International Journal of Future Engineering Innovations

Towards Smarter Climate Monitoring: CO_2 Emission Predictions through Machine Learning

Kamatham Venkata Nagamalleswari [1*], Yarram Tanuja [2], Kallam Srikanth Reddy [3], Chinthareddy Vishnu Kanth Reddy [4]
[1-4] Student, Department of Information Technology, Kallam Haranadhareddy Institute of Technology, Chowdavaram, Guntur, Andhra Pradesh, India

* Corresponding Author: **Kamatham Venkata Nagamalleswari**

Article Info

ISSN (online): 3049-1215
Volume: 02
Issue: 02
March-April 2025
Received: 02-02-2025
Accepted: 28-02-2025
Page No: 56-60

Abstract
The development of effective prediction models for emissions monitoring and control is essential due to the growing amounts of CO_2 emissions from automobiles, which greatly contribute to environmental pollution as well as global warming. Conventional techniques for calculating CO_2 emissions frequently depend on empirical calculations and static presumptions, which might not adequately represent the intricate connections between emission rates and vehicle attributes. This paper presents a machine learning-based method to forecast CO_2 emissions based on important vehicle metrics including engine size (L), cylinder count and fuel use (L/100 km) in order to solve Environmental problem. In order to refine the dataset before training several models, such as Linear Regression, Random Forest, Support Vector Regression (SVR) and K-Nearest Neighbors (KNN), the study uses exploratory data analysis (EDA) and data preprocessing. The Random Forest model was the best option for real-time predictions, with the greatest degree of precision (0.9609), according to performance assessment using the R2 score. In order to provide customers with an easy-to-use and efficient tool for making decisions in the automotive and environmental policy sectors, a Streamlit-based user interface (UI) was created that enables users to enter car specs and instantaneously get CO_2 emission forecasts. This paper demonstrates how machine learning may improve pollution control methods encourage environmentally friendly transportation and support legal compliance.

DOI: https://doi.org/10.54660/IJFEI.2025.2.2.56-60

Keywords: CO_2 Emissions, Data Preprocessing, Machine Learning, Prediction Modelling, Random Forest, Real-Time Prediction and Vehicle Emissions

1. Introduction

Significant ecological and health issues are raised by the sharp rise in CO_2 emissions from automobiles, which is a primary cause of environmental degradation and climate change [1]. The growth of international transportation has made it increasingly difficult for governments, legislators and automakers to monitor and regulate emissions. Historically, mathematical models based on engine size and fuel consumption have been used to predict CO_2 emission levels; however, these approaches frequently fall short of capturing the nonlinear interactions between various vehicle parameters and their impact on the environment. Using machine learning (ML) can help design sustainable transportation regulations by offering a more precise, data-driven method of anticipating emissions.

Current techniques for predicting CO_2 emissions are based on regulatory frameworks and static models that employ predetermined emission factors derived from vehicle specifications. These techniques produce less accurate forecasts because they ignore changes in vehicle performance over time. Furthermore, a lot of conventional methods don't have an interactive system that allows users to enter particular vehicle characteristics and get real-time feedback on emissions levels [2]. By examining patterns in huge datasets, machine learning algorithms, on the opposite hand, can learn from data dynamically and increase forecast accuracy. Even though there are ML-based models, many of them are difficult to utilize or obtain for real-time applications.

By creating a machine learning-based CO_2 emission forecast system with a streamlit user interface for real-time user involvement our research aims to close this gap. This study assesses many machine learning models such as Linear Regression, Random Forest, Support Vector Regression (SVR) and K-Nearest Neighbors (KNN) by examining a dataset of automobile emissions in order to determine which prediction is the most successful. The Random Forest model was chosen for deployment because it showed the highest accuracy (R2 = 0.9609). The finished system is a useful tool for policy planning, car manufacture and consumer awareness since it enables users to enter important vehicle information, such as engine size, cylinder count and consumption of fuel (L/100 km) and receive immediate emission predictions.

