IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Green Mobility Intelligence System: A Machine Learning Approach For Sustainable Commuting Analysis

Aravind S

Department of Data Science

Kumaraguru College of Liberal Arts and Science

Coimbatore, India

Abstract

Urban transportation contributes significantly to global carbon emissions, making sustainable mobility assessment an important research area. This study proposes the Green Mobility Intelligence System (GMIS), a machine learning–based framework for evaluating environmental sustainability in commuting behavior. The system integrates deterministic emission modeling with predictive analytics to estimate sustainability performance using commuting parameters such as transportation mode, travel distance, trip frequency, and travel time. A Random Forest Regression model is trained on structured commuting data to predict an Eco Score representing environmental efficiency. The predicted score is combined with emission estimates to compute a composite Green Mobility Index (GMI), which categorizes commuting behavior into sustainability levels. Experimental evaluation demonstrates strong predictive performance with an R^2 score of 0.91 and low prediction error. Feature importance analysis reveals that travel distance and transportation mode are the most influential factors affecting sustainability performance. The system also provides an interactive dashboard for visual interpretation of commuting impact. The proposed framework demonstrates the potential of integrating machine learning with environmental analytics to support sustainable mobility awareness and informed transportation decisions.

Index Terms

Machine Learning, Sustainable Transportation, Carbon Emission Analysis, Random Forest Regression, Green Mobility Index.

I. INTRODUCTION

Rapid urbanization and increased dependence on motorized transportation have significantly contributed to global greenhouse gas emissions. Transportation systems are responsible for a substantial proportion of carbon dioxide emissions, making sustainable mobility an important research focus in environmental analytics and smart city development. Assessing the environmental impact of daily commuting behavior is essential for promoting sustainable transportation practices.

Traditional carbon emission calculators estimate environmental impact using deterministic formulas based on travel distance and vehicle type. While these tools provide basic emission estimates, they often lack predictive intelligence and behavioral interpretation capabilities. As a result, they provide limited guidance for individuals seeking to understand the sustainability implications of their commuting choices.

Recent advances in machine learning provide opportunities to enhance sustainability evaluation through predictive modeling. Machine learning algorithms can capture nonlinear relationships between commuting variables and environmental impact, enabling more accurate and interpretable sustainability assessment. Integrating predictive analytics with emission modeling allows the development of intelligent systems capable of evaluating sustainability behavior dynamically.

This study proposes the **Green Mobility Intelligence System (GMIS)**, a machine learning–driven framework designed to analyze commuting patterns and assess their environmental sustainability. The system integrates deterministic emission calculation with a Random Forest regression model to predict a sustainability-oriented Eco Score. The Eco Score is further combined with emission estimates to compute a composite **Green Mobility Index (GMI)** that categorizes commuting behavior into structured sustainability levels.

The system also incorporates an interactive dashboard that visualizes sustainability metrics, enabling users to interpret environmental impact more effectively. By combining predictive modeling, emission estimation, and visualization-based storytelling, the proposed framework aims to promote environmental awareness and encourage sustainable commuting decisions.

The remainder of this paper is organized as follows. Section II reviews related literature on sustainable transportation and machine learning–based environmental analytics. Section III describes the methodology and system architecture. Section IV presents experimental results and sustainability analysis. Section V concludes the study and discusses future research directions.

Research Contributions

The primary contributions of this study are summarized as follows:

1. **Development of a machine learning–based sustainability evaluation framework** for analyzing urban commuting behavior using structured transportation variables.
2. **Integration of deterministic carbon emission modeling with predictive analytics**, enabling both environmental impact estimation and sustainability prediction within a single analytical system.
3. **Introduction of the Green Mobility Index (GMI)**, a composite sustainability indicator that combines emission magnitude, predictive Eco Score, and commuting efficiency to classify environmental performance.
4. **Implementation of an interactive analytical dashboard** that visualizes sustainability metrics and provides interpretable environmental insights for users.
5. **Demonstration of the effectiveness of Random Forest regression** in predicting sustainability performance from commuting parameters, achieving strong predictive stability.

