Carbon-Wise: Toward Intelligent Estimation of Personal Carbon Footprints

Karanam Praneethanjanee[1], Muthoju Harshini[2], Neerudi Sravani[3], Dr. D. Radhika[4]
[1,2,3]*B. E in CSE, Stanley College of Engineering & Technology for Women, Abids, India*
[4]*Associate Professor, CSE, Stanley College of Engineering & Technology for Women, Abids, India*

***Abstract*—Climate change mitigation requires effective tools that enable individuals to measure and understand their personal carbon footprint. While large industrial emissions are widely monitored, emissions generated through daily lifestyle activities remain less visible to individuals. This paper presents Carbon-Wise, an AI-powered carbon footprint tracking system designed to estimate and analyze individual carbon emissions based on lifestyle patterns. The proposed system integrates emission-factor based carbon estimation with machine learning regression models to predict monthly carbon emissions. A hybrid dataset combining survey responses and synthetic lifestyle data was constructed to improve model generalization. The dataset includes features related to electricity consumption, transportation habits, dietary patterns, household size, flight frequency, and renewable energy usage. Two regression models Linear Regression and Random Forest were trained and evaluated using standard metrics including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and coefficient of determination (R^2). Experimental results demonstrate that the proposed approach achieves reliable prediction performance while providing interpretable insights into the lifestyle factors contributing to carbon emissions. The Carbon-Wise platform further provides interactive visual analytics and sustainability recommendations to encourage environmentally responsible behavior. The system demonstrates how machine learning and data-driven insights can support individual-level climate awareness and sustainable lifestyle choices.**

***Index Terms*—Carbon footprint prediction, machine learning, sustainability analytics, emission estimation, regression modeling, environmental monitoring.**

I. INTRODUCTION

Advancements in artificial intelligence and data-driven analytics offer new opportunities to develop intelligent systems capable of estimating carbon emissions from lifestyle behaviors. By combining machine learning models with emission factor calculations, it is possible to provide more accurate and personalized emission estimates.

In this work, we present Carbon-Wise, an AI-powered carbon footprint tracking platform designed to estimate individual carbon emissions based on lifestyle data. The system integrates a machine learning prediction model with a web-based platform to analyze user inputs and generate personalized emission insights. A hybrid dataset constructed from survey responses and synthetic lifestyle data is used to train predictive models capable of estimating monthly carbon emissions with improved accuracy.

The proposed system aims to bridge the gap between environmental awareness and actionable insights by providing individuals with an accessible tool for monitoring and reducing their carbon footprint.

II. RELATED WORK

Climate change research has consistently emphasized the need to measure, monitor, and reduce carbon emissions at both global and individual levels. The Intergovernmental Panel on Climate Change (IPCC) highlights that limiting global temperature rise requires systematic tracking of greenhouse gas emissions across sectors. The IPCC's assessment reports stress the importance of data-driven mitigation strategies, reinforcing the need for technological systems that translate climate science into actionable solutions for individuals and communities.

Government organizations such as the Department for Environment, Food and Rural Affairs (DEFRA) have developed standardized greenhouse gas conversion factors used in carbon accounting and sustainability

reporting. These emission factors provide reliable estimates of carbon dioxide equivalents (CO_2e) associated with activities such as electricity consumption, transportation, and food production. Such standardized emission factors form the foundation for many existing carbon footprint calculators.

The concept of the carbon footprint was formally defined by Wiedmann and Minx as the total greenhouse gas emissions caused directly and indirectly by an individual, organization, event, or product. Their work established the theoretical basis for measuring environmental impact at multiple scales. Subsequent research by Pandey et al. examined various carbon footprint calculation approaches and emphasized the importance of activity-based emission estimation methods.

Recent studies have explored the application of machine learning techniques for environmental modeling and emission prediction. Regression models have been widely used to estimate carbon emissions from behavioral and consumption data. Machine learning approaches enable the discovery of relationships between lifestyle variables and emission patterns, improving prediction accuracy compared to static calculation models.

