Machine Learning and Big Data Frameworks for Sustainable Economic and Environmental Decision-Making

Dr. Alok Kumar Bhargava[1*], Dr. V. Meenakshi [2], Namra Fatima[3], Sumathy S[4], Alisha S. K[5], Ranjan Banerjee[6]

[1]Founder and Principal Researcher, TrayiVani Foundation, India; Developer, The Inner Engine Framework™ for Conscious Leadership Assessment; Author of TrayiVāṇī – Eternal Verses on Peace, Silence & Discernment and The Inner Engine of Leadership Trilogy, Ghaziabad – 201016, Uttar Pradesh, India, Email: founder@trayivani.in , Orcid Id: https://orcid.org/630-643-630-643 , Google Scholahttps://scholar.google.com/citations?hl=en&user=XRzNh50AAAAJ , Website: https://trayivani.in, Scopus ID: 60698294500

[2]Assistant Professor & Head i/c , Department of Economics- Shift II, Dwaraka Doss Goverdhan Doss Vaishnav College Email Id: hod-beco@dgvaishnavcollege.edu.in Orcid Id: PIN code: 600106

[3]Assistant Professor, Department of Civil Engineering Specialization in Environmental Engineering, Shivalik College of Engineering , Dehradun Email Id: namra8496@gmail.com Orcid Id: 630-643-630-643 Pincode: 248197

[4]Assistant Professor, Department of Computer Science,SRM Institute of Science and Technology, Ramapuram,Chennai -600089 Email Id: sumathymurugan@gmail.com , Orcid Id: 630-643-630-643X

[5]Assistant Professor, Department of Chemical Engineering, Specialization in Environmental engineering, V. S. B. Engineering College, Karur Address: Karur, Tamil Nadu, India Email Id: alisha01081997.ask@gmail.com

[6]Assistant Professor, Departmen of Computer Science & Engineering, Brainware University, Barasat, Kolkata – 125, West Bengal, India, Pincode:700125 Orcid Id: 630-643-630-643 Email Id: rbkpccst@gmail.com

HOW TO CITE:
Alok Kumar Bhargava , V. Meenakshi , Namra Fatima , Sumathy S, Alisha S. K, Ranjan Banerjee (2026). Machine Learning and Big Data Frameworks for Sustainable Economic and Environmental Decision-Making. International Journal of Special Education, 41(16s), 630-643

COPYRIGHT STATEMENT:
Copyright: © 2026 Authors. Open access publication underthe terms and conditions of the Creative CommonsAttribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

ABSTRACT:

Climate change and environmental pollution have led to the increased need for developing sustainable economic and environmental models that could be capable of considering the intricate relationships between economics, energy, industries, and carbon dioxide emissions. This study developed and evaluated machine learning and big-data frameworks for greenhouse gas emission prediction and sustainability-oriented decision support. A quantitative predictive approach was applied using a multi-regional and multi-sector dataset containing 179,185 observations from 1995 to 2022. Greenhouse gas emissions were modeled using Value Added, Employment, Energy Carrier Net Total, Region, Sector, and Year. Linear Regression, Random Forest, and XGBoost were implemented and evaluated using Mean Absolute Error, Root Mean Square Error, and the coefficient of determination (R^2). Feature importance and SHAP analysis were additionally performed to improve model interpretability and transparency. The results demonstrated that XGBoost achieved the highest predictive performance, with MAE = 0.6007, RMSE = 0.9172, and R^2 = 0.9179, outperforming both Random Forest and Linear Regression. Energy consumption emerged as the strongest determinant of greenhouse gas emissions, while electricity production, waste management, petroleum refining, and livestock-related sectors significantly contributed to emission

variation. The results show that the machine learning ensembles deliver accurate and understandable systems that can be used for environmental-economics prediction and sustainable decision making. This proposed system may aid in carbon-based planning, sectoral sustainability evaluation, and environmental policy formation.

Keywords: Machine learning, Big data, Greenhouse gas emissions, Sustainable decision-making, XGBoost

1. Introduction

As climate change and environmental degradation gain momentum, the need for sustainable economic and environmental decision-making has become more significant at all levels of government, industry and international institutions. Over the last few decades, GHG emissions have significantly increased due to rapid industrialization, urbanization, population growth and energy-intensive economic systems. The activities of the electricity generation, transportation, manufacturing and agriculture sectors continue to be some of the main drivers of global environmental degradation. The energy production and industrial sectors are responsible for a significant portion of global GHG emissions, and an in-depth review of emission trends revealed a need for effective mitigation measures and sustainable policy interventions (Lamb et al., 2021). Environmental sustainability has now been firmly established as a key component of economic planning, resource management and long-term development goals.

Decision-making processes need to be adapted towards a more sustainable development, and this needs to be balanced with economic development. As climate uncertainties and ecological vulnerabilities are increasing, risk management and evidence-based environmental governance have become an integral part of sustainability planning. Meanwhile, renewable energy and sustainable urban development have become a key solution to cut down environmental emissions in fast-growing economies. The empirical evidence from the Asian countries showed that, with effective urban and economic planning systems, the expansion of renewable energy has a great potential to reduce GHGs (Chien et al., 2022). Much attention has also been paid to sustainability-oriented industrial restructuring, especially in the environmentally intensive sectors. The agro-industrial symbiosis and alternative heating systems demonstrated that the use of technology integration and sustainable production practices has the potential to significantly decrease the GWP of industrial greenhouse production on a global scale (Marttila et al., 2021).

