40	0.13	0.37	0.13
LASSO (lambda.min)	0.40	0.13	0.37	0.13
LASSO (lambda.1se)	0.43	0.14	0.38	0.13
Random Forest	0.15	0.04	0.09	0.05
Gradient Boosting	0.17	0.07	0.09	0.07

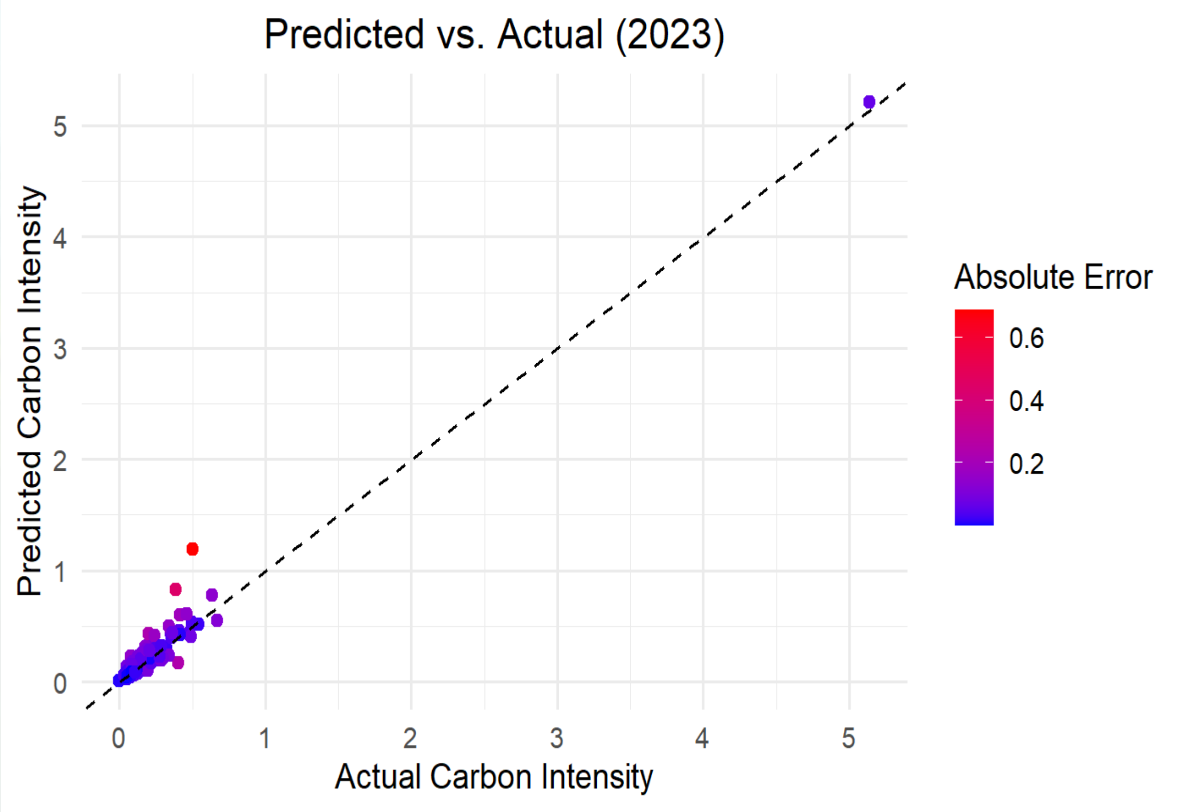
- **The Random Forest model** is selected because

(1) Its superior predictive performance;

(2) Its flexibility to capture nonlinearities and interactions among predictor variables;