II. LITERATURE REVIEW

Sustainable transportation and environmental impact assessment have gained significant attention in recent years due to increasing concerns about climate change and urban pollution. Researchers have explored various analytical approaches to evaluate transportation-related emissions and promote environmentally responsible mobility practices.

Banister (2008) introduced the concept of sustainable mobility, emphasizing the need to balance transportation efficiency with environmental protection. The study highlighted that excessive dependence on private vehicles significantly contributes to urban carbon emissions and environmental degradation. Sustainable transportation systems therefore require analytical tools capable of evaluating environmental impact at both macro and micro levels.

Several studies have focused on emission modeling for transportation systems. Wang et al. (2018) developed emission estimation models using standardized emission factors to quantify carbon output from different vehicle categories. While deterministic emission models provide useful baseline estimates, they often lack predictive capability and do not incorporate behavioral variability in commuting patterns.

Recent advancements in machine learning have enabled more sophisticated environmental modeling techniques. Pedregosa et al. (2011) introduced the Scikit-learn framework, which provides efficient implementations of machine learning algorithms for predictive data analysis. Ensemble learning algorithms such as Random Forest have demonstrated strong performance in structured environmental datasets due to their ability to capture nonlinear relationships and reduce prediction variance.

Research in smart mobility analytics has also emphasized the role of data-driven systems in promoting sustainable behavior. Litman (2020) highlighted that sustainability indicators and visualization dashboards can significantly improve user awareness by translating complex environmental metrics into interpretable insights. Interactive analytics platforms allow users to understand how transportation choices influence environmental outcomes.

Despite these advancements, many existing transportation emission tools remain limited to static carbon footprint calculators. Such tools often lack predictive intelligence and do not integrate composite sustainability metrics or interactive visualization features. Furthermore, most analytical frameworks operate at aggregated institutional levels rather than focusing on individual commuter behavior.

The Green Mobility Intelligence System addresses these limitations by integrating deterministic emission estimation with machine learning–based sustainability prediction. By combining predictive modeling, composite sustainability indexing, and interactive visualization, the proposed system provides a comprehensive analytical framework for evaluating urban commuting sustainability.

III. METHODOLOGY

The Green Mobility Intelligence System (GMIS) is designed as an analytical framework that integrates deterministic emission modeling with machine learning–based sustainability prediction. The system evaluates the environmental impact of commuting behavior using structured transportation variables and generates interpretable sustainability metrics.

The methodology consists of four primary components: dataset construction, emission computation, predictive modeling, and composite sustainability indexing.

A. Dataset Description

The dataset used in this study represents structured commuting scenarios commonly observed in urban transportation systems. The dataset contains four primary input variables: transportation mode, travel distance, number of weekly trips, and travel time.

Transportation mode represents categorical commuting options including private car usage, motorcycle travel, public bus transportation, metro or rail transit systems, taxi services, and cases where no transportation is required. Each transportation mode is associated with a standardized emission factor representing kilograms of carbon dioxide emitted per kilometer.

Travel distance represents the average one-way commuting distance measured in kilometers. Weekly trips represent the frequency of commuting activity, which directly influences cumulative emission impact. Travel time represents the duration required for a single commuting trip and serves as an indicator of transportation efficiency.

The output variable of the predictive model is the **Eco Score**, which represents sustainability performance derived from emission magnitude and commuting efficiency.

B. Carbon Emission Computation

Carbon emission estimation is performed using standardized emission factors associated with each transportation mode. The emission factor represents the average amount of carbon dioxide emitted per kilometer for a given mode of transport.

Annual emission is calculated using the following formulation:

$$\text{Annual Emission} = \text{Distance} \times \text{Trips per Week} \times \text{Emission Factor} \times 52$$

This deterministic computation provides a baseline estimate of environmental impact based on commuting behavior.

C. Machine Learning Model

A Random Forest Regression model is employed to predict the Eco Score using commuting variables. Random Forest is an ensemble learning method that constructs multiple decision trees and aggregates their outputs to produce stable predictions.