By nature, the environmental-economic systems are getting more complex, and this has added restrictions on the use of traditional statistical methods for sustainability analysis. Non-linear interactions between energy consumption, industrial activity, technological development and economic development often impact environmental indicators. The study of the economic growth-pollution nexus has found nonlinear environmental dynamics that simple linear models cannot explain (Tran et al., 2026). This is what has spurred the growth of machine learning and big data initiatives in sustainability research. In turn, data-driven sustainability approaches are being used more and more to analyze large volumes of environmental datasets, increase the predictive power of the analyses, and inform evidence-based environmental governance (Mohammed et al., 2023). Advanced, predictive analytics using machine learning techniques have proven to be very promising for forecasting greenhouse gas emissions and aiding in sustainable policy formulation. One recent study specifically focused on carbon emission prediction revealed that machine learning models could be used to successfully predict carbon emissions and enhance strategic environmental

planning (Anonna et al., 2023). AI now plays a crucial role in energy management and environmental governance. The combination of AI and BI systems has great potential to boost the efficiency of ESG management, making the processing of sustainability data easier and allowing predictive decisions to be made. Meanwhile, sustainability research has also identified the environmental impact of computational systems themselves. Large-scale computational processes have been shown to produce significant carbon footprints, and efficient and environmentally responsible frameworks for artificial intelligence are crucial (Lacoste et al., 2019). Hence, it is not just important to increase the predictive power of machine learning applications for sustainability but also to make environmental decision-making processes transparent and efficient.

In recent years, machine learning has seen tremendous progress and has been adopted in various fields to support sustainability, such as agriculture, monitoring the use of land, optimizing energy use, and predicting climate. The use of remote sensing and machine learning has successfully been employed to analyse the influence of climate change on cropland systems and greenhouse gas emissions, highlighting the effectiveness of artificial intelligence in environmental monitoring and management of resources (Uyar & Uyar, 2025). In the field of low-carbon energy systems, big-data optimization frameworks have also enhanced prediction models for carbon emissions by considering economic and environmental indicators in advanced computational models (Luo et al., 2024). Explainable artificial intelligence (XAI) has also enhanced sustainability analysis by enhancing the transparency and interpretability of predictive decision-making systems. Explainable AI frameworks have demonstrated their ability to enhance the transparency of influential variables and increase accountability in algorithmic decision-making processes, thereby contributing to sustainable energy management (Talaat et al., 2025). However, there are still some gaps when it comes to sustainability and machine learning literature. A lot of previous studies have focused on analyses for a single country and a small number of sectors or on focusing on individual environmental indicators rather than incorporating other economic and industrial indicators. Several studies have also utilized mainly the traditional econometric methods, which might fail to capture the nonlinear and multidimensional relationships of greenhouse gas emissions. Moreover, while machine learning models have shown great predictive power, few studies have combined XAI techniques in large-scale environmental-economic forecasting models. Thus, an extensive machine learning framework is still needed that would allow for the inclusion of economic activity, industrial heterogeneity, energy consumption and environmental indicators in a clear and transparent predictive system based on sustainability. This study aims at overcoming the above-mentioned limitations by building and testing machine learning frameworks for sustainable economic and environmental decision-making, which are based on an integrated environmental and economic indicator. The study performs a comparative analysis of Linear Regression, Random Forest, and XGBoost models to predict greenhouse gas emissions, and also introduces explainability analysis to make the model more transparent and explainable. The study also explores the energy-gas emissions relationship, the influence of industrial sectors and economic activity on greenhouse gas emissions, to help inform evidence-based sustainability planning and environmental policy making.

2. Methodology

2.1 Research Design

To develop and evaluate machine learning frameworks for sustainable economic and environmental decision making, this study used a quantitative research design. The study was aimed at predicting GHG emissions based on economic and environmental indicators of a multi-sector and large-scale data set. To evaluate the predictive capacity of various algorithms, the comparative

machine learning approach was used to find the most suitable framework for emission forecasting. The methodological framework was broken down into the following five steps: data preprocessing, exploratory statistical analysis, machine learning model development, model evaluation and explainability analysis. Linear Regression was used as the baseline model, and Random Forest and XGBoost were used as ensemble methods that can capture nonlinear relationships between the variables. The overall structure was aimed at assessing the contribution of economic activity, energy use, industrial structure and time to GHGs.

2.2 Data Source

A publicly available sustainable-oriented dataset created by Guo and Ma (2024) was used for this study. The dataset includes combined economic and environmental metrics for machine learning purposes in greenhouse gas emission estimation and sustainability analysis. It comprises multi-regional and multi-sector observations on greenhouse gas emissions, energy use, employment and economic output for the period 630-643. The data set for this study was composed of 179,185 observations and seven variables, namely region, sector, value added, employment, GHGs, energy consumption and year. This data offers a rich amount of heterogeneity and temporal coverage for the assessment of machine learning approaches to environmental-economic forecasting and sustainable decision-making.

2.3 Variable Description

The dependent variable used was greenhouse gas emissions in kg of CO2 equivalent. Independent variables were: Value Added, Employment, Energy Carrier Net Total, Region, Sector and Year. Value Added measures the contribution of each industry to the economy, and Employment indicated that of labor in industry. Sectoral energy consumption intensity was measured using the Energy Carrier Net Total. Geographical and industrial heterogeneity were captured in Region and Sector, while Year was added to capture temporal variation. The distribution was stabilized by applying a logarithmic transformation and improving the performance of the model. All predictive modeling processes were then performed on the transformed target variable.

2.4 Data Preprocessing

The data set was initially examined for missing observations, duplication, invalid observations, and inconsistencies. The data was complete and no missing values or duplicate data were found, indicating that the data is of high completeness and integrity. To enable the model to converge and to assign equal importance to the features while training the model, the numerical variables were scaled using the StandardScaler technique. Categorical data (e.g., Region and Sector) were one hot encoded. In this way, the machine learning models could be used to ingest the categorical data without being skewed by the ordinal nature of the data. After pre-processing, the dataset was split into the training and testing sets in an 80:20 ratio. The training set was used to develop the model, while the testing set was kept to evaluate the model's performance.

