Advanced Programming 2025

Green Economy Project: Forecasting CO_2 Emissions from Macroeconomic Variables.

Final Project Report

Natalia Wyszatycki
natalia.wyszatycki@unil.ch
Student ID: 22421911

HEC Lausanne, University of Lausanne

January 2026

Abstract

Forecasting carbon dioxide (CO_2) emissions is essential for assessing progress toward climate targets and informing green economic policies. This project investigates whether a small set of macroeconomic variables like real gross domestic product (GDP), unemployment rate and inflation can effectively predict France's annual CO_2 emissions, and identifies which factor is most influential. Using annual data from 1991 to 2024, we compare a range of regression models, including linear methods (Ordinary Least Squares, Ridge, and Lasso) and non-linear tree-based models (Random Forest, Gradient Boosting, and XGBoost). Models are trained on data from 1991–2015, validated on 2016–2020, and evaluated on a post-2020 test period (2021–2024) to reflect a realistic forecasting setting. Results show that regularized linear models and in particular Ridge regression substantially outperform more complex models on the validation set (R^2 0.77) and exhibit minimal overfitting. In contrast, tree-based models severely overfit the training data and generalize poorly. Despite strong validation performance, all models fail on the test period, highlighting the COVID-19 pandemic as a major structural break. Feature importance and SHAP analyses consistently identify GDP as the dominant predictor of emissions.

Keywords: data science, Python, machine learning, CO_2 emissions prediction, macroeconomic analysis

Contents

1 Introduction 3

2 Research Question and Literature Review 3
2.1 Economic Drivers of CO_2 Emissions . . . 3
2.2 Emissions Forecasting Methods . . . 4

3 Methodology 4
3.1 Data Description . . . 4
3.2 Approach . . . 5
3.3 Implementation . . . 6

4 Results 8
4.1 Experimental Setup and Overall Performance . . . 8
4.2 Overfitting Analysis . . . 8
4.3 Feature Importance . . . 9
4.4 SHAP Analysis . . . 9
4.5 Visualisation: Ridge model . . . 10

5 Discussion 11
5.1 Model Performance and Generalization . . . 11
5.2 Structural Break and Post-2020 Forecast Failure . . . 11
5.3 Feature Importance and Economic Structure . . . 11
5.4 Scope and Reliability of the Results . . . 11

6 Conclusion and Future Work 12
6.1 Main Empirical Insights . . . 12
6.2 Implications and Future Directions . . . 12

References 13

A Additional Figures 14

B Code Repository 15

C AI Tools Documentation 15

1 Introduction

Climate change mitigation requires accurate monitoring and forecasting of carbon dioxide (CO_2) emissions. Under international agreements such as the Paris Agreement, countries have committed to reducing emissions, making it essential to understand how emissions evolve in response to economic activity. Historically, CO_2 emissions have been closely linked to economic growth: higher levels of output have generally coincided with higher emissions. This relationship poses a major challenge for sustainable development, as economic expansion tends to increase emissions in the absence of structural changes such as cleaner energy systems or efficiency improvements.

In this context, forecasting CO_2 emissions based on macroeconomic indicators can provide valuable insights for policymakers. Reliable forecasts allow governments to assess whether current economic trajectories are compatible with climate objectives and to identify when additional interventions may be required. However, emissions forecasting is particularly challenging when data availability is limited and when economic–environmental relationships are subject to sudden disruptions.

This project focuses on forecasting France's annual CO_2 emissions using a small set of macroeconomic variables and a limited historical dataset. Using annual data from 1991 to 2024, the study examines whether real GDP, the unemployment rate and inflation can explain and predict variations in emissions. Two main objectives guide the analysis: First, the project compares the predictive performance of models of varying complexity, from simple linear regressions to advanced non-linear machine learning methods. Second, it seeks to identify which macroeconomic variable is the most influential driver of emissions, using both traditional feature importance measures and modern interpretability tools.

This report is structured as follows. Section 2 reviews related work and literature on emissions forecasting and economic drivers. Section 3 presents the data and methodological approach. Section 4 reports and visualizes the empirical results. Section 5 discusses the findings, challenges and limitations. Section 6 concludes and outlines directions for future research.

2 Research Question and Literature Review

2.1 Economic Drivers of CO_2 Emissions

The relationship between economic activity and CO_2 emissions has been extensively studied in environmental economics. A central conceptual framework is the Environmental Kuznets Curve (EKC), which posits an inverted-U relationship between income per capita and environmental degradation. While some empirical studies find partial support for this hypothesis, comprehensive reviews such as Shahbaz and Sinha (2019) conclude that results are highly context-dependent. What remains robust across most studies is that, in the absence of strong mitigation policies, GDP and emissions tend to remain positively correlated.

