

Carbon Footprint Estimator & Reducer

Mrs. Jageti. Padmavathy¹, S.Radhika²

¹ Assistant Professor, Dept of Computer Science and Engineering,

G. Narayanamma Institute of Technology and Science, Hyderabad, India – 500090

² Assistant Professor, Dept of Computer Science and Engineering,

G. Narayanamma Institute of Technology and Science, Hyderabad, India – 500090

1. Abstract

The Carbon Footprint Estimator and Reducer is a complete online sustainable tool that has been built with the implementation of machine learning to facilitate easy and tailored assessment of carbon footprint. The tool offers a user interface, allowing users to input data in four diverse fields, namely Transportation, Electricity, Diet, and Waste. The users are provided an opportunity to calculate and evaluate carbon emission in a thorough and convenient manner. Machine learning models were trained in each domain to accurately calculate CO₂ emission based on user actions. The application provides total emissions, along with emissions per domain, by utilizing interactive Streamlit dashboards, thus making it very effective for monitoring and comparison purposes. Secondly, the application also predicts future carbon emissions by utilizing current usage patterns and even provides information regarding Scope 1, Scope 2, and Scope 3 emissions. Finally, for ease of usability, the application integrates chatbot using Gemini LLM API, which produces customized, human-readable instructions regarding the reduction of carbon emissions within the environment. The successful development of this application is one of the best ways to prove the combined efficiency of machine learning, data visualization, and AI-based instructions related to the encouragement of sustainable lifestyle habits.

2. Introduction

The increase in the amount of greenhouse gases and carbon dioxide is one of the world's biggest problems right now. Global warming is largely due to everyday activities such as transportation, energy consumption, food choices and waste generation. Another issue is that most carbon footprint calculators do not provide immediate feedback on your changes to your behavior over time and do not provide guidance in an interactive way. Further, most calculators only have very limited capabilities.

The project uses innovative machine learning technology to analyze individual behavior and calculate total carbon emissions from all three levels of scope (1, 2 and 3) for each individual. Users are provided with transport, energy use, dietary habits and waste (from all three scopes of emissions) to provide a complete overall view of an individual's carbon footprint. Predictive analytics will then provide an accurate estimate of carbon emissions and provide users with intelligent, action-oriented ways to reduce their environmental impact. In addition, a Gemini API-powered chatbot will offer users real-time, personalized suggestions and assist them with communicating through an interactive, friendly, human-like manner. The application will also help to provide users with a projection of their future carbon emissions, based upon their current habits, so that they can make informed, proactive decisions to lead a more environmentally sustainable lifestyle.

2. Literature review

[1] **Ahmed, S., Rahman, T., & Zhou, L. (2022)** Compared multiple ML models such as regression, neural networks, and hybrid approaches for carbon emission prediction. Their study showed that ML models effectively identify emission sources and patterns, but many lacked real-world validation and relied on static datasets.

[2] **Brown, R., & Li, Y. (2024)** employed Advanced Temporal Models to analyze emission trends over time using smart meter and transportation data, ultimately achieving a better level of accuracy in their validation processes than traditional models; they recommend Federated Learning to allow for secure and collaborative training.

[3] **Chen, Y., & Kumar, R. (2021)** found that SVM and Decision Trees could be used to estimate regional carbon footprints, however, accuracy was dependent upon the amount of hazards in each area, so standardizing and having them within the same geographic region was critical in their analysis.

[4] **Davis, M., Ortega, S., & Lin, P. (2022)** They studied synthetically generated data to compensate for data scarcity and privacy issues. While enabling model generalisation, synthetically generated datasets required proper validation against real-world circumstances.

[5] **Garcia, R., Lopez, D., & Martins, A. (2021)** Used Random Forest, Linear Regression and Gradient Boosting to forecast emissions. They stressed that understanding how to interpret their models is necessary, decision-makers in their efforts to reduce their carbon footprints.