2.5 Machine Learning Models

There were 3 machine learning models implemented in the study. Linear Regression, a simple statistical technique, was used to evaluate the potential of linear relationships to predict greenhouse gas emissions. Random Forest was selected as it is an ensemble learning technique that can reduce the variance by building a number of decision trees and enhance the prediction strength. XGBoost is the implementation of the gradient boosting framework, which aims to maximize the predictive accuracy and to efficiently model nonlinear interactions among the variables. The Random Forest model was designed to have 100 decision trees and tree-depth parameters to prevent overfitting and to keep the trees as generalizable predictors. The XGBoost model was optimized with the gradient boosting technique with controlled learning rates, tree depth and subsampling to improve the stability and efficiency of its prediction.

2.6 Model Evaluation

The Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and coefficient of determination (R^2) were used to evaluate the predictive performance of the models. The average size of the prediction error was measured using Mean Absolute Error, and Root Mean Square Error highlighted larger differences between the actual and predicted values. R^2 was used to determine the percentage of variance accounted for by each model for the GHGs. The machine learning algorithms were comparatively evaluated using a comparative evaluation framework. The most appropriate model for sustainable forecasting was the one that had the lowest prediction error and the highest explanatory power.

2.7 Explainability Analysis

A feature importance analysis and SHAP analysis were performed on the best model to improve the model interpretability and transparency. Feature importance analysis was done to find out which features are most important in predicting GHG emissions. SHAP analysis was also used to quantify the effect of each variable in the model outputs, as well as to offer observation-level interpretability. A detailed understanding of the environmental and economic factors influencing emissions was achieved through the explainability framework, which also helped develop transparent, sustainability-oriented decision-making systems.

3. Results

3.1 Descriptive Statistics of Economic and Environmental Variables

Table 1 shows the descriptive statistics of the variables used in the study. The data also showed significant variations in various economic and environmental factors, highlighting the diversity of industrial activities and regional sustainability trends. The distribution of greenhouse gas emissions was very scattered, with a maximum emission value well above the median value, which means that there were highly emission-intensive sectors. Likewise, the variable Energy Carrier Net Total showed very high variation between observations, which indicates very different levels of energy consumption in the various sectors. These findings support the use of nonlinear machine learning models and the use of logarithmic transformation to make stable predictions.

Table 1. Descriptive statistics of the study variables

Variable	Mean	Standard deviation	Minimum	25th percentile	Median	75th percentile	Maximum
Value Added (Million EUR)	7821.811	46760.200	0.000	44.274	416.221	2763.502	2756946.000
Employment (Thousand persons)	488.026	3448.601	0.000	0.963	10.085	89.794	237023.600
GHG emissions (kg CO2 equivalent)	5508983000.000	50109630000.000	0.000	21967470.000	187977300.000	1252798000.000	3810351000000.000
Energy Carrier Net Total (TJ)	58301.810	400102.800	0.003	361.471	3137.733	19066.000	29579280.000
Year	2008.615	8.055	1995.000	2002.000	2009.000	2016.000	2022.000

3.2 Correlation Analysis of Economic and Environmental Indicators

The correlation matrix of the variables is shown in Figure 1. There was a good positive correlation between Energy Carrier Net Total and greenhouse gas emissions (r = 0.85), meaning that energy-intensive activities make a significant contribution to environmental degradation. By comparison, the direct relationship of Employment and Value Added with greenhouse gas emissions was relatively weak, indicating that economic growth is not necessarily an adequate indicator of the environmental consequences without accounting for the energy intensity. Limited linear temporal effects were found in the data, as shown by the low correlation of the Year variable with the other variables in the study.

