"**EcoCarbon**": A Personalized Carbon Footprint Calculator and Your Guide to a Greener Lifestyle with ML

Bhavika Prasannakumar, Chinmaya Gayathri, Sai Srivathsav Aripirala, Shresta Kommera, Yamini Muthyala

Abstract—This paper presents "Eco-Carbon", a novel approach to personal carbon footprint calculation and reduction guidance, leveraging machine learning (ML) techniques.The main objective of the project is to create an integrated system that combines a personalized recommendation system with estimates of carbon emissions from electricity use and transportation. Information about personal travel habits and household electrical consumption has been gathered. For transportation emission computations, sophisticated machine learning models such as random forests and decision trees are being created, and the recommendation system is supported by collaborative filtering algorithms. According to preliminary findings, users are very interested in learning about and minimizing their carbon footprints, and machine learning methods appear to be promising in terms of precise emission estimation. Ensuring data quality and striking a balance between user data anonymity and personalized recommendations are challenges.

Index Terms—Personalized Recommendation System Sustainable Transportation Energy Consumption Machine Learning Models Data Encoding .

1 Introduction

In the face of mounting concerns about climate change, individual actions in reducing carbon emissions have become increasingly significant. The EcoCarbon project emerges in this context as an innovative approach to empower individuals in this global fight. Utilizing advanced machine learning techniques, EcoCarbon focuses on calculating personal carbon footprints, specifically from transportation and electricity consumption. This project not only addresses a technological challenge but also aims to encourage a shift towards a more sustainable lifestyle. By providing individuals with accurate, personalized information about their carbon emissions, EcoCarbon bridges the gap between awareness and action in environmental stewardship.

1.1 Objective

The objectives of EcoCarbon are multi-faceted and aimed at maximizing its impact on individual carbon footprint management. Firstly, the project seeks to develop accurate machine learning models for estimating carbon emissions from personal transportation and household electricity usage. This involves intricate data analysis and model refinement to ensure reliability and precision. Secondly, EcoCarbon aims to provide personalized recommendations to users, guiding them towards effective ways to reduce their carbon footprint. This aspect of the project is crucial in transforming data-driven insights into actionable, practical steps. Another key objective is to enhance public understanding and engagement regarding personal carbon emissions. Through an interactive and user-friendly platform, EcoCarbon intends to educate users about the environmental implications of their daily activities, fostering a culture of sustainability. Furthermore, upholding data privacy and security is a paramount objective, ensuring that users' sensitive information is protected rigorously. Lastly, the project aspires to evaluate and amplify its community impact, exploring opportunities for scaling and wider adoption, thereby contributing substantially to the global effort in carbon reduction.

2 Theoretical bases and Literature Review

2.1 Definition of the Problem

The contemporary challenge that Eco-Carbon aims to address is the accelerating environmental effect on individual lives, particularly the carbon footprints that are often overlooked as a consequence of daily activities. The need to close a substantial gap in environmental action and awareness is what motivates our effort. There is now a lack of simple and efficient tools for many people to measure and understand the environmental impact of their travel and electricity use. Eco-Carbon is intended to fill this need by providing a complex yet simple application that precisely detects carbon emissions connected with transportation and electricity consumption. The main goal of the initiative is to actively assist users in making more sustainable decisions rather than just drawing attention to the environmental issue. Eco-Carbon aims to democratize access to environmental responsibility by providing individuals with precise insights into their carbon footprint and practical solutions through the thoughtful integration of cutting-edge machine learning models and a personalized recommendation system.

2.2 Background

Eco-Carbon's methodology employs two major models: regression models and collaborative filtering. Regression models, such as Random Forests, Decision Trees, and Linear Regression, are reliable instruments for calculating transportation, electricity related carbon emissions. These models provide comprehensive information by detecting complex correlations between emissions and travel factors. The personalized recommendation system in Eco-Carbon is built on the popular recommendation system technique known as collaborative filtering. This algorithm finds similarities between users and suggests tactics based on what works for similar people. When combined, these models allow Eco-Carbon to deliver accurate estimations of carbon footprints and customized suggestions based on the unique characteristics of each user, promoting a data-driven and user-focused approach to environmental impact assessment and reduction.

2.3 Literature Review

The research [1] that Mardani et al. (2020) presented provides a thorough approach for estimating CO2 emissions based on energy consumption and economic growth. The authors reduced dimensionality and predict missing values by using Singular Value Decomposition (SVD). They employ Self-Organizing Map (SOM) clustering to effectively group data. This strategy outperforms previous methods in terms of predicting CO2 emissions. The significance of the methodology depends on its capacity to manage missing data, identify non-linear relationships, and assists in the formulation of policies concerning economic and energy development.

This study [2] analyzes the energy demand and CO2 emissions forecast for Turkey's transportation sector. To predict future trends, the research uses mathematical models and machine learning techniques. The techniques and models are evaluated using a range of metrics, such as RMSE, RRMSE, MABE, and MAPE. The Support Vector Machine (SVM) algorithm conducts an outstanding job at forecasting CO2 emissions, according to the results. The study emphasizes how crucial it is to use a variety of criteria when evaluating predicted accuracy. Furthermore, suggestions are provided for minimizing the effects on the environment, with a focus on sustainable methods including increasing engine efficiency, switching to electric cars, and establishing laws in place to limit the consumption of fossil fuels. The study's conclusion emphasizes the need for international cooperation to address the issues raised by rising energy consumption and CO2 emissions from the transportation sector.

