|  |  |  |
| --- | --- | --- |
| A book with a flag and a car on it  Description automatically generated Optimizing Vehicle Choices: A Data-Driven Approach to Fuel Consumption and CO2 Emissions in Canada | | |
|  |  |  |
| Project PLAN | | |
| Group-5  Algorithm Avengers | | | |

|  |  |  |
| --- | --- | --- |
| A car driving on a road with a city in the background  Description automatically generated | |  | | --- | | TABLE OF CONTENTS |   [Problem / Analytical Statement 2](#_Toc352468337)  [1. PROJECT SIGNIFICANCE 3](#_Toc1584413838)  [2. ANTICIPATED OUTCOMES 4](#_Toc168374433)  [3. COURSE RELEVANCE 5](#_Toc233568995)  [4. ASSUMPTIONS 5](#_Toc1080677264)  [5. CONSTRAINtS 6](#_Toc1510190974)  [Timeline and Deliverables 7](#_Toc1394548837) |

|  |
| --- |
| Problem / Analytical Statement |

**Updated Problem Statement**

The automotive industry continues to face increasing pressure to enhance fuel efficiency and minimize environmental impact. However, consumers still lack comprehensive insights into the long-term financial and carbon tax implications of their vehicle choices. While existing tools like **EnerGuide for Vehicles** and the **Fuel Consumption Ratings Tool** provide basic fuel efficiency data, they fail to offer predictive cost analysis and personalized recommendations based on real-world factors.

To bridge this gap, our project will advance into a **machine learning-powered dashboard** capable of estimating **annual fuel costs, carbon tax impact, and fuel efficiency scores** with greater precision. Instead of classification-based evaluation metrics, we will assess model performance using **regression metrics**, including **R²** (Coefficient of Determination)**,** Mean Squared Error **(MSE), and** Root Mean Squared Error **(RMSE)** to ensure robust, reliable and data-driven predictions. This approach will provide data-driven insights that empower consumers to make more sustainable and cost-effective vehicle choices based on quantifiable financial and environmental impacts.

**Justification**

The lack of predictive cost analysis in existing fuel efficiency tools leaves consumers without the necessary insights to make informed decisions. Our machine learning approach fills this gap by dynamically estimating fuel costs, emissions, and carbon tax implications, offering a personalized, data-driven decision-making tool. Since our dataset involves continuous numerical predictions, regression metrics like **R², MSE, and RMSE** are more appropriate than classification-based accuracy, ensuring a more precise and reliable model evaluation.

R² helps explain how well our model accounts for variance in fuel costs, while MSE and RMSEminimize prediction errors, improving the trustworthiness of cost analysis. Ultimately, our dashboard empowers consumers with transparent financial and environmental insights, guiding them toward fuel-efficient and eco-friendly vehicle choices that align with both economic savings and sustainability goals.

|  |
| --- |
| PROJECT SIGNIFICANCE |

This research is important because it addresses a crucial problem that affects consumers and regulatory agencies alike: CO2 emissions and car fuel economy. The initiative intends to optimize carbon tax estimates, improve government regulations on fuel efficiency standards, and give consumers relevant insights by utilizing analysis of data and machine learning.

In addition to saving money for both individuals and businesses, lowering fuel use and CO2 emissions promotes environmental sustainability. The project will show how data-driven approaches may be used practically to enhance decision-making in the automobile industry. It will also give manufacturers useful information to improve car designs and make sure they comply with regulations. This project promotes greener mobility options by bridging the gap between access to data and actionable insight through the integration of predictive analytics into an enhanced EnerGuide dashboard.

|  |
| --- |
| ANTICIPATED OUTCOMES |

#### **1. On Empowering Consumers with Data-Driven Insights**

* The Consumers will also receive **personalized estimates** such as fuel costs, emissions, and carbon taxes, helping them make informed vehicle choices.
* The dashboard will be **real-time comparisons** of different vehicle models.

#### **2. Enhanced Accuracy in Cost and Emissions Predictions**

* The use of **regression-based machine learning models** will enable **more precise predictions** of fuel expenses and emissions.
* By integrating **real-world factors** such as fuel price changes, driving patterns, and regional tax rates, the tool will provide **more accurate insights** than conventional static tools.

#### **3. Supporting Policy Development with Improved Carbon Tax Estimates**

* The system will offer dynamic projections for carbon tax costs, assisting policymakers in refining regulations.
* Insights from the dashboard can help governments **improve fuel efficiency policies** and introduce incentives for eco-friendly vehicle adoption.

#### **4. Helping Automakers Improve Vehicle Efficiency**

* Automakers will gain access to **detailed analytics** on vehicle efficiency, enabling improvements in design and compliance with emission standards.
* This tool may also highlight **consumer trends** related to fuel economy, guiding manufacturers in developing more sustainable vehicles.

#### **5. Real-World Application of Machine Learning in Automotive Analysis**

* This project will also explain **how predictive analytics can be used for decision-making** in the automotive industry.
* By converting **raw data into actionable insights**, it will also showcase the **practical benefits of AI-driven solutions** for fuel efficiency analysis.

