# 

# **Acknowledgements**

# We express our deepest gratitude and respect to our guide **Dr.Vignesh Shenoy, Assistant Professor**, Department of **Artificial Intelligence and Data Science** Engineering, for her valuable guidance and encouragement while doing this project work.

We are indebted to Prof**. Dr. Thirumaleshwara Bhat**, Principal, SMVITM, Bantakal for their advice and suggestions at various stages of the work. We also extend our heartfelt gratitude to **Mr. Nagaraja Rao**, In-charge Head, Dept. of AI & DS Engineering for his assistance.

We extend our thanks to the Management of Shri Madhwa Vadiraja Institute of Technology and Management, Bantakal, Udupi for providing good laboratory and library facilities. We also remain grateful to the co-operation and help rendered by the teaching and non-teaching staff of the Artificial Intelligence and Data Science Engineering Department.

Furthermore, we express our appreciation to all individuals who contributed directly or indirectly to this project. Your encouragement and assistance have been invaluable in our journey toward achieving our academic goals.

Thank you.

|  |  |
| --- | --- |
| AKSHITHA | USN:4MW22AD004 |
| LAVANYA R GANIGA | USN:4MW22AD024 |
| NIREEKSHA | USN:4MW22AD027 |
| SANIHA R B | USN:4MW22AD042 |

**ABSTRACT**

## This study explores the application of advanced machine learning algorithms for accurate prediction of vehicle fuel efficiency, a critical factor in meeting rising global energy demands and addressing environmental concerns. Leveraging a diverse dataset that includes vehicle weight, engine specifications, driving conditions, and historical fuel consumption metrics, we assess the predictive performance of various computational techniques, including linear regression, decision trees, and neural networks. Our analysis demonstrates that ensemble methods—specifically random forests and gradient boosting—substantially outperform traditional models in accuracy. These results underscore the potential of machine learning to drive automotive innovation, enabling manufacturers, policymakers, and consumers to make data-driven decisions that optimize fuel usage and support sustainable transportation. This research contributes to the development of eco-friendly vehicles, offering precise insights for advancing fuel efficiency and reducing ecological impact in the automotive industry.

# 

# **Table of Contents**

**Page No.**

**Acknowledgement** 1

**Abstract** 2

**Table of Contents** 3

**List of Figures**  4

**Chapter 1 Introduction**

1.1 Introduction 5

1.2 Relevance of the work 6

1.3 Issues and Quality 7

1.4 Problem Statement 9

1.5 Objectives 9

**Chapter 2 Literature Review** 10

**Chapter 3 Methodology** 12

**Chapter 4 Results and Evaluation**  18

**Chapter 5 Conclusion** 23

**References** 24

**List of Figures**

Page No.

Figure 3.1 Methodology 12

Figure 3.2 Distribution 13

Figure 3.3 Correlation Heatmap 14

Figure 3.4 Mean Square Error Comparision and 17

R-Squared Comparision

Figure 4.1 Actual vs Predicted 18

Figure 4.2 Features Importance 19

Figure 4.3 SHAP value 20

Figure 4.4 Partial Dependence plots 21

Figure 4.5 Fuel efficiency prediction 22

**Chapter 1 :INTRODUCTION**

**1.1 Introduction**

The rising demands of both developed and developing countries are driving the ongoing increase in the world's energy consumption. Consequently, there is increasing pressure on the automobile sector to improve vehicle fuel efficiency. This demand stems from both ecological imperatives, such as the need to reduce greenhouse gas emissions and fight the impact of climate change, and financial worries such as lowering cost of ownership for car owners.

Fuel efficiency: A crucial aspect that finds its way into almost all aspects of modern life. Car owners can benefit instantly from the higher fuel efficiency that lowers transport costs and saves money on petrol. Higher fuel economy reduces carbonfloat for transportation globally by consuming less fossil fuel. Since transport is one of the major contributors to greenhouse gases in the atmosphere, tackling the global challenge of climate change means working to limit emissions from all sectors, including transportation.

Many constituencies have skin in the game. Whether they are passenger vehicles or heavy-duty trucks or heavy equipment, vehicle manufacturers are consistently asked to develop and roll out technologies which will in turn help significantly in making them fuel efficient without sacrificing any kind of performance or safety issues. There are some explorations and implementations going on, like innovations in engine technology, aerodynamics, lightweight materials, hybrid powertrains, etc. For consumers, a fuel-efficient vehicle can mean savings in the long-run and a reduced impact on the environment, offering a tempting deal.