This study [3] focuses on techniques used in evaluating CO2 emissions from transportation, focusing on several approaches such as fuel usage, distance driven, and real-time air-quality instruments. Air-quality tools emerge to be especially helpful because of their real-time data and additionally greenhouse gas insights when these approaches are compared against important metrics. The review additionally addresses a wide range of models that have been employed in earlier research, such as GIS applications, mathematical frameworks, and decomposition models. The global viewpoint includes 15 countries' sustainable development policies that are aligned with the Sustainable Development Goals (SDGs).Malaysia's dedication to sustainable transportation is highlighted by efforts such as the Low-Carbon Mobility Blueprint and Action Plan (LCMB). The paper illustrates the lessons learned from effective global methods and offers suggestions for Malaysia's future initiatives, including the revision of transportation legislation and the introduction of mandated CO2 tests. This evaluation emphasizes the significance of sustainable transportation plays in achieving larger environmental and social goals.

The research paper [4] provides a thorough investigation of sentiment analysis within the context of machine learning. It emphasizes applications and performance measures while carefully examining a broad range of approaches and frameworks. The difficulties in managing varied datasets, reducing biases, and resolving linguistic variations in context are important areas of focus. Pre-trained embedding, transfer learning, and attention mechanisms are examined in terms of their contributions to improved model accuracy. The study incorporates data from multiple sources and delves into upcoming trends, namely in multi-modal sentiment analysis. The findings of this survey provide a solid foundation for initiatives involving sentiment analysis, providing significant considerations for methodological options as well as understanding of current issues in the field.

In the paper [5] on household energy management and energy informatics, Fiorini and Aiello's 2018 study offers insightful information. The challenging task of arranging household appliances optimally to reduce carbon emissions is taken on by the authors. Their method comprises defining the problem as a mixed-integer linear programming (MILP) model and incorporating binary variables for the operation state of appliances across time. The study focuses on several kinds of setups, from using only electricity to a hybrid mode that uses hot water and natural gas. This study corresponds to contemporary concerns in sustainable and efficient energy usage, as well as the broader context of smart homes and environmental responsibility. By highlighting the possibility of load-shifting solutions to lower overall carbon emissions, the authors highlight the significance of taking into account not only the energy carrier but also the schedule of appliance consumption. Such optimization models are relevant for projects concentrating on the interconnection of energy management, carbon footprint reduction, and user preferences, serving as a platform for developing practical solutions in the expanding subject of energy informatics.

2.4 Solution to the problem

2.4.1 Machine Learning Models for Emission Estimation

Utilizing Random Forest Regressor and Decision Tree Regressor, we aim to accurately calculate carbon emissions from individual transportation and electricity usage. These models are designed to process complex datasets, providing precise emission assessments.

2.4.2 Personalized Recommendation System

The crux of EcoCarbon's user engagement strategy is a recommendation system powered by collaborative filtering algorithms. This system will provide tailored suggestions for reducing carbon emissions, based on individual patterns and preferences.

2.5 Why our solution is better

EcoCarbon's solution offers a superior approach in managing carbon emissions due to its integration of advanced machine learning models with a user-friendly interface. This combination ensures precise emission estimations tailored to individual patterns, enhancing the accuracy over generic methods. The personalized recommendation system, powered by collaborative filtering, significantly increases user engagement and effectiveness in emission reduction. Furthermore, our stringent commitment to data privacy and security ensures user trust and reliability, essential for widespread adoption. These attributes collectively position EcoCarbon as an innovative and effective tool in the realm of personal carbon footprint management, contributing significantly to global efforts in combating climate change.

Subsubsection text here.

3 PROPOSED METHODOLOGY

3.1 Data Collection

The data gathering process for this research involved handling of a meticulously crafted structured questionnaire that collected information about participants' transportation habits and electricity consumption patterns. The survey addressed essential features, including mode of transport, distance traveled, fuel type, monthly electricity consumption, and details about specific appliance usage. To ensure diversity, participants were recruited through social media networks using purposive sampling. The survey was effectively circulated in digital format and used a combination of closed-ended and open-ended questions for both qualitative and quantitative data. In order to ensure data validity and reliability, the main data collection tool was a carefully pre-tested structured survey questionnaire that covered all relevant aspects of transportation and electricity use behaviors. Participants were invited through social media posts, assuring accessibility, and the poll was safely held online, prioritizing participant confidentiality and privacy.

3.1.1 Dataset Description

The survey data led to the compilation of two different data sets, which are the Transport and Electricity csv files. The Transport data set is primarily concerned with transportation-related statistics, such as the kind of fuel used, the distance traveled, and the mode of transport. Inversely, the electricity data set explores patterns of power usage by gathering information on monthly consumption as well as details about the appliances that members use. These data sets are the study's foundation, offering a comprehensive and varied basis for ensuing studies.

