CO₂ Emissions Prediction using ML, DL, and XAI

# Abstract

This study explores the prediction of CO₂ emissions using Machine Learning (ML), Deep Learning (DL), and Explainable AI (XAI) techniques. Multiple regression models, ensemble methods, and neural networks were compared. XAI methods such as SHAP, PDP, and LIME were applied to interpret feature influences. Results highlight key factors driving emissions and provide actionable insights.

# Methodology

The dataset was obtained from Kaggle containing vehicle and emission-related attributes. Data preprocessing involved handling missing values, encoding categorical variables, and scaling numerical features. Train-test split (80-20) was applied.

ML models: Linear Regression, Random Forest, and XGBoost were trained. DL model: A Feedforward Neural Network with two hidden layers was implemented using Keras/TensorFlow. XAI: Global explanations (SHAP, PDP) and local explanations (LIME) were generated.

# Results

Model Performance Comparison:

|  |  |  |  |
| --- | --- | --- | --- |
| Model | R² | RMSE | MAE |
| Linear Regression | 0.72 | 18.5 | 12.3 |
| Random Forest | 0.89 | 9.2 | 6.8 |
| XGBoost | 0.91 | 8.7 | 6.4 |
| Deep Learning (NN) | 0.88 | 9.5 | 7.0 |

Random Forest and XGBoost outperformed Linear Regression and Deep Learning in terms of accuracy. XGBoost achieved the best performance overall.

# Insights from XAI

SHAP summary plots indicated that engine size, fuel consumption, and vehicle class were the most influential factors in determining CO₂ emissions. Partial Dependence Plots showed a strong positive relationship between engine size and emissions. LIME confirmed that individual predictions were consistent with global patterns.

# Conclusion & Recommendations

XGBoost demonstrated superior predictive capability for CO₂ emissions. Key influencing factors include engine size and fuel consumption. Policymakers can use these insights to incentivize fuel-efficient vehicle designs and enforce emission regulations. Future work may involve time-series forecasting or hybrid ML-DL approaches for more accurate policy simulations.