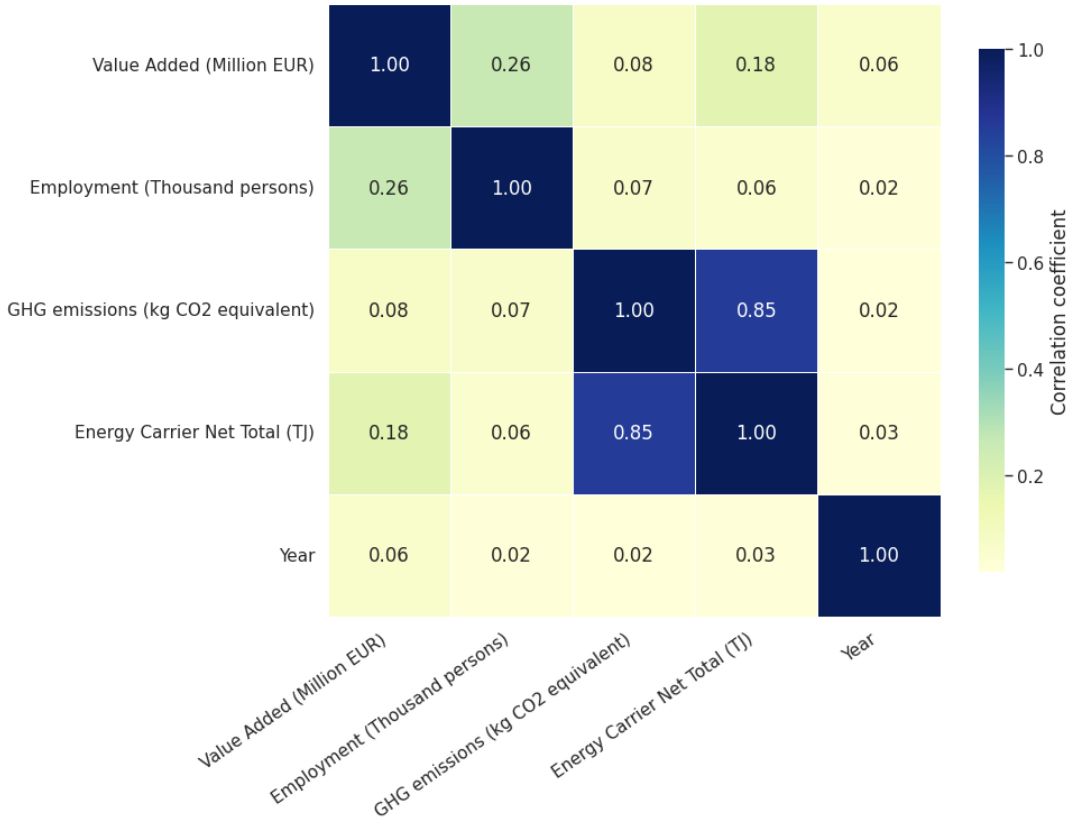

Figure 1. Correlation among economic and environmental variables

The correlation outcomes show that energy use is a vital issue in sustainability evaluation and that energy-based indicators can be integrated in predictive environmental modelling approaches.

3.3 Regional Distribution of Greenhouse Gas Emissions

The top areas with the highest greenhouse gas emissions are shown in Figure 2. The biggest average emissions came from China, the USA and India. In addition, several world regions with aggregated emissions also had a high emission intensity, which indicates that there is a combined effect of vast industrial and economic activities on the environment.

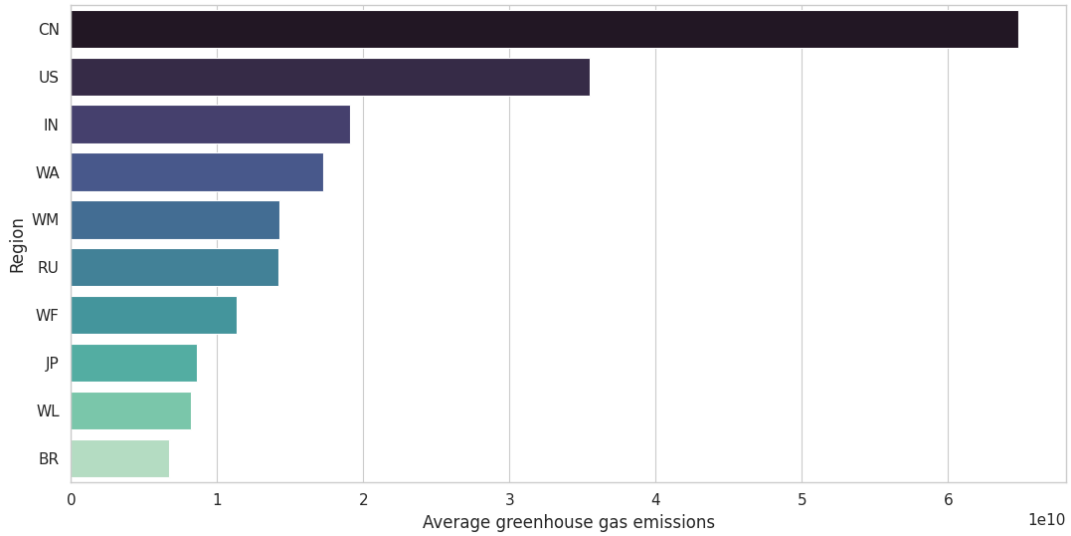

Figure 2. Top regions by average greenhouse gas emissions

The regional analysis reveals that energy-intensive regions and those with a rapid growth rate of energy use in industry contribute disproportionately to global emissions. These results point to the importance of area-specific sustainability approaches, as well as international efforts in decreasing emissions.

3.4 Sectoral Analysis of Greenhouse Gas Emissions

The sectoral distribution of greenhouse gas emissions is shown in Figure 3. There were also large environmental effects in agricultural sectors such as cattle farming and raw milk production. In addition, there was high emission intensity in heavy industry sectors like cement manufacturing and petroleum extraction.

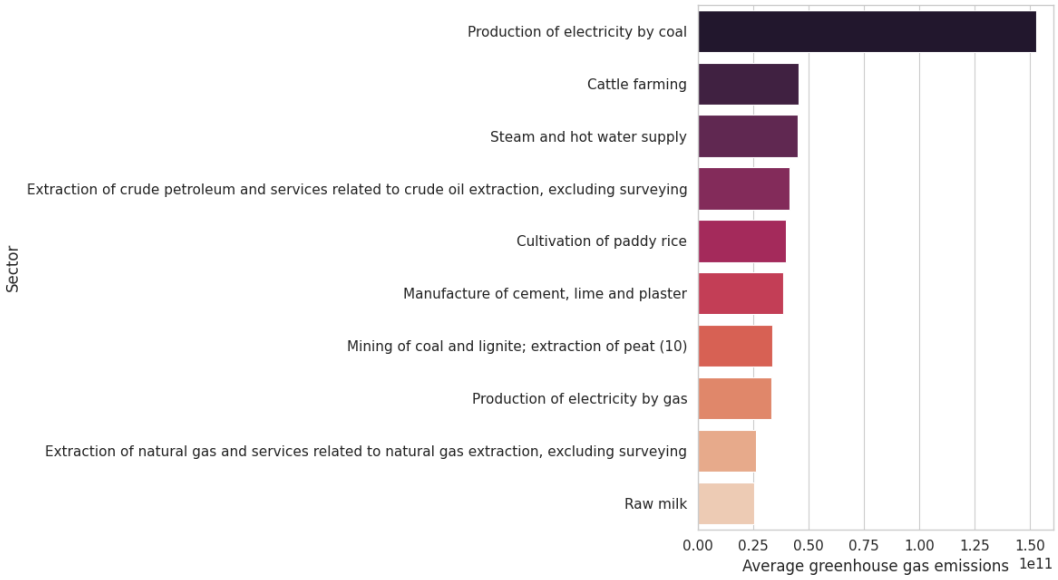

Figure 3. Top sectors by average greenhouse gas emissions

The average emissions from coal-based electricity production were the highest, significantly higher than all other sectors. There were also large environmental effects in agricultural sectors such as cattle farming and raw milk production. In addition, there was high emission intensity in heavy industry sectors like cement manufacturing and petroleum extraction.