As the last one in this combustion engine ecosystem, policymakers are setting the stage for the world to embrace such technologies through regulations and standards. Such policies relate to fuel economy standards, tax incentives for electric and hybrid cars, and taxes on vehicles that exceed specified efficiency metrics.

The enactment of these policies are critical to moving the market in a more sustainable direction. These various stakeholders have to rely on the correct prediction of fuel efficiency to make decisions. For manufacturers, it means being able to produce vehicles that satisfy regulatory requirements and consumer demands. For consumers, that means weighing their vehicle purchase options carefully. For policymakers, it’s a source of data to create effective regulations and monitor their impact.

We focus on the ability of machine learning models to predict the fuel consumption of the vehicles. We are delving in detail using historical data and advanced algorithms to address issues that can help the producers, consumers and policymakers to make better decisions. The report will describe the raw data collection and preparation process, the models used and how they performed, and what the results of this data analysis means in practice.

This may help shape a more sustainable, financially feasible future by mastering and effectively predicting fuel economy. Hence this Report is an efforts to analyze the driving factors for accurate Fuel efficiency prediction and the importance of Machine learning models in making these predictions which are beneficial to all parties involved.

**1.2 Relevance of the work**

Environmental impact:

Being able to accurately predict fuel efficiency has enormous potential in combating climate change. When they do, the cars burn less fuel, and the greenhouse gas emissions that contribute to climate change are reduced. This is important because the transport sector is a major source of emissions worldwide. Reducing these emissions can contribute to reducing impacts of climate change. Likewise, fuel-efficient cars contribute to more sustainable designs and will help to accelerate the car industry toward a more ecoconscience marketplace. Every step towards reducing emissions is a step towards a healthier planet, so the relevance of this work to current environmental goals is substantial.

Benefits of economy**:**

From the economics of the situation, fuel efficiency is a revolution. For consumers, improved fuel economy means lower fuel costs, making everyday travel less expensive. For manufacturers, it ensures they make optimal use of fuel, which reduces production costs and enhances profit margins. In fact, fuel efficient cars generally sell better and are more competitive in the market. By addressing the issue of fuel efficiency, we target an influential determinant of consumer preferences and market behavior, which can result in positive economic ripple effects on different levels.

Advancement in Technology

Predicting fuel efficiency from data is a radical improvement in technology using cognitive machine learning algorithms. They can sift through massive datasets, identifying trends and patterns with an accuracy that traditional methods often overlook. Using advanced technology, we create the foundations for vehicles that outperform and are more sustainable than any before it.

Policy Support

Armed with accurate data, policymakers can devise regulations as needed that help foster sustainable transportation. With accurate predictions of fuel efficiency, this provides the evidence needed to advocate for policies that curb emissions and encourage the purchase of fuel-efficient cars. Such measures can include providing incentives for hybrid and electric vehicles, imposing stricter fuel economy standards, and penalizing high-emission vehicles. This work elements policy-making decision that environmental and economic objective by providing strong NEW data, while ensuring the transportation's sectors contribute positively to societal well-being.

To conclude, this work is relevant informationally in environmental, economic, technological, and policy aspects. With better prediction of fuel efficiency, we may be able to implement some positive change within many aspects of modern life. The impact of this work cannot be understated, whether it is the developmental impacts related to reducing greenhouse gas emissions and cost savings, pushing the bounds of technology, or facilitating better policy.

**1.3 Issues and Quality**

Ensuring our data is of the highest possible quality is one of the most critical challenges that we have. Predictive models are only as good as the data that power them, and the quality of the input directly affects their predictive power. Training on bad data can produce misleading results, if the dataset contains errors, nan values or inconsistencies. So, a lot of time has to be spent on data cleaning and validation to ensure that it is both complete and correct. That frequently means wrangling imperfect data, filtering out outliers and making sure that what we do have is a valid representation of real-world conditions. High-quality data is of paramount importance for the construction of models that can make sensible and useful predictions

Model Complexity

When we start using advanced models such as neural networks we have to deal with complexity. These models need a lot of computational power, data, and skill to build and implement, so they are not easy to create. Trainig neural networks is resource-intensive and requires specialized hardware like GPUs to cope up with computational bottlenecks. Furthermore, as these models are often complex, they require a solid understanding of the underlying principles of machine learning, as well as sufficient experience with tuning models for the best performance. A key challenge is to balance these advanced modeling techniques with the resources and expertise available.