2.Related Work

This section provides a review of existing approaches, highlighting their advantages and limitations of the work.

Farahzadi *et al*. proposed that Innovative innovations [3] like machine learning and artificial intelligence can assist lower carbon dioxide emissions, which are largely caused by the building industry. Five clusters were found through a content examination of 78 papers: optimized, making choices and solution-based platforms, on-site cars and equipment, energy consumption and life cycle evaluation, sustainable components and materials design/production, and real-world monitoring. The study offers insights into clever methods used in the building sector to reduce CO_2 emissions, offering insightful suggestions for further study.

Meng *et al*. purpose of this work [4] is to forecast future carbon dioxide emissions using four ARIMA-based SARIMA prediction models (SARIMAX). The models are predicated on the COVID-19 pandemic's peak period. Global total emissions of carbon dioxide are predicted by the study to be 2022–2027 for the near future, 2022–2054 for the future, and 2022–2072 for the long future. In terms of forecasting previous emissions, present COVID-19-related emissions, and future emissions, the study's Machine Learning (ML) approach has demonstrated excellent agreement to the IPCC model. These forecast findings can be utilized to create appropriate policies for lowering CO_2 emissions worldwide. For more accurate predictions, future studies should incorporate more external influence variables.

Bhatt *et al*. examined that in the twenty-first century [5], carbon dioxide emission is a big problem since they are causing global warming, melting Antarctic ice caps, polar wildlife extinction and coastal flooding. Determining hazard levels and milestones, including the risk threshold and point of no return, is essential to the fight against climate change. By 2047, the threshold of 500 ppm will be achieved, according to historical statistics, necessitating a decrease rate of 6.37% and an upsurge rate of 23.38%. These emissions are caused by socioeconomic variables such as the population, greenhouse gasses, and combustion industry. To create action plans and hasten the transition to renewable energy sources, more research is required.

3. CO_2 Emission predictions through machine learning

Using important characteristics including engine size (L), cylinder count, and consumption of fuel (L/100 km), this study proposes a machine learning-based method for forecasting CO_2 emissions from automobiles. exploratory data analysis (EDA) and data preliminary processing, which includes feature selection, addressing missing values and normalizing the information for improved model performance, are the first steps in the methodology [6]. The R2 score is used to train and assess a number of machine learning models, such as K-Nearest Neighbors (KNN), Random Forest, Support Vector Regression (SVR) and Linear Regression. Random Forest was chosen for real-time predictions since it had the highest accuracy (0.9609) among them. The development of a streamlit-based UI, or user interface, allows customers to enter car specifications and obtain real-time CO_2 emission forecasts.

- **Developing a reliable system for predicting Co_2 emissions:**
  Improve on conventional estimation techniques by developing a machine learning model that can precisely forecast CO_2 emissions based on vehicle attributes.
- **Perform Exploratory Data Analysis (EDA) to identify trends:**
  Analyze the dataset thoroughly to find trends and connections between engine size, consumption of fuel, number of cylinders, and CO_2 emissions, offering information on important affecting factors.
- **Apply data preprocessing techniques for enhanced model performance:**
  To make sure the dataset is clean and appropriate for machine learning models, apply data preparation techniques such as feature scaling, addressing missing values, and encoding categorical variables.
- **Compare multiple machine learning models to identify the best performer:**
  To find the most accurate and effective model, train and assess various machine learning models, including Linear Regression, Random Forest, SVR and KNN. Then, compare the model's performances using the R2 score.
- **Select and optimize the best model for real-time predictions:**
  Select the best model (Random Forest) and adjust its hyperparameters to optimize accuracy and guarantee accurate real-time CO_2 emissions forecasts.
- **Develop an interactive streamlit UI for user-friendly predictions:**
  Make the software accessible to a broad spectrum of users by designing a streamlit-based UI (user interface) that allows users to enter important vehicle attributes and obtain anticipated CO_2 emissions instantaneously.