Despite these advancements, many existing carbon footprint applications focus primarily on displaying emission values rather than providing predictive insights or personalized recommendations. Additionally, several platforms lack integration between machine learning analytics and user-friendly interfaces that enable individuals to interpret their environmental impact effectively.

The Carbon-Wise system addresses these limitations by integrating emission factor calculations with machine learning-based prediction models and interactive visual analytics. By combining scientific carbon accounting principles with modern data-driven analytics, the proposed system provides a comprehensive platform for monitoring, analyzing, and reducing personal carbon emissions.

III. DATASET AND DATA PREPARATION

Accurate carbon emission prediction requires a dataset that represents diverse lifestyle behaviors. To improve model training and prediction robustness, a hybrid dataset was constructed by combining multiple data sources.

The initial dataset was collected through a structured survey conducted using an online questionnaire. The survey gathered information about electricity usage patterns, transportation habits, dietary preferences, household size, and travel behavior. Several responses were initially recorded as categorical ranges (for example, electricity consumption ranges). During preprocessing, these categorical values were converted into numerical midpoint values to enable machine learning training.

To further improve dataset diversity and model generalization, an additional synthetic dataset containing approximately 3000 records was generated. This dataset was created using realistic lifestyle ranges and emission equations based on standardized emission factors. Small Gaussian noise was introduced during dataset generation to avoid perfect correlations between features and emission outputs.

After preprocessing and feature alignment, the datasets were merged into a unified dataset containing more than 3000 records. This combined dataset was used for training and evaluating the machine learning models used in the Carbon-Wise system.

IV. PROPOSED METHODOLOGY

The Carbon-Wise system employs a hybrid approach that combines emission factor–based carbon estimation with machine learning regression techniques to predict individual carbon emissions based on lifestyle behavior. The overall methodology consists of several stages including data collection, feature engineering, emission estimation, dataset construction, and machine learning model training.

The first stage of the methodology involves collecting lifestyle-related data from users. The collected attributes include electricity consumption, transportation habits, dietary patterns, household characteristics, flight frequency, and renewable energy usage. These attributes represent key lifestyle activities that contribute to personal carbon emissions. Data was obtained from two primary sources: survey responses collected through an online questionnaire and synthetic datasets generated using realistic lifestyle ranges.

The second stage involves feature preprocessing and standardization. Survey responses containing

categorical ranges were converted into numerical midpoint values to ensure compatibility with machine learning algorithms. Missing values were handled using median-based imputation, and all datasets were standardized to contain a consistent set of features.

In the third stage, carbon emissions were estimated using standardized emission factors published by environmental agencies. These emission factors represent the amount of carbon dioxide equivalent produced per unit of activity.

Energy emissions were calculated using electricity consumption and electricity emission factors:

a) Energy = Electricity Consumption × Electricity Emission Factor × (1 − Renewable Energy Ratio) Transportation emissions are calculated based on travel distance and mode of transportation.
b) Transport = (Car Distance × Car Emission Factor) + (Bus Distance × Bus Emission Factor) + (Bike Distance × Bike Emission Factor) Food-related emissions are estimated using dietary patterns.
c) Food = (Vegetarian Meals × Veg Emission Factor) + (Non-Vegetarian Meals × Non-Veg Emission Factor) Air travel emissions are calculated using the annual number of flights taken by the user.
d) Flight Emissions = Flights per Year × Flight Emission Factor The total carbon footprint is calculated as the sum of emissions from all categories.
e) Total Emissions = Energy + Transport + Food + Flight

In the next stage, the calculated emission values were used as the target variable for machine learning model training. A hybrid dataset was constructed by combining the original synthetic dataset, survey responses, and an additional synthetic dataset generated using realistic lifestyle distributions. Small Gaussian noise was introduced during synthetic data generation to prevent perfect correlations between features and emission outputs.