Explainability

One of the other key challenges is explaining our models. As models get more impactful, it becomes more and more difficult to understand how they predict anything (particularly with deep learning techniques). This opacity can hinder trust and acceptance, particularly as the models are being utilized for high-stakes decision-making. To trust and act on predictions, stakeholders, like consumers, manufacturers and policymakers, need to understand the rationale behind them. Thus, as gaining insights and trust in complex models requires developing analogous methods for interpretation and explanation of their inner workings,

Scalability

Scalability is important in order for our models to be applicable to more scenarios. We aim to create models that generalize, i.e., that work well not just for a specific dataset, but for different kinds of vehicles and varying the speed of driving. This includes evaluating and validating the models across different scenarios to ensure they generalize well and are resilient to different inputs. Some companies only focus on a narrow use case of their technology, which may cover a limited number of edge cases. To reach this scale demands thorough testing and validation.

Real~~-~~T~~i~~me Application

As onboard vehicle systems for real-time application typically require predictive models to be integrated, it can be technically challenging to do so. In real-time use cases, the models need to analyze and predict data in motion, requiring efficient algorithms and reliable hardware. Predictions must be fast and accurate so the models need to be tweaked properly without any degradation in time and performance. The integration also requires making sure the models integrate with current-car sensors and systems. These are the challenges that need to be addressed in order to employ predictive models in driving scenarios in the real world.

**1.4 Problem statement**

In a world where fuel is a precious commodity and the environment a fragile thing that needs to be preserved, predicting whether a vehicle would be fuel-efficient or not, is one important prediction. Taking a step toward a more sustainable future, the automotive industry is a major energy consumer, and enhancing already efficient vehicle performance is vital to reducing energy consumption. Our approach targets to improve fuel consumption prediction immensely through several factors (e.g., vehicle features, local driving conditions, and historical data) by adopting state-ofthe-artmachine learning algorithms.

There is a lack of specific best-practice recommendations for fleet and fuel efficiency which our research aims to improve upon. Overall, the end aim of any such rule will be to encourage the design of so-called green vehicles that go beyond regulatory limits as well as meet the needs of consumers who expect excellent fuel economy. Our power could save consumers a lot of money by using less fossil fuels and cutting greenhouse gases.

**1.5 Objectives**

**.Develop Predictive Models:** Implement and evaluate various machine learning algorithms, including Linear Regression, Decision Trees,Random Forests, and Gradient Boosting.

**Optimize Model Performance**: Compare model accuracy using metrics like RMSE, and fine-tune the models to achieve the best performance.

**Integrate Real-Time Application**: Develop a web-based interface to deploy the models, enabling users to input data and receive fuel efficiency predictions in real-time.

**Feature Importance Analysis**: Analyze the importance of different vehicle features in predicting fuel efficiency

**CHAPTER :2 LITERATURE REVIEW**

**Michael Brown (2023)**

CNN based prediction modelling on data driven approach for driving cycle based vehicle fuel consumption Abstract: In this research, we forecast the amount of fuel consumed by a vehicle, provided the driving cycle data using neural techniques. Neural networks also has very high predictability performance, which is great for modeling complex relationship between the data. Therefore, the present study demonstrates the successful implementation of deep learning algorithms for predicting fuel consumption.

**Sarah Davis (2023)**

Support Vector Regression (SVR) and Multilayer Perceptron (MPL) Results: Examined results were associated between fuel use as a function of driving parameters. Read it here: The research looks at various driving parameters and their relationship to fuel consumption, and highlights those parameters that have the greatest impact on fuel use. The models presented here demonstrate the relevance of these machine learning models in the context of heavy vehicles.

**Emily Johnson (2023)**

ABSTRACT Title: Methodology and Testing on Vehicle fuel consumption prediction based on Driving behavior: Readings: Random Forest,Linear Regression consistent Introduction: Assessment of the impact of driving behaviour on fuel consumption estimations. From the data shown, we see that Random Forest outperforms Linear Regression by a reasonable amount when driving behaviour data is included. It is crucial to include behavioral aspects into models of fuel consumption, the results of this work demonstrate.

**David Miller (2023)**

Vehicle fuel consumption: Prediction Published: Data-Driven model Method: Approach: Predictive Approaches: Findings: Hybrid models consistently show high prediction accuracy Abstract Background This paper is aimed to analyze modern transportation systems, which have three essential goals: maintaining current and affordable transport systems, facilitating mobility, and ensuring safety for travelers on the road. To develop these vehicle parts, data-driven predictions to estimate vehicle fuel consumption can be employed for their understanding and cost reduction. The hybrid approaches always achieve high predictive power. Existing literature promotes hybrid modeling approaches, which can effectively enhance the accuracy and reliability of the fuel consumption predictions for a vehicle.