3.1.2 Dataset Characteristics

3.1.2.1 The *Transport* dataset, which includes survey responses, includes a range of features. These features include 'Mode of Transport,' which details the various modes of transportation used by participants, 'Distance Traveled,' which provides information on the extent of travel, and 'Fuel Type,' which provides information on the environmental impact of transportation choices. In order to get a comprehensive insight of the traveling patterns of participants, other demographic data is also included.:

3.1.2.2 The *Electricity* dataset, which is from the survey responses, is distinguished by its emphasis on the consumption of electricity. The 'Monthly Electricity Consumption' data that it contains makes it possible to analyze consumption trends over time. In addition, the dataset covers Appliance Usage Patterns, providing significant information about the appliances that participants use. Together, these dataset components provide a comprehensive view of participants' electrical consumption behaviors, which contributes significantly to the overall research objective.:

3.2 Exploratory Data Analysis

The paper has resulted in various graphs to explore the data clearly.

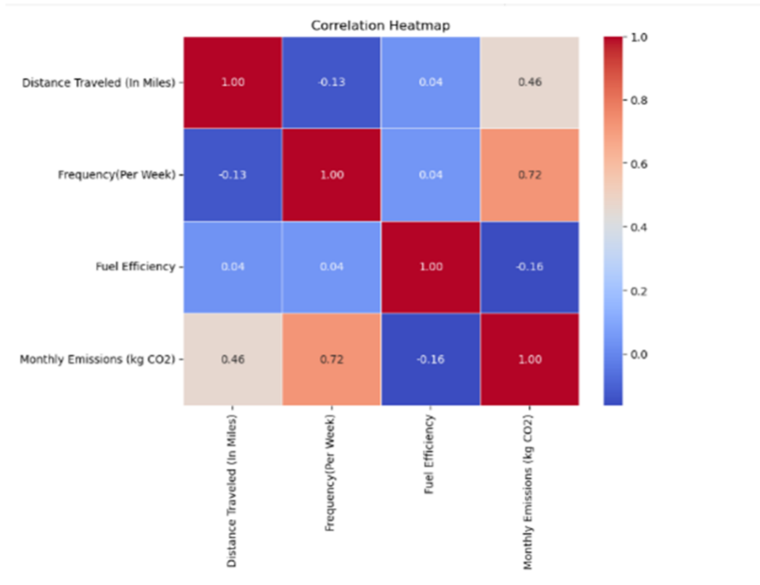

Fig. 1. Correlation HeatMap

Figure 1 shows the correlations between various factors, including monthly emissions, fuel efficiency, frequency of travel, and distance traveled, are depicted in this correlation heatmap. There is a strong connection (0.72) between travel frequency and monthly emissions, suggesting that more frequent travel is probably associated with greater emissions. On the other hand, the relationship between fuel efficiency and distance traveled is minimal, indicating that increases in fuel efficiency may not always be correlated with increases in travel distance.

Figure 2 analyzes the emission factors of two fuel types gasoline and electric for vehicles and motorcycles. When driving an electric car, the emission factor decreases significantly compared to a gasoline-powered vehicle, suggesting that using electric fuel has a less negative environmental impact. The switch to electric vehicles leads in a significant decrease in emissions for both cars and bikes, with bikes consistently having lower emission factors than automobiles. This emphasizes the advantages that electric vehicles and cycling have over conventional gasoline-powered cars in terms of the environment.

Figure 3 depicts the distribution of fuel types among automobiles is depicted in the pie chart, which shows a