The dataset was then divided into training and testing subsets using an 80:20 split. Two regression algorithms were evaluated: Linear Regression and Random Forest Regression. Linear Regression was used due to its interpretability and ability to illustrate how individual lifestyle variables influence carbon emissions. Random Forest Regression was used to capture nonlinear relationships between features and emission values.

The trained models were evaluated using standard regression metrics including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2 score). The model with the best prediction performance on the testing dataset was selected for integration into the Carbon-Wise system.

This methodology enables the system to estimate carbon emissions accurately while providing interpretable insights into the lifestyle behaviors that contribute most significantly to an individual's environmental impact.

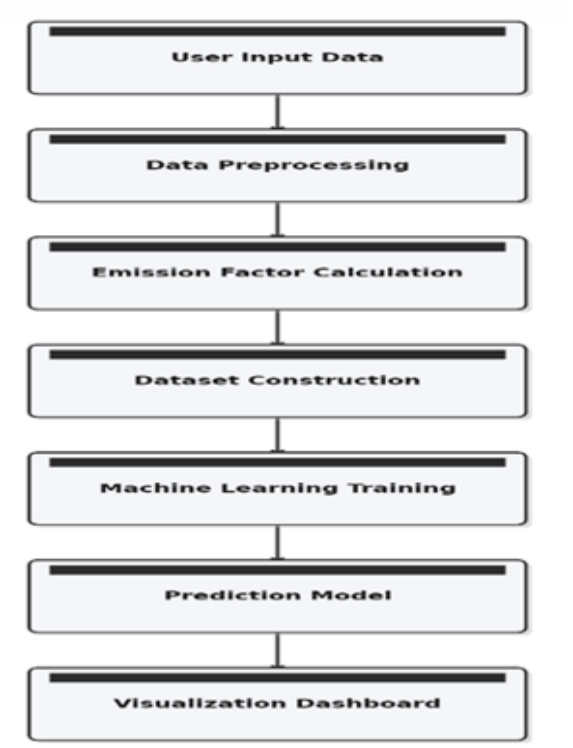

Fig. 1. Methodology Flow Diagram

V. SYSTEM ARCHITECTURE

The Carbon-Wise system follows a modular architecture designed to integrate data collection, emission estimation, machine learning prediction, and visualization into a unified platform. The architecture consists of four primary components: the frontend interface, backend processing layer, machine learning prediction module, and cloud database.

At the user interaction level, the system provides a web-based interface developed using React. This interface allows users to enter lifestyle information such as electricity consumption, transportation habits, dietary patterns, household size, flight frequency, and renewable energy usage. The interface is designed to guide users through structured input fields, ensuring that the required information for carbon emission estimation is collected efficiently.

The frontend communicates with the backend through RESTful API requests. The backend layer is implemented using the FastAPI framework, which handles data validation, request processing, and integration with the machine learning model. FastAPI was selected due to its high performance, asynchronous capabilities, and ability to efficiently handle API-based communication.

Once the user submits lifestyle data, the backend processes the inputs and performs emission factor calculations. These calculations estimate emissions generated from energy consumption, transportation activities, food habits, and air travel. The processed input features are then passed to the machine learning prediction module.

The machine learning module is responsible for predicting the user's carbon footprint using regression models trained on a hybrid dataset consisting of survey responses and synthetic lifestyle data. The trained model receives the processed feature inputs and generates an estimated carbon emission value.

All emission records and user data are stored in a cloud-based Firestore database. The database maintains historical emission records, enabling long-term tracking and trend analysis. This allows users to monitor changes in their carbon footprint over time and observe how lifestyle modifications influence emission levels.

The system also includes a visualization module that presents emission data through interactive dashboards. The dashboard displays category-wise emission distribution, historical trends, and prediction insights using graphical charts and analytics. These visualizations help users interpret complex environmental data in a simple and accessible manner.