[6] **Johnson, A., & Patel, R. (2020)** Using machine learning and data compiled from smart devices, created a real-time feedback system for individuals to gain feedback on their own estimated carbon footprint.

[7] **Lee, H., Park, J., & Silva, M. (2020)** used ensemble models to analyze patterns of urban energy consumption and transportation, demonstrating that predicting these urban energy use patterns in regions with minimal availability of data could be achieved through transfer learning.

[8] **Nguyen, T., Roy, S., & Abbas, K. (2023)** used different transfer learning techniques can improve prediction accuracy in regions with limited data availability. They observed that locally fine-tuned models were more successful at capturing local patterns.

[9] **Smith, J., Carter, B., & Wilson, M. (2019)** conducted earlier research with data for emission predictions, using linear regression, decision trees, and ensemble methods to make predictions. They discovered some general patterns of emission, but were hindered by lack of data.

[10] **Tanaka, R., & Roy, P. (2023)** they examined methods to explain machine learning outputs. The authors used SHAP as a method of interpreting their ML output. They concluded that increased transparency and explainability of ML outputs.

3. Objectives

- To create a Carbon Footprint Estimator and Reducer that will enable an accurate calculation of a user's carbon footprint based on the user-provided information and data regarding their transportation, energy usage, diet, and waste disposal patterns.
- To create a personalized recommendation tool that assists users in making lifestyle modifications that would enable them to reduce their carbon footprint. This tool would allow both individual users and groups of users to work together to reach their goals for sustainability.

- To develop a method for predicting future carbon footprints based on the carbon emissions from the current activities of users. This prediction will allow users to plan ahead and make optimal choices for sustainable living.
- To create an artificial intelligence (AI) chatbot that will be available 24/7 to provide real-time support and responses to all questions and inquiries; provide customized recommendations and advice regarding sustainable living; and engage users in ongoing dialogue about ways to be more environmentally responsible and aware.

4. Existing System

Currently available methods for calculating one's carbon footprint are based primarily on manual calculations and/or models that use universally used static emission factors. For example, the EPA Carbon Footprint Calculator requires the user to manually input several parameters pertaining to their household energy usage, transportation patterns, and waste management practices into the calculator, which will produce a set of generalised values about what an average American earns in gross emissions. While this does give greater awareness about the impact of lifestyle choices, it is still limited in that the values will not necessarily be representative of how each individual person's choices directly affect overall emissions levels. Another form of existing solutions for emissions estimation is the use of mobile apps, such as the Joro App. These mobile application platforms estimate emissions based on an analysis of the financial transaction records of their users. While automated collection of this type of data provides for improved efficiencies, there are serious concerns regarding the privacy and security risks associated with giving developers access to sensitive banking information of their users. Another form of existing solutions for emissions estimation is the use of mobile apps, such as the Joro App.

These mobile application platforms estimate emissions based on an analysis of the financial transaction records of their users. While automated collection of this type of data provides for improved efficiencies, there are serious concerns regarding the privacy and security risks associated with giving developers access to sensitive banking information of their users. In summary, these types of applications are focused primarily on providing an estimate of an individual's emissions; however, they do not help to provide any information about trends in how people choose to live, nor do they assist in forecasting the impact of future emissions based on how people make decisions – or not.

5. Proposed System

The Carbon Footprint Estimator and Reducer is a web application that utilizes machine learning techniques to provide users with a complete analysis of their carbon emissions resulting from transportation, electricity, food, and waste on a daily basis. The application collects user data To predict and forecast accurate emissions, the system performs data preprocessing for the collected data and applies multiple machine learning models, such as CatBoost and Linear Regression. In addition, the system incorporates an LLM enabled chatbot powered by Gemini, which provides users with recommendations on how they can adopt sustainable behaviours in their daily lives. Furthermore, the system predict their future associated emissions based on their current behaviours and provides users with actionable insights that will allow them to minimise their carbon footprint. Overall, the Carbon Footprint Estimator and Reducer is an effective, efficient, scalable, user-friendly application that assists users in making environmentally-friendly decisions.

