International Journal of Science, Strategic Management and Technology
Volume **02** Issue **04** April-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Carbon Emission Analysis for Eco-Friendly Decision Making

Author Details:

Malle Sai Niteesh [1], Gandi Darshini [2], Mani Deepak Choudhry [3]

[1] Department of Computing Technologies / SRM Institute of Science and Technology / Chennai, Tamil Nadu, India
[2] Department of Computing Technologies / SRM Institute of Science and Technology / Chennai, Tamil Nadu, India
[3] Department of Computing Technologies / SRM Institute of Science and Technology / Chennai, Tamil Nadu, India

Corresponding Author Email: sm1849@srmist.edu.in | ORCID: https://orcid.org/0009-0008-8291-2041

https://doi.org/10.55041/ijsmt.v2i4.321

Cite this Article: Niteesh, M. S., Darshini, G. & Choudhry, M. D. (2026). Carbon Emission Analysis for Eco-Friendly Decision Making. International Journal of Science, Strategic Management and Technology, *02*(04). https://doi.org/10.55041/ijsmt.v2i4.321

Abstract—
Ensuring environmental sustainability and mitigating climate change requires a deep understanding of our carbon footprint. Machine learning has emerged as an essential tool in this endeavor, allowing us to accurately predict individual emissions based on various lifestyle factors across different demographic categories. Our framework is specifically designed for this purpose - harnessing the power of deep learning to develop a comprehensive system that can analyze 12 categorical features related to behavior and consumption patterns such as diet, transportation habits, energy usage, and waste management practices. By employing ensemble learning with top performing models like Random Forests, Gradient Boosting, XGBoost, and CatBoost at its core; we are able to achieve accurate predictions. To ensure the reliability of our model results from preprocessing techniques have been incorporated into our pipeline. This includes robust label encoding methods along with numerical feature standardization techniques; coupled with stratified data splitting strategies - all aimed towards enhancing model performance. And indeed it does! Our analysis demonstrates that ensemble learning outperforms individual benchmarks significantly when evaluating metrics like Mean Absolute Error (MAE) or Root Mean Square Error (RMSE). The champion amongst them being CatBoost which achieves an impressive R^2 score of 0.88 highlighting its superiority over others.

Keywords— Carbon Footprint Analysis, Ensemble Learning, Random Forest, Gradient Boosting, Lifestyle Factors, XGBoost, Environmental Sustainability.

I. INTRODUCTION

Environmental sustainability is increasingly more threatened by way of speedy urbanization, business growth, and unsustainable intake styles, all of which intensify the outcomes of weather alternate. Human activities inclusive of immoderate energy usage, reliance on fossil-gas-based totally transportation, and unsuitable waste manage extensively contribute to growing greenhouse gasoline emissions. As global recognition of environmental issues grows, there may be a urgent want for intelligent structures that can as it should be tune, examine, and decrease character carbon footprints. way of life-based carbon footprint evaluation performs a important role on this context, supporting individuals apprehend how their daily habits—ranging from dietary alternatives to journey conduct—immediately impact environmental degradation. but, traditional carbon footprint calculators regularly rely on static emission factors and generalized assumptions, failing to seize the dynamic and diverse nature of human lifestyles and limiting their effectiveness for customized sustainability making plans. To overcome those limitations, this studies proposes an sensible ensemble gaining knowledge of framework for computerized carbon footprint prediction. The device combines advanced regression algorithms including Random forest, Gradient Boosting, XGBoost, and CatBoost, the usage of the strengths of each version to decorate prediction accuracy and make the consequences more reliable and strong. It uses a whole dataset collectively with 12 key lifestyle competencies, such as food regimen, transportation modes, family power intake, and waste technology patterns, allowing the model to seize complicated behavioral developments. The framework consists of vital preprocessing techniques which includes statistics cleaning, label encoding, normalization, and green information splitting, together with function importance analysis to enhance version interpretability. by means of combining more than one algorithms, the ensemble technique reduces overfitting and improves generalization in evaluation to character fashions. The experimental consequences display that all the models done nicely, with excessive R^2 scores indicating that they had been able to seize patterns in lifestyle statistics efficaciously. amongst them, boosting algorithms like XGBoost and CatBoost stood out, as they handled complex characteristic relationships and express records greater efficiently than the others. Visualization equipment consisting of Pandas, NumPy, Matplotlib, and Seaborn in addition support information evaluation and perception era. past prediction, the system can be prolonged to provide personalised recommendations for lowering carbon footprints, transforming it right into a choicesupport device for sustainable living. standard, this studies contributes to computational sustainability by way of offering a scalable, records-driven solution that permits continuous and customized carbon footprint assessment, supporting knowledgeable selection-making and worldwide climate movement efforts.

