Product Carbon Footprint Evaluation Tool using Machine Learning - Automobile Industry

CH. Sandhya Rani[1], A. Ananya[2], M. Pranati Sai[3], K. Akshitha Reddy[4], G. Lakshmi Srujana[5]

[1]Assistant Professor, Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India

[2,3,4,5] Students, Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India

[1] ch4sandhya@gmail.com, [2]ananya.arramalla@gmail.com, [3]pranati.manthena@gmail.com,[4] akshithareddyk23@gmail.com,[5]sruj.gl@gmail.com

Abstract

The car industry has a large share in the global carbon emissions, caused by the sophisticated and energy-demanding production activities to produce vehicles. The objective of this project is to develop a Product Carbon Footprint Evaluation Tool designed to measure the carbon footprint and compare each step of the automobile lifecycle, including material sourcing, production, distribution, and end-life disposal. The tool has been prototyped in the form of a web application and uses machine learning algorithms such as SVM and Random Forest to estimate and analyze emissions at every stage, while allowing either user input in terms of production parameters, and utilized in scenarios which can be predefined. The specifics of this tool is that users can compare scenarios so that the environmental impact of the different materials or production process can be evaluated. For example, should a user want to compare the use of aluminium with the use of steel, or production process with high energy versus low energy, based on the information available they can make more informed choices towards long term sustainable outcomes. Ultimately, the tool provides a data-driven reference providing automobile manufacturers, policymakers, and researchers with an opportunity to make more sustainable choices when producing, and or regulating automobile production practices, thus enabling global ambitions to almost reduce the carbon footprint for industrial production.

Keywords:

Product Carbon Footprint, Automobile Production, Machine Learning Algorithms, Environmental Impact Assessment, Sustainable Manufacturing, Emission Estimation Tool, Industrial Carbon Reduction.

1. INTRODUCTION

In a world where sustainability is increasingly becoming a global priority, understanding the environmental attributes of products is critical to both consumers and companies. Growing industries and increasing public consciousness indicates a real demand for practical tools that other facilitate carbon emissions assessments to reduce impact throughout the product lifecycle.

Effective analysis of carbon footprints can enable responsible production and consumption while contributing to the global struggle against climate change. This project will build a tool that utilizes machine learning analysis to provide the carbon footprint of products based on information provided by users, such as materials used, production, or transportation processes. After providing this and other relevant inputs, the tool will then compute and provide reliable insights about the carbon footprint in relation to the inputs considered.

The tool will utilize advanced algorithms to drive adaptive reasoning, in ways that are more robust than simply estimating carbon footprints through traditional methods. The tool will also provide scenario comparisons and personalized recommendations, enabling the user to understand which materials or processes are greener alternatives. The actionable insights will provide an avenue for businesses to optimize supply chains, while individuals can entertain more eco-friendly options. Overall, this project will provide a step towards environmental sustainability by providing knowledge and tools to take effective actions to reduce carbon footprints.

2. LITERATURE SURVEY

This paper is centered on significant references, where the authors have employed innovative approaches to CO_2 emission analysis and sustainable design.

Silvio Lang and colleagues (2024) [1] applied Life Cycle Analysis (LCA) and machine learning capabilities in the MINDFUL framework to streamline the evaluation of CO_2 emissions and Product Carbon Footprint (PCF) calculations. The paper proposes requirements for a simplified editorial tool describing its methodology and the software architecture, with an objective of efficiency and usability.

Ashkan Safari and colleagues (2024) [2] offered a technique to create predictive modeling for emissions of gaseous pollutants (i.e., CO_2) in the automotive sector using machine learning. Their research ultimately seeks to improve accuracy of CO_2 emission predictions by examining data from 46 automotive brands, resulting in accurate predictions with substantial correlation coefficients between predicted and actual CO_2 emissions.

In the study by Kwaku Boakye, Kevin Fenton, and Steve Simske (2023) [3], they studied machine learning algorithms to predict CO_2 emissions based on a dataset of historic production variables from cement manufacturing, especially examining a case study at the Union Bridge Plant, Heidelberg Materials, Maryland. The authors recognize the limit of calculating CO_2 emissions using the IPCC method and utilize other approaches, including advanced machine learning and artificial intelligence, to better predict emissions during the calcination of cement manufacturing, using sensitivity analysis to determine significant manufacturing parameters.

A multi-layer perceptron neural network with backpropagation was developed by Wisthoff et al (2016) [4] to connect the LCA impacts of 37 case study products to attributes of those products. A search tree of sustainable design knowledge and a weighting system allow designers in the early stages of design to identify attributes that have the highest environmental impact, leading to a more sustainable product redesign.