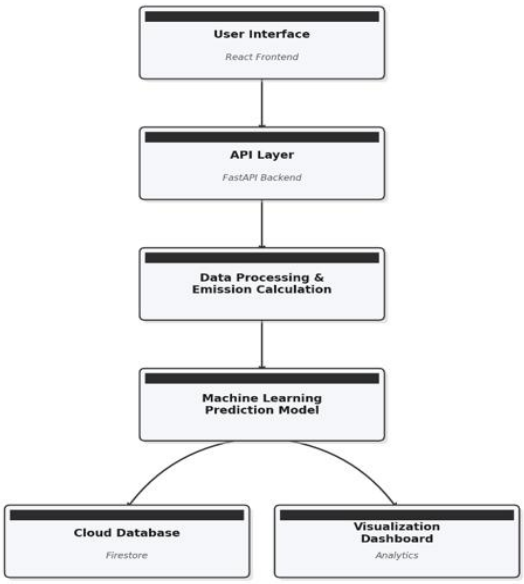

Fig. 2. System Flow Diagram

The overall data flow begins when the user enters lifestyle information through the frontend interface. The data is transmitted to the backend API, where emission calculations and prediction models are applied. The resulting emission estimates are stored in the database and displayed to the user through the visualization dashboard.

This architecture enables seamless communication between system components while ensuring scalability, modularity, and efficient data processing for carbon emission prediction and sustainability analysis.

VI. MODEL TRAINING AND IMPLEMENTATION

The carbon emission prediction model was trained using a hybrid dataset constructed from survey responses and synthetically generated lifestyle data. The combined dataset contained approximately 3000 records and included features related to electricity consumption, transportation distance, dietary habits, household size, flight frequency, and renewable energy usage.

The dataset was divided into training and testing subsets using an 80:20 split to ensure reliable model evaluation. Two regression algorithms were used for experimentation: Linear Regression and Random Forest Regression.

Linear Regression was selected for its interpretability and ability to illustrate how individual lifestyle variables influence carbon emissions. Random Forest Regression was also evaluated because of its ability to capture nonlinear relationships between input features and emission values.

Model performance was evaluated using three standard regression metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2 score). These metrics provide a comprehensive evaluation of prediction accuracy and model reliability.

The trained model was serialized and stored using Python's pickle library to allow integration with the Carbon-Wise backend system for real-time prediction.

VII. RESULT AND DISCUSSION

The performance of the proposed carbon emission prediction system was evaluated using the testing dataset obtained from the hybrid dataset containing survey responses and synthetically generated lifestyle data. Two regression models, namely Linear Regression and Random Forest Regression, were trained and compared to determine the most effective model for carbon emission prediction.

Fig. 3 illustrates the relationship between actual carbon emission values and the predicted values generated by the machine learning model. The scatter distribution indicates a strong correlation between predicted and actual emission values, demonstrating that the model is capable of accurately capturing the relationship between lifestyle behaviors and carbon emissions. This visualization is useful for understanding how closely the predicted values align with real emission values and for identifying potential prediction deviations.

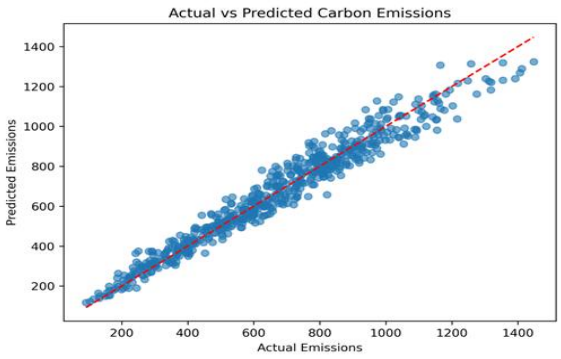

Fig. 3. Actual vs Predicted Plot

Fig. 4 presents the comparison between training and testing performance of the evaluated models using standard regression metrics. The graph shows that the testing performance closely follows the training performance, indicating that the model generalizes well to unseen data and does not suffer from significant overfitting.