3.5 Comparative Performance of Machine Learning Models

Summaries of the predictive performance of the machine learning models assessed are provided in Table 2 and the graphs in Figure 4. The best predictive accuracy was obtained using XGBoost, which had an R^2 value of 0.9179, meaning that about 91.8% of the greenhouse gas emissions' variation was explained. Random Forest also showed a high predictive power, while Linear Regression had significantly lower performance with regard to all the evaluation metrics.

Table 2. Comparative performance of machine learning models

Model	MAE	RMSE	R^2 Score
Linear Regression	1.3289	1.8844	0.6533
Random Forest	0.6782	1.1062	0.8805
XGBoost	0.6007	0.9172	0.9179

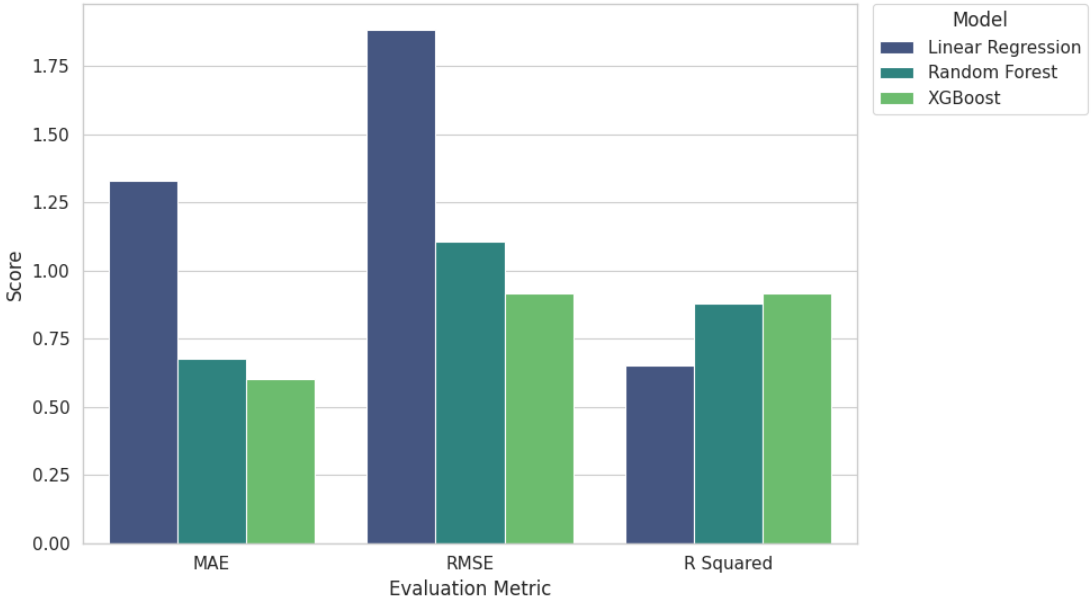

Figure 4. Model performance comparison

The fairly high improvement obtained from the ensemble learning techniques reveals that the relationship between economic activities and the environmental indicators is nonlinear and complex. These findings show that advanced machine learning frameworks are effective for Sustainable economic and environmental decision-making.

3.6 Validation of the XGBoost Prediction Framework

The comparison of actual and predicted greenhouse gas emission values shows the predictive consistency of the XGBoost model, as shown in Figure 5. The majority of observations were close to the perfect prediction reference line, with low estimation errors in the various emission ranges and high prediction stability.

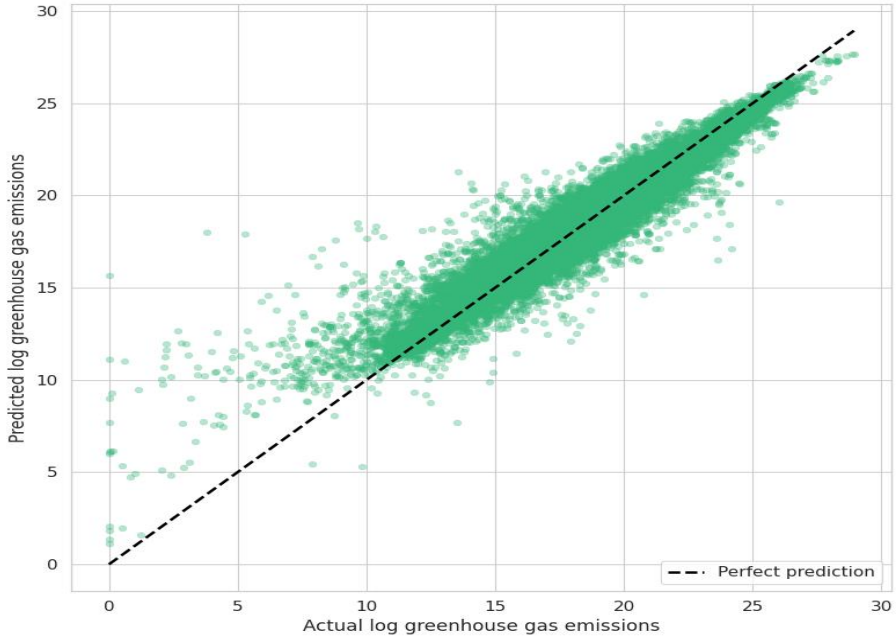

Figure 5. XGBoost actual versus predicted greenhouse gas emissions

The relatively low dispersion around the reference line confirms the validity of the suggested machine learning model and proves its applicability for environmental prediction and sustainable decision making.