According to Prashant Kumar Singh and Prabir Sarkar (2022) [5], they introduced a tool based on artificial neural networks (ANN) to assist designers in incorporating sustainability into product design. The tool optimizes the design parameters - which it assesses by looking at the carbon footprint, cost, efficiency, and user-friendliness - in the same optimization process rather than separately.

Dhyan R et al. (2024) [6] analyzed the carbon emissions of vehicles through machine learning and deep learning methods investigating data on fuel consumption and CO_2 emissions. They trained models using a dataset of vehicle characteristics. Both Random Forest and CNN have the highest predictive performance. The research aims to gain a better understanding of prediction and provide insights into improving automotive-related emissions.

Vishwanadham Mandala and colleagues (2024) [7] researched the impact of AI and ML on the reduction of carbon emissions in the automotive industry. The focus of the investigation concerns AI-enabled approaches to energy use optimization, electric vehicle charging management, emissions reduction, and smart traffic systems.

Gökalp Çınarer et al. (2024) [8] applied a machine learning approach to historical data (energy consumption, vehicle kilometers, population, and GDP) in order to predict CO_2 emissions in the transport sector of Türkiye. Some model comparisons demonstrated strong correlations, and energy consumption was identified as the best predictor of CO_2 emissions. In some cases, XGBoost performed better than the other models for CO_2 emissions predictions, which supports the promise of advanced modeling approaches for enhancing CO_2 emissions predictions.

In their 2023 paper, David Tena-Gago and colleagues [9] created a UWS-LSTM model that accurately predicts carbon dioxide (CO_2) emissions from hybrid vehicles while optimizing the model structure for real-time use on low-powered IoT devices. The model achieved 97.5% accuracy, while outperforming other traditional and advanced machine learning models by addressing the non-linear CO_2 fluctuations typically associated with hybrid vehicle powertrain switchovers in real-world applications.

An SVM model was developed by Chairul Saleh, Nur Rachman Dzakiyullah and Jonathan Bayu Nugroho (2024) [10] to forecast carbon (CO_2) emissions using energy consumption measures, like coal burning and electrical energy consumption. The model was developed and evaluated using data from an alcohol industry organization. The research used standard cross-validation and allocated 90% of the data for training and 10% for testing. The study selected the optimal parameters to use in the SVM model after trying several combinations, and determined the best parameters were C = 0.1 and Epsilon = 0. The results indicated a Root Mean Square Error (RMSE) of 0.004. The authors concluded that the SVM model's capability to provide factual forecasting information would benefit executive managers in tracking CO_2 emissions and making efficient business decisions.

3. METHODOLOGY

In this chapter, the method implemented to develop the Product Carbon Footprint Evaluation Tool using Machine Learning in the Automobile Sector is discussed. This process is described as multi- phase: covering data acquisition, pre- processing, model building, tool development, and analysis. Each phase will be examined to justify the accuracy, reliability, and user- friendliness of the system in place.

A. System Architecture

The system is implemented as a web application allowing users to input production parameters to acquire a real-time prediction of the carbon footprint. The architecture includes:

- User Interface (Frontend): A web-based interface through which users can enter data like materials consumed, manufacturing processes, energy usage, and transportation.
- Backend Processing: A machine learning engine that takes the input data, applies trained models, and produces carbon footprint estimates.
- Database: A repository for pre-defined industry data, emission factors, and user input.
- ML Model Integration: Integration of SVM (Support Vector Machine) and Random algorithms to accurately assess emissions.

B. Data Collection

The dataset used in this project includes industry-standard emission data from a variety of sources, such as:

- Life Cycle Assessment (LCA) databases that provide emission factors for materials and production processes.
- Publicly available datasets from environmental agencies and automobile manufacturers.
- Research case studies and studies on carbon footprint analysis in car manufacturing.

	Make	Model	Vehicle Class	Engine Size(L)	Cylinders	Transmission	Fuel Type	Fuel Consumption City (L/100 km)	Fuel Consumption Hwy (L/100 km)	Fuel Consumption Comb (L/100 km)	Fuel Consumption Comb (mpg)	CO2 Emissions(g/km)
0	ACURA	ILX	COMPACT	2.0	4	AS5	Z	9.9	6.7	8.5	33	196
1	ACURA	ILX	COMPACT	2.4	4	M6	Z	11.2	7.7	9.6	29	221
2	ACURA	ILX HYBRID	COMPACT	1.5	4	AV7	Z	6.0	5.8	5.9	48	136
3	ACURA	MDX 4WD	SUV - SMALL	3.5	6	AS6	Z	12.7	9.1	11.1	25	255
4	ACURA	RDX AWD	SUV - SMALL	3.5	6	AS6	Z	12.1	8.7	10.6	27	244

Table - 1. Table depicting the dataset used

C. Data Preprocessing

In order to ensure our forecast is precise and trustworthy, we go through data preparation to clean the data shown in Table.1 by the following steps performed in an order as show in Fig.1.