6. System Architecture

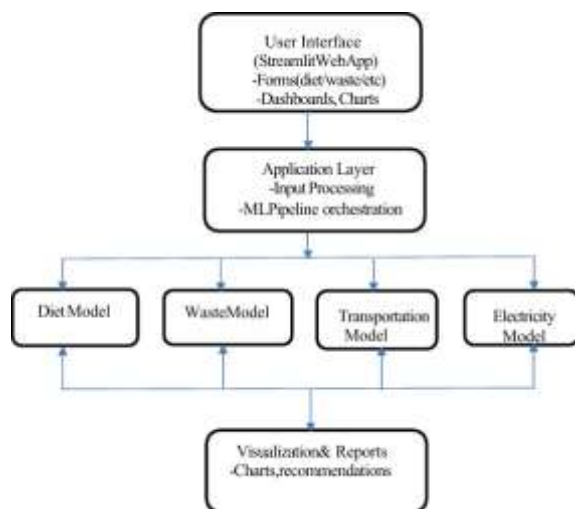


Fig 1 System Architecture

The design involves the use of a User Interface with Streamlit functionality for data entry and dashboarding, Application Level logic with the handling of ML tasks, a number of domain-specific models varying from diet to electricity, and Visualization Level logic to produce output and recommendations.

The Carbon Footprint Monitoring System aims to enhance the carbon emission measurement with smarter, more efficient, and faster processes by employing machine learning, predictive models, and data analysis capabilities. Consumers and businesses are now able to accurately determine their carbon footprint and the impact they have on the environment through the use of data-driven insights. The Carbon Footprint Monitoring System consists of a number of core

components that work together to allow users to analyze and measure emissions, keep track of their historical impact, and give users access to actionable inform. System construction on four layers and its responsibilities.

User Interaction Layer: A simple interface over the web built using Streamlit which enables users to submit their day's activity details like travel, electricity, food habits and waste

Processing Layer: The processing analog of the system, where ML models process user inputs to predict CO₂ emissions for a domain.

Data Storage Layer: The structured storage data warehouse that contains datasets, trained ML models and user emissions history to monitor progress.

Integration & Recommendation Layer: This layer connects the models to Gemini API (LLM) offering personalized recommendations and integrate chatbot to decrease emissions and be more sustainable.

7. Algorithms

CatBoost regressor was selected as the main prediction model because it can handle numerical and categorical columns naturally without larger preprocessing. It employs an automatic pre-processing of categorical input features like transport mode, diet type, waste category and appliance usage based on the boosting of decision trees to predict the target variables more accurately. Performance of the model was assessed by RMSE and R² score.

The linear regression method was used as a benchmark model for the estimate of carbon emissions based on user activity by establishing a linearity relationship between user activity variables and total CO₂ Emissions. Its simplicity and interpretability do not obscure the way in which individual factors influence emission levels, making it potentially applicable for comparison purposes. Performance was evaluated based on RMSE and R².

The Decision Tree Regressor was used to model the non-linear associations between user activities and carbon emissions. It exhibited fair performance—especially in transportation and waste-related predictions, but it was sensitive to data changes (such as noise) and had a propensity to overfit the model for smaller datasets.

The Random Forest (RF) Regressor was adopted to improve the stability and accuracy by combining numerous decision trees through ensemble learning. It was effective in mitigating the overfitting and demonstrated good power for electricity and waste emission.

The **XGBoost Regressor** was applied to obtain high prediction accuracy, especially Dataset for data in transportation and electricity. Its gradient boosting method naturally controls the bias–variance trade-off and can distribute missing values and outliers, although some attention is needed to not overfit tuning its hyper-parameters.

Linear SVR was also used for prediction of carbon emission on lower dimensional datasets including dietary. It was efficient in processing high-dimensional features but achieved scant applicability with larger or complex datasets. The model required higher computational efforts.