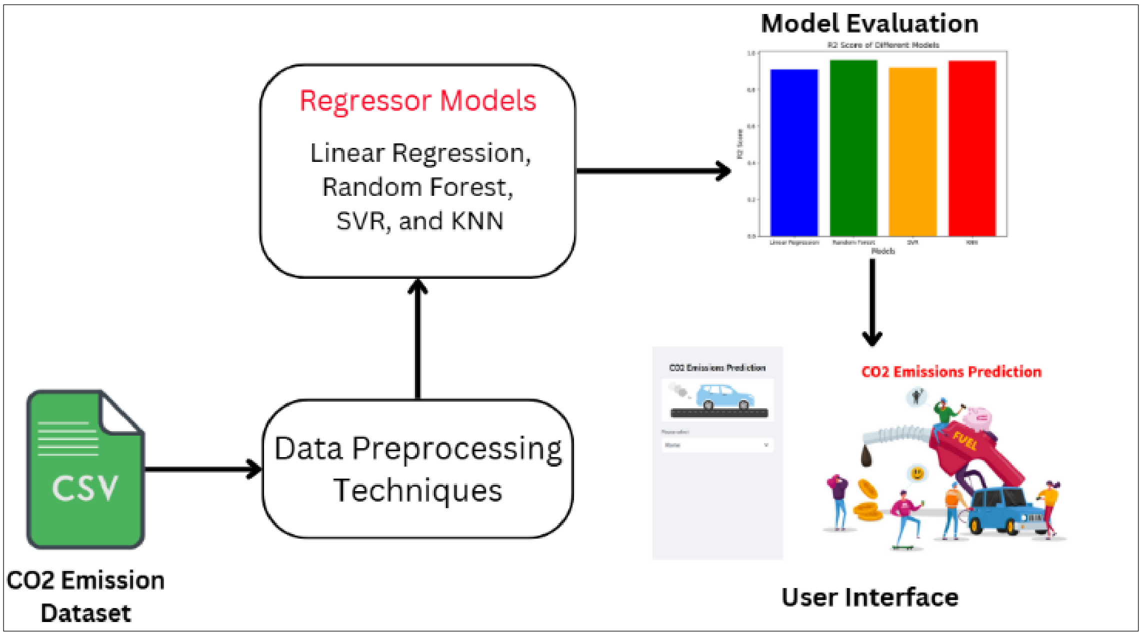

Fig 1: Architecture of Regressor models for CO_2 Emission Detection

Linear Regression:

Based on an exponential connection between the input factors (engine size, cylinders, and fuel usage) and the output (CO_2 emissions), the straightforward yet powerful supervised learning process known as linear regression is utilized to forecast CO_2 emissions [7]. By reducing the discrepancy between the actual and anticipated values, it fits the data in a straight line. Despite having an R2 score of 0.9074, linear regression's capacity to identify intricate patterns in the data is constrained by its assumption of linearity.

Random Forest:

To increase accuracy and robustness, Random Forest, an ensemble learning technique, constructs several decision trees and aggregates their predictions [8]. It effectively manages both linear and nonlinear interactions, which makes it ideal for predicting CO_2 emissions. With an R2 score of 0.9609, Random Forest beat the other models in this study, making it the best option for real-time forecasts because of its accuracy and versatility in handling various vehicle features.

Support Vector Regression (SVR):

SVR is a potent regression algorithm that can handle irregular patterns in data by fitting the optimal function within a specified margin using a hyperplane [9]. SVR effectively captures intricate relationships between characteristics of inputs and CO_2 emissions through the use of kernel functions SVR was a strong candidate in this investigation with an R2 score of 0.9193, although it cost more to compute than Random Forest.

K-Nearest Neighbors (KNN):

A non-parametric technique called KNN averages the results of the K closest points of data in the feature space to forecast CO_2 emissions [10]. It performs well with little datasets and is quite interpretable, but as the dataset size increases, it becomes computationally costly. KNN was a competitive model in this study with an R2 score of 0.9561. Nevertheless, it needed the K value to be adjusted for best results and was marginally not as precise than Random Forest.