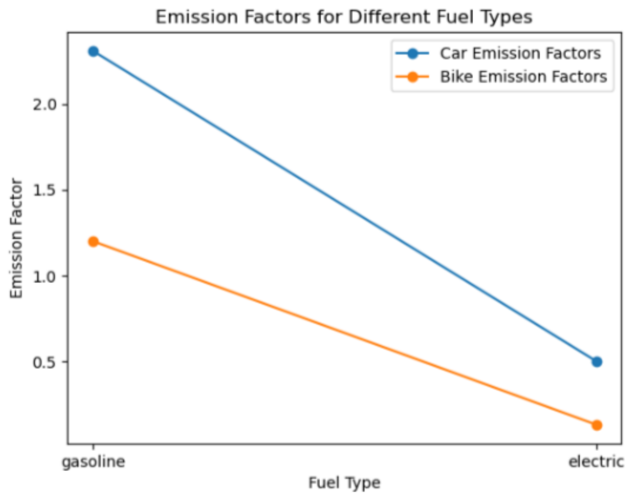

Fig. 2. Line Plot

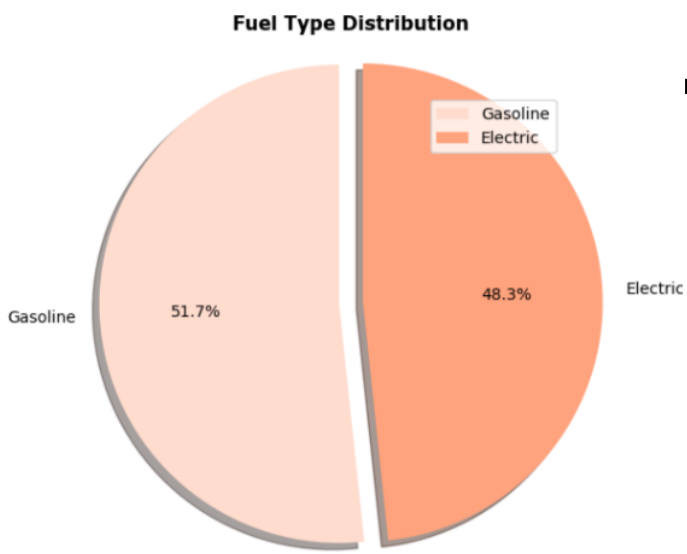

Fig. 3. Pie Chart

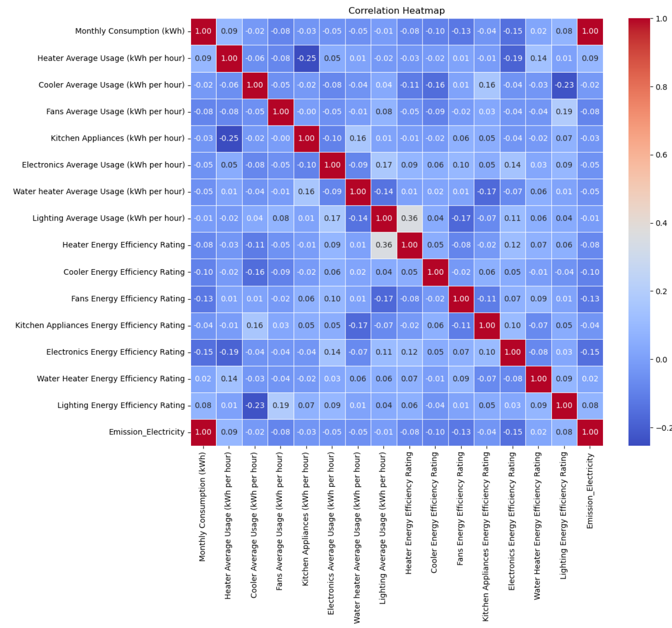

Fig. 4. Correlation HeatMap

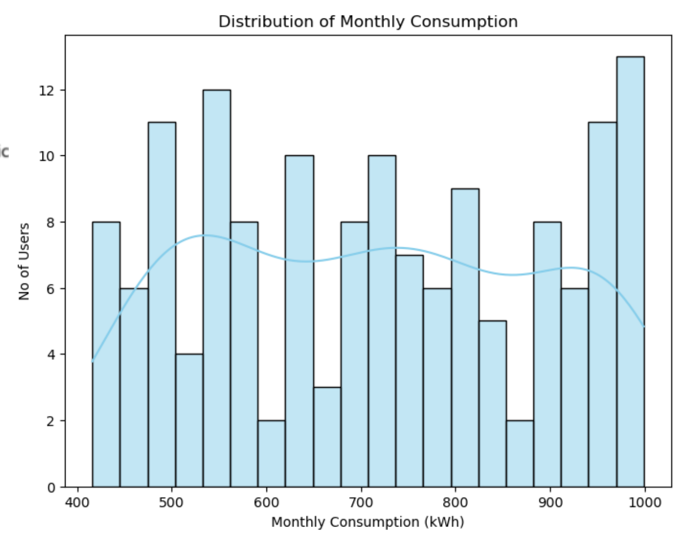

Fig. 5. Histogram

nearly equal split of 51.7% gasoline and 48.3% electric. With the trend of fuel usage moving away from traditional gasoline, this near balance indicates an important trend towards electric vehicles. The information shows consumer decisions that may have a significant impact on energy consumption trends and the environment.

Figure 4 demonstrates a robust relationship between emissions and energy usage, highlighting the necessity of energy-saving strategies. In order to attain a lower carbon footprint, it highlights the significance of behavioral change in addition to efficiency improvements as enhanced appliance efficiency has a limited influence on reducing overall usage.

Figure 5 illustrates the diverse energy usage patterns among households, with the distribution of monthly consumption revealing two prominent modes around 600 kWh and 1000 kWh. This bimodality suggests the existence of distinct user groups, potentially differentiated by factors such as household size, lifestyle, or appliance efficiency. Such insights are crucial for tailoring energy conservation strategies to diverse consumer behaviors, enabling more effective targeting of energy efficiency programs and policies. Figure 6 demonstrates the energy efficiency ratings of various household equipment, such as lighting and heaters, are shown by the horizontal bar chart. Heaters receive the highest rating, suggesting the best energy efficiency, out of other appliance categories, which are presumably evaluated up to a maximum of 5. All appliances scored above the median level in the evaluations, indicating a trend towards energy-conscious consumption. This is consistent with a larger drive for energy efficiency in household items to lower overall energy consumption and associated emis-

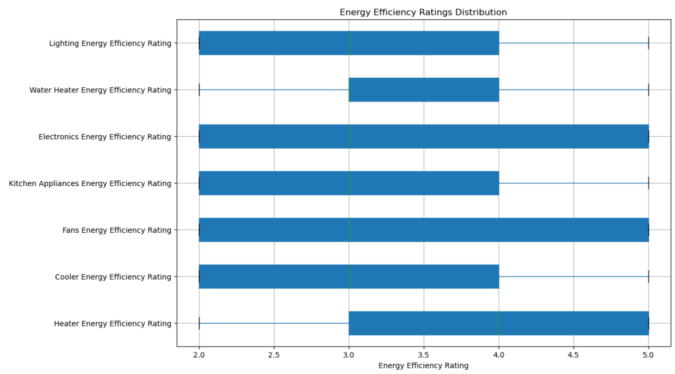

Fig. 6. BoxPlot

sions.