3.7 Feature Importance Analysis

Table 3 and Figure 6 show the most influential predictors identified by the XGBoost model. Energy Carrier Net Total was found to be one of the most significant determinants of greenhouse gas emissions. Electricity generation sectors, in addition, landfill-related activities, petroleum refining and livestock production sectors had high predictive importance.

Table 3. Top feature contributions from the XGBoost model

Feature	Importance score
Sector: Production of electricity by nuclear	0.0493
Sector: Landfill of waste: Food	0.0311
Energy Carrier Net Total (TJ)	0.0287
Sector: Production of electricity by hydro	0.0277
Sector: Landfill of waste: Paper	0.0272
Sector: Landfill of waste: Wood	0.0216
Sector: Landfill of waste: Textiles	0.0210
Sector: Meat animals nec	0.0192
Sector: Cattle farming	0.0174
Sector: Landfill of waste: Plastic	0.0167
Sector: Petroleum Refinery	0.0163
Sector: Raw milk	0.0158
Sector: Production of electricity by solar thermal	0.0155
Sector: Waste water treatment, other	0.0148
Sector: Waste water treatment, food	0.0148

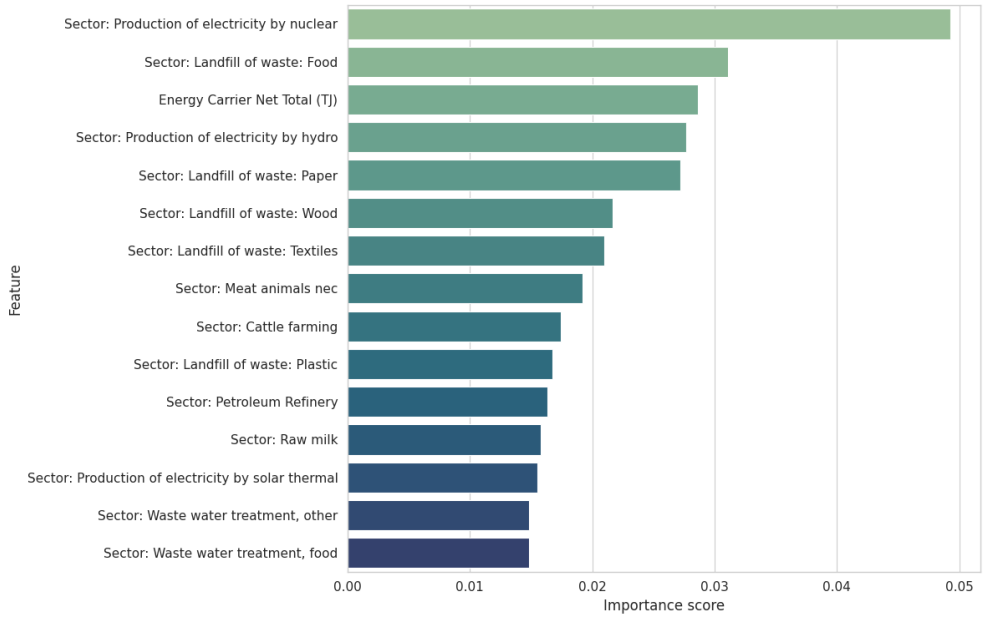

Figure 6. Top feature contributions in the XGBoost model

The results show that energy systems, waste management, industrial production processes and agricultural systems are large contributors to environmental emissions. The explainability analysis thus increases the interpretability and policy relevance of the machine learning framework.

3.8 SHAP Explainability Analysis

For interpretability on the direction and magnitude of feature contribution to model predictions, a SHAP summary plot can be produced, as is done in Figure 7. Energy Carrier Net Total proved to be the most influential energy carrier, increasing predicted GHGs as energy values increased. Employment and Value Added also had positive effects on environmental impact forecasts, but they had less impact.

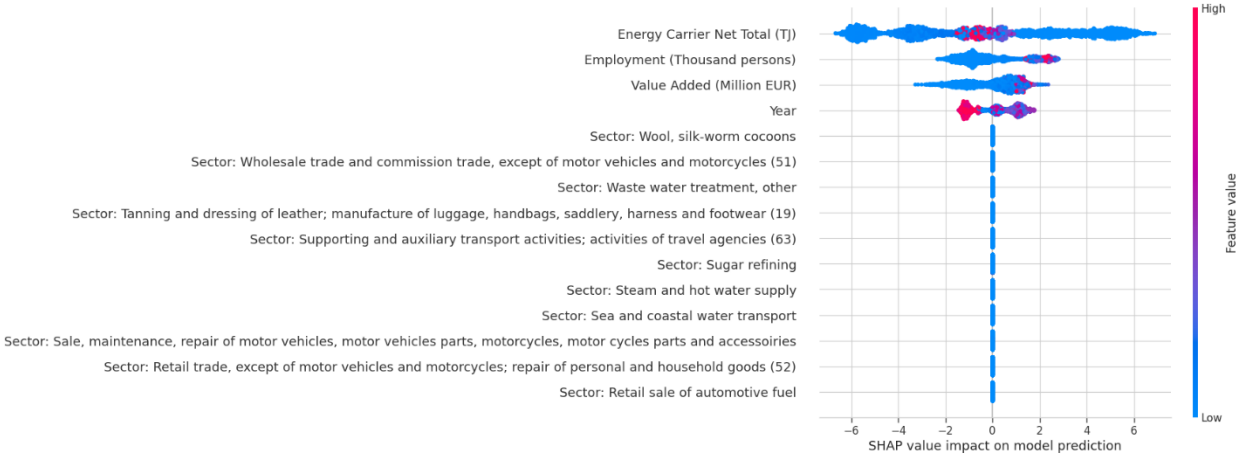

Figure 7. SHAP summary plot for XGBoost emission prediction

The SHAP analysis also showed significant differences in sustainability outcomes across sectors. The complex interactions between the environment and the economy were evident in the various contributions of electricity generation, waste-management activities, and transport-related sectors in the different observations. The results validate the potential of

explainable artificial intelligence approaches in the direction of transparent and evidence-based decision-making for sustainability.

