GreenMiles: Utilizing Deep Learning to Analyze Vehicular Carbon Emission Trends

Rohan Arunkumar, Dr. Ronald E. McNair Academic High School, Jersey City, USA
rohan.arunkumar07@gmail.com

Abstract

The escalating impacts of climate change have notably influenced many global problems, including economic loss and environmental disasters. Global warming is caused by the buildup of heat-trapping greenhouse gases in the atmosphere, which has increased the temperature of the planet at a rate of roughly 0.15 to 0.20°C per decade. Transportation contributes to roughly 25% of global carbon dioxide (CO_2) emissions, and vehicles alone are responsible for nearly 75% of that share. Despite these alarming statistics, most consumers are unaware of how vehicle configurations affect CO_2 output. To address this gap, I developed a set of deep learning models that analyze patterns in vehicle-related carbon emissions using an official dataset from the Canadian government. The models identified which vehicle settings (such as fuel type and transmission) are most strongly associated with high emissions. After testing, the best-performing model was deployed on a user-friendly web application, where consumers can input different vehicle parameters and receive predicted CO_2 emission levels. By allowing users to explore how different vehicle configurations affect carbon emissions, the application also functions as an educational tool, raising awareness of environmental impacts through direct interaction with AI. Unlike existing tools that provide only general estimates, this web application allows real-time, interactive predictions tailored to specific vehicle configurations, filling a gap in consumer-facing technology. The tool is currently live and has the potential for future integration into vehicle checkout systems to help consumers make more climate-conscious decisions.

Introduction and Literature Review

The Problem with CO_2 Emissions from Vehicles

Climate change has greatly influenced many global problems, including economic loss, increased rates of natural disasters, and other negative environmental impacts. One of the most dangerous factors of climate change is global warming, as it directly affects weather patterns and natural disasters. Global warming is caused by two main factors, the buildup of heat-trapping greenhouse gas in the atmosphere and the hole in the Earth's ozone layer (which is caused by ozone-depleting substances). This leads to measurable increases in the amount of UV-B radiation reaching the surface (Zhang et al., 2012, 118-119).

One such greenhouse gas, carbon dioxide (CO_2), is commonly found in the air and is a heavy contributor to global warming. UCAR states that "Carbon dioxide levels in our atmosphere have thus risen about 40% since the start of the Industrial Revolution" (*Carbon Dioxide | Center for Science Education*, n.d.). Globally, transport accounts for around a quarter of CO_2 emissions, with leading sources being vehicles and planes, wherein vehicles account for nearly three quarters of the greenhouse gas emissions that come from transport. According to research scientist Brian C. MacDonald, "In addition to CO_2, motor vehicles emit combustion byproducts that lead to air quality problems and exert short-lived effects on climate. Relevant pollutants include carbon monoxide (CO), nitrogen oxides (NOx), volatile organic compounds

(VOCs), and black carbon" (McDonald et al., 2014, 5283). Therefore, air pollution from vehicles needs to be minimized through public outreach, which is the main focus of this project.

Current Solutions

Current methods to reduce the production of CO_2 from vehicles include public action and air quality monitoring, one example being the LEV program in California (Ravi et al., 2023, 2). However, the issue with aiming to lower CO_2 production is that there is no employed technology that can analyze patterns in emissions nor is there an online tool that civilians can use to help lower emissions on their end through consumer actions such as buying vehicles. The current method of monitoring air quality and only setting an emission standard for each state is not effective. While research studies exist, most public-facing tools still lack the interactivity and predictive capabilities this project introduces, especially for individual consumers making real-time decisions. More consumers need to understand the numerical impact of their carbon footprints.

Deep Learning Applications

One technology that can achieve pattern sensing is deep learning, a subset of machine learning models that can understand patterns at a deeper level (Mathew et al., 2021, 599-601). Although other scientists have previously investigated the correlation between CO_2 concentration and vehicles, only a few have experimented forecasting with deep learning models. One analysis was done by research scientist Fatih Gurcan, who explored the performance of machine learning, deep learning, and ensemble learning models to forecast CO_2 emissions of fuel vehicles. Gurcan trained these models on an open source dataset provided by the Canadian government, where the input columns include the type of vehicle, brand of vehicle, fuel consumption, etc. and the output column is the vehicle's CO_2 emissions. Gurcan used machine learning performance metrics such as RMSE (Root Mean Square Error), R^2 (Coefficient of Determination), and the model runtime to conclude that ensemble models are the best performers in terms of speed and accuracy (Gurcan, 2024, 4-6). However, the deep learning model used in this project (MLP) was chosen for its ability to capture nonlinear relationships between input features and CO_2 emissions, which simple models like KNN cannot easily detect. The performance gains show the importance of capturing those patterns. To expand on Gurcan's approach, I test additional regression deep learning models and create features that users can interact with while learning about the importance of minimizing emissions.