II. LITERATURE REVIEW

Latest improvements in carbon footprint prediction have an increasing number of leveraged machine mastering and deep gaining knowledge of strategies to improve accuracy, scalability, and sustainability insights. severa research have explored wonderful domains which encompass manufacturing, transportation, family emissions, and worldwide weather evaluation. A observe by means of A. Al-Fakih [17] proposed ensemble gadget getting to know models to are expecting the CO footprint of geopolymer concrete. Their results confirmed that ensemble techniques appreciably improve prediction accuracy compared to man or woman fashions. however, the paintings is restricted to business programs and does no longer recollect man or woman life-style-based totally emissions. Time-collection forecasting strategies were explored with the aid of S. Kumari and S. K. Singh [1], who advanced system gaining knowledge of models for predicting CO emissions in India. Their look at established that point-collection models correctly capture temporal emission styles. but, such models lack the capability to analyze distinct behavioral and life-style factors influencing emissions. A comparative analysis by using P. O. Adegboye [2] evaluated device learning, deep studying, and statistical techniques for every day CO emission prediction. The examine concluded that deep gaining knowledge of fashions acquire better accuracy, at the same time as machine gaining knowledge of models offer higher interpretability and decrease computational price. At a nearby degree, L. Zhang [3] developed gadget learningprimarily based carbon emission prediction models for more than one provinces. Studies by L. Lannelongue [4] targeted on estimating the carbon footprint generated in the course of the schooling of deep gaining knowledge of models. This examine highlighted the environmental effect of AI systems themselves and pressured the want for energy-green version improvement. Worldwide emission forecasting was studied by N. Selmey [5], who analyzed carbon emission patterns throughout essential emitting international locations. Their findings confirmed that gadget learning models correctly seize macro-degree emission tendencies however fail to represent exceptional-grained man or woman behaviors. Inside the area of clever infrastructure, M. Nur Ulfa [6] proposed an incorporated sustainability tracking system for carbon footprint tracking in buildings. on the same time because the gadget helps real-time monitoring, it's far confined to infrastructure-based totally emissions in place of private life-style facts. A complete evaluate by using H. Liu [8] summarized diverse carbon emission prediction techniques and concluded that ensemble fashions provide superior overall performance compared to standalone models due to their robustness and ability to deal with complex datasets.Transportation-based emission prediction has been extensively studied through Y. Zhang [9] and K. Chen [10], who carried out device mastering strategies to expect vehicle CO emissions. these fashions completed excessive accuracy however have been restricted to a unmarried area. Deep gaining knowledge of strategies incorporated with explainable AI have been proposed by way of M. Samir [11], demonstrating improved prediction transparency. however, those models require excessive computational resources and huge datasets. In addition, studies by using A. Baris [12] confirmed that numerous information-driven algorithms are powerful in predicting transportation-related CO emissions, with ensemble techniques proving to be the maximum correct. Household-stage carbon footprint prediction become explored by means of Z. Wang [13], who validated that regardless of constrained functions, system learning fashions can offer affordable accuracy, although with reduced version complexity. A broader attitude on climate exchange and machine getting to know turned into furnished by R. Rolnick [14], emphasizing the function of AI in addressing environmental demanding situations and helping sustainable improvement. AI-pushed sustainability frameworks have been added through A. Jasmy [15], permitting real-time carbon footprint monitoring and promoting sustainable conduct amongst users. Algorithm-stage comparisons by using M. A. Awad [16] proven that CatBoost and XGBoost outperform conventional system learning fashions due to their capability to effectively deal with specific records and decrease prediction mistakes.