Beyond GDP, labor market indicators such as unemployment are often used as proxies for economic slack. Higher unemployment typically reflects reduced industrial activity and lower energy demand, which may translate into lower emissions. Inflation, by contrast, has a less direct theoretical link to emissions and is rarely identified as a primary driver, though it may capture broader macroeconomic regimes or energy price dynamics.

Beyond GDP, other macroeconomic variables have been considered. Unemployment is often used as a proxy for economic slack, with higher unemployment associated with lower industrial activity and energy demand and thus lower emissions. Inflation is less directly linked to emissions and is rarely identified as a primary driver, though it may capture broader economic or policy regimes, such as energy price shocks. Overall, prior literature suggests that GDP is the dominant predictor of emissions, with labor market conditions providing complementary information and inflation playing a limited role.

2.2 Emissions Forecasting Methods

Traditional emissions forecasting relies on econometric time-series models such as ARIMA, cointegration and error-correction models, which impose specific functional forms and assume stable long-run relationships. More recently, machine learning approaches have been introduced as flexible alternatives. Bennedsen et al. (2021) combine Lasso-based variable selection with dynamic factor models to forecast U.S. CO_2 emissions using hundreds of predictors, while Algieri et al. (2023) show that large macroeconomic and climate datasets can improve forecasting accuracy when sufficient data are available.

However, methodological research highlights important trade-offs between model flexibility and generalization. Tibshirani and Friedman (2009) emphasize that regularization techniques such as Ridge and Lasso are particularly effective in low-sample settings, where complex models are prone to overfitting. This insight is especially relevant for country-level studies using annual data.

Finally, several studies emphasize the difficulty of forecasting emissions during exogenous shocks. Events such as the COVID-19 pandemic represent structural breaks that invalidate historical patterns, leading to large forecast errors regardless of model sophistication. In such cases, model failure may itself provide useful information about regime changes.

3 Methodology

3.1 Data Description

This study uses annual data for France from 1991 to 2024. The target variable is total annual CO_2 emissions from fossil fuels and industry, measured in metric tons of CO_2. Emissions data are sourced from the Global Carbon Project, accessed via Our World in Data, ensuring consistency with internationally reported statistics.

The core explanatory variables consist of three macroeconomic indicators:

- **Real GDP** (constant USD), sourced from the World Bank's World Development Indicators (indicator NY.GDP.MKTP.KD). GDP is expressed in constant prices to remove inflation effects and capture real economic activity.
- **Unemployment rate** (of labor force), from the World Bank (indicator SL.UEM.TOTL.ZS), which reflects labor market conditions and economic slack.
- **Inflation** (annual change in CPI), also from the World Bank (indicator FP.CPI.TOTL.ZG), used as a proxy for macroeconomic stability and price dynamics.

All series were retrieved programmatically through official APIs and merged by calendar year. CO_2 emissions were downloaded as a CSV file from Our World in Data and filtered to retain observations for France only. World Bank indicators were accessed through their REST API, allowing for automated updates and reproducibility. After merging, the final dataset is complete for the period 1990–2024, with no major missing values.

During exploratory analysis, additional variables were initially tested, including the share of renewable energy in final energy consumption (from Our World in Data) and the Industrial Production Index (IPI) from Eurostat. These variables were ultimately excluded from the final specification. The renewable energy share did not improve predictive performance, while the IPI was found to be highly correlated with GDP and introduced multicollinearity without adding explanatory power. The final model therefore focuses on the three core macroeconomic indicators for parsimony and robustness.

The final dataset contains 34 yearly observations and is largely complete. No major data quality issues were identified.

3.2 Approach

Data Preprocessing and Split

To reflect a realistic forecasting setting and avoid look-ahead bias, the data were split chronologically: Training set: 1990–2015; Validation set: 2016–2020; Test set: 2021–2024. The validation period was used for model selection and hyperparameter tuning, while the test set was held out entirely for final evaluation. This split deliberately excludes the COVID-19 shock from model training, allowing us to assess how models generalize to an unprecedented structural break.

Given the small sample size, this approach trades statistical efficiency for interpretability and realism. Traditional cross-validation was not used, as shuffling time-series data would violate temporal dependence. All explanatory variables were standardized (mean zero, unit variance) for models involving regularization to ensure comparability across scales. Tree-based models, which are scale-invariant, were trained on the same standardized inputs for consistency.

Hyperparameters Optimization

Hyperparameters are optimized using validation-set performance. For linear models, the regularization parameter alpha is selected to maximize validation R^2. For tree-based models, a grid search is performed over a restricted hyperparameter space, with the objective of improving validation performance while limiting overfitting. Given the small sample size, the search space is deliberately constrained to avoid fitting noise.