8. Datasets

a. Vehicle CO₂ Emissions Dataset

Records: 7,385, Features: Make, Model, Vehicle Class, Engine Size, Cylinders, Transmission, Fuel Type, Fuel Consumption (City/Hwy/Comb), CO₂ Emissions (g/km), Purpose: Assess carbon footprint from transportation.

b. Food Product Emissions Dataset

Records: 43 ,Features: Food Product, Land Use Change, Feed, Farm, Processing, Transport, Packaging, Retail, Total Emissions, Global Average GHG Emissions, Unit of GHG Emissions ,Purpose: Estimate dietary carbon footprint.

c. Electricity Generation Carbon Emission Dataset

Records: 5,000 ,Features: Fuel Type, Plant Capacity, Plant Age, Efficiency, Capacity & Load Factors, Weather & Maintenance Factors, Region, Grid Demand, Operating Temperature, Distance to Grid, Generated, CO₂ Emissions Electricity ,Purpose: Evaluate carbon footprint from electricity production. Average GHG Emissions, Unit of GHG Emissions ,Purpose: Estimate dietary carbon footprint.

d. Electricity Generation Carbon Emission Dataset

Records: 5,000 ,Features: Fuel Type, Plant Capacity, Plant Age, Efficiency, Capacity & Load Factors, Weather & Maintenance Factors, Region, Grid Demand, Operating Temperature, Distance to Grid, Generated, CO₂ Emissions Electricity ,Purpose: Evaluate carbon footprint from electricity production.

e. Carbon Emission Wastage Dataset

Records: 10,000 ,Features: Personal and Lifestyle factors(body type, sex, diet, shower frequency, transport), activity patterns (travel, grocery, waste production), consumption habits (clothing, screen time), household behavior (appliances, recycling, cooking), Emissions Carbon ,Purpose: Analyze carbon footprint from lifestyle and waste behavior.

9. Results

The Results section contains all of the outputs from each of the Carbon Footprint Estimator and Reducer Modules. In addition to presenting all predicted carbon emissions through transportation, diet, electricity and waste, there are also visual dashboards and analytics for the results for each Module.

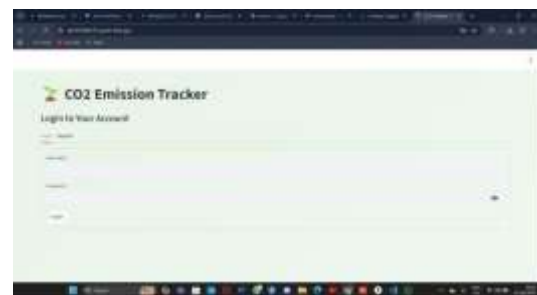


Fig 2 Home Page

The CO₂ Emission Tracker's home page, which includes a user login interface for safe system access, is depicted in **Fig. 2**.

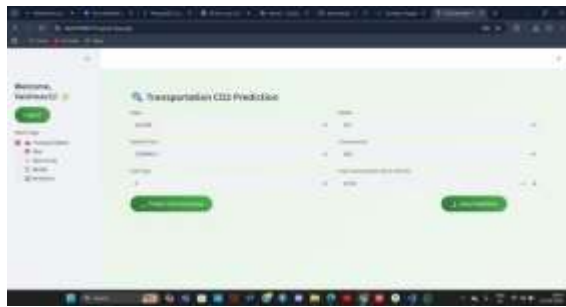


Fig 3 Transportation CO2 Prediction

The Transportation CO2 Prediction module, which calculates carbon emissions based on fuel type, travel mode, and distance, is depicted in **Fig. 3**.



Fig 4 Transportaion Emission Results

The Transportation Emission Results (**Fig. 4**) display the estimated daily CO2 emission value based on the user's transportation inputs.