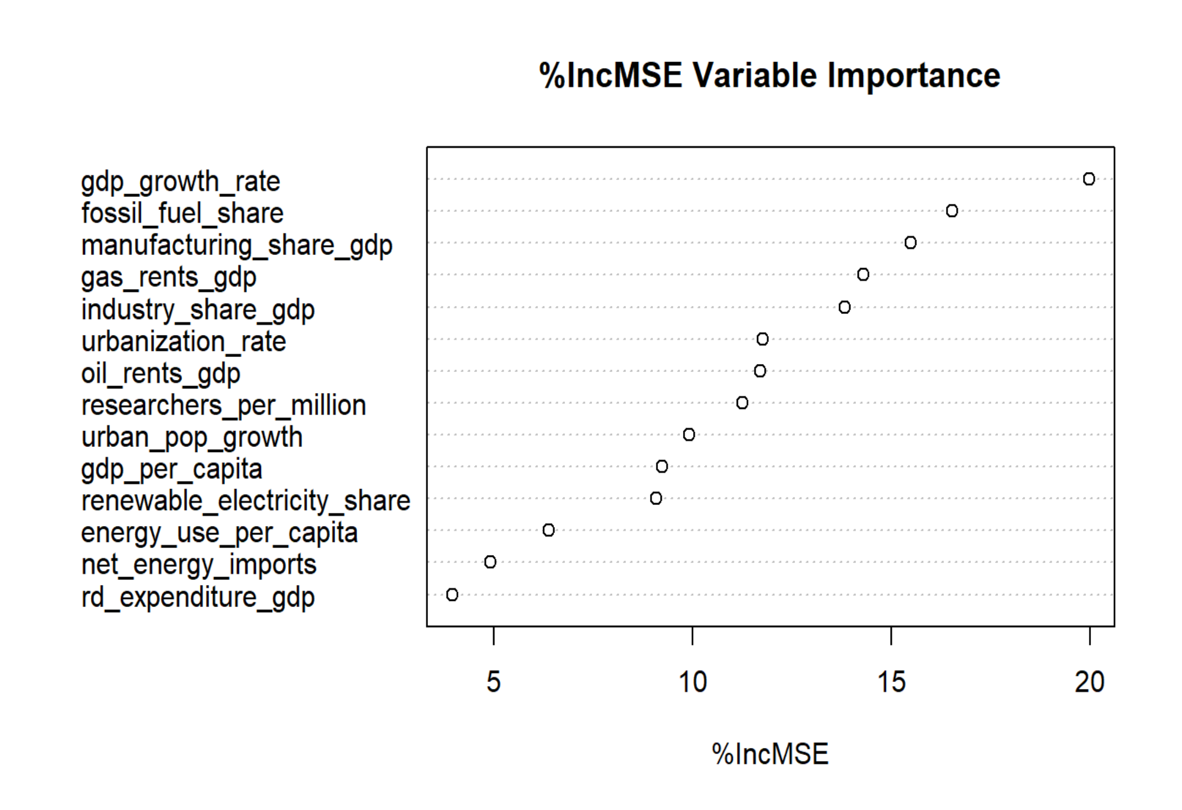
(3) Its ability to explore the importance of each predictor variable

III. Main Results



- Using the Random Forest model, we obtain a Test RMSE of 0.09 and a Test MAE of 0.05.
- Given that the actual 2023 carbon intensity of GDP has a standard deviation of 0.38, the model demonstrates strong predictive performance.

III. Main Results



- In terms of variable importance measured by %IncMSE in the Random Forest model, the top five predictors are **GDP growth rate, fossil fuel share, manufacturing share of GDP, gas rents as a share of GDP, and industry share of GDP.**

IV. Main Open Questions or Concerns

- **Short-term Prediction**: The model relies on historical patterns to generate forecasts, and in this study, the training data cover the period 2000–2022 with only 2023 used as the test year. As a result, the model is suitable primarily for short-term forecasting, such as predicting the following year, rather than projecting long-term structural changes in carbon intensity.
- **Drivers of Carbon Intensity:** Although the model identifies key predictors of carbon intensity, predictive importance does not imply causal influence. Further research is needed to determine whether these variables also represent effective levers for reducing carbon intensity.