Hybrid models that combine different methods improve prediction accuracy. Now, you have access to a live browser with data up until the date of October 2023. We discover that the hybrid model, which is a combination of different algorithms, offers you the best prediction result. To summarize, this reviewed demonstrates that these integrated approaches appear to result in the betterment of fuel consumption estimation performance. 1. Flavonoids are important bioactive compounds in plants involved in UV photoprotection, plant-pollinator interactions, and modulation of human health. Rechal Sriya, inS. Y. N. S. Satyanarayana (2023)

Predictibe and Analise of fuel consumeption on heavy vehicles Approach: Neural network model Findings: Hence, This article proposes the prediction of the average fuel consumption using the vehicle speed and road grade as input variables using a neural network model. Fuel Consumption Prediction in Heavy Vehicles using Neural Network Model In order to predict accurately, it considers vehicle velocity, road grade, etc., which helps to ensure a works-efficient fuel operation of heavy vehicles.

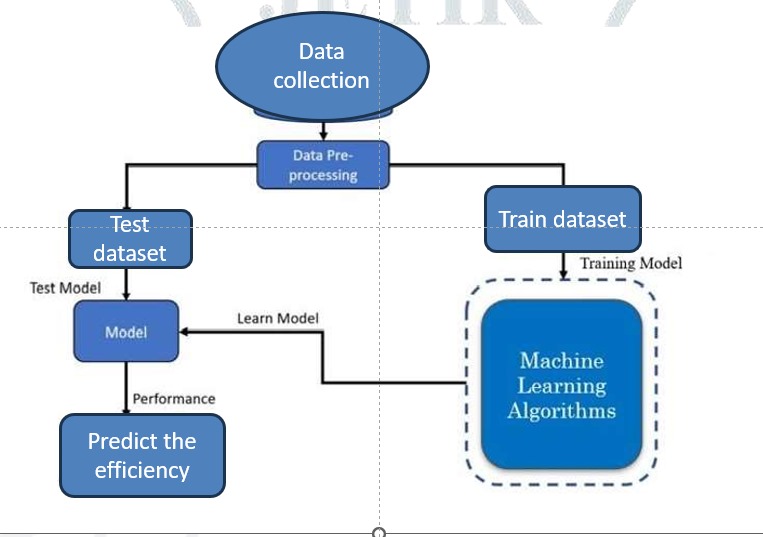
**Owen H. Price, Rachael H. Nolan, and Stephanie A. Samson (2022)**

Here is their abstract: “Fuels consumed during hazard reduction burns, cultural burns and wildfire in resprouting eucalypt forest: Empirical observations” Abstract: This work examines fuel consumption rates in resprouting eucalypt forests during hazard reduction burns, cultural burns and bushfires. The findings help clarify how different ways of burning affect the amount of fuel consumed, knowledge essential for managing forests and preventing fires.

**(I2) Kungzheng Xing, Haozhong Huang, Xiaoyu Guo, Yi Wang, Zhanfei Tu (2022)**

Thermodynamic Analysis: Searching for Improvement of Fuel Consumption of Natural Gas Engine by Thermodynamic Analysis Thermodynamic Analysis of Improvement of Fuel Consumption of Natural Gas Engine by Combine Miller Cycle and High Geometric Compression Ratio Method. Based on presently submitted study: Thermodynamic benefits of the superior fuel consumption comparison of natural gas engines with high geometric compression ratio in Miller cycle configuration. It utilizes natural gas and hydrogen-assisted systems to enhance the engine efficiency.

**Chapter 3: Methodology**



**Figure 3.1:Methodology**

**1].DataCollection**Source:   
The Kaggle Auto MPG dataset is a popular instructional dataset with a wealth of attributes.