Models and Algorithms

Six regression models were implemented to compare predictive performance across different levels of complexity:

1. Ordinary Least Squares (OLS): A baseline linear regression without regularization, offering full interpretability but potentially prone to overfitting.
2. Ridge Regression: A linear model with L2 regularization, designed to reduce variance and improve generalization when predictors are correlated.
3. Lasso Regression: A linear model with L1 regularization, which can shrink coefficients and perform implicit feature selection.

4. Random Forest Regressor: An ensemble of decision trees capable of capturing non-linear relationships and interactions. Tree depth and number of trees were limited to control overfitting.

5. Gradient Boosting (XGBoost): A boosting-based tree ensemble trained sequentially to reduce prediction errors. Conservative hyperparameters and early stopping were used to limit complexity.

6. Gradient Boosting (scikit-learn): A second boosting implementation included as a robustness check against model-specific behavior.

Evaluation Metrics

Model performance was assessed using multiple complementary metrics:

- R^2: Measures the proportion of variance explained and is used to diagnose overfitting by comparing training and validation scores.
- MAE (Mean Absolute Error): Indicates the typical absolute prediction error in CO_2 units.
- RMSE (Root Mean Squared Error): Penalizes large errors more heavily than MAE.
- MAPE (Mean Absolute Percentage Error): A scale-free metric allowing comparison across time periods.

Feature Importance and Interpretability

To identify the most influential economic drivers of CO_2 emissions, two interpretability approaches were used:

- Model-based importance: Absolute standardized coefficients for linear models, and built-in feature importance scores for tree-based models, normalized to percentages.
- SHAP values: Model-agnostic Shapley Additive Explanations were computed to quantify each feature's contribution to predictions. Mean absolute SHAP values were used to compare importance across models and to assess the direction of each feature's effect.

Using both methods ensures robust and consistent interpretation, especially when comparing linear and non-linear models.

3.3 Implementation

The entire project is implemented in Python 3.11 using a modular structure. Core dependencies include `pandas` for data handling, `scikit-learn` for most machine-learning models, `xgboost` for boosted trees, and `shap` for explainability. Reproducibility is ensured by fixing random seeds and by providing an `environment.yml` file listing all package versions.

```
project/
|-- main.py
|-- src/
|   |-- data_loader.py
|   |-- models.py
|   |-- hyperparameters_optimization.py
|   |-- evaluation.py
|   |-- plotting.py
|-- data/
|   |-- processed/
|       |-- france_1991_2024.csv
|-- results/
|-- environment.yml
```

The `data-loader` module handles data retrieval from external APIs, cleaning and dataset construction. The `models.py` module defines and trains the different regression models used in the study. The `hyperparameters-optimization.py` module performs systematic tuning of model hyperparameters. Model evaluation is implemented in the `evaluation.py` module. The `plotting.py` module generates all figures. Finally, the `main.py` script orchestrates the full workflow, ensuring reproducibility and consistent execution.

A non-trivial aspect of the codebase is that it explicitly maintains two parallel feature pipelines: one with year-numeric (used for forecasting performance) and one without year-numeric (used for interpretation) to prevent a time proxy from dominating economic signals in feature attribution. Concretely, the training/evaluation stage uses the "with-year" standardized design matrix because the year index can legitimately improve short-horizon predictive fit in a small sample, while the interpretability stage recomputes scaling and refits models on the "no-year" matrix so that feature importance and SHAP explain macroeconomic drivers rather than a near-deterministic trend.

Hyperparameter tuning is implemented through a dedicated optimization module. For linear models, regularization parameters are selected via a logarithmic grid search maximizing validation R^2 on a fixed hold-out validation set. For tree-based models, a GridSearchCV procedure combined with a time-series cross-validation scheme (TimeSeriesSplit) is employed. Candidate hyperparameter configurations are ranked jointly by their cross-validated validation performance and their train–validation generalization gap, explicitly penalizing overfitting and favoring robust models in a low-sample, temporal setting.

```python
def optimize_linear_models(X_train_scaled, y_train, X_val_scaled, y_val):
...
    ridge_alphas = np.logspace(-5, 2, 15)
    best_ridge_r2 = -np.inf
    best_ridge_alpha = 0.1

    for alpha in ridge_alphas:
        ridge_model = Ridge(alpha=alpha, random_state=42)
        ridge_model.fit(X_train_scaled, y_train)
        val_r2 = ridge_model.score(X_val_scaled, y_val)
        if val_r2 > best_ridge_r2:
            best_ridge_r2 = val_r2
            best_ridge_alpha = alpha
```

Listing 1: Linear Models Tuning

```python
def optimize_tree_models(X_train, y_train, X_train_without_year=None):
...
    cv_results = pd.DataFrame(rf_search.cv_results_)
    cv_results['overfitting_gap'] = cv_results['mean_train_score'] - cv_results['mean_test_score']
    best_idx = cv_results.sort_values(['mean_test_score', 'overfitting_gap'],
        ascending=[False, True]).index[0]
```