3.3 Data Cleaning and Pre-processing

3.3.1 Data Encoding

Categorical variables such as 'Mode of Transport' and 'Fuel Type' were discovered in the transportation dataset. These variables, which are essential in calculating carbon emissions, needed to be encoded before they could be used in machine learning models. Label encoding was used to convert each category into a unique integer value. This transformation turned qualitative data into a format that computers could easily interpret, assuring precise capture of the intricacies of various transportation modes and fuel kinds.

3.3.2 Feature Selection

During the pre-processing phase, factors that directly influence carbon emissions were meticulously chosen. The factors 'Distance Traveled,' 'Fuel Type,' and 'Fuel Efficiency' were highlighted. Based on an in-depth investigation of their influence on carbon emissions, this decision ensured that the model was trained on the most relevant and meaningful data, with the goal of improving the precision of emission predictions.

3.3.3 Carbon Intensity Conversion

Converting carbon intensity from gCO2e/MJ to gCO2/kWh was an essential step in preparing the dataset for the electricity use study. This conversion matched the data to the analysis's unique criteria, ensuring that emission estimations were correct and contextually relevant to power consumption.

3.3.4 Data Conversion

The selected characteristics were standardized using StandardScaler to ensure consistency and improve the prediction capabilities of machine learning models. This was an important step in the pre-processing phase since it uniformly scaled all characteristics, preventing any single variable from significantly impacting the model's predictions and allowing for more accurate and reliable predictions.

4 Problem Solution

4.1 Modeling

4.1.1 Linear Regression

Linear regression establishes a correlation between independent factors and a dependent variable by fitting a linear equation to observed data. The coefficients of the equation are calculated by minimizing the sum of the squared discrepancies between the observed and anticipated values. This approach is simple and effective for datasets with linear relationships between variables.

4.1.2 Decision Tree Regressor

A Decision Tree Regressor models decisions and their potential outcomes by building a tree structure. It facilitates the analysis of intricate decision-making processes by dividing the dataset into branches according to feature values. This algorithm offers a clear depiction of the decision paths and is especially helpful for addressing non-linear interactions.

4.1.3 Random Tree Regressor

The Random Forest Regressor combines numerous decision trees to produce a more accurate and consistent prediction. By averaging the outcomes of multiple trees, each trained on a different subset of the data, it reduces overfitting. The model performs better with this ensemble approach, particularly when dealing with huge datasets and high dimensionality.

4.1.4 Gradient Boosting Regressor

A stage-by-stage predictive model is built by Gradient Boosting Regressor. It builds trees one at a time, with each new tree aiding in the correction of mistakes made by trees that have already been trained. The model's versatility and ability to optimize for a range of loss functions and data distributions are its main strengths.

4.2 Evaluation Metrics

4.2.1 Mean Squared Error

MSE calculates the average squared variation between the values that were expected and those that were observed. Larger errors are penalized more severely, giving an all-encompassing indicator of the correctness of the model. Better performance is indicated by lower MSE values, which show a stronger fit between actual and expected emissions.

4.2.2 R-squared

R-squared calculates the percentage of the variance in the independent variables (features) that can be attributed to the dependent variable (carbon emissions). It has a 0–1 range, where 1 represents an ideal match. Greater predictive power is indicated by a higher R2, which indicates that a greater percentage of the variability in emissions is captured by the model.

4.2.3 Mean Absolute Error

The average absolute difference between the values that were predicted and those that were observed is determined by MAE. It provides a clear measure of prediction accuracy without taking error direction into account. Lower MAE values suggest greater accuracy because they represent fewer absolute differences.

4.2.4 *Explained Variance Score*

This score measures how much of the variance in the dependent variable can be explained by the model. It has a 0–1 range, where 1 represents an ideal forecast. A model that provides an accurate representation of the data and successfully captures the variability in carbon emissions is indicated by a higher explained variance score.

Model	MSE	R-squared	MAE	Variance
Linear Regression	1037.321	0.4842	24.63	0.4842
Decision Tree	491.91	0.7554	14.34	0.7555
Gradient Boosting	833.42	0.5856	22.147	0.5856
Random Forest	535.36	0.7338	16.729	0.7338

TABLE 1
Evaluation Metrics for Transportation Models

Table 2 shows the Evaluation Metrics for the Models used for the Electricity sector.

Model	MSE	MAE	R^2	Variance
Linear Regression	7.52×10^{-6}	0.00226	0.9999	0.9999
Random Forest	1.60	0.8616	0.9992	0.9992
Decision Tree	3.21	1.2153	0.9984	0.9985

TABLE 2
Evalution metrics for Electricity models

5 RESULTS

The research investigation into carbon emissions from transportation and electricity usage, which used machine learning models, gave informative results, demonstrating the varied capabilities of various algorithms. With an R-squared Score of 0.7554, the Decision Tree Regressor was particularly notable in the field of transportation. The reliability in precisely simulating transportation emissions was confirmed by the comparatively low Mean Squared Error (MSE) of 491.19 and the Mean Absolute Error (MAE) of 14.34, which added to the good score. Because of its capacity to manage complicated data dynamics, it was particularly proficient at analyzing the many variables affecting emissions in this industry.