- Data Cleaning: We get rid of all of the incomplete and duplicate data so we can have the data cleansed and trustworthy.
- Feature Selection: We identify the relevant and most impactful data points, which include material type, the energy consumed in production, the distance from the production site to the service location, to help provide us the most relevance in our analysis.
- Normalization & Scaling: Standardizing numerical values to maintain consistency for training ML models.
- Handling Missing Data: Using interpolation methods or default values to address missing attributes.

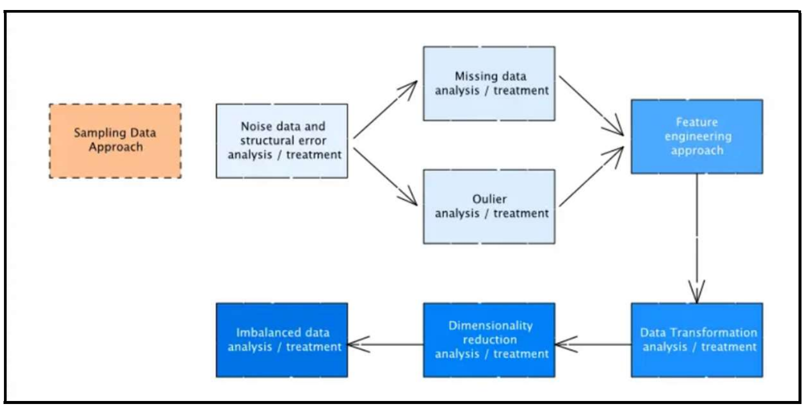

Figure-1. Flowchart depicting preprocessing steps

D. Machine Learning Model Selection

To calculate carbon emission at various stage of production, we use two separate models:

- Support Vector Machine (SVM): Applied for high-dimensional data with complex relationships, ensuring better generalization in footprint estimation.
- Random Forest Algorithm: Used for decision-tree-based analysis, allowing multi-factor comparisons and improved predictive accuracy.

These models learn from historical production data—and we continuously refine them to ensure the highest accuracy.

E. Web-based Tool Deployment

This web app is developed using the following technologies:

- Frontend: React.js / Angular.js for an interactive UI.
- Backend: Handle ML model requests and computation -> Use flask or Django for that.
- Database: MySQL/PostgreSQL to save emission factors and previous users queries.
- API Integration: So that external environmental databases can update emission data in real time.

F. Scenario Comparison Feature

A major functionality of the tool is Scenario Comparison, which enables users to assess the environmental impact of various manufacturing possibilities. The user can examine:

- Material-based comparisons (e.g., Aluminum vs. Steel).
- Manufacturing process variations (e.g., Renewable energy-based production vs. Conventional energy).
- Transportation impact (e.g., Air transport vs. Sea transport for raw material sourcing).

G. Model Evaluation & Performance Metrics

The models are evaluated across various metrics to ensure correctness and rigor:

- Mean Absolute Error (MAE) – Measures the average error in predictions. Our model had a Mean Absolute Error of 3.57.
- Root Mean Square Error (RMSE) – Evaluates how well the model predicts unseen data. RMSE was calculated to be 9.22 for our model, indicating Some predictions are more off, but overall errors are controlled.
- R^2 Score (Coefficient of Determination) – Determines the proportion of variance explained by the model. Our stacked model produced an R^2 Score of 0.98, meaning it explains 98% of the variance in CO_2 emissions, indicating high accuracy.
- Plot between Predicted vs. Actual values – Used to visually assess classification accuracy. The plot for our model is shown in Fig.2 below:

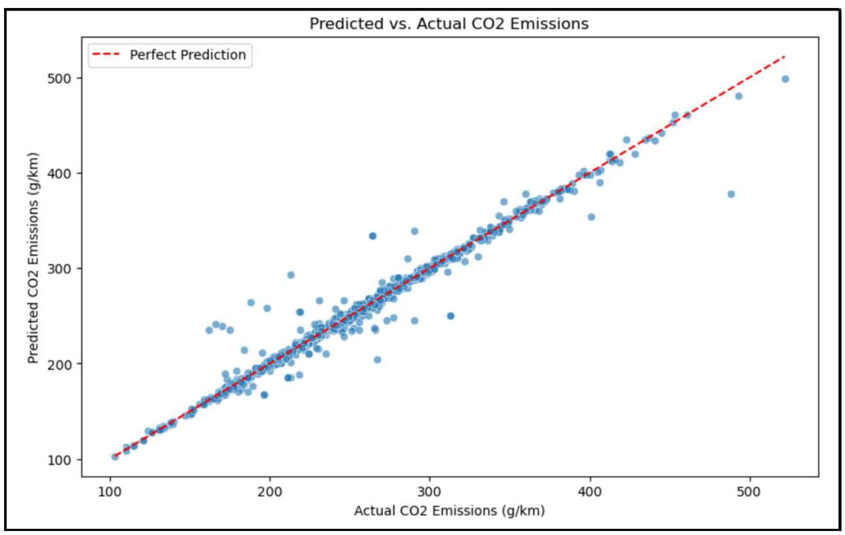

Figure-2. Predicted vs. Actual values plot

H. Deployment and Testing

The developed system has undergone thorough testing, which includes:

- Unit Testing: Ensuring each component functions as expected.
- Integration Testing: Verifying the interaction between frontend, backend, and ML model.
- User Testing: Collecting feedback from industry professionals and researchers to improve usability.

4. CONCLUSION

The Product Carbon Footprint Assessment Tool using Machine Learning is a sophisticated, data-driven solution to assess carbon emissions in the automotive industry. The – Machine Learning algorithms SVM and Random Forest provide accurate estimates of carbon footprints for automobiles at different stages of production: material procurement, manufacturing, transportation, and disposal. Users can use its scenario comparison function to assess the environmental consequences of employing various materials, sourcing energy from different providers, or engaging in alternative manufacturing methods - which will support producers and policymakers in making more sustainable decisions. The tool advances the accuracy of these footprint determinations from what is typically available based on LCA (life-cycle assessment) databases, real emissions data from industry practices, and machine learning models. Furthermore, as an online tool, the way this online platform is designed allows for a variety of users (i.e., automakers, researchers, and policy-makers) to access it. Additionally, this toolkit is related to any global sustainability goals as it promotes sustainable decision making in the automotive manufacturing industry.

5. FUTURE WORK

While the current implementation is informative, there are several areas where the tool can be further enhanced:

- Enhanced Model Performance: Enhance the accuracy of the models by integrating deep learning models (e.g. LSTMs and XGBoost), and use real-time IoT based data.
- Greater Applications: Expand the scope of the tool to other industries (e.g. electronics and construction) and build a mobile app for accessibility and convenience.
- Live Data Integration: Use APIs for a real time carbon footprint, and incorporate Blockchain technology for audited emissions and emissions tracking.
- Improved User Interaction: Offer personalized reports, interactive dashboards, and scenarios to offer a more engaging user experience.
- Alignment to Policy: Ensure that the tool is line with government policies and work with regulatory agencies to standardize carbon footprint assessments.

These improvements will transform the tool into a comprehensive decision-support system for sustainable manufacturing.

6. REFERENCES

[1] Stephanie Martinez, “Vehicles and Emissions”, M.S. thesis, California State University, San Bernardino, CA, USA, 2022.

[2] WorldAutoSteel, “Life Cycle Assessment: Good for the Planet, Good for the Auto Industry”, Apr. 2023.

[3] G. Asaithambi, M. Treiber, and V. Kanagaraj, “Life Cycle Assessment of Conventional and Electric Vehicles”, in International Climate Protection, M. Palocz-Andresen, D. Szalay, A. Gosztom, L. Sípos, and T. Taligás, Eds.Cham: Springer, 2019.

[4] Lindita Bushi, “EDAG Silverado Body Lightweighting Final LCA Report”, Aluminum Association, USA, Aug. 2018.

[5] Green NCAP, “LCA: How Sustainable is Your Car?”, Nov. 16, 2022. [Online]. Available: https://www.greenncap.com/press-releases/lca-how-sustainable-is-your-car/

[6] Silvio Lang, Bastian Engelmann, Andreas Schiffler, and Jan Schmitt, “A simplified machine learning product carbon footprint evaluation tool”, Cleaner Environmental Systems, Volume 13, 100187, June 2024.

[7] Dhyan R, Helen K Joy, Sridevi R, Electa Alice Jayarani A, Vanusha D, “Machine Learning and Deep Learning Analysis of Vehicle Carbon Footprint”, International Journal of Environmental Impacts, June 2024.

[8] Y. Jin, A. Sharifi, Z. Li, S. Chen, S. Zeng, and S. Zhao, “Carbon emission prediction models: A review”, Apr. 2024.