Listing 2: Tree Models Tuning

4 Results

4.1 Experimental Setup and Overall Performance

All models were fitted using the optimized hyper-parameters reported in the methodology section. Table 1 below summarizes the training, validation, and test results for each model. We report the coefficient of determination (R^2), mean absolute percentage error (MAPE) and root mean squared error (RMSE). The table shows that Ridge regression performs best on the validation set (R^2 0.765), closely followed by OLS and Lasso (R^2 0.539). The tree-based models (Random Forest, XGBoost and Gradient Boosting) fit the training data extremely well but exhibit highly negative R^2 on the validation set. These results indicate severe overfitting: ensemble methods learned noise in the small training sample and failed to generalize.

Table 1: Model performance across training, validation, and test sets

	Train			Validation			Test		
Model	R^2	MAPE	RMSE	R^2	MAPE	RMSE	R^2	MAPE	RMSE
OLS	0.840	0.023	10.31	0.539	0.037	13.84	-0.083	0.061	18.41
Ridge	0.836	0.024	10.45	0.765	0.025	9.88	-0.435	0.062	21.19
Lasso	0.840	0.023	10.31	0.539	0.037	13.83	-0.083	0.061	18.41
Random Forest	0.898	0.018	8.24	-2.236	0.100	36.66	-13.911	0.237	68.31
Gradient Boosting	0.999	0.002	0.76	-1.047	0.076	29.16	-11.759	0.218	63.18
XGBoost	0.998	0.002	1.19	-1.887	0.092	34.63	-15.637	0.251	72.15

Importantly, no model achieved a positive R^2 on the test period (2021–2024). Even the best-performing linear models show negative R^2 (–0.083 for OLS and Lasso; –0.435 for Ridge), indicating that the models could not extrapolate through the COVID-19 shock. Ensemble models performed catastrophically (test R^2 –12 to –26), confirming that their overfitting during training left them unable to handle the structural break. Despite this, linear models produced smaller errors (MAPE 6 percent) compared with the ensembles, whose test errors exceeded 20 percent.

4.2 Overfitting Analysis

To diagnose overfitting, we compared training and validation R^2 scores. OLS and Lasso had moderate gaps (0.30), suggesting some overfitting. Ridge achieved a much smaller gap (0.07), indicating better generalisation. In contrast, Random Forest, XGBoost and Gradient Boosting exhibited enormous gaps (3–8 R^2 points).

Table 2: Train–Validation R^2 Gap by Model

Model	OLS	Ridge	Lasso	Random Forest	XGBoost	Gradient Boosting
R^2 Gap	0.302	0.071	0.301	3.134	2.885	2.046

Thus, the ensembles memorised the training data but failed to capture patterns that persisted into the validation period. We conclude that regularised linear models provide the best bias–variance trade-off for this small dataset.

4.3 Feature Importance

The feature-importance plot derived from model coefficients and tree importances shows a consistent ranking: GDP (real constant USD) is by far the most predictive variable, contributing about 55 percent of the explanatory power in linear models and even more in Random Forest and Gradient Boosting (92 percent and 85 percent respectively).

The unemployment rate is the next most important feature, contributing roughly 38 percent in linear models; its negative coefficient indicates that high unemployment (economic slack) reduces emissions. Inflation has minimal influence (7 percent), suggesting that price changes do not materially drive emissions once GDP and unemployment are accounted for. XGBoost is an outlier: it assigns greater importance to inflation (39 percent), likely reflecting spurious patterns captured during overfitting.