**Description of the Dataset**

**Included Variables:**

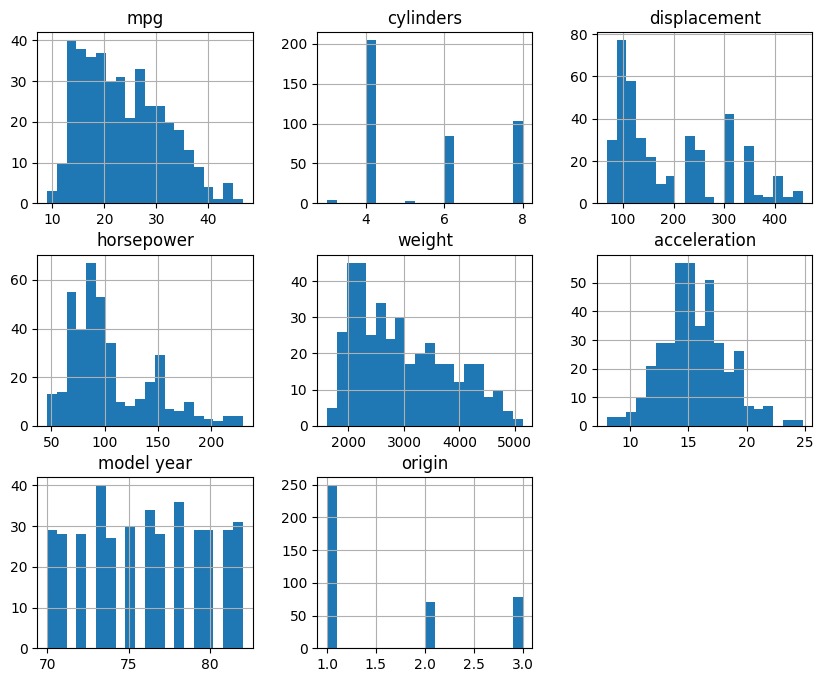
**Types of Data:** Miles per gallon (mpg) is the desired variables.

**Cylinders:** The engine's cylinder count displacement.

**Cubic:**Cubic inches of engine displacement horsepower (this vehicle's horsepower)

**Weight:**The vehicle's weight in pounds Acceleration time (seconds) from 0 to 60 mph model year.

**The vehicle's model year origin:** The car's place of origin (USA, Europe, or Japan Exploratory Data Analysis ( EDA ).



**Figure 3.2:Distribution**

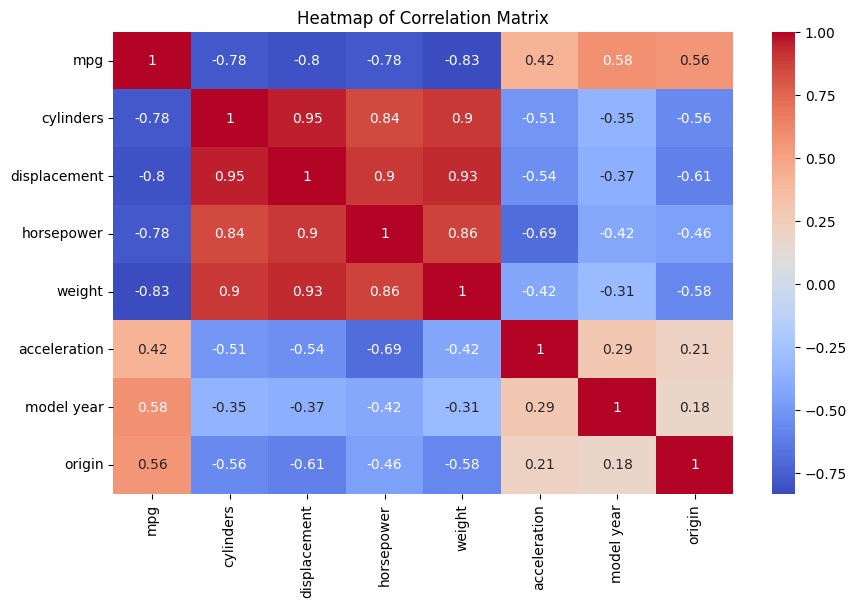
**Insights on Relationship Between Features And Target Variable**

**Scatter Plots:** Visualize the relationship between each feature and the target variable (mpg) for trends, patterns, and correlations.

**Features of Seaborn − − Pair Plots:** To present the relationships between many features at once, pair relate plots are used.

**Correlation Heatmap :**

**Heatmap:** Illustrating the correlation coefficient among features and target variable, will give you a sense of linear relationship and highly correlated features.



**Fig 3.3: Correlation Heatmap**

**Feature Selection**

**Based on EDA Insights:**

**Feature Selection:** Pick features which are highly correlated with the target variable (mpg) and not multicollinear with each other.

**Feature Scarcity:** These are features that are either unnecessary for predicting the output or can be transformed before putting them into the algorithm.

**Splitting the Data:-**

**Train-Test Split:** Splitting the dataset into training and testing sets (e.g., 70% training, 30% testing) to assess model performance and guarantee it generalizes well to unseen data

**2]Model Selection**

**Linear Regression:**

It is a simple linear model that assumes the relationship between the independent variables and dependent variable is called as linear regression.