Proposed Solution

To minimize CO_2 emissions from vehicles, I propose the idea of creating deep learning regression models that can visualize patterns in what vehicle settings lead to high air pollution. The model is trained using an open-source dataset with vehicular CO_2 emission data provided by the government of Canada (*Fuel Consumption Ratings - Open Government Portal*, 2019). The model is accessible through a free web application with a simple, user-friendly interface. In the future, it could be integrated into checkout systems used by vehicle companies, providing real-time feedback for consumers. The model's insights can help drivers adjust their behaviors or vehicle settings to reduce CO_2 emissions.

This solution uses modern technology to make patterns between CO_2 levels and vehicle usage more accessible. Furthermore, it empowers vehicle owners with information on how to minimize their environmental impact, contributing to the larger effort to combat climate change.

Methods and Procedure

Preliminary Analysis

- The dataset was downloaded and its structure, key features, and accompanying documentation were reviewed.
- An initial exploratory analysis was conducted by computing descriptive statistics and generating visualizations to understand data distributions and identify potential outliers.
- The final list of input features used in the model includes: transmission type, engine size, fuel type, number of cylinders, vehicle class, fuel consumption (city and highway), and combined fuel consumption. These features were selected based on relevance and availability in the dataset.

Data Preprocessing and Feature Engineering

- The dataset was cleaned by handling missing values, removing duplicates, and correcting any data inconsistencies.
    - The “Natural Gas” fuel type was removed as it only had one row of data.
- Feature engineering was performed by selecting and transforming relevant features using tools such as Scikit-learn to enhance model performance.
- The cleaned dataset was split into training (80%), validation (20%), and test (20%) sets to make sure the model performs well on unseen data.

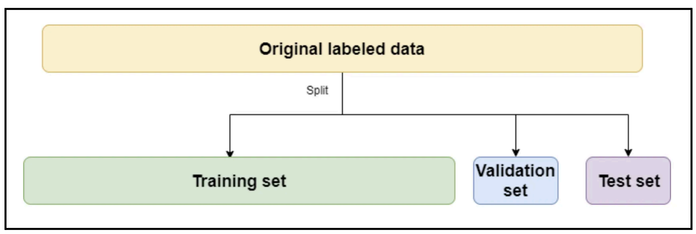

Figure 1: Dataset Train-Test-Validation Split

Model Development

- Baseline regression models were established using Scikit-learn. The models used included RandomForestRegressor, Multi-Layer Perceptron, and K-Nearest Neighbors.

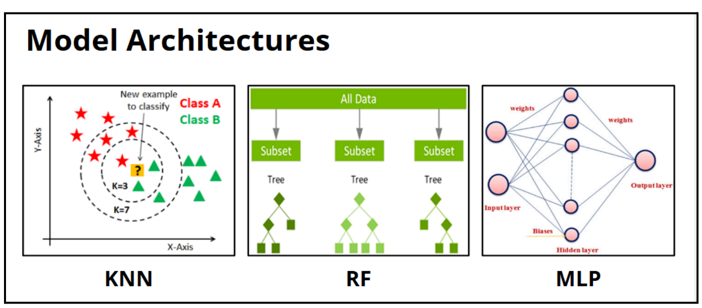

Figure 2: Comparison of Model Architectures

- The models were developed within Google Colab, which provides free, cloud-based CPU access for model training. They were trained using the cleaned training dataset, and the training process was monitored via loss and performance metrics (RMSE, R^2).

- The trained models were evaluated on the validation and test sets to assess its predictive accuracy and generalizability. When required, the model's performance was optimized with re-training and adjusting parameters.
- The MLP architecture consisted of an input layer matching the number of selected features, followed by two hidden layers with 64 and 32 neurons respectively, each using the ReLU activation function. The final output layer used a linear activation function appropriate for regression. The model was compiled using the Adam optimizer with a learning rate of 0.001 and was trained using the mean squared error (MSE) loss function. Early stopping and dropout were applied to prevent overfitting. The MLP's deep layered structure allowed it to learn nonlinear relationships (between vehicle specifications and CO_2 emissions) more effectively than the other models tested. Its superior accuracy on both validation and test sets led to its deployment in the final web application.