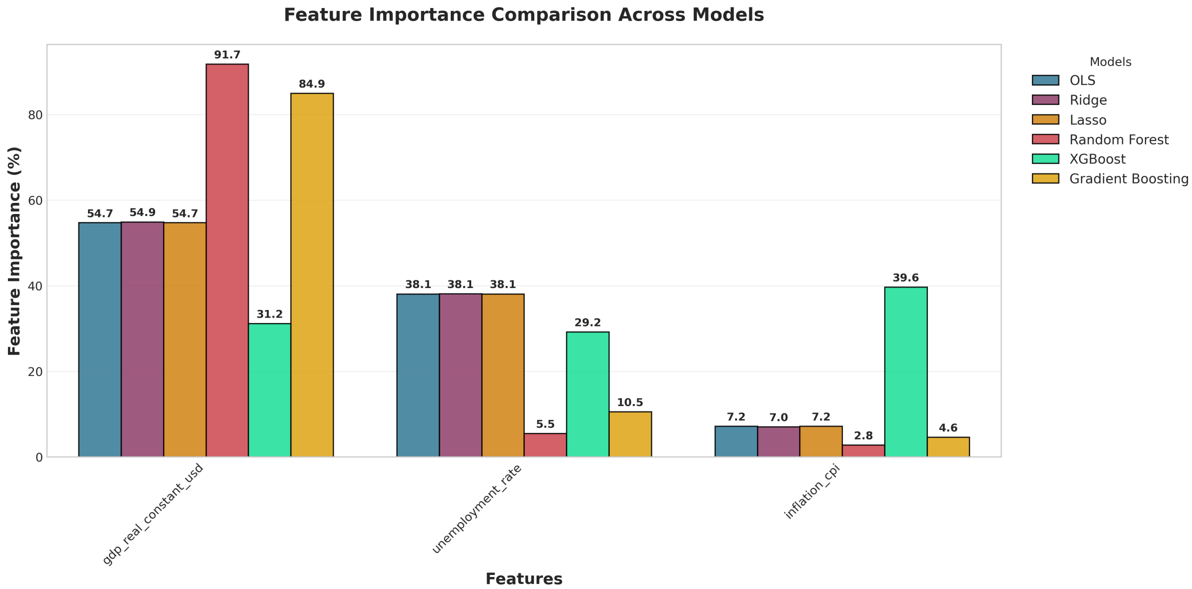

Figure 1: Feature Importance Plot

4.4 SHAP Analysis

To provide a model-agnostic perspective, we computed SHAP (Shapley Additive Explanations) values across observations to quantify the relative contribution of each predictor to model predictions. Compared to the model-specific feature importance measures reported earlier, the SHAP-based results exhibit greater alignment across models. The results display a clear and stable ranking of features. Real GDP accounts for approximately 56 percent of the total contribution to the models' predictions, followed by the unemployment rate at around 37 percent, while inflation represents a substantially smaller share, close to 6 percent. This ordering is consistent across models, even XGBoost.

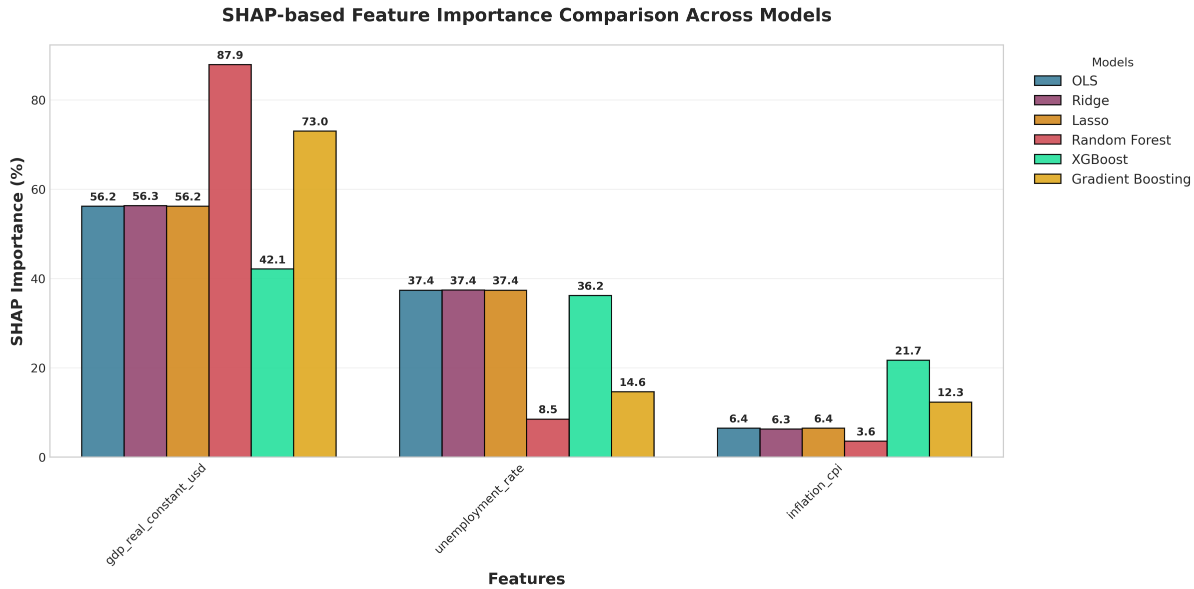

Figure 2: SHAP Summary Plot

4.5 Visualisation: Ridge model

Figure 3 illustrates the Actual vs Predicted CO_2 Emissions Ridge plot. The black line shows actual emissions, while the red crosses depict Ridge's predictions. During the training and validation periods, the Ridge model tracks the downward trend reasonably well, missing only some short-term fluctuations. In the test period, however, the model cannot anticipate the sharp drop in 2020 or the partial rebound thereafter. This explains the negative R^2 in the test set; the pandemic constitutes an unprecedented shock outside the model's experience.