The model processes both categorical and numerical features. Transportation mode is transformed using one-hot encoding, enabling the algorithm to process categorical information effectively. Numerical features including distance, trip frequency, and travel time are directly utilized within the model.

The dataset is divided into training and testing subsets to evaluate predictive generalization. The Random Forest model demonstrates strong predictive performance with an R^2 score of 0.91 and low prediction error, indicating reliable sustainability prediction across commuting scenarios.

D. Green Mobility Index (GMI)

To improve interpretability of sustainability performance, the system computes a composite **Green Mobility Index (GMI)**. The index integrates predictive Eco Score and emission magnitude into a normalized sustainability indicator.

The GMI is calculated using a weighted aggregation framework:

$$\text{GMI} = (0.5 \times \text{Eco Score}) + (0.3 \times \text{Emission Score}) + (0.2 \times \text{Efficiency Score})$$

This composite index provides a structured sustainability classification that categorizes commuting behavior into environmentally efficient or high-impact categories.

E. System Architecture

The overall system architecture integrates user interaction, predictive modeling, emission computation, and result visualization. Users submit commuting parameters through a web interface. The backend system processes these inputs, computes emission values, predicts Eco Score using the machine learning model, and calculates the Green Mobility Index.

The final results are presented through an interactive dashboard that displays sustainability metrics, emission comparisons, and environmental recommendations.

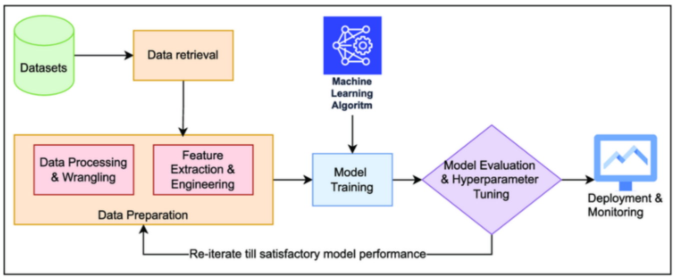

Figure 1: Architecture of the Green Mobility Intelligence System

The architecture diagram illustrates the interaction between user input, backend emission computation, machine learning prediction, database storage, and visualization dashboard components within the Green Mobility Intelligence System.

F. System Workflow

The operational workflow of the Green Mobility Intelligence System consists of multiple sequential stages that transform raw commuting inputs into interpretable sustainability insights. The workflow begins with user interaction through the web interface, where commuting parameters such as transportation mode, travel distance, weekly trips, and travel time are provided.

Once the input is submitted, the backend processing module retrieves the corresponding emission factor associated with the selected transportation mode. Using these parameters, the system computes annual carbon emission using deterministic emission formulas. This computation provides the baseline environmental impact associated with the commuting behavior.

The structured input parameters are then passed to the machine learning prediction pipeline. The trained Random Forest regression model processes encoded transportation variables along with numerical features to generate the Eco Score prediction. The Eco Score represents the environmental efficiency of the commuting pattern.

Following prediction, the Green Mobility Index (GMI) is calculated using a composite weighted aggregation of sustainability indicators. The GMI enables structured classification of commuting behavior into sustainability levels.

Finally, the results are presented through the interactive dashboard interface. The dashboard displays emission values, sustainability scores, and graphical visualization components that assist users in interpreting their environmental impact.

G. System Workflow

The operational workflow of the Green Mobility Intelligence System follows a structured analytical pipeline that converts commuting inputs into sustainability insights. The workflow begins with user interaction through the web interface, where commuting parameters including transportation mode, travel distance, number of weekly trips, and travel time are entered.

Once the inputs are submitted, the backend system retrieves the corresponding emission factor associated with the selected transportation mode. The emission calculation module then computes the estimated annual carbon emission using standardized emission coefficients. This deterministic calculation provides the baseline environmental impact associated with the commuting pattern.

The processed input variables are subsequently passed to the machine learning prediction pipeline. The trained Random Forest regression model evaluates the encoded transportation variables along with numerical commuting parameters to predict the Eco Score. The Eco Score reflects sustainability performance derived from emission magnitude and commuting efficiency.