4. Results and Discussion

The R2 score, a gauge of how closely the projected values match the actual emissions data, was used to assess the efficacy of predictive machine learning models for CO_2 emissions prediction. With an R2 score of 0.9609, Random Forest outperformed the other models tested, demonstrating its impressive capacity to capture intricate correlations between engine size, cylinder count, fuel consumption, and CO_2 emissions. With an R2 score of 0.9561, K-Nearest Neighbors (KNN) came in second, demonstrating how well it could identify regional trends in the data. With R2 scores of 0.9193 and 0.9074, respectively, Support Vector Regression (SVR) and Linear Regression both did rather well. Although linear regression offered a simple method, its accuracy was constrained by its linearity assumptions. Even though SVR was more accurate than linear regression, it was computationally demanding and needed kernel parameters to be carefully adjusted.

One of the study's main conclusions is that when it comes to predicting CO_2 emissions, ensemble learning methods Random Forest in particular perform noticeably better than conventional regression models. This is mostly because Random Forest aggregates predictions from several decision trees, making it resistant against overfitting and able to manage nonlinear dependencies. Despite its effectiveness, the KNN model was quite sensitive to the K value selection and needed a lot of fine-tuning to function at its best. Furthermore, compared to Random Forest, SVR was less feasible for real-time applications due to its high computational cost, even if it produced results that were competitive.