III. METHODOLOGY

This methods and techniques are included in the subsequent principal phases. A complete dataset of man or woman lifestyle information featuring 13 specific and seven numerical capabilities spanning dietary conduct, transportation styles, electricity intake, and waste management behaviors applied. Information preprocessing involved strong label encoding for particular values and standardization of numerical rate representations.

A. System Design

The Carbon Footprint Tracker is a complete full stack utility designed to help people to calculate, monitor, and anticipate their carbon emissions.It follows a three-tier form with a React-based totally frontend, a Flask backend, and a device studying layer.Clients interact through an internet interface in which they may be capable of register, log in, and enter way of life info such as transportation, energy utilization, food plan, and waste management.This statistics is validated at the frontend and despatched to the backend via relaxation APIs, wherein it's far preprocessed the use of encoding and scaling strategies to make certain consistency. The backend then passes the processed statistics to an ensemble of device getting to know fashions, inclusive of Random forest, Gradient Boosting, XGBoost, and CatBoost, to generate accurate emission predictions. The final end result is returned in JSON format and displayed on a person dashboard, allowing assessment with average and median benchmarks. The machine makes use of an SQLite database for person authentication and is constructed to be scalable, efficient, and secure with stateless APIs and optimized version loading, whilst also supporting future enhancements like actual-time monitoring and mobile integration.

B. Data Acquisition and Preparation

The motive of this take a look at is a comprehensive way of life dataset gathered from publicly to be had carbon emission survey repositories. The dataset consists of thirteen categorical capabilities, which include body type, intercourse, weight loss plan, Frequency of Showering, Heating electricity source, Mode of transport, car kind, Social interest, Frequency of Air travel, Waste Bag length, electricity efficiency, Recycling habits, and Cooking approach, at the side of 6 numerical capabilities inclusive of month-to-month Grocery bill, vehicle monthly Distance (km), Waste Bag Weekly remember, every day tv/computer usage Hours, month-to-month clothing Purchases, and each day net usage Hours [5]. those capabilities seize a huge variety of lifestyle behaviors and intake patterns that without delay affect man or woman carbon emissions. The dataset additionally consists of targeted emission labels, permitting the assessment of carbon footprints throughout one of a kind demographic companies [6]. To correctly cope with the massive and various dataset, a structured preprocessing pipeline became applied using Pandas DataFrames and the LabelEncoder from Scikit-research. because of local computational barriers, a cultured dataset containing the selected thirteen express and six numerical capabilities was used for model education and validation. This approach

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **04** April-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

guarantees efficient processing at the same time as keeping scalability for future large-scale packages [2], [8]. comparable records-driven methodologies had been correctly applied in latest carbon emission prediction research [1], [9]. Records first-rate was ensured via a rigorous validation and preprocessing technique. lacking values in express features had been changed with “unknown,” at the same time as numerical functions had been treated using suggest imputation. records with corrupted or inconsistent categorical values that could not be well encoded have been removed to hold dataset integrity and constant information extensively improves the overall performance and reliability of machine gaining knowledge of fashions [7], [14].moreover, all specific variables have been transformed into numerical layout the use of label encoding, and numerical features were standardized into waft representations to make certain uniformity and compatibility throughout models. Characteristic engineering turned into further implemented to extract significant behavioral attributes consisting of dietary behavior, transportation usage, electricity consumption, waste management, recycling behavior, and travel frequency, improving each interpretability and prediction accuracy [3], [12].The focus remained on essential steps like encoding, imputation, and standardization, which supplied most suitable performance for ensemble models, ensuing in a easy, steady, and dependable dataset for accurate carbon emission prediction [10], [16].