Fig. 4. R^2 score comparison

Fig. 5 shows the feature importance analysis generated from the trained model. This plot highlights the contribution of different lifestyle variables toward carbon emission prediction. Electricity consumption, transportation distance, and flight frequency were identified as the most influential factors affecting an individual's carbon footprint.

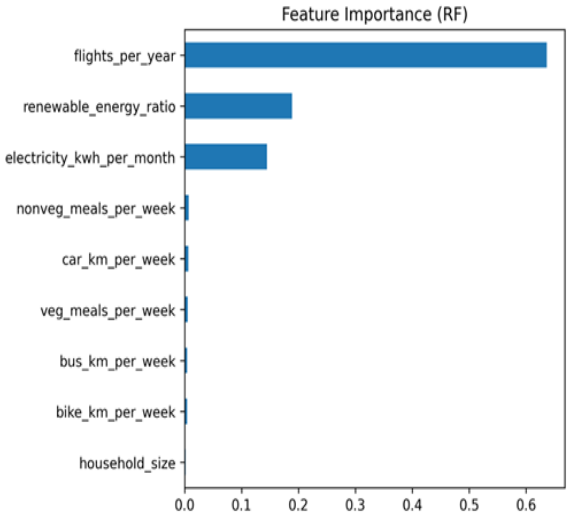

Fig. 5 Feature Importance Plot

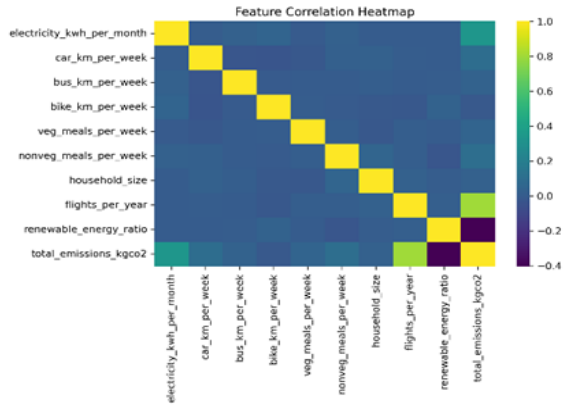

Fig. 6 presents the correlation heatmap of the dataset features.

The heatmap illustrates the relationships between input variables and the target emission value. Strong correlations between certain lifestyle factors and total emissions confirm the suitability of these variables for predictive modeling.

To quantitatively evaluate the performance of the models, three standard regression evaluation metrics were used: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2 score). MAE measures the average magnitude of prediction errors, RMSE penalizes larger prediction errors, and the R^2 score indicates how well the model explains the variance in the dataset.

Model	MAE	RMSE	R^2 Score
Linear Regression	53.89	69.13	0.936
Random Forest	44.20	57.09	0.957

Table summarizes the evaluation results obtained for both Linear Regression and Random Forest Regression models.

The Random Forest model achieved higher predictive performance compared to Linear Regression due to its ability to capture nonlinear relationships between lifestyle features and emission values. However, Linear Regression provides higher interpretability and allows easier understanding of how individual features influence emission predictions.

Among the evaluated models, the Random Forest regression model demonstrated the best performance with the highest R^2 score and the lowest error values. The model achieved an R^2 score of 0.957, indicating that approximately 95.7% of the variance in carbon emission values is explained by the model.

The improved performance of Random Forest can be attributed to its ability to capture nonlinear relationships between lifestyle features and emission values. Therefore, the Random Forest model was selected as the final prediction model for the Carbon-Wise system.