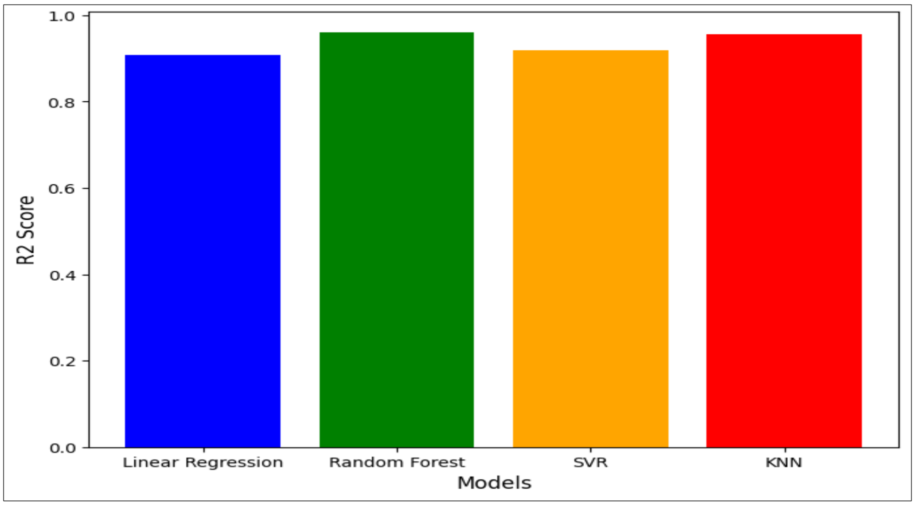

Fig 2: R2 Score of Different Models

Fig 3: Visualizations Page in User Interface

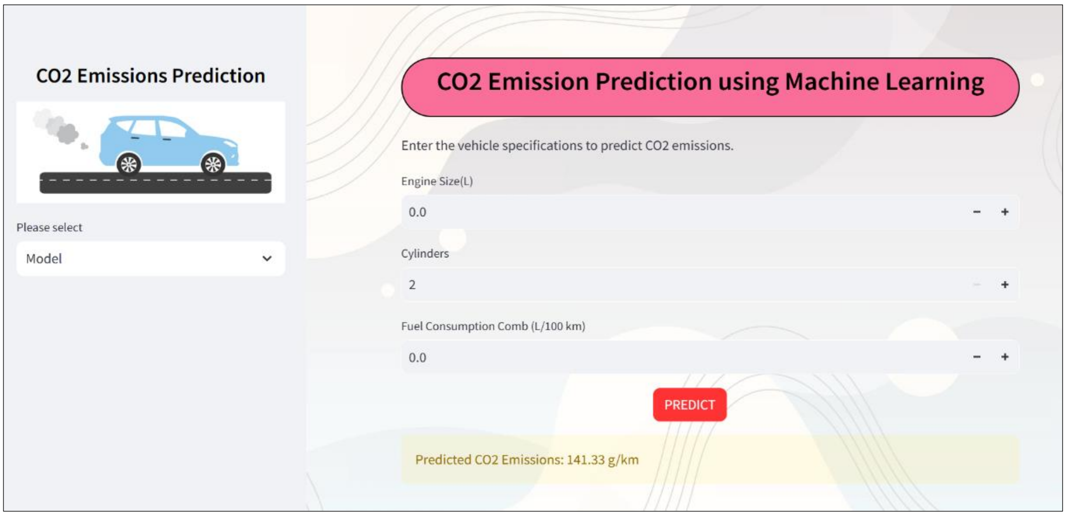

Fig 4: CO_2 Emission Page in User Interface

A Streamlit-based user interface (UI), which enables users to input engine size, cylinder count, and fuel use to generate real-time CO_2 emission forecasts, was created in order to make the simulation accessible for real-world applications. This interactive technology offers a simple and effective way to monitor the environment, make policy, and raise consumer awareness. By providing instantaneous insights into car emissions, the user interface (UI) improves user engagement. This information can be especially helpful for regulatory bodies, automakers, and environmentally aware consumers. This study closes the gap between theoretical emissions models and workable, data-driven solutions by using machine learning for immediate time prediction, ultimately supporting efforts to reduce emissions and promote sustainable mobility.

5. Conclusion

Using important criteria including engine size, cylinder count, and fuel consumption, this study offers an effective machine learning-based method for forecasting CO_2 emissions from automobiles. Several predictive models were trained and assessed following the application of data preparation techniques and Exploratory Data Analysis (EDA). With the highest accuracy (R2 = 0.9609) among them, Random Forest is the most dependable model for predicting emissions in real time. By enabling users to enter car specs and receive real-time emission estimations, the Streamlit-based UI, or user interface, further improves the system's accessibility. The study shows that ensemble learning approaches outperform conventional regression methods in predicting CO_2 emissions. This research supports regulatory decision-making and environmental awareness by offering an interactive, data-driven solution. This study can be extended in the future by adding deep learning methods, real-time sensor data, and a smartphone application for more accessibility.

6. References

1. Piguet E. Linking climate change, environmental degradation, and migration: An update after 10 years. Wiley Interdisciplinary Reviews: Climate Change. 2022;13(1): e746.
2. Liu Z, *et al*. Near-real-time carbon emission accounting technology toward carbon neutrality. Engineering. 2022;14:44–51.
3. Farahzadi L, Kioumarsi M. Application of machine learning initiatives and intelligent perspectives for CO_2 emissions reduction in construction. Journal of Cleaner Production. 2023;384:135504.
4. Meng Y, Noman H. Predicting CO_2 emission footprint using AI through machine learning. Atmosphere. 2022;13(11):1871.
5. Bhatt H, *et al*. Forecasting and mitigation of global environmental carbon dioxide emission using machine learning techniques. Cleaner Chemical Engineering. 2023;5:100095.
6. Ahsan MM, *et al*. Effect of data scaling methods on machine learning algorithms and model performance. Technologies. 2021;9(3):52.
7. Ayaz İ. Forecasting CO_2 emissions with machine learning methods: Türkiye example and future trends. NATURENGS. 2024;5(2):82–87.
8. Salman HA, Kalakech A, Steiti A. Random forest algorithm overview. Babylonian Journal of Machine Learning. 2024;69–79.
9. Ahmed HU, *et al*. Support vector regression (SVR) and grey wolf optimization (GWO) to predict the compressive strength of GGBFS-based geopolymer concrete. Neural Computing and Applications. 2023;35(3):2909–2926.
10. Faruque MO, *et al*. A comparative analysis to forecast carbon dioxide emissions. Energy Reports. 2022;8:8046–8060.