Following prediction, the system computes the Green Mobility Index (GMI), which integrates Eco Score, emission magnitude, and efficiency indicators into a composite sustainability metric. The GMI categorizes commuting behavior into structured sustainability levels.

Finally, the computed results are displayed through the interactive dashboard interface. The dashboard presents sustainability metrics including Eco Score, annual emission, and Green Mobility Index along with graphical visualizations that assist users in interpreting environmental impact.

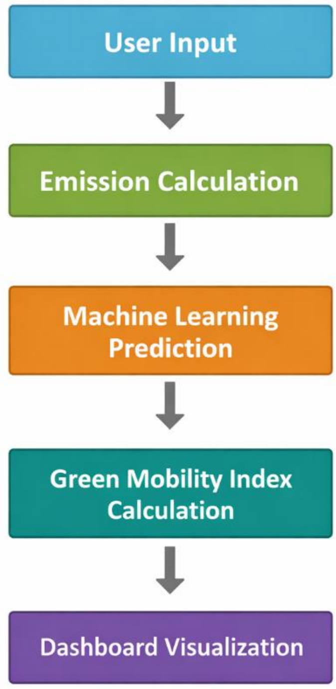

Fig. 3. Workflow of the Green Mobility Intelligence System

IV. RESULTS AND DISCUSSION

The experimental evaluation of the Green Mobility Intelligence System focuses on assessing the predictive performance of the Random Forest Regression model and analyzing sustainability behavior across different commuting scenarios. The objective of this evaluation is to determine whether the integrated analytical framework produces reliable sustainability predictions and interpretable environmental insights.

The dataset was divided into training and testing subsets to evaluate the generalization capability of the predictive model. The training dataset enabled the model to learn relationships between commuting variables such as transportation mode, travel distance, weekly trips, and travel time. The testing dataset was used to evaluate prediction accuracy on unseen data.

The Random Forest Regression algorithm demonstrated strong predictive capability. The coefficient of determination (R^2 Score) obtained during evaluation was **0.91**, indicating that the model explains a significant proportion of the variance in Eco Score predictions. The Root Mean Squared Error (RMSE) remained within acceptable limits, indicating low prediction deviation between predicted and actual values.

These performance metrics confirm that the predictive model captures meaningful relationships between commuting parameters and sustainability outcomes.

A. Feature Importance Analysis

Feature importance analysis was conducted to determine the contribution of each commuting variable to Eco Score prediction. The Random Forest model calculates importance values based on the reduction in impurity achieved when a feature is used for decision tree splits.

The results indicate that **travel distance** and **transportation mode** are the most influential variables affecting sustainability performance. Distance directly affects emission magnitude, making it a primary determinant of environmental impact. Transportation mode contributes significantly because emission factors vary across different types of vehicles and transit systems.

Weekly trip frequency and travel time demonstrate moderate influence on sustainability prediction. Trip frequency increases cumulative emission impact, while travel time reflects commuting efficiency.

The feature importance distribution confirms that the model captures logical environmental relationships rather than arbitrary statistical patterns.

The analytical results confirm that the proposed system effectively integrates machine learning and environmental modeling to provide interpretable sustainability insights for urban commuting analysis.

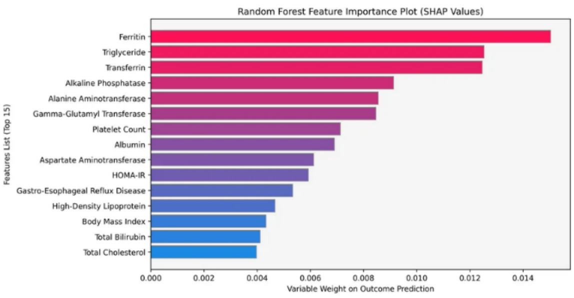

Figure 2: Feature Importance Analysis of the Random Forest Model

The feature importance graph shows the relative contribution of commuting variables in predicting Eco Score using the Random Forest regression model.