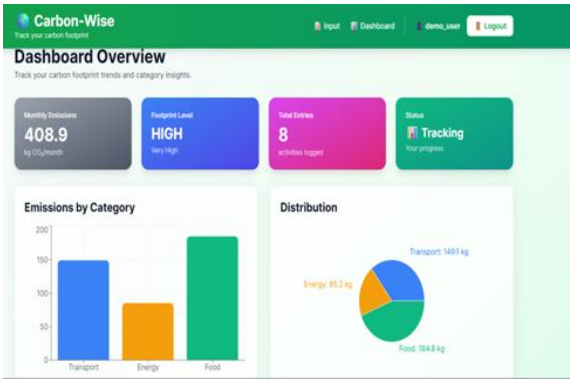

Fig. 7. Footprint Analytics Dashboard

VIII. CONCLUSION

This paper presented Carbon-Wise, an AI-powered carbon emission tracking platform designed to estimate and analyze individual carbon footprints using lifestyle data. By integrating emission factor calculations with machine learning regression models, the system provides users with accurate and interpretable insights into their environmental impact. Experimental evaluation demonstrates that machine learning techniques can effectively model the relationship between lifestyle behaviors and carbon emissions. The use of a hybrid dataset combining survey responses and synthetic data improved prediction robustness and model performance.

The Carbon-Wise system highlights the potential of data-driven sustainability platforms in promoting environmental awareness and encouraging individuals to adopt more sustainable lifestyle choices.

IX. FUTURE WORK

Future improvements to the Carbon-Wise platform include incorporating additional lifestyle variables such as waste generation, water consumption, and regional energy grid factors. Integration with real-time smart meter data could further improve emission estimation accuracy.

The system could also be extended with recommendation algorithms that provide personalized

suggestions for reducing carbon emissions. In addition, deploying the platform as a mobile application could improve accessibility and encourage broader adoption among individuals and institutions.

Appendix Appendix A
Survey Questionnaire

To collect lifestyle-related data for the Carbon-Wise system, a structured survey questionnaire was designed and distributed using Google Forms. The objective of the survey was to gather information about individual lifestyle behaviors that contribute to carbon emissions.

The questionnaire collected responses related to key factors such as monthly electricity consumption, transportation habits including car and bus travel distances, dietary patterns including vegetarian and non-vegetarian meal frequency, household size, and air travel frequency. These variables represent common activities that significantly influence personal carbon footprints.

The collected responses were used as part of the dataset for training and evaluating the machine learning models developed in this study. During the preprocessing stage, categorical responses obtained from the survey were converted into numerical midpoint values to ensure compatibility with regression-based machine learning algorithms.

The complete survey questionnaire used for data collection can be accessed through the following link: Survey Form

REFERENCES AND FOOTNOTES

A. References

[1] Department for Environment, Food & Rural Affairs (DEFRA), "UK Government greenhouse gas conversion factors for company reporting," 2023.

[2] Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2021: The Physical Science Basis*. Cambridge, U.K.: Cambridge Univ. Press, 2021.

[3] Food and Agriculture Organization (FAO), "Greenhouse gas emissions from the food system," FAO Statistical Report, 2020.

[4] T. Wiedmann and J. Minx, "A definition of carbon footprint," in *Ecological Economics Research Trends*, 2008.

[5] D. Pandey, M. Agrawal, and J. S. Pandey, "Carbon footprint: Current methods of estimation," *Environmental Monitoring and Assessment*, vol. 178, no. 4, 2011.

[6] Chakraborty and S. Roy, "A machine learning approach for carbon emission prediction," *Int. J. Environmental Science and Technology*, 2020.

[7] Y. Shan, J. Liu, and D. Guan, "CO_2 emissions from energy consumption: A review of methods and data sources," *GHG Measurement Journal*, 2019.

[8] Google DeepMind, "Gemini: A frontier multimodal foundation model," Technical Report, 2024.

[9] W. Gao *et al*., "Lifestyle carbon footprint estimation tools: A comparative review," *Journal of Cleaner Production*, 2022.

[10] Sarkar and R. Singh, "AI-based decision support systems for sustainable living," *Sustainable Computing: Informatics and Systems*, 2021.

B. Footnotes

[1] This work was carried out as part of the B.E. Computer Science Engineering program at Stanley College of Engineering and Technology for Women, Hyderabad, India.