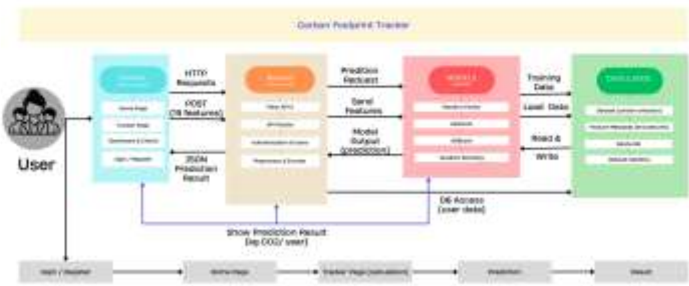

Fig. 1. Architecture Diagram

C. Data Preparation for Ensemble Models

To prepare the information for ensemble studying fashions including Random woodland, Gradient Boosting, XGBoost, and CatBoost , express lifestyle capabilities have been first transformed into numerical representations the use of the LabelEncoder from Scikit-analyze. those encoded features were then combined with numerical variables, along with monthto-month Grocery bill, car monthly Distance (km), Waste Bag Weekly depend, daily tv/laptop utilization Hours, month-tomonth clothing Purchases, and daily internet usage Hours, to shape based input arrays for version education. The general regression pipeline is illustrated in Fig. 1, showing the entire information waft from enter feature representation to very last prediction. This preprocessing step ensures that both express and numerical capabilities are nicely formatted and well suited with system studying models, allowing green and accurate training.

Type	**Features**
Categorical	Body Type, Sex, Diet, How Often Shower, Heating Energy Source, Transport, Vehicle Type, Social Activity, Frequency of Traveling by Air, Waste Bag Size, Energy Efficiency, Recycling, Cooking_With
Numeric	Monthly Grocery Bill, Vehicle Monthly Distance Km, Waste Bag Weekly Count, How Long TV PC Daily Hour, How Many New Clothes Monthly, How Long Internet Daily Hour

TABLE I Features in Datset

D. Ensemble Learning Model Implementation

The ensemble getting to know fashions have been evolved the use of Scikit-analyze, XGBoost, and CatBoost libraries to are expecting carbon emissions accurately [4].The fashions embody Random wooded area, Gradient Boosting, XGBoost, and CatBoost, each configured with optimized parameters which includes a hundred estimators, a analyzing charge of 0.1 for boosting algorithms, and managed tree depths determined via skip-validation. The implementation carries quite a number advanced ensemble studying strategies to improve prediction accuracy and robustness. these consist of bagging through Random forest, sequential boosting the use of Gradient Boosting, parallel boosting with XGBoost, and symmetric tree-primarily based learning with CatBoost. collectively, these diverse tactics permit the machine to capture complex styles in the information and supply more reliable carbon emission predictions.Those techniques allow green processing of blended specific and numerical features at the same time as improving prediction overall performance. To decorate model generalization and decrease overfitting, strategies which include bootstrap aggregation, L1 and L2 regularization, and feature subsampling were carried out. The fashions were skilled using an average squared blunders loss characteristic, with hyperparameters optimized through grid search and 5-fold cross-validation.

