**Model Description**

**Input:** The model inputs include various vehicle characteristics excluding 'Fuel Consumption':

* Year (all entries are from the year 2000)
* Make (manufacturer)
* Model
* Vehicle Class (e.g., SUV, sedan)
* Engine Size (in litres)
* Cylinders (number of)
* Transmission (type)
* Fuel Type (e.g., gasoline, diesel)

**Output:** The model outputs a prediction of the CO2 emissions of the vehicle, measured in grams per kilometre.

**Model Architecture:** Gradient Boosting Machine (GBM) was utilised for its robustness and effectiveness in handling tabular data and its ability to model complex nonlinear relationships. The final model used 150 trees, a learning rate of 0.2, and a max depth of 4.

**Performance**

The model was evaluated using Mean Squared Error (MSE) and R-squared (R²) metrics:

* **Initial Gradient Boosting MSE:** 378.58
* **Optimized Gradient Boosting MSE:** 262.29 after hyperparameter tuning.
* **R-squared (R²):** Achieved nearly 95% showing high explanatory power.

**Limitations**

* **Data Year Limitation:** The model only includes data from the year 2000, which may not generalise well to modern vehicles with newer technologies and emissions standards.
* **Exclusion of Fuel Consumption:** While excluding this feature avoids direct inference, it might overlook its practical importance in predicting emissions.

**Trade-offs**

* **Model Complexity vs. Interpretability:** The GBM model, while powerful, is less interpretable than simpler models like linear regression. This might make it harder for stakeholders to understand how input features relate to predictions.
* **Accuracy vs. Generalisability:** The high performance on the dataset might not perfectly translate to real-world performance due to the controlled conditions of the data collection and the potential non-representation of all types of vehicles.