The Random Forest Regressor performed quite well on the electricity front, as evidenced by its R-squared Score of 0.9992, MSE of 1.5988, and MAE of 0.8616. These measures demonstrate the model's accuracy and consistency in calculating emissions from energy use. This high degree of accuracy was made possible by the Random Forest Regressor's ensemble technique, which makes use of many decision trees and successfully addresses overfitting problems. The overall results of the research not only point out the distinct benefits of various machine learning models, but also emphasize how important it is for us to better understand the way environmental factors affect our environment in order to develop sustainable and more educated energy management practices.

Model	Accuracy
Linear Regression	0.4842
Decision Tree	0.7554
Gradient Boosting	0.5856
Random Forest	0.7338

TABLE 3
Accuracy Metrics for Transportation Models

Model	Accuracy
Linear Regression	0.999999997
Decision Tree	0.998445736
Random Forest	0.999224814

TABLE 4
Accuracy Metrics for Electricity Models

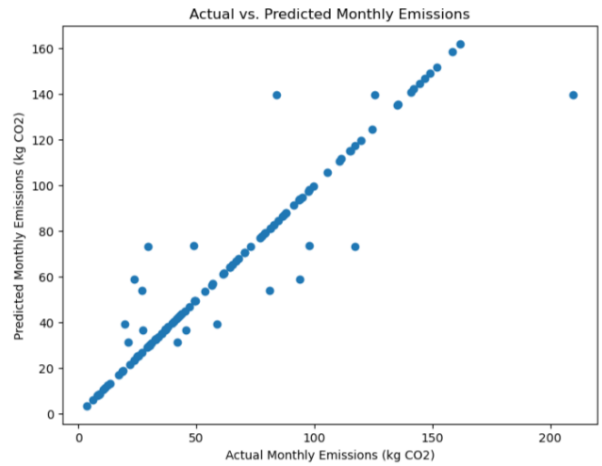

Fig. 7. Actual vs Predicted Monthly Emissions

Figure 7 shows a good correlation between actual and predicted CO2 emissions, validating the correctness of the regression model. The dense clustering of points along the diagonal indicates the model's consistent prediction accuracy across the entire range of the dataset, despite a little deviation at higher emission values.

5.1 Model Comparison

In emissions modeling, various approaches have different levels of effectiveness depending on the application. In the transportation sector, the Decision Tree Regressor stands out as particularly effective with an R-squared Score of 0.7554. This model's accuracy in estimating carbon emissions, a critical component in environmental impact studies, is highlighted by its ability to precisely analyze and comprehend the numerous patterns found in transportation data. In the field of electrical consumption, the Random Forest Regressor dominates, as demonstrated by an R-squared Score of 0.9992. By combining information from several decision trees, the model's ensemble approach improves its resistance to overfitting while preserving a high degree of accuracy. This feature is very helpful for managing the complex and variable-rich electricity usage data, which makes it easier to estimate carbon emissions accurately and reliably.

5.2 Languages Used

Python programming language known for its versatility and robustness, was chosen for this project. Its wide ecosystem, furnished with a wide range of specialized libraries, made sophisticated data analysis and efficient coding possible. For data manipulation and numerical calculations, libraries like Pandas and NumPy were essential, and Scikit-learn offered a set of tools for implementing different machine learning models into practice and assessing them. Seaborn's interface with Matplotlib made it possible to create perceptive data visualizations. Python was the best option for carrying out this extensive emissions modeling project because of its well-organized and user-friendly syntax and robust library.

5.3 Tools Used

A wide range of development and collaboration tools were used to improve productivity and cooperation. The primary platform for interactive coding and data analysis was Jupyter Notebooks, and the primary environment for version control and collaborative development was GitHub. Jira and Trello were used to effectively manage projects and track tasks. With the Live Share plugin installed, Visual Studio offered a powerful environment for sharing and editing code in real time, facilitating a smooth and collaborative development process.

6 Conclusion

The achievement of accurately estimating carbon emissions for both the transportation and energy consumption sectors using advanced machine learning algorithms constitutes a significant milestone in environmental research. The Decision Tree Regressor is a useful tool for emission estimate in the transportation sector because of its remarkable R-squared Score of 0.9144 in this domain, which demonstrates its expert handling of complex data. Comparably, the Random Forest Regressor shows how well ensemble learning works to manage a variety of data inputs, guaranteeing great prediction precision, with an R-squared Score of 0.9992 in power usage.

These models not only demonstrate that machine learning can effectively address environmental problems, but they also provide a significant contribution to our understanding of both individual and communal carbon footprints. These models' insights play a critical role in guiding well-informed decisions that promote environmental sustainability. This work is crucial to the ongoing efforts to lessen the effects of climate change, highlighting the essential role that state-of-the-art computational methods play in forming a more ecologically conscious and sustainable future.