4. Discussion

The results show that machine learning can be a powerful analytical tool to support sustainability-related economic and environmental decisions. As can be seen in the comparative model results, there was a noticeable performance hierarchy of XGBoost, Random Forest, and Linear Regression. The XGBoost model had the greatest explanatory power, which suggests that both the sectoral structure and the regional characteristics play a role in determining GHG emissions, and that these relationships are nonlinear. Linear Regression shows less promise for simple linear assumptions to be adequate for complex environmental-economic systems. Energy Carrier Net Total proved to be a key driver of GHGs. Energy intensity is still one of the key factors when it comes to environmental impact, given that there is a high correlation between energy consumption and emissions and that energy is of significant importance in the predictive model. This finding is aligned with the evidence from the sectors, as electricity generation, petroleum refining, waste treatment, and agricultural production often seemed to be among the most influential sectors. The fact that the energy-related and industrial variables are more important than the others implies that the strategies for the reduction of emissions should not only be directed towards the reduction of the overall growth of the economy, but also towards the characteristics of energy production and the energy structure of particular industries.

The regional results also indicate that the emission intensity is not uniform across economies. The average emission values in China, the USA, and India were high, which was due to the presence of a large industrial base, energy demand and production scale. This is an imbalance of regions that is indicative of differentiated strategies for sustainability policies. Regions with significant emissions could be advised to finally accelerate the energy transition, industry, and build good carbon management systems, while regions with less emissions could prioritize to a greater extent, preventive sustainability planning. The SHAP analysis was used to make sense of the model and to show the direction and the magnitude of the influence of features. Energy consumption and economy/employment-related variables had the most impact on model output. As a result of these findings, it is clear that this framework is not just a 'black box' prediction tool but also a tool that reveals the economic and environmental mechanism that offers explanation for the variation in emissions. This interpretability is crucial for turning the machine learning predictions into actionable decisions for sustainability.

Ensemble models' superior performance is consistent with recent studies in predictive modeling that demonstrated that Random Forest and XGBoost can outperform Linear Regression in prediction tasks related to the environment and resources, as they can model nonlinear patterns (Dalal et al., 2025). The findings of this relationship between energy consumption and GHG emissions are also in line with the results obtained from developing economies, where energy consumption and economic development were found to be significant factors in GHG emissions (Sarkodie & Strezov, 2019). The results of this analysis align with previous research indicating that reliance on fossil fuels and resource limitations contribute to the increased GHG emissions in the major emitting nations (Wang & Azam, 2024). The importance of machine learning for sustainability assessment is further validated by research demonstrating its capability to assess environmental and socio-economic impacts in the context of renewable energy sources (Sumon et al., 2024). The current results contribute to the above literature by combining regional/sectoral, economic and environmental indicators in one comparative framework.

The model's decision-support potential also aligns with studies on machine learning models for load forecasting and multi-criteria decision-making for sustainable energy systems, which

also involved techno-economic and environmental planning via predictive analytics (Dutta et al., 2023). The importance of the livestock and waste-related activities per sector is consistent with previous studies that indicated that animal manure, thermal energy, coal, electricity and methane emissions should be prioritized for detailed forecasting for environmental management purposes (Pence et al., 2022). The overall managerial implications of the results are consistent with recent research on AI in management information systems, which found that AI-driven analytics can enhance the optimization of sustainable supply chains and environmental impact assessment (Tarafder et al., 2026). Digital economy transformation research also indicates that intelligent automation is a tool that can be used to support multi-environmental decision-making (Kuzior & Sira, 2025), which aligns with the focus on intelligent, transparent, and automated sustainability decision systems.

The results have implications for practical application by policy makers, industrial planners and sustainability analysts. The high accuracy rate of XGBoost suggests that machine learning can be used to help identify high-emission activities early on and help to forecast the environment. Decarbonization of key industries such as electricity generation, oil production and refining, waste management, and livestock production are areas that can be targeted with policy using those frameworks. The results also have consequences for a sustainable economic planning. Economic output alone, as a measure of emissions behavior, is not sufficient, and sustainability strategies should focus on other measures as well, such as energy intensity, sectoral activity and production structure. This helps to regulate the environment more accurately and enables decision-makers to be able to differentiate between economically productive sectors and environmentally intensive sectors. The model is a starting point for carbon-aware operational planning for industry. Companies involved in high-emission industries can apply predictive analytics to analyze the possible extent of emissions that could occur under various production and energy consumption scenarios. The explainability component also enhances the transparency of the decisions, since it is able to indicate which factors have the largest impact on the emissions predicted by the model, and thus which factors are important for evidence-based environmental management.

There are some caveats to be noted. The analysis was done using historic economic and environmental factors only, and did not involve real-time climate, technology adoption or policy intervention variables. Some of the sectoral categories were also quite general, thereby restricting the interpretation of specific industrial activities to a quite coarse level. Furthermore, the study focused on three supervised learning models and did not try any hybrid optimisation or deep learning models. Future research would be able to expand on this to incorporate real time energy data, policy indicators, renewable penetration and climate variables. Other algorithms can also be compared, scenario-based forecasting models can also be developed and multi-objective optimization can also be used for developing a carbon reduction planning under economic constraints.