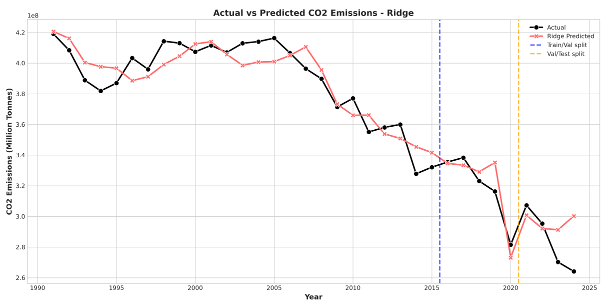

Figure 3: Prediction timeline for Ridge

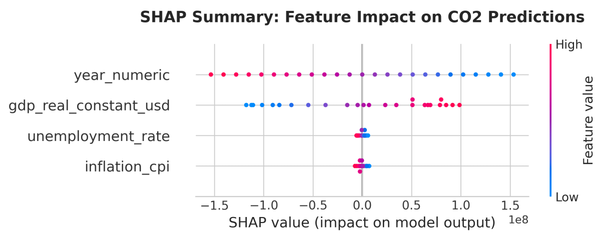

Figure 4: SHAP summary for Ridge

Figure 4 illustrates SHAP summary plot for the Ridge model. Points represent individual years; colour denotes feature value (pink for high, blue for low). GDP has the largest SHAP values, reinforcing that economic output is the dominant driver of emissions predictions.

The SHAP scatter demonstrates that high GDP values yield positive SHAP scores (increasing predicted emissions), while high unemployment values yield negative SHAP scores (reducing emissions). Inflation yields SHAP values clustered near zero, confirming its limited predictive role.

5 Discussion

5.1 Model Performance and Generalization

From an empirical standpoint, the most robust result concerns the relative performance of the different model classes. Regularized linear models clearly outperform tree-based models on the validation set. This outcome is both statistically and economically intuitive. With a small number of macroeconomic predictors and limited annual observations, the relationship between economic activity and emissions appears largely smooth and monotonic. In such settings, linear models with mild regularization are well suited to extract the dominant signal while controlling variance. Ridge regression, in particular, achieves a favorable bias–variance tradeoff and delivers the most stable out-of-sample performance prior to 2020.

By contrast, the poor validation performance of tree-based models indicates that their flexibility is not supported by the data. Despite achieving extremely high training R^2 values, these models fail to generalize, suggesting that they fit idiosyncratic patterns rather than economically meaningful relationships. From an economic perspective, this is unsurprising: with annual data, it is unlikely that complex non-linearities or high-order interactions between GDP, unemployment, and inflation can be reliably identified. These results reinforce a key methodological insight: greater model complexity does not imply better performance when informational content is limited.

5.2 Structural Break and Post-2020 Forecast Failure

The sharp deterioration in performance on the test set (2021–2024) constitutes a central finding. All models produce negative R^2 values in the post-2020 period, reflecting a pronounced structural break associated with the COVID-19 pandemic. Emissions declined abruptly due to lockdowns and behavioral constraints not captured by standard macroeconomic indicators, and the subsequent recovery followed a non-standard trajectory. As a result, models trained on pre-2020 data are forced to extrapolate beyond their domain of validity. In this context, negative test performance signals not only predictive failure, but the breakdown of historical relationships under an exogenous shock.

5.3 Feature Importance and Economic Structure

Feature importance and SHAP analyzes provide complementary insights into the structure learned by the models. Across all specifications, real GDP consistently emerges as the dominant predictor of emissions, accounting for roughly half or more of total importance. This finding is consistent with established economic reasoning, as higher output is typically associated with increased energy use. Unemployment plays a secondary but meaningful role, generally exerting a negative effect, while inflation contributes little explanatory power. These results should nevertheless be interpreted cautiously: given the small sample size and correlation among macroeconomic variables, the estimated contributions reflect correlations rather than structural causal effects.

5.4 Scope and Reliability of the Results

Overall, the results indicate that the project successfully identifies stable empirical relationships in historical data, while also exposing the limits of purely data-driven forecasting in the presence of regime shifts. Linear models perform well under normal conditions, but all approaches struggle when confronted with unprecedented shocks. Consequently, the findings are best viewed as descriptive and diagnostic rather than as a basis for precise long-term prediction.

6 Conclusion and Future Work

6.1 Main Empirical Insights

This study set out to assess whether a small set of macroeconomic indicators (real GDP, unemployment and inflation) can meaningfully forecast France's annual CO_2 emissions and to identify which factors dominate the emissions–economy relationship. The empirical results point to a clear and robust conclusion: in a low-dimensional, small-sample setting, regularized linear models outperform more complex machine learning alternatives. Ridge regression, in particular, achieves the most stable validation performance by balancing bias and variance, while tree-based models consistently overfit and fail to generalize.