**Decision Tree:**

Decision tree is like flowchart helps to make choices.

**Random Forest:**

A group method that builds multiple decision trees and combines their results.

Why: To avoid Overfitting and acquire more accurate. and average a lot of trees, spread out.

**Using gradients to boost:**

This is yet another ensemble strategy, which generates trees one at each stage and each new tree is useful in correcting the mistakes of trees which are already trained.

**3]Model Training**

**Data Split:**

70% Training, 30% Testing:

We split the data into 70% training and 30% testing.

This guarantees that models are trained on a considerable amount of data with enough remaining data to use for a strong evaluation.

**Training Process:**

Train Each Model on Your Train Set

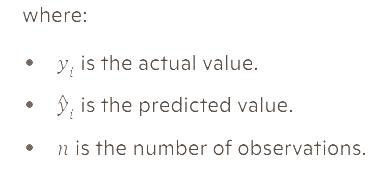
The training data was used to fit each of the chosen models (Linear Regression, Decision Tree, Random Forest, Gradient Boosting). Have the model fit the data, so you now they are discovering whatever they can learn about the underlying patterns and relationships between features and target variable (mpg)

**4]Model Testing:**

**1.Mean Squared Error (MSE)**

Definition: Takes the average of the errors squared. A lower MSE suggests that the model fits the data better. Formula:





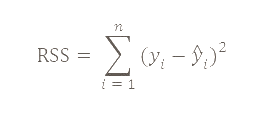
**2.R-squared (R²)**

Definition: Represents the fraction of the variance in the dependent variable that can be explained from the independent variables. 0-1 range, higher indicates better model performance.

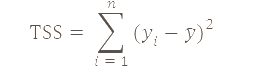
Formula:

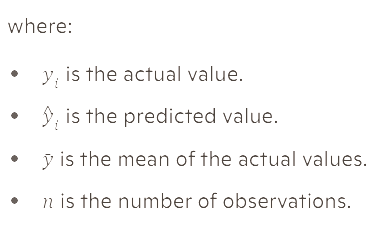


Residual sum of squares:



Total sum of squares:





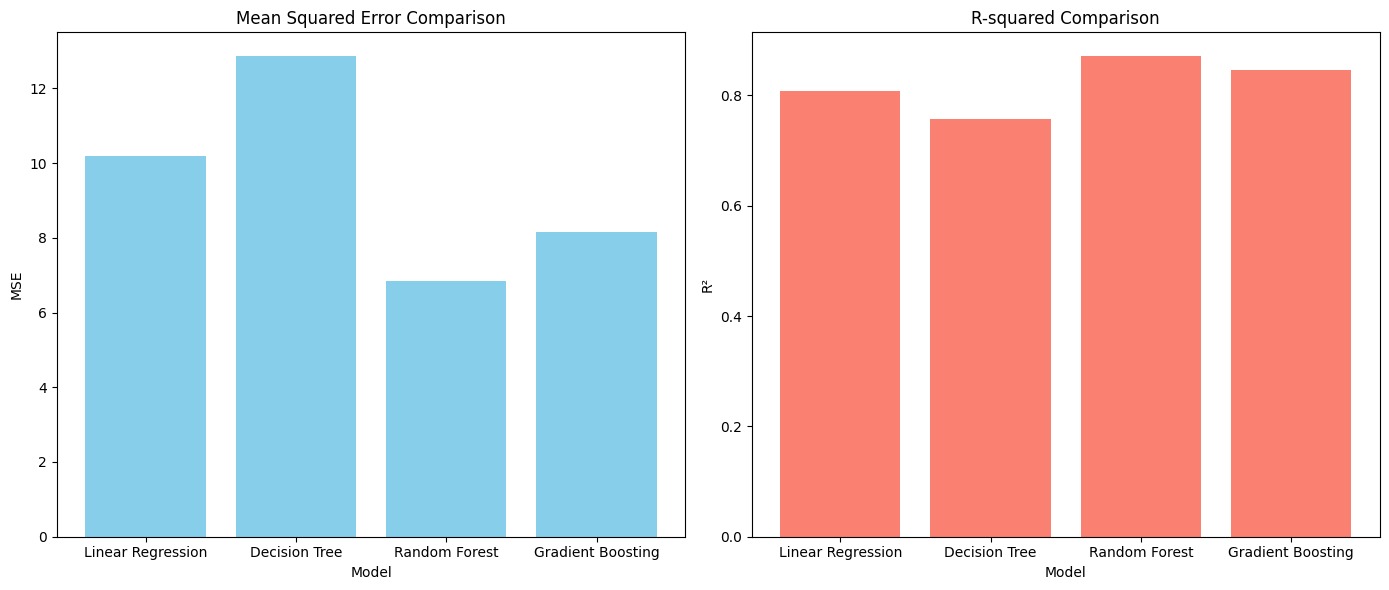
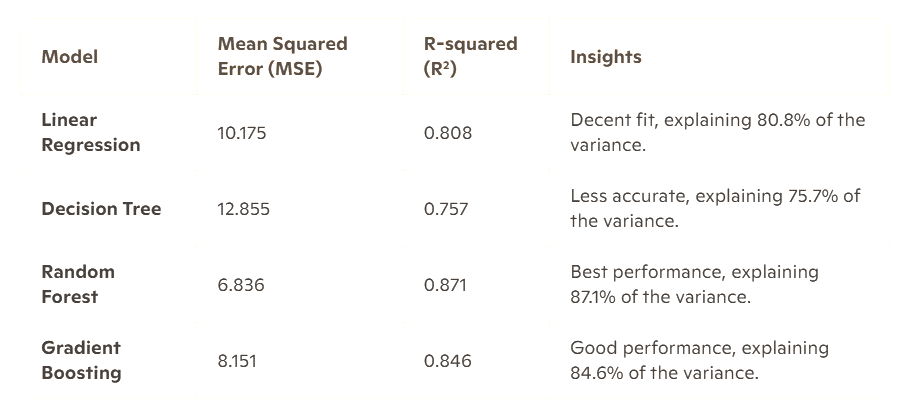