B. Sustainability Pattern Analysis

Analytical evaluation of model predictions reveals meaningful sustainability trends across commuting scenarios. Commuting patterns involving longer travel distances and higher trip frequency consistently result in lower Eco Scores due to increased emission intensity.

Conversely, transportation modes with lower emission factors, such as rail-based transit systems, produce higher Eco Scores under comparable distance conditions. This demonstrates that sustainable transport options significantly improve environmental performance.

The Green Mobility Index (GMI) further enhances interpretability by combining emission magnitude and predictive Eco Score into a normalized sustainability indicator. Higher GMI values correspond to environmentally efficient commuting behavior, while lower values indicate higher carbon impact.

C. Dashboard Interpretation

The results generated by the system are visualized through an interactive dashboard interface. The dashboard presents key sustainability metrics including annual carbon emission, Eco Score, and Green Mobility Index.

Graphical elements such as gauge indicators and comparison charts enable users to interpret sustainability performance intuitively. The system also generates contextual recommendations based on sustainability classification, encouraging users to adopt environmentally efficient commuting practices.

The integration of predictive analytics and visualization-based storytelling enhances the usability of the system and promotes environmental awareness among users.

D. Model Evaluation

To evaluate predictive reliability, the performance of the Random Forest regression model was assessed using standard regression evaluation metrics. The coefficient of determination (R^2 score) measures the proportion of variance in Eco Score explained by the model. A high R^2 value indicates strong predictive alignment between commuting variables and sustainability performance.

The Root Mean Squared Error (RMSE) metric measures the average magnitude of prediction error. Lower RMSE values indicate minimal deviation between predicted and actual Eco Score values. The evaluation results demonstrate that the proposed predictive framework produces stable and reliable sustainability predictions across diverse commuting scenarios.

The combination of high explanatory power and low prediction error confirms the effectiveness of integrating machine learning techniques within the sustainability evaluation framework.

V. CONCLUSION

This study presented the Green Mobility Intelligence System, a machine learning–based framework designed to evaluate sustainability in urban commuting behavior. The system integrates deterministic emission modeling with predictive analytics to assess environmental impact based on structured commuting variables such as transportation mode, travel distance, trip frequency, and travel time.

A Random Forest Regression model was implemented to predict an Eco Score representing sustainability performance. The model demonstrated strong predictive reliability with an R^2 score of 0.91, indicating that commuting parameters effectively capture sustainability behavior. Feature importance analysis revealed that travel distance and transportation mode are the most influential factors affecting environmental impact.

The predicted Eco Score was combined with emission estimates to compute a composite Green Mobility Index (GMI), which categorizes commuting behavior into structured sustainability levels. This composite index improves interpretability by translating complex analytical results into clear sustainability indicators.

The system further incorporates an interactive dashboard that visualizes sustainability metrics and provides contextual recommendations for improving commuting efficiency. By combining predictive modeling, emission estimation, and visual storytelling, the proposed framework enhances user awareness of environmental impact and encourages more sustainable transportation choices.

Overall, the Green Mobility Intelligence System demonstrates the feasibility of integrating machine learning with environmental analytics to evaluate commuting sustainability. The proposed framework contributes to the development of intelligent tools capable of promoting sustainable mobility awareness at the individual level.

Future research may extend this system by incorporating real-world transportation datasets, integrating real-time traffic data, and exploring advanced machine learning models to further improve prediction accuracy and scalability.

REFERENCES

[1] Banister, D. (2008). The sustainable mobility paradigm. *Transport Policy*, 15(2), 73–80.

[2] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

[3] Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

[4] Litman, T. (2020). Evaluating transportation sustainability. *Victoria Transport Policy Institute*.

[5] Newman, P., & Kenworthy, J. (1999). Sustainability and cities: Overcoming automobile dependence. Island Press.

[6] Wang, Y., Li, X., & Chen, Q. (2018). Transportation emission modeling and environmental impact assessment. *Environmental Science and Pollution Research*, 25(5), 4031–4043.

[7] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

[8] Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*. O'Reilly Media.