Web Application Development

- A user-friendly web application was developed using the Python "Streamlit" module, allowing users to interact with the model by inputting vehicle data and visualizing predicted CO_2 emissions. The best performing deep learning model was integrated into the Streamlit application and backend processes were developed for real-time predictions.
- In addition to the page where users can try out the deep learning model, there are also pages with background information for consumers, including statistics and the feature engineering of the model (what vehicle settings contribute most to CO_2 emissions).
- The current version of the web application has default values pre-filled for convenience; however, I acknowledge this may skew perception of outputs. Future updates will include more realistic defaults and optional blank inputs for more accurate user-driven exploration.
    - Feature engineering revealed that ethanol gas reduced CO_2 emissions by 12%.
    - Fuel consumption was found to be directly proportional to CO_2 emissions.

Results

The best model was determined by its test accuracy. After testing different configurations of KNN, the best model had a test RMSE of 5.984 and test R2 of 0.989, both of which were the worst of the three models. The best Random Forest model had a test RMSE of 3.654 and test R2 of 0.995. Although the ratings are exceptionally high, Random Forest was slightly under MLP. The best MLP architecture (see "Model Development") had a test RMSE of 3.375 and test R2 of 0.996. MLP performed best in terms of accuracy, which is understandable due to its ability to learn nonlinear relationships.

In conclusion, the best model was MLP, and the worst model was KNN. Therefore, MLP was used in the web application as the base model for predictions (the "predict" tab).

This confirms the value of using MLP over simpler models, especially for capturing subtle nonlinear interactions that other regressors may miss, supporting its inclusion despite being more complex.

Table 1: Accuracy Results

Model	Val RMSE	Test RMSE	Val R^2	Test R^2
KNN	5.677	5.984	0.990	0.989
RF	3.048	3.654	0.997	0.995
MLP	2.979	3.375	0.997	0.996

Conclusion/Discussion

Reducing CO_2 emissions is crucial for mitigating climate change. Lower emissions not only reduce greenhouse gases in the atmosphere, thereby alleviating global warming, but they also improve local air quality, addressing pollution concerns that affect both the environment and public health (West et al., 2013, p. 885).

The results of this study demonstrate that machine learning regression models can effectively predict CO_2 emissions based on key vehicle parameters such as fuel type and miles traveled. MLP ended up being the best model due to its multiple layers and connections, allowing it to generalize better. The high accuracy of my trained models supports my hypothesis that regression models can serve as a reliable source of information for unaware consumers looking to buy vehicles. This project shows how machine learning solutions can address environmental challenges by informing consumers on how to reduce vehicle emissions.

Though this tool is still in early stages, it has strong potential to raise awareness and influence more climate-conscious consumer behavior through education and engagement. It can serve as a foundation for future outreach or integration into decision-making platforms.

References

Carbon Dioxide | Center for Science Education. (n.d.). UCAR Center for Science Education. https://scied.ucar.edu/learning-zone/how-climate-works/carbon-dioxide

Fuel consumption ratings - Open Government Portal. (2019). Canada Open Government. https://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64#wb-auto-6

Gurcan, F. (2024). Forecasting CO2 emissions of fuel vehicles for an ecological world using ensemble learning, machine learning, and deep learning models. *PeerJ Comput. Sci*. https://peerj.com/articles/cs-2234.pdf

Mathew, A., Amudha, P., & Sivakumari, S. (2021). Deep Learning Techniques: An Overview. *Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020*, 599-608. https://www.researchgate.net/profile/Munish-Sabharwal/publication/341652471

McDonald, B. C., McBride, Z. C., & Martin, E. W. (2014). High-resolution mapping of motor vehicle carbon dioxide emissions. *J. Geophys. Res. Atmos*, *119*, 5283-5298. https://agupubs.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/2013JD021219

Ravi, S. S., Osipov, S., & Turner, J. W. G. (2023). Impact of Modern Vehicular Technologies and Emission Regulations on Improving Global Air Quality. *Atmosphere*, *14*(7), 1-20. https://www.mdpi.com/2073-4433/14/7/1164

West, J. J., Smith, S. J., Silva, R. A., Naik, V., Zhang, Y., Adelman, Z., Fry, M. M., Anenberg, S., Horowitz, L. W., & Lamarque, J.-F. (2013). Co-benefits of mitigating global greenhouse gas emissions for future air quality and human health. *Nature climate change*, *3*(10), 885-889. https://pmc.ncbi.nlm.nih.gov/articles/PMC4051351/

Zhang, X. F., Zhang, S. Y., Hu, Z. Y., Yu, G., Pei, C. H., & Sa, R. N. (2012). Identification of connection units with high GHG emissions for low-carbon product structure design. *Journal of Cleaner Production*, *27*, 118-125. https://www.researchgate.net/profile/Zhang-Xiufen/publication/370260121