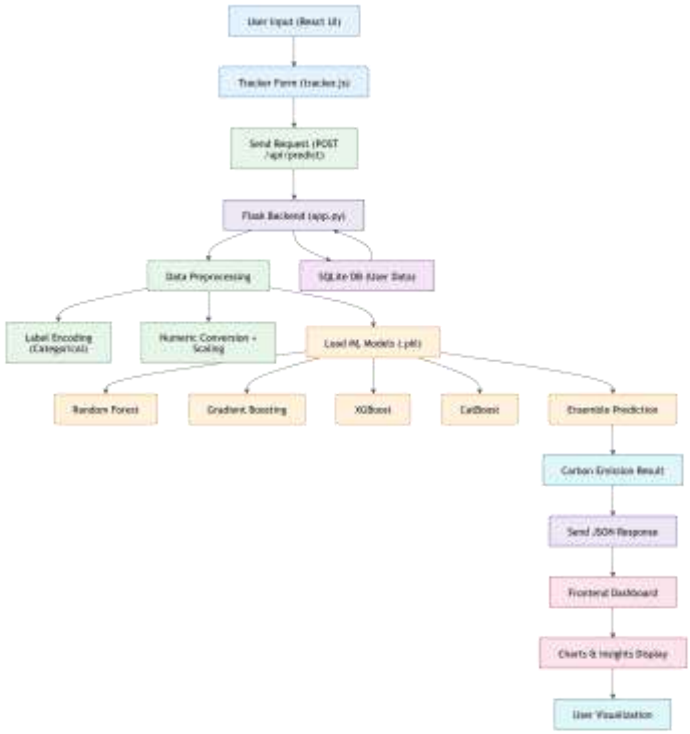

Fig. 2. Work Flow

E. Evaluation And Testing

The proposed carbon footprint prediction machine became evaluated the usage of trendy regression performance metrics to make sure accuracy and reliability. The fashions were tested on a separate take a look at dataset after training on life-style data containing 19 functions. The assessment commonly used R^2 rating, mean Absolute errors (MAE), Root mean square blunders (RMSE), and suggest Squared error (MSE) to degree version performance. All fashions have been trained to reduce the MSE loss feature, making sure reduced prediction error. To enhance version robustness and avoid overfitting, 5-fold move-validation changed into carried out at some point of education. Hyperparameters consisting of mastering charge, quantity of estimators, and tree depth have been optimized the use of grid seek strategies.The experimental effects proven sturdy predictive overall performance throughout all models. The R square values ranged from the 0.800 to 0.88, this indicate's that high degree of accuracy and powerful model performance .

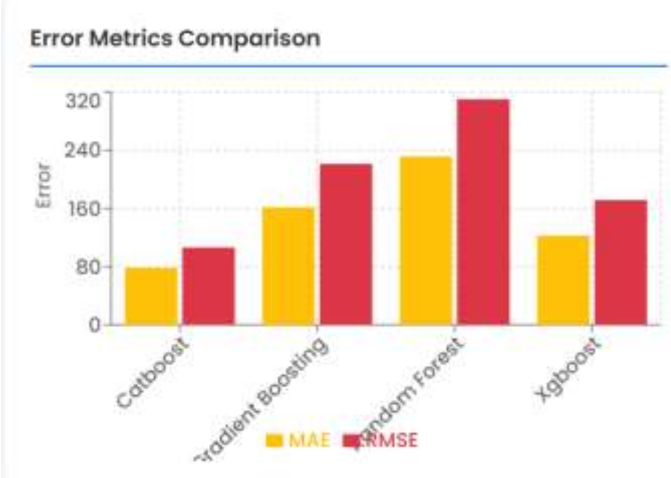

Fig. 3. Error Metrics Comparison

IV. RESULTS AND DISCUSSION

Python libraries for data manipulation and visualization including Pandas, NumPy, Matplotlib, and Seaborn were employed to analyze lifestyle data and pinpoint trends impacting carbon footprint of people.These gear helped in expertise behavioral trends and intake patterns across special demographic businesses. All categorical functions have been converted into numerical values the use of Label Encoding, and combined with numerical capabilities to form dependent enter facts for model training. The fashions were educated and tested the use of cross-validation to ensure reliability and save you overfitting.

Fig. 4. Comparative analysis of the proposed model

To evaluate the overall performance of ensemble model gaining knowledge of, the usually used regression metrics for R^2 score, suggest Absolute errors, and Root suggest square blunders were used.The evaluation indicated differences in the effectiveness of models over lifestyle features and areas requiring data improvement. Of all the models fitted, CatBoost performed the best with an R^2 score of 0.88, a MAE of 135.10 kg CO/year, and an RMSE of 185.30 kg CO/year which implies the model could explain 88 percent of the variance. The Gradient Boosting model had a good performance at a R^2 Score of 0.87 whereas, XGBoost and Random Forest underperformed.