Fig 5 Diet CO2 Prediction

Fig. 5 displays the Diet CO2 Prediction module, which calculates daily carbon emissions based on food consumption patterns and dietary choices.

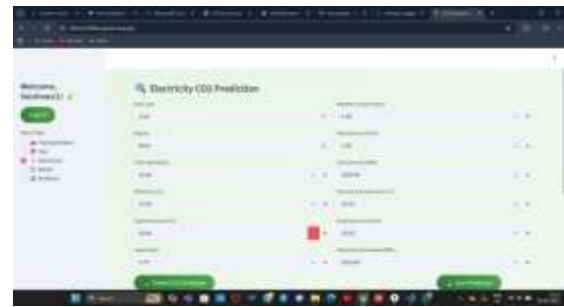


Fig 6 Electricity CO2 Prediction

Fig. 6 captures the Electricity CO2 Prediction module, responsible for calculating carbon emissions using electricity consumption and source information.

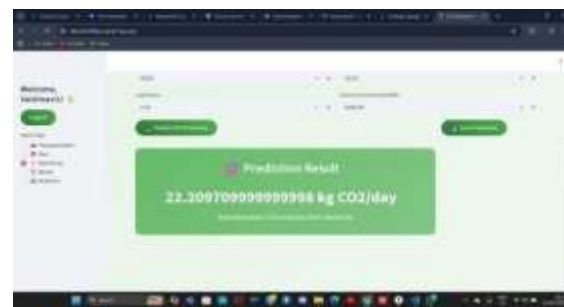


Fig 7 Electricity Emission Results

In the Electricity Emission Results shown in **Fig.7**, the everyday emission of CO2 is estimated based on the every house electricity use.



Fig 8 Waste CO2 Prediction

Fig. 8 illustrates the Waste CO2 Prediction module. This module predicts the carbon emissions produced from various kinds of wastes disposed through various methods.



Fig 9 Waste Emission Results

Fig. 9 above is showing Waste Emission Results with predicted CO₂ emissions daily produced from waste disposal practices conducted by the user.



Fig 10 Analytics Dashborad

In **Fig. 10**, there is an analytics dashboard with a list of key statistics such as total users and total CO₂ footprint. Also shown are results of predictions concerning user activities.



Fig 11 Analytics

Fig. 11 shows some interactive charts visualizing carbon emissions distribution across categories for supporting comparisons and insights.



Fig 12 Overall Analytics

Fig. 12 depicts an overall analytics that highlights top users based on different carbon footprint metrics. This enables the comparison of performance and assessment of impacts.



Fig. 13 depicts the AI-powered chatbot module, which provides personalized recommendations and guidance to users based on their carbon emission data.

10. Performance Measures

A range of performance metrics in each of the four domains, i.e., transportation, diet, waste, and electricity, were used to evaluate the precision and reliability of the carbon footprint prediction models.

- **Root Mean Square Error(RMSE):** The root mean squared error, or RMSE, is a measure of the average difference between the actual and predicted emission values. Higher prediction accuracy is indicated by a lower RMSE.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

- **R2 Score:** The coefficient of determination (R2 score) shows how much the independent variables account for the variation in the target variable. A better fit between the data and the model is indicated by a value closer to 1.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

1. ELECTRICITY DATASET:

Model	RMSE value	R2 score
CatBoost Regressor	459.01	0.79
XGB Regressor	483.00	0.77
Random Forest Regressor	505.02	0.75
Linear Regression	675.37	0.55
Decision Tree Regressor	729.72	0.48
Linear SVR	744.47	0.46

Table 1 Electricity Dataset

2. DIET DATASET:

Model	RMSE value	R2score
Linear Regression	2.61	0.99
Linear SVR	13.95	0.79
Decision Tree Regressor	22.06	0.47
XGB Regressor	22.21	0.46
Random Forest Regressor	22.25	0.45
Cat Boost Regressor	23.46	0.40