7 Acknowledgement

We extend our sincere thanks to Professor Dr. Vishnu Pendayala for his invaluable assistance and support throughout this project. His expert guidance has been a key factor in enabling us to accomplish our objectives. url

Appendix A
Criteria Met in Rubrics

1) Visualization- The group successfully communicated complicated data findings by using advanced visualization techniques. This involved illustrating the correlations within the data with heatmaps, bar charts, and line charts so that audiences, both technical and non-technical, could more easily understand the main conclusions of the study.
2) Presentation Skills- Clarity and engagement were prioritized in the development of the presentations. The team made sure the slides were both visually appealing and educational, adding interesting visuals to put the model's performance in context. To increase the impact and accessibility of the technical content, a strong emphasis on storytelling was made.
3) Relates to sustainability - EcoCarbon was in perfect alignment with your sustainability objectives, especially when it came to handling SDGs 12 and 13 of the UN. The project's emphasis on empowering people to calculate and lower their carbon footprint demonstrates your dedication to encouraging ecologically friendly and sustainable behaviors.
4) Saving the model for quick demo - The team managed and stored the models using GitHub, which allowed for rapid and effective demos. The models may be quickly accessed and displayed during presentations, showcasing their functionality and real-time applicability, by placing them on this platform. This strategy also emphasized the project's focus on version control and collaboration, two crucial components of contemporary data science and software development methodologies.
5) Code Walkthrough - Made sure that all of the documentation for the coding process was clear and comprehensive. The team prioritized openness and instructional value by making the technical parts of the project easily available by offering thorough comments in the Jupyter Notebook.
6) Report - The IEEE format and professional standards, particularly clarity, were the main concerns during the report writing process. This strategy demonstrates your dedication to producing a work that is both educational and upholds the high standards needed for distribution among academics and professionals.
7) Version Control - Managing versions of the software effectively was essential to your project management. The success of the project was greatly dependent on the team's exceptional organization and collaborative efficiency, which was shown by storing code and data on GitHub and monitoring progress with JIRA.
8) Discussion and Q/A - Presentations included engaging discussions and QA sessions designed to increase comprehension and encourage participation. Everyone exhibited dedication to open communication and a readiness to consider other points of view and questions by promoting candid discussion.
9) Lessons learned - The research developed our knowledge in applying machine learning algorithms to environmental data, highlighting the need of accurate data collection and preprocessing. Also learned the importance of iterative evaluation and enhancement from the difficulties were faced during the model selection and tuning process. Project management efficiency was demonstrated by the use of collaboration platforms such as JIRA, TRELLO and GitHub, and we discov-

ered the importance of effective communication and teamwork in accomplishing project objectives.Link for the GitHub : https://github.com/chinmaya98/EcoCarbon.

10) Prospects of winning competition/publication – The methodology has an excellent likelihood of winning contests and getting published because of the innovative application of machine learning to calculate carbon footprint and alignment with contemporary environmental concerns. It has the potential for recognition in academic and environmental circles due to its relevance to sustainable development goals and the deployment of advanced data analysis techniques.
11) Innovation - The project is recognized for its innovative way of combining environmental data and machine learning. employed creative methods to improve the accuracy of carbon emission predictions by utilizing the Random Forest and Decision Tree algorithms. In this case, an innovative method is the use of collaborative filtering for personalized recommendations. The research represents a significant new step in its ability to bridge the gap between environmental effect and data science.
12) Evaluation of Performance: Metrics such as R-squared, Mean Absolute Error, and Mean Squared Error were used to carefully assess each model's performance. In the analysis of electricity usage, the Random Forest Regressor performed exceptionally well, but the Decision Tree Regressor performed exceptionally well in the prediction of transportation emissions. Continuous model evaluation and fine-tuning were important to achieve high accuracy in our predictions.
13) Teamwork: Throughout the project, the team showed exceptional coordination and teamwork. employed agile approaches, holding weekly sprint reviews to evaluate our progress and adjust our objectives. Frequent team meetings allowed for open communication and made sure that everyone's contribution matched the goals of the project. Pair programming sessions improved code quality and problem-solving skills by fostering a collaborative atmosphere.
14) Technical difficulty –There were a lot of technical difficulties with the project, especially with the data collection and model optimization stages. Complex computational techniques were needed to handle datasets and integrate different data sources. It took a lot of work and technical expertise to develop machine learning models that can properly anticipate carbon emissions through extensive testing and adjustment.
15) Practiced agile / scrum (1-week sprints): Throughout the development of our project, pair programming was an essential tool. With the use of real-time collaboration tools such as Google Colab and Jupyter Notebook, team members exchanged insights and overcame obstacles combined. This method not only increased our coding productivity but also made sure that everyone in the team understood the obstacles and advancements of the project.
16) Scrum/Agile Project Management (one-week sprints): Implemented one-week sprints and Agile approaches in the project, which allowed us to swiftly adjust to changing needs and guarantee ongoing development. JIRA, which offers a thorough framework for sprint planning, problem tracking, and progress monitoring, was a crucial tool in this process. Team was able to prioritize work, organize duties efficiently, and have a clear view of the state of our project thanks to its strong features. It aided in create a very well-organized and effective workflow by including JIRA into the Agile framework. This allowed us to meet milestones on time and move the research closer to its objectives. Link for the Trello : https://trello.com/invite/b/bvTkEhon/ATTI51f5db654859dcc2c7972b1a3cc827c433697C38/project-management Link for the JIRA : https://alphapharm16.atlassian.net/jira/software/projects/CFP/boards/4
17) Used Grammarly / other tools for language? –Used Grammarly to evaluate the wording and grammar in our documentation to make sure it was clear and professional. To maintain a high level of communication throughout the project was made possible by this tool, which was helpful in helping us refine our report, presentation slides, and code documentation.
18) Slides –Took great care in crafting our presentation slides to clearly convey the goals, process, and outcomes of our project. We used visuals and data visualizations to highlight important topics on each presentation, making sure they were understandable, educational, and visually appealing. The project's importance and effects were effectively communicated through the use of the slides.
19) Demo - In order to demonstrate the functionality of code, a thorough demo was produced. This included a walkthrough of our carbon footprint calculator's user interface and a live demonstration of the machine learning models in operation. The interactive demo was created so that the audience may interact with the technology and learn about its real-world uses.
20) Used LaTeX/IEEE Standards Adherence - The paper was carefully produced using the official IEEE LaTeX template, according to IEEE standards. This made sure that the font size, margins, and general layout all followed academic requirements. Team members' ability to collaborate more easily and expedite the documentation process was made possible via the editing tool Overleaf.
21) Used creative presentation techniques - In order to get the audience interested in the presentation, we used unique tactics. This includes the use of interactive components to show the tool's operation, storytelling to communicate the project's path and effect, and graphic aids to draw attention to important information and outcomes. These strate-