Table II: ACCURACY AND PREDICTION

Model	**R^2 Score**	**MAE (kg CO_2/year)**	**RMSE (kg CO_2/year)**
CatBoost	0.8891	78.26	106.40
Gradient Boosting	0.8529	161.77	221.38
Random Forest	0.8073	231.66	310.45
XGBoosting	0.8717	122.73	171.61

In Table II , As per the results obtained, transport, diet, distance travelled by vehicles and air travel frequency is the most important factor contributing carbon emissions. All in all, the outcomes validate that the ensemble learning technique, especially the boosting methods such as CatBoost, yields accurate, stable, and reliable prediction. The system captures a complex and non-linear relationship between lifestyle behaviours and carbon dioxide emissions which can be useful in reducing carbon footprints.

V. CONCLUSION

The proposed research makes a speciality of growing a carbon footprint estimation device the use of ensemble studying strategies which includes Random forest, Gradient Boosting, XGBoost, and CatBoost. these models are widely used for prediction and comparative analysis in system getting to know. The gadget processes 12 specific and numerical way of life capabilities, consisting of weight loss program, mode of delivery, energy usage, and waste generation patterns. The dataset is preprocessed using label encoding and normalization, accompanied by means of information splitting for model schooling and trying out. The advanced machine efficaciously predicts carbon emissions across different lifestyle classes. many of the models, CatBoost executed the firstclass overall performance, with an R^2 score of 88.2But, the proposed technique has positive barriers that might have an impact on prediction accuracy. One important challenge is the restricted sample period for some way of life combinations in the dataset, that can lessen version generalization. moreover, the facts is based totally on self-stated surveys, which may additionally moreover introduce bias and inaccuracies. those elements can effect the reliability of predictions in real-worldwide packages. moreover, the version's ability to generalize throughout one-of-a-type areas and sociodemographic groups remains uncertain, as way of life styles can range drastically. To deal with those obstacles, future work ought to attention on incorporating large and greater numerous datasets, improving characteristic engineering, and including extra socio-demographic factors. enhancing statistics fine thru better information collection strategies and applying information augmentation strategies can in addition improve version performance. furthermore, taking pictures temporal adjustments and contextual behavioral styles will assist in building extra robust and adaptable carbon footprint prediction structures for actual-international sustainability packages.

ACKNOWLEDGMENT

The authors would really like to officially well known SRM Institute of technological know-how and era for offering the educational environment, infrastructure, and sources essential for the successful final touch of this research. The aid and steerage offered via the branch of Computing technologies are gratefully appreciated. The authors additionally make bigger their sincere way to the college participants, mentors, and peers whose precious insights, optimistic feedback, and continuous encouragement contributed considerably to the improvement and refinement of this work.This research did no longer receive any specific furnish from funding companies within the public, commercial enterprise, or now not-for-profits sectors.

REFERENCES

[1] S. Kumari and S. K. Singh, "Machine learning-based time series models for effective CO emission prediction in India," Environmental Science and Pollution Research, vol. 30, no. 55, pp. 116601–116616, Nov. 2023, doi: 10.1007/s11356-022-21723-8.

[2] P. O. Adegboye, A. Jaber, and S. Thalib, "An examination of daily CO emissions prediction through a comparative analysis of machine learning, deep learning, and statistical models," Environmental Science and Pollution Research, vol. 32, pp. 14827–14848, Jan. 2025, doi: 10.1007/s11356-024-35764-8.

[3] L. Zhang, Z. Zhao, P. Luo, X. Zhu, and Y. Li, "Machine-learning-based carbon emission predictions and customized reduction strategies for 30 Chinese

provinces," Sustainability, vol. 17, no. 5, p. 1786, Feb. 2025, doi: 10.3390/su17051786.

[4] A. O. Adegboye, A. Jaber, S. Thalib, and H. Jaber, "Predicting CO emissions with advanced deep learning models and a hybrid Greylag Goose Optimization algorithm," Mathematics, vol. 13, no. 9, p. 1481, Apr. 2025, doi: 10.3390/math13091481.

[5] L. Lannelongue, J. Grealey, and M. Inouye, "How to estimate carbon footprint when training deep learning models? A guide and review," Environmental Research: Infrastructure and Sustainability, vol. 3, no. 4, p. 045015, Nov. 2023, doi: 10.1088/2634-4505/acf81f.