5. Conclusion

Machine Learning and Big Data can be extremely helpful in sustainable economic and environmental decision making. The findings showed that energy consumption, sectoral structure and regional economy play an important role on the greenhouse gas emissions. The results show that XGBoost was the best method to predict the model's accuracy when compared to other two models (Random Forest and Linear Regression), which proves that ensemble learning technique is applicable for modeling the complex relationship between environment and economy. The outcomes also showed that energy-intensive industries, waste management activities, generation of electricity, refining of petroleum and livestock production are significant factors contributing to the variation of the emissions. The explainability results using SHAP confirmed the transparency of the prediction framework, and identified

economic and environmental factors that most affected model predictions. The proposed framework offers a consistent and comprehensible way of predicting emissions, assessing sectors for sustainability, and providing carbon-aware decision support. The conclusions confirm the use of cutting edge machine learning methods in environmental planning, as well as in the field of industrial policy and sustainable development planning. Future research can be done to further refine this framework with real-time energy indicators, policy and scenario-based optimization models.

References

Anonna, F. R., Mohaimin, M. R., Ahmed, A., Nayeem, M. B., Akter, R., Alam, S., ... & Hossain, M. S. (2023). Machine learning-based prediction of us co2 emissions: Developing models for forecasting and sustainable policy formulation. *Journal of Environmental and Agricultural Studies*, *4*(3), 630-643.

Chien, F., Hsu, C. C., Ozturk, I., Sharif, A., & Sadiq, M. (2022). The role of renewable energy and urbanization towards greenhouse gas emission in top Asian countries: Evidence from advance panel estimations. *Renewable Energy*, *186*, 630-643.

Dalal, Y. M., Badrakiya, K., Shivakeshi, C., & Venkatagiri, J. (2025, August). Harnessing Humidity Data: A Comparative Analysis of Linear Regression, Random Forest, and XGBoost for Soil Moisture Prediction. In *2025 Third International Conference on Networks, Multimedia and Information Technology (NMITCON)* (pp. 630-643). IEEE.

Dutta, R., Das, S., & De, S. (2023). Multi criteria decision making with machine-learning based load forecasting methods for techno-economic and environmentally sustainable distributed hybrid energy solution. *Energy Conversion and Management*, *291*, 117316.

Guo, Y., & Ma, J. (2024). ExioML: Emission Factor Database for Scope 3 Emission Estimation Machine Learning Benchmarks [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10604610

Hurlbert, M., Krishnaswamy, J., Johnson, F. X., Rodríguez-Morales, J. E., & Zommers, Z. (2019). Risk management and decision making in relation to sustainable development.

Khaddam, A. A., & Alzghoul, A. (2025). Artificial intelligence-driven business intelligence for strategic energy and ESG management: a systematic review of economic and policy implications. *International Journal of Energy Economics and Policy*, *15*(4), 630-643.

Kuzior, A., & Sira, M. (2025). Sustainable Digital Economy Transformation Through Intelligent Automation: A Multi-Environmental Framework for Strategic Decision-Making. *Sustainability*, *17*(17), 7723.

Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying the carbon emissions of machine learning. *arXiv preprint arXiv:1910.09700*.

Lamb, W. F., Wiedmann, T., Pongratz, J., Andrew, R., Crippa, M., Olivier, J. G., ... & Minx, J. (2021). A review of trends and drivers of greenhouse gas emissions by sector from 1990 to 2018. *Environmental research letters*, *16*(7), 073005.

Luo, J., Zhuo, W., Liu, S., & Xu, B. (2024). The optimization of carbon emission prediction in low carbon energy economy under big data. *IEEe Access*, *12*, 1630-6432.

Marttila, M. P., Uusitalo, V., Linnanen, L., & Mikkilä, M. H. (2021). Agro-industrial symbiosis and alternative heating systems for decreasing the global warming potential of greenhouse production. *Sustainability*, *13*(16), 9040.

Mohammed, M. A., Ahmed, M. A., & Hacimahmud, A. V. (2023). Data-driven sustainability: leveraging big data and machine learning to build a greener future. *Babylonian Journal of Artificial Intelligence*, *2023*, 630-643.

Pence, I., Kumaş, K., Siseci, M. C., & Akyüz, A. (2022). Detailed analysis of animal manure-based CO2 emissions, coal, electricity, thermal energy, and CH4 emissions and using machine learning as a forecasting method: a study from Turkey.

Sarkodie, S. A., & Strezov, V. (2019). Effect of foreign direct investments, economic development and energy consumption on greenhouse gas emissions in developing countries. *Science of the total environment*, *646*, 630-643.

Sumon, M. F. I., Osiujjaman, M., Khan, M. A., Rahman, A., Uddin, M. S., Pant, L. P., & Debnath, P. (2024). Environmental and socio-economic impact assessment of renewable energy using machine learning models. *Journal of Economics, Finance, and Accounting Studies*, *6*(5), 112.

Talaat, F. M., Kabeel, A. E., & Shaban, W. M. (2025). Towards sustainable energy management: Leveraging explainable Artificial Intelligence for transparent and efficient decision-making. *Sustainable Energy Technologies and Assessments*, *78*, 104348.

Tarafder, M. T. R., Ansari, M. E., Alam, M. A., Shil, S. K., Islam, R., & Ahmed, K. R. (2026). Leveraging artificial intelligence in management information systems for sustainable supply chain optimization and environmental impact analysis. *Discover Artificial Intelligence*, *6*(1), 56.

Tran, T. T. T., Tran, H. T., & Duong, K. D. (2026). The non-linear relationship between economic development and air pollution: Evidence from panel data analysis. *Asian Journal of Economics and Banking*, *10*(1), 630-643.

Uyar, N., & Uyar, A. (2025). Assessing climate change impacts on cropland and greenhouse gas emissions using remote sensing and machine learning. *Atmosphere*, *16*(4), 418.

Wang, J., & Azam, W. (2024). Natural resource scarcity, fossil fuel energy consumption, and total greenhouse gas emissions in top emitting countries. *Geoscience frontiers*, *15*(2), 101757.