A second key insight concerns the limits of forecasting under structural disruption. All models perform poorly on the post-2020 test period, yielding negative R^2 values. Rather than reflecting a simple modeling failure, this outcome highlights the breakdown of historical macroeconomic relationships during the COVID-19 shock. Emissions dynamics during this period were driven by constraints and behavioral changes that are not captured by standard macroeconomic aggregates. In this sense, the loss of predictive accuracy is itself informative, signaling a regime change rather than noise or misspecification.

Finally, feature importance and SHAP analyses consistently identify real GDP as the dominant correlate of emissions, with unemployment playing a secondary role and inflation contributing little explanatory power. These findings are aligned with economic intuition and existing literature, while also illustrating that the models capture correlation patterns rather than structural causal effects.

6.2 Implications and Future Directions

The results of this project suggest several promising directions for further development. Methodologically, future work could move beyond static relationships and explicitly model time variation or regime shifts. Approaches such as rolling-window estimation, state-space models, or regime-switching frameworks may be better suited to environments characterized by policy changes or external shocks. These methods could help distinguish between normal-cycle dynamics and exceptional periods.

Additional empirical extensions could also enhance the explanatory content of the models. Incorporating energy-specific variables, such as energy prices, sectoral energy demand, or the electricity generation mix, would allow for a more direct link between economic activity and emissions. Exploring higher-frequency data, where available, could further improve the understanding of short-run dynamics and adjustment processes.

From an applied perspective, the framework developed here could be repurposed for scenario-based analysis rather than point forecasting. Instead of predicting emissions mechanically, the models could be used to assess how alternative macroeconomic trajectories or policy environments translate into emissions outcomes. Such an approach would strengthen the relevance of the analysis for climate policy design while preserving the transparency and interpretability that proved valuable in this study.

References

1. Algieri, B., Lecca, P., & Reiter, O. (2023). *Forecasting energy-related CO_2 emissions using large information sets.* Energy Economics, 118, 106499.

2. Bennedsen, M., Hillebrand, E., & Koopman, S. J. (2021). *Modeling, forecasting, and nowcasting U.S. CO_2 emissions using many predictors.* Energy Economics, 96, 105151.

3. Shahbaz, M., & Sinha, A. (2019). *Environmental Kuznets curve for CO_2 emissions: A literature survey.* Journal of Economic Studies, 46(1), 106–168.

4. Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer Series in Statistics. Springer.

5. Global Carbon Project. (2024). *Global Carbon Budget.* Available at: `https://www.globalcarbonproject.org`

6. Our World in Data. (2024). *CO_2 emissions dataset.* Available at: `https://ourworldindata.org/co2-and-greenhouse-gas-emissions`

7. World Bank. (2024). *World Development Indicators.* Available at: `https://databank.worldbank.org/source/world-development-indicators`