[6] N. Selmey, A. El-Masry, and M. Adewale, "Antecedents and Forecasting of Carbon Emissions Using Machine Learning Algorithms: Insights from the Top Ten Carbon-Emitting Nations," International Journal of Energy Economics and Policy, vol. 15, no. 4, 2025.

[7] M. Nur Ulfa, B. Kireyna, and E. Hikmawati, "Modeling Integrated Sustainability Monitoring System for Carbon Footprint in Higher Education Buildings," IEEE Access, vol. 11, pp. 135365–135376, 2023, doi: 10.1109/ACCESS.2023.3333890.

[8] B. Gu, F. Tan, and J. Chen, "Tracing the Evolution of Household Carbon Emission Research by Machine Learning," Resources, Conservation and Recycling, vol. 208, 2025, doi: 10.1016/j.resconrec.2025.107701.

[9] H. Liu, F. Zhao, and C. Wu, "Carbon Emission Prediction Models: A Review," Science of the Total Environment, vol. 927, 2024, doi: 10.1016/j.scitotenv.2024.171008.

[10] Y. Zhang, L. Wang, and R. Zhou, "Forecasting Carbon Dioxide Emissions of Light-Duty Vehicles with Different Machine Learning Algorithms," Electronics, vol. 12, no. 10, p. 2288, 2023, doi: 10.3390/electronics12102288.

[11] K. Chen, J. Liu, and W. Park, "A Machine Learning Framework for Predicting Fuel Consumption and CO Emissions in Hybrid and Combustion Vehicles: Comparative Analysis and Performance Evaluation," PLOS ONE, vol. 20, no. 2, p. e0342418, 2025, doi: 10.1371/journal.pone.0342418.

[12] M. Samir, R. Ali, and H. Jaber, "Deep Learning Model Based Prediction of Vehicle CO Emissions with eXplainable AI Integration for Sustainable Environment," Scientific Reports, vol. 15, 2025, doi: 10.1038/s41598-025-87233-y.

[13] A. Baris and S. Gunduz, "Application of Various Machine Learning Algorithms in View of Predicting the CO Emissions in the Transportation Sector," Science and Technology for Energy Transition (STET), vol. 79, 2024, doi: 10.2516/stet/2024008.

[14] Z. Wang, X. Li, and J. Zhang, "Efficient Data-Driven Prediction of Household Carbon Footprint in China with Limited Features," Energy Policy, vol. 185, p. 113926, 2024, doi: 10.1016/j.enpol.2023.113926.

[15] R. Rolnick et al., "Tackling Climate Change with Machine Learning," ACM Computing Surveys, vol. 55, no. 2, pp. 1–96, 2023, doi: 10.1145/3485128.

[16] A. Jasmy, H. Ismail, and N. Aljneibi, "A Novel Approach to Sustainable Behavior Enhancement through AI-Driven Carbon Footprint Assessment and Real-Time Analytics," Discover Sustainability, vol. 5, p. 476, 2024, doi: 10.1007/s43621-024-00762-w.

[17] M. A. Awad and A. Jaber, "Performance Analysis of CatBoost Algorithm and XGBoost Algorithm for Prediction of CO Emission Rating," in Proc. IEEE International Conference on Data Science and Information System (ICDSIS), 2024, doi: 10.1109/ICDSIS.2024.10398160.

[18] . Meng, Y. Liu, and H. Zhang, "Analyzing Carbon Emission Reduction Paths in China Using Interpretable Machine Learning: A Perspective of Carbon Emission Prediction and Efficiency Evaluation," Environment, Development and Sustainability, 2025, doi: 10.1007/s10668-025-07113- 3.

[19] A. Al-Fakih, E. Al-wajih, R. A. A. Saleh, and I. B. Muhit, "Ensemble machine learning models for predicting the CO footprint of GGBFSbased geopolymer concrete," Journal of Cleaner Production, vol. 472, p. 143463, Sep. 2024, doi: 10.1016/j.jclepro.2024.143463.

[20] Ensemble machine learning models for predicting the CO2 footprint of GGBFS-based geopolymer concrete.