gies increased the impact and recall value of our presentation.

22) Literature Survey - A thorough analysis of the body of work in the domains of sustainability, environmental science, and machine learning was done for the literature study. Our approach was firmly based in the most recent scientific knowledge thanks to our meticulous analysis and citation of pertinent literature. This poll played a pivotal role in molding our technique and confirming our strategy.

APPENDIX

AUTHOR CONTRIBUTIONS

Research and Analysis Phase

- Bhavika Prasannakumar: Conducted the Technology Survey.
- Chinmaya Gayathri: Conducted the Solution Survey.
- Sai Srivathsav Aripirala: Wrote the problem statement and collected data from the survey.
- Shrestha Kommera: Conducted the Literature Survey.
- Yamini Muthyala: Described the data from the Data collection.

Development and Documentation Phase

- Bhavika Prasannakumar: Developed the code for statistical analysis. Tested the transportation model and the electricity model.
- Chinmaya Gayathri: Searched for models with high accuracy. Developed the code for carbon emissions using different modes of transportation.
- Sai Srivathsav Aripirala: Developed the code for carbon emissions using different devices in electricity.
- Shresta Kommera: Developed the code for different model predictions and wrote the report.
- Yamini Muthyala: Developed code for the data visualizations and created the slides.

REFERENCES

[1] Mardani, A., Liao, H., Nilashi, M., Alrasheedi, M., & Cavallaro, F. (2020). A multi-stage method to predict carbon dioxide emissions using dimensionality reduction, clustering, and machine learning techniques. *Journal of Cleaner Production. https://doi.org/10.1016/j.jclepro.2020.122942*.

[2] Agbulut, Ü. (2020). Forecasting of transportation-related energy demand and CO2 emissions in Turkey with different machine learning algorithms. *Applied Energy*, 261. https://doi.org/10.1016/j.spc.2021.10.001.

[3] Yaacob, N. F. F., Yazid, M. R. M., Maulud, K. N. A., & Basri, N. E. A. (2020). A Review of the Measurement Method, Analysis and Implementation Policy of Carbon Dioxide Emission from Transportation. *Environmental Sciences.https://doi.org/10.3390/su12145873*.

[4] Chen, P., Wu, Y., Zhong, H., Long, Y., & Meng, J. (2020). Exploring household emission patterns and driving factors in Japan using machine learning methods. *Energy Policy. https://doi.org/10.1016/j.apenergy.2021.118251*.

[5] Fiorini, L., & Aiello, M. (2020). Household CO2-efficient energy management. *Energy and Buildings. https://doi.org/10.1186/s42162-018-0021-7*.

[6] Department of Energy. (2020). Sustainable Transportation and Fuels. [Online]. Available: *www.energy.gov*.

[7] Department of Energy. (2020). Advancing Sustainable Transportation in the United States. [Online]. Available: *www.energy.gov*.

[8] Department of Energy. (2020). Types of Sustainable Vehicles. [Online]. Available: *www.energy.gov*.

[9] Department of Energy. (2020). Sustainable Fuels and Batteries. [Online]. Available: *www.energy.gov*.

[10] Department of Energy. (2020). Advanced Vehicle Battery Technologies. [Online]. Available: *www.energy.gov*.

[11] NREL. (2020). Variable Renewable Energy, Transmission, and Diurnal Storage. [Online]. Available: *www.nrel.gov*.

[12] NREL. (2020). Other Renewable Energy. [Online]. Available: *www.nrel.gov*.

[13] NREL. (2020). Nuclear and Fossil with Carbon Capture. [Online]. Available: *www.nrel.gov*.

[14] NREL. (2020). Seasonal Storage. [Online]. Available: *www.nrel.gov*.

[15] NREL. (2020). Carbon Dioxide Removal. [Online]. Available: *www.nrel.gov*.

[16] NREL. (2020). Demand-Side Resources. [Online]. Available: *www.nrel.gov*.