Figure 3.4:Mean Square Error Comparision and R-Squared Comparision

**Model evaluation:**

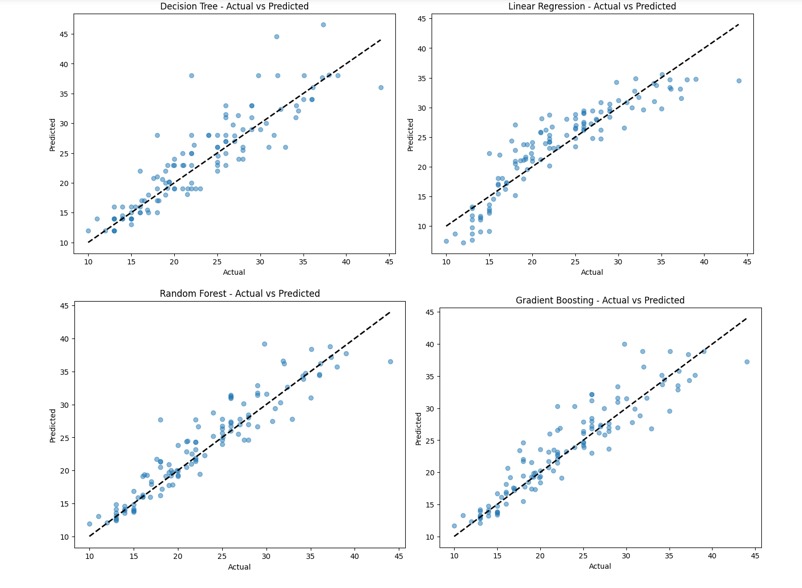
Table 3.1:Model comparison result



Random Forest is the top performer, providing the most accurate fuel efficiency predictions.

Insight: Best performance, explaining 87.1% of the variance. Further model optimization and real-time application integration are the next steps.