Table 2 Diet Dataset

3. TRANSPORTATION DATASET:

Model	RMSE value	R2 score
CatBoost Regressor	5.86	0.9898
Random Forest Regressor	6.69	0.9867
XGB Regressor	7.29	0.9842
Decision Tree Regressor	10.58	0.9668
Linear Regression	20.59	0.8743
Linear SVR	66.06	0.8743

Table 3 Transportation Dataset

4. WASTAGE DATASET:

Model	RMSE value	R2 score
CatBoost Regressor	551.61	0.9942
XGB Regressor	711.88	0.9903
Random Forest Regressor	718.26	0.9901
Decision Tree Regressor	1258.30	0.9698
Linear Regression	4350.06	0.6396
Linear SVR	7314.27	-0.0187

Table 4 Waste Dataset

11. Conclusion

The Carbon Footprint Estimator and Reducer is a useful, data driven option to address one of the most pressing challenges of our time addressing our environmental footprint through a reliable carbon footprint estimation. This is a convenient and flexible tool that allows the individual and institutions to easily track, record, and reduce greenhouse gas emissions, while saving time and effort. The tool could be useful for the individual as well as the institutional, collecting data from machine learning algorithms, interactive data data analytics, visualization, and providing estimates of carbon emissions from a variety of activities with precision. The dashboard-based interface of the platform increases usability through providing real-time insights, trends, and tracking of progress. In short, the project enhances environmental awareness and users to take

measurable action towards sustainability. The use of IoT devices and intelligent sensors for collecting real-time data will allow users to monitor their energy, water, and transportation consumption, which assists in establishing an accurate emissions profile of the area in which they live. The emissions data collected will allow an

expansion of the emissions dataset through the addition of regional and industry-related data, while simultaneously allowing them to use methods such as carbon credits or purchasing renewable energy to offset their emissions. Machine learning will provide the user with the ability to forecast future emissions and develop sustainability strategies. Through the use of gamification such as badges and achievements, users can be motivated to engage in environmentally friendly behaviour.

12. References

- Ahmed, S., Rahman, T., & Zhou, L. "A Review of Machine Learning and Statistical Approaches for Carbon Emission Prediction," *Environmental Data Analytics Journal*, 2022.
- Brown, R., & Li, Y. "Temporal Deep Learning Models for Carbon Emission Forecasting Using LSTM and Transformer Networks," *Journal of Smart Energy Systems*, 2024.
- Chen, Y., & Kumar, R. "Carbon Footprint Assessment Across Regional Energy Systems Using SVM and Decision Trees," *International Journal of Sustainable Energy Analytics*, 2021.
- Davis, M., Ortega, S., & Lin, P. "Synthetic Data Generation and Fairness Evaluation for Carbon Emission Forecasting," *Journal of Environmental Informatics Research*, 2022.
- Garcia, R., Lopez, D., & Martins, A. "Interpretable Machine Learning Models for National and Regional Carbon Emission Forecasting," *Sustainability Modeling and Policy Review*, 2021.
- Johnson, A., & Patel, R. "Behavior-Aware Machine Learning for Real-Time CarbonFootprint Estimation," *Journal of Smart Environmental Systems*, 2020.
- Lee, H., Park, J., & Silva, M. "ML-Based Prediction of Urban Energy Use and Transportation Emissions Using Random Forest and Gradient Boosting," *Urban Computing and Sustainability Journal*, 2020.
- Nguyen, T., Roy, S., & Abbas, K. "Transfer Learning for Carbon Emission Prediction in Data-Limited Regions," *International Journal of Green AI Systems*, 2023.
- Smith, J., Carter, B., & Wilson, M. "Machine Learning Models for Carbon Emission Estimation Across Lifestyle Sectors," *Journal of Climate Informatics*, 2019.
- Tanaka, R., & Roy, P. "Explainable AI Techniques for Carbon Emission Modeling Using SHAP and Rule-Based Systems," *Environmental AI Transparency Review*, 2023.