8. Eurostat. (2024). *Industrial Production Index (STS).* Available at: `https://ec.europa.eu/eurostat`

A Additional Figures

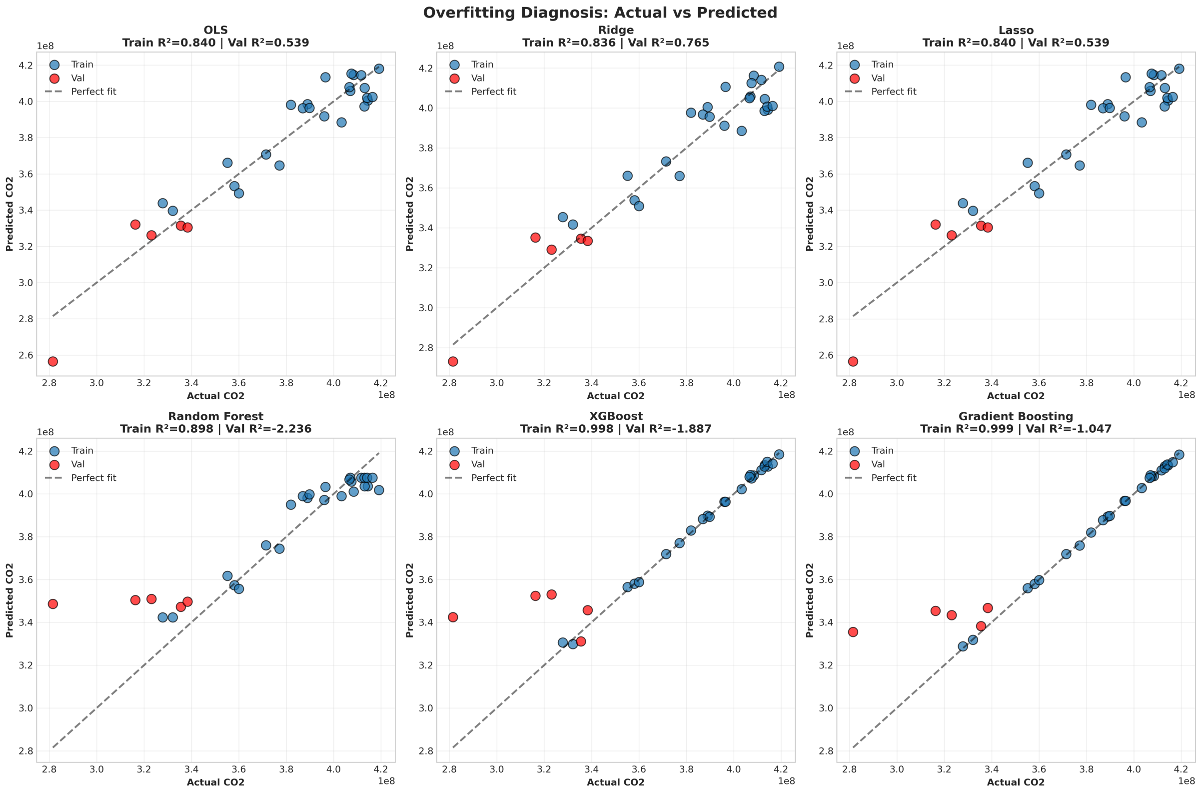

Figure 5: Overfitting diagnosis based on the train–validation R^2 gap across models. Large gaps indicate severe overfitting, particularly for tree-based models.

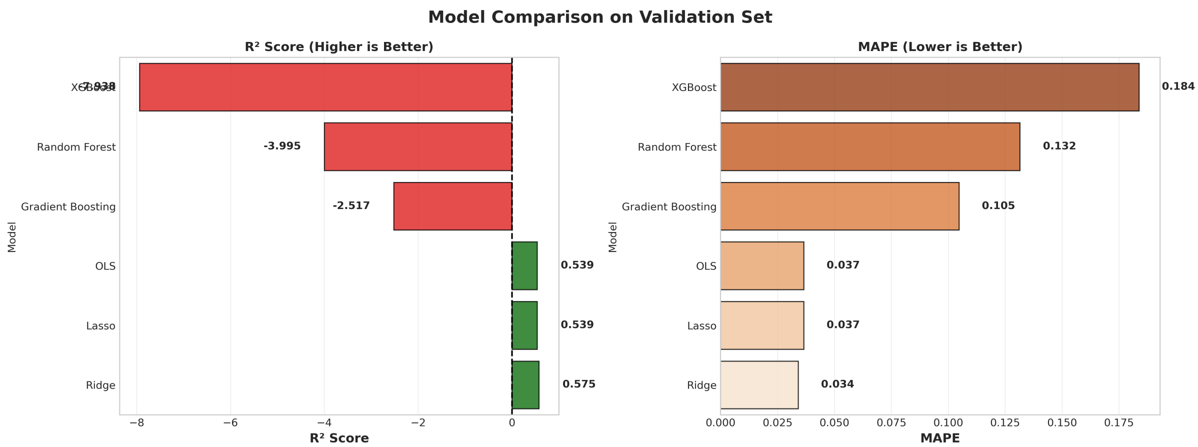

Figure 6: Comparison of R^2 performance and MAPE on validation set for all models. Linear models generalize better than tree-based models.

B Code Repository

GitHub Repository: `https://github.com/natalia23wy/green_economy_project.git`

- **Repository structure**: The repository is organized into modular Python scripts handling data loading, model definition, hyperparameter optimization, evaluation and visualization as shown above.

- **Installation**: The computational environment can be recreated using Conda:

```
conda env create -f environment.yml
conda activate green_economy
```

- **Reproducibility**: All results can be reproduced by running the main pipeline:

```
python main.py
```

  This command executes the full workflow, from data collection to model evaluation and figure generation.

C AI Tools Documentation

Use of ChatGPT

ChatGPT was used as a learning and support tool throughout the project.

- It was consulted to clarify machine learning and econometric concepts, including overfitting, regularization and the bias variance tradeoff.
- Used to discuss hyperparameter optimization strategies, with a focus on improving validation R2 while limiting overfitting. The final optimization logic, model selection criteria and implementation were designed and coded independently.
- Supported improvements in code structure and readability. All architectural and implementation decisions were made independently.
- Used to rewrite explanations in clear academic English. Final wording choices were made independently.
- Used for LaTeX-related questions, including tables, figures, captions and layout formatting.

Use of GitHub Copilot

- GitHub Copilot was used occasionally during coding.
- Used for syntax completion and boilerplate suggestions.

All scientific and technical decisions were made independently.