DS 3000 Project: Predicting the Impact of Individualized Car Usage on Climate Change

By: Alyssa Benjamin, Anoushka Bhatia, Jai Gollapudi, Monisha Kapadia

ABSTRACT

This project aims to employ ML techniques to analyze the impact of individual car usage in the U.S. on pollution and the environment. By examining carbon emissions data, we will identify key factors affecting greenhouse gas emissions from vehicles. Using regression techniques, we will reveal trends in emissions and assess policy effectiveness. Our findings can guide legislation, policy, and individual actions to mitigate the negative effects of car pollution.

INTRODUCTION

Definition of the Problem

According to the United States Environmental Protection Agency (EPA), the transportation sector is responsible for the largest share of greenhouse gas emissions in the United States, accounting for about 27% of the total emissions in 2020.

The problem of car pollution is complex, as it involves not only the environmental impact of emissions but also the socioeconomic factors that contribute to car dependency and the lack of viable alternatives in many areas.

Overall, this project seeks to shed light on the negative consequences of heavy car usage and traffic congestion in America and to provide insights that can inform policies and individual behavior changes to reduce these impacts.

Motivation & Objectives

As members of Generation Z, we have all grown up with the warning of climate change looming over our heads. The U.S., in particular, heavily relies on cars for transportation, over alternatives such as public transportation, biking, and walking. As mere college students, we cannot control how much pollution/harmful gas corporations omit. Therefore, we chose the topic of car pollution because we wish to live in a world where the air & nature's environment is healthy.

Another motivation for picking this topic is because we knew there would be robust, reliable data sets that we could choose from to conduct our analysis. Environmental datasets are easy to obtain from environmental agencies/government websites.

The objectives for our project were to:

1. Analyze relationships between vehicle features and their carbon dioxide ouput
2. Be able to predict a car's output by leveraging Machine Learning regression methods
3. Raise awareness about the environmental impact of cars and the importance of reducing carbon emissions

RELATED WORKS

By building upon the work of many researchers who have explored this topic, we aim to offer new and valuable insights to the existing body of knowledge. Listed below are a few analyses similar to ours:

1. CO2 Emission by Vehicles (Kaggle)
2. Car Fuel & Emissions (Kaggle)
3. International Greenhouse Gas Emissions (Kaggle)

METHODOLOGY

Data Acquisition

For our project, we utilized a dataset from the United States Department of Energy. It entails fuel economy data on a tested vehicles at the EPA from 1984 to 2024.

Data Cleaning & Exploration

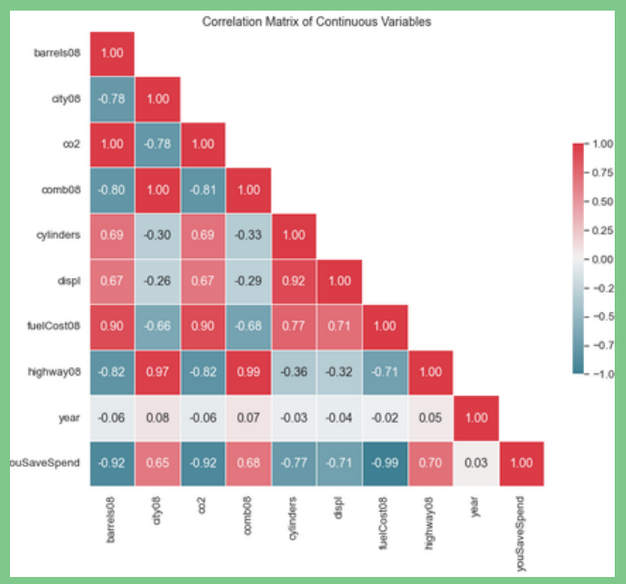
To better prepare the data for our analysis, we assigned the appropriate data types, performed mean and mode imputation on numerical and categorical columns respectively, and removed null values and columns outside the scope of our project. The features we choose to focus on were:

- **car make and model**
- **the year the model was released**
- **the EPA determined vehicle class size**
- **the fuel type used by the vehicle**
- **miles per gallon**
- **the annual fuel cost**
- **the number of cylinders**
- **their EPA determined greenhouse gas score,**
- **engine displacement**
- **the drive axle type**
- **the transmission type**
- **the savings vs spending cost of a car over a 5 year span,**
- **the annual petroleum consumption of the fuel type,**
- **the EPA composite gasoline-electricity and highway MPGe for hybrid plug ins.**

To ensure that we chose the most relevant features for our ML models, we also considered the domain knowledge of environmental science. By combining our analysis of the correlation matrix with expert knowledge, we were able to identify the most relevant features and build models that can accurately predict CO2 emissions. This approach helped us to build a more robust and reliable model that can be used to support efforts to reduce carbon emissions from vehicles.

In our EDA, we also explored different exploratory questions such as "What is the average CO2 emissions per vehicle class?" and "What are the top 10 Car Models with highest CO2 Emissions?"

CORRELATION MATRIX



RESULTS & EVALUATION

Our project's goal was to better understand how individual car usage affects climate change, particularly in terms of CO2 emissions.We compared 3 different ML algorithms: RandomForestRegressor, Support Vector Regression, and GradientBoostingRegressor. We adjusted each model's hyper-parameters to find the best possible performance, and then we tested how well they made predictions on new data. Here's how each models performed:

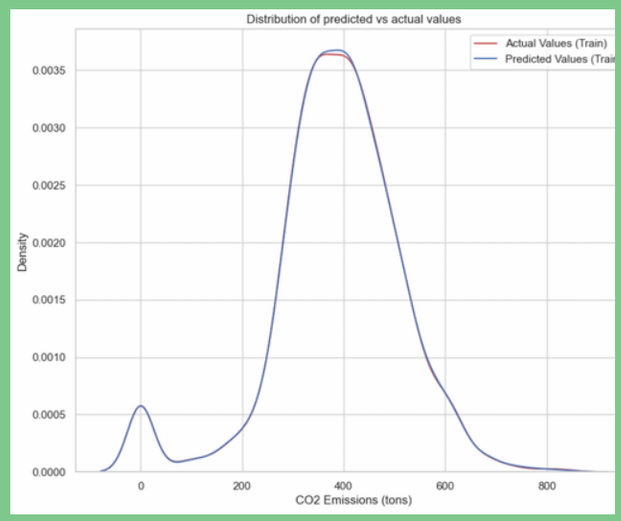
Model	R^2 (approx)	Mean Square Error (approx)
Random Forest Regressor	0.9988	19.5249
Support Vector Regressor	0.9977	36.9387
Gradient Boosting Regressor	0.9985	24.1121

Hyper-parameters: n_estimators : [10, 50], max_depth : [5, 10]
Cross validation: 5 (GridSearchCV)

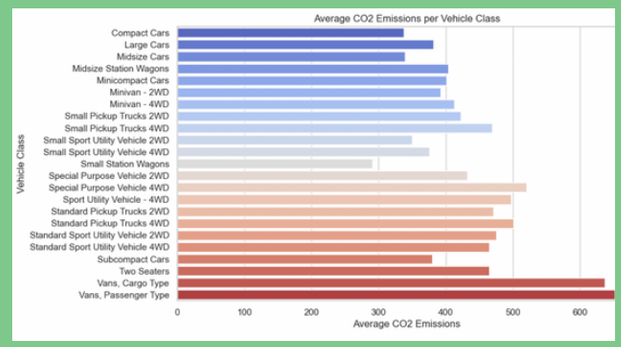
The RandomForestRegressor model performed the best, with the lowest MSE and the highest R2, meaning it can explain 99.88% of the variation in CO2 emissions based on features affecting individual car usage.

In conclusion, we recommend using the RandomForestRegressor model to predict the individualized impact of CO2 emissions from vehicles. This model will help us better understand the effects of car usage on climate change and make more informed decisions about reducing our carbon footprint.

DIST. OF ACTUAL VS PRED VALUES



C02 EMMISSIONS PER VEHICLE CLASS



IMPACTS

Transportation, specifically cars, contribute significantly to climate change, with direct and indirect impacts on the environment and human society. As the global population rises, the number of cars on the road is expected to increase, **exacerbating the issue**. Cars contribute to air pollution and depletion of the ozone layer, as well as the depletion of finite resources such as oil and the destruction of ecosystems. Traffic congestion worsens the problem, increasing emissions of pollutants and negatively impacting public health. The World Health Organization estimates that air pollution is responsible for 6.7 million premature deaths worldwide each year, with traffic-related pollution being a major contributor. Moreover, traffic congestion has significant economic costs, estimated at $166 billion in the United States in 2021. To address car pollution, a multi-faceted approach is needed, including promoting public transportation, encouraging alternative modes of transportation, and increasing the availability of electric and hybrid vehicles. **The project's findings can inform policies and individual behavior changes to reduce the impacts of heavy car usage and traffic congestion, ultimately contributing to a more sustainable future.**

CONCLUSION

In conclusion, our project has demonstrated the potential of machine learning techniques to assess and predict the individualized impact of CO2 emissions from vehicles in the United States. By comparing three different regression algorithms – RandomForestRegressor, Support Vector Regression, and GradientBoostingRegressor – we found that the RandomForestRegressor model performed the best, with the lowest MSE and the highest R2 score. This model can explain 99.88% of the variation in CO2 emissions based on features affecting individual car usage.

Utilizing this RandomForestRegressor model, we can better understand the effects of car usage on climate change and make more informed decisions on policies and individual behavior changes to reduce our carbon footprint. Our findings highlight the importance of promoting public transportation, encouraging alternative modes of transportation, and increasing the availability of electric and hybrid vehicles to address the complex issue of car pollution.

By informing future legislation, policy, and individual actions, this project contributes to the global effort to mitigate the adverse effects of car pollution on the environment and human health. The insights gained can help guide the transition towards a more sustainable and eco-friendly transportation system, ultimately contributing to a cleaner, greener future for all.

REFERENCES

Dataset :
https://www.fueleconomy.gov/feg/download.shtml
https://www.fueleconomy.gov/feg/ws/index.shtml#vehicle
Emmissions Data and Statistics:
https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions#transportation

https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

https://static.tti.tamu.edu/tti.tamu.edu/documents/mobility-report-2021.pdf