

From National Goals to Industry Action: AI-Driven Forecasting of India’s Carbon Emissions

Motivation. Accurately predicting carbon emissions is vital for global climate goals and for companies working to meet ESG (Environmental, Social, and Governance) standards. This study evaluates traditional time series models against machine learning techniques using Climate TRACE emissions data for India [1], covering seven sectors: waste, manufacturing, fossil fuels, transportation, power, agriculture, and buildings. While models like Prophet are interpretable and capture trends and seasonality effectively, they struggle with missing data, a common issue in ESG reporting. Machine learning methods, such as XGBoost, can handle missing values and model complex interactions but may not fully capture underlying emission dynamics. To address this, we introduce a hybrid approach combining Physics-Informed Neural Networks (PINNs) with classical and machine learning models to improve prediction accuracy across sectors.

Method. We used Climate TRACE emissions data for India (Jan 2021–May 2025), cleaned, aggregated monthly, and handled missing values with distribution-aware imputation. For mixed-type columns containing both numeric values and qualitative tags (“high,” “very high”), ordinal mappings were applied before imputation. Lag features of past emissions were engineered for predictive modeling. Classical models (ARIMA, SARIMA, Holt-Winters, Prophet) and machine learning models (Random Forest, Gradient Boosting, XGBoost, SVR) were trained on Jan 2021–Dec 2023 data, with forecasts evaluated on Jan 2024–May 2025 using Mean Absolute Percentage Error (MAPE). We introduce a hybrid approach which first trains a PINN to learn the main emission patterns, then models the residual errors using XGBoost or Prophet to capture patterns the PINN missed.

Results. Standalone PINNs capture the main emission trends but have higher error (MAPE: 8.18%). The hybrid approach reduces errors significantly: PINN + Prophet (1.70%) outperforms both standalone PINN and Prophet (2.33%), while PINN + XGBoost (4.33%) improves over PINN-only. Classical models such as Holt-Winters (2.76%) and SARIMA (3.67%) performed moderately, and machine learning baselines like XGBoost (2.72%) and Gradient Boosting (2.72%) were competitive. The results highlight that combining physics-informed modeling with ML residual correction can capture both underlying patterns and complex interactions, especially in datasets with missing values.

Impact. This hybrid modeling framework demonstrates how national-level emissions data can be leveraged for industry-level forecasting, enabling organizations to anticipate emission trends, plan targeted reductions, and strengthen ESG compliance. By integrating PINNs with machine learning or classical models, the work underscores the game-changing potential of AI and machine learning to advance carbon transparency and accelerate the shift toward sustainable development.

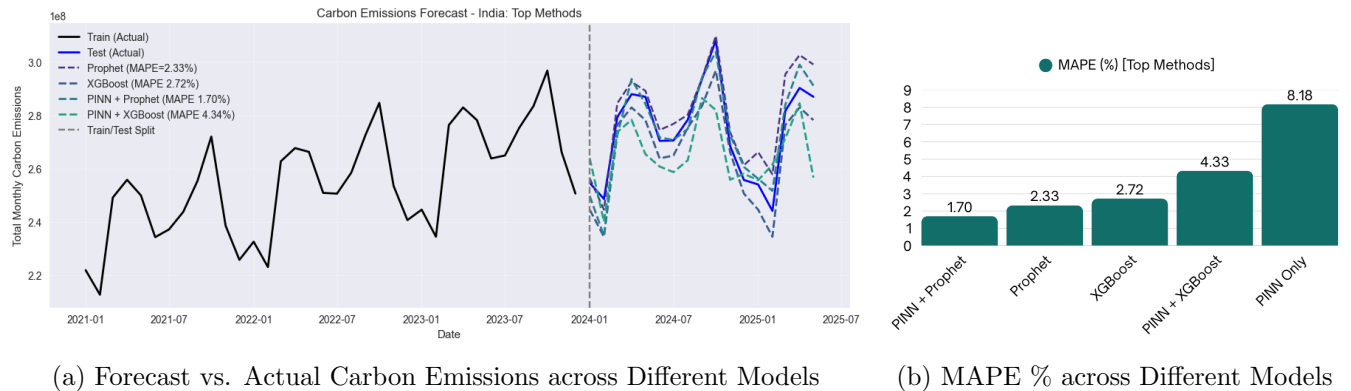


Figure 1: Total Carbon Emission Forecast - India

References

- [1] Climate TRACE. Climate trace, data downloads. <https://climatetrace.org/data>.