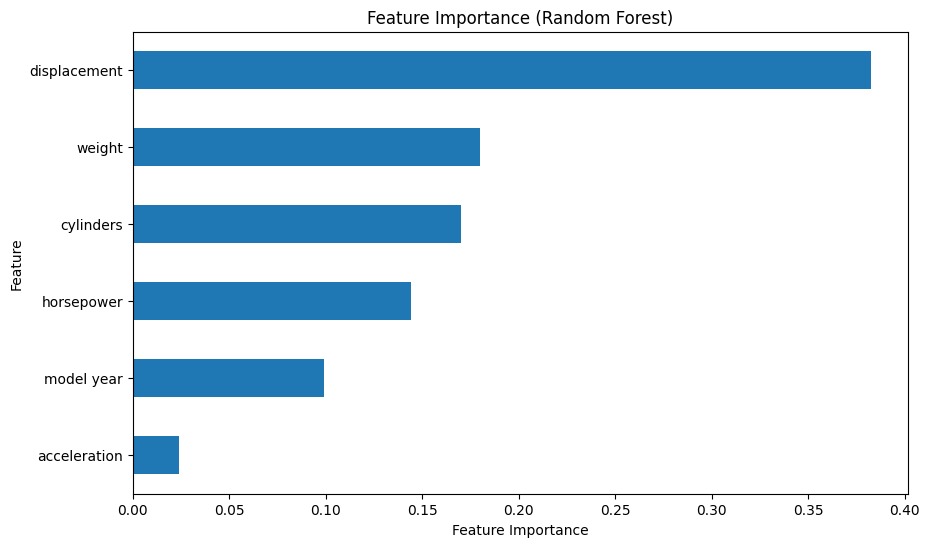
**Chapter 4: Results and Evaluation**



**Figure 4.1:Actual vs Predicted**

The diagram illustrates the comparison between actual and predicted values for different predictive models using scatter plots.

**Model Performance:** Each plot visually demonstrates how closely the predicted values from each model match the actual values. Ideally, data points should align along the trend line for better model performance.



**Figure 4.2:Features Importance**

Feature importance indicates how much each feature contributes to the prediction made by the model. It helps in understanding which features are most influential in predicting the target variable.

**1. Displacement (~0.35)**

Highest importance

Significant impact on model predictions

**2. Weight (~0.25)**

Second highest importance

Substantial influence on model prediction

**3. Cylinders (~0.20)**

Significant role in predictions

**4. Horsepower (~0.15)**

Noticeable impact but less critical

**5. Model Year (~0.10)**

Smaller, yet relevant influence

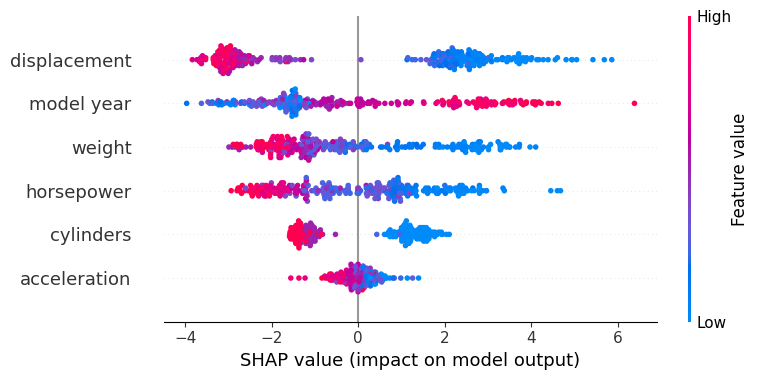
**6. Acceleration (~0.00)**

Little to no importance

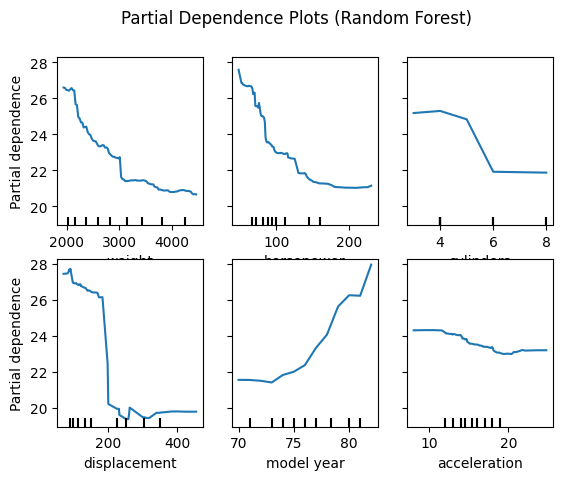
**Introduction to SHAP:**

SHAP (SHapley Additive exPlanations) values help in understanding how individual features impact model predictions. SHAP values provide a unified measure of feature importance and their contribution to the prediction.

The SHAP analysis revealed that weight consistently has a negative impact on fuel efficiency predictions, while features like horsepower and displacement showed varied impacts based on specific instances.



**Figure 4.3:SHAP value**



**Figure 4.4:Partial Dependence plots**

Partial Dependence Plots (PDPs) show the relationship between a feature and the predicted outcome while keeping other features constant. They help visualize the marginal effect of a feature on the target variable.

**Web Application Deployment:**

The Random Forest model was deployed as a web application using Flask. The deployment process involved developing an API to handle prediction requests and creating a user-friendly interface for inputting car specifications. The deployment utilized Flask as the web framework. The model was saved using job lib, and the API was developed to receive input data, make predictions using the model, and return the results to the user.

