Project Proposal

Team Members:

1. Akhilesh Akoju
2. Kianoush Ramezani Shabolaghi
3. Lokesh Reddy Kuppireddy
4. Sai Akhil Sadula
5. Suraj Hanumanthapura Chandra

Title:

Predicting CO2 Emissions of Vehicles Using Real-Time Telematics Data

Project Overview:

This project aims to create a predictive model for CO2 emissions based on vehicle characteristics. The primary focus is on leveraging machine learning to accurately predict CO2 emissions, which is essential for environmental planning and regulatory compliance. We will perform data pre-processing, feature engineering, model selection, training, and evaluation to achieve our objectives.

Research Questions:

1. Can we develop a machine learning model to accurately predict the CO2 emissions of vehicles using real-time telematics data?
2. What are the most important factors that influence CO2 emissions?
3. How can we use our model to help reduce CO2 emissions from the transportation sector?
4. Which features have the strongest correlation with CO2 emissions?
5. How can we assess the environmental impact of vehicles more effectively using predictive modelling?

Dataset:

We plan to use the following dataset in our research project:

- Car Fuel & Emissions 2000-2013
- Link: https://data.world/amercader/car-fuel-emissions-2000-2013
- Size: 7.54 MB
- Samples: 45512

Results of Basic EDA:

During the preliminary EDA, we found the following key results:

1. The dataset contains 45119] samples, with each sample representing a vehicle's information.
2. Our target variable is "CO2 emissions," which is a continuous variable.
3. We have identified **5** relevant continuous features, including engine capacity, urban metric, extra urban metric, combined metric, and noise level.
4. The dataset shows a portion of missing values, which we plan to handle through imputation or removal based on their impact on the analysis.
5. We have identified [number of outliers some of the outliers in the dataset, particularly in the "engine capacity" feature. We will carefully address these outliers to ensure they do not significantly affect our predictions
6. There is a moderate to strong positive correlation between the "engine capacity" and "CO2 emissions." The correlation with other features will be explored further in the analysis.
7. The preliminary EDA has shown that the dataset is suitable for our research questions and for building a predictive model for CO2 emissions.

Supervised Task:

We want to construct a machine learning model that forecasts car CO2 emissions using real-time telemetry. Supervised CO2 emission prediction using telematics data. And Project tasks include dataset cleaning, missing value management, and outlier treatment. Engineering features: Add needed features. A regression model can forecast CO2 emissions. Select and train dataset models. Model evaluation should employ regression metrics (MAE, MSE, R-squared). Interpretation is Calculate vehicle factors' CO2 emissions using models. Get the best CO2 emission model. Write a complete project report including findings. Show, assess, and propose model performance.

Conclusion:

Our research may boost transportation emissions research. Real-time telematics data may be used to estimate vehicle CO2 emissions using machine learning. We are excited to use our methodology to reduce transportation CO2 emissions. This study improves CO2 emission projections to address an environmental concern. We will evaluate and model a credible dataset to give environmental academics, regulators, and the automotive industry with important information.