#### **6. Encouraging Environmentally Friendly Transportation Choices**

* The dashboard will also promote fuel efficiency and sustainability of fuel efficiency and sustainability, helping reduce carbon footprints.
* By providing **clear cost-benefit insights,** Providing clear cost-benefit insights will encourage consumers to choose fuel-efficient vehicles, contributing to a greener future.

|  |
| --- |
| Course RELEVANCE |

This project aligns with multiple courses we’ve taken, including:

* **DATA 1200: Introduction to Data Analytics** (for the development of the model, analyzing the model, and its evaluation using metrics like RMSE, R², and MSE).
* **DATA 1201: Data Collection** (for cleaning and pre-processing the selected dataset before analysis and implementation).
* **DATA 1204: Statistics I** (to ensure the statistical significance of the model’s results and conduct exploratory data analysis).
* **DATA 2204: Statistics II** (for developing algorithms for evaluation, analyzing vehicle efficiency trends, and performing feature selection and one-hot encoding).
* **DATA 2205: Visualizations, Leadership** (for the implementation of an interactivedashboard showcasing vehicle fuel efficiency, CO₂ emissions, and carbon tax impact).

We will apply tools and techniques such as Python for model implementation using scikit-learn, regression analysis, feature engineering, and data visualization techniques to derive insights for consumers, policymakers, and manufacturers.

|  |
| --- |
| ASSUMPTIONS |

1. **Data Accuracy:** The dataset offers precise and trustworthy statistics on fuel usage and vehicle emissions from 2015 to 2025. We can rely on its accuracy because information originates from the open database of the Government of Canada. This guarantees that the models and suggestions we provide are grounded on actual data, which makes them trustworthy and pertinent.
2. **Stable Regulatory Policies:** To guarantee precise tax estimates, current and anticipated carbon tax rates and emission rules stay constant throughout the project's duration.
3. **Consistency in Consumer Behavior**: Throughout the study period, we assumed that trends in fuel consumption, driving behaviors, and car purchase patterns are constant.
4. **Standardized Driving Distances:** The annual distance traveled used to calculate the carbon tax is the same for all cars to provide fair comparisons. All the annual computations will be predicted on the premise that the average annual driving distance is 20,000 kilometers.
5. **Data Privacy and Compliance:** The dataset complies with ethical and regulatory requirements for privacy and data use. It is made publicly available via a government-run website, guaranteeing that it is usable by the public and conforms with all applicable data protection laws.
6. **Cost of Fuel:** We are assuming that fuel prices will remain constant throughout the project analysis. As of February 3, 2025, according to Ontario.ca, the fuel prices in Canada are as follows (prices in cents per liter):

|  |  |  |
| --- | --- | --- |
| **Fuel Type** | **Code** | **Ontario (¢/L)** |
| **Diesel** | D | 170.2 |
| **E85** | E | 160.1 |
| **Regular Gasoline** | X | 154.9 |
| **Premium Gasoline** | Z | 184.1 |

*Referred from the official Ontario Motor Fuel Prices website.*

*Link: <https://www.ontario.ca/motor-fuel-prices/>*

|  |
| --- |
| Constraints |

1. **Fuel Price Fluctuations**
   * **Type:** Soft constraint
   * **Assumption:** Market Stability
   * **Description:** The fuel prices used in the calculations are based on current government estimates, which may change due to market conditions, impacting cost-related analysis.
2. **Regulatory Constraints**
   * **Type:** Hard constraint
   * **Assumption:** Compliance with Canadian Standards
   * **Description:** The model must adhere to Canadian environmental and fuel efficiency regulations, ensuring realistic CO₂ emissions and fuel consumption calculations.
3. **Feature Engineering Limitations**
   * **Type:** Soft constraint
   * **Assumption:** Data Consistency
   * **Description:** The calculation of derived metrics like Fuel Efficiency Score, Annual Carbon Tax, and Fuel Cost per Passenger-Kilometer depends on predefined formulas and assumptions, which may not cover all real-world variations.
4. **Data Access**

* **Type:** Hard constraint
* **Assumption:** Data Completeness
* **Description:** The accuracy and reliability of the model are dependent on the quality and sufficiency of the dataset. If data is missing or biased, the model's predictions may not generalize well.

|  |
| --- |
| Timeline and Deliverables |

|  |  |  |
| --- | --- | --- |
| **DELIVERABLE** | **DUE DATE** | **DESCRIPTION** |
| **Project Plan** | January 31, 2025 | Initial problem statement, methodology, and expected outcomes. |
| **Data Collection and Cleaning** | February 19, 2025 | Gather, clean, and prepare the dataset for analysis. |
| **Preliminary Model** | March 18, 2025 | Build and test a preliminary model that predicts fuel costs, emissions, and carbon taxes. |
| **Final Model** | April 4, 2025 | Create a final model for fuel costs, emissions, and carbon taxes prediction. |
| **Final Report** | April 12, 2025 | Create a final report that includes the final model, analysis, and actionable insights. |
| **Final Presentation** | Week of Apr 9th and 16th | Present the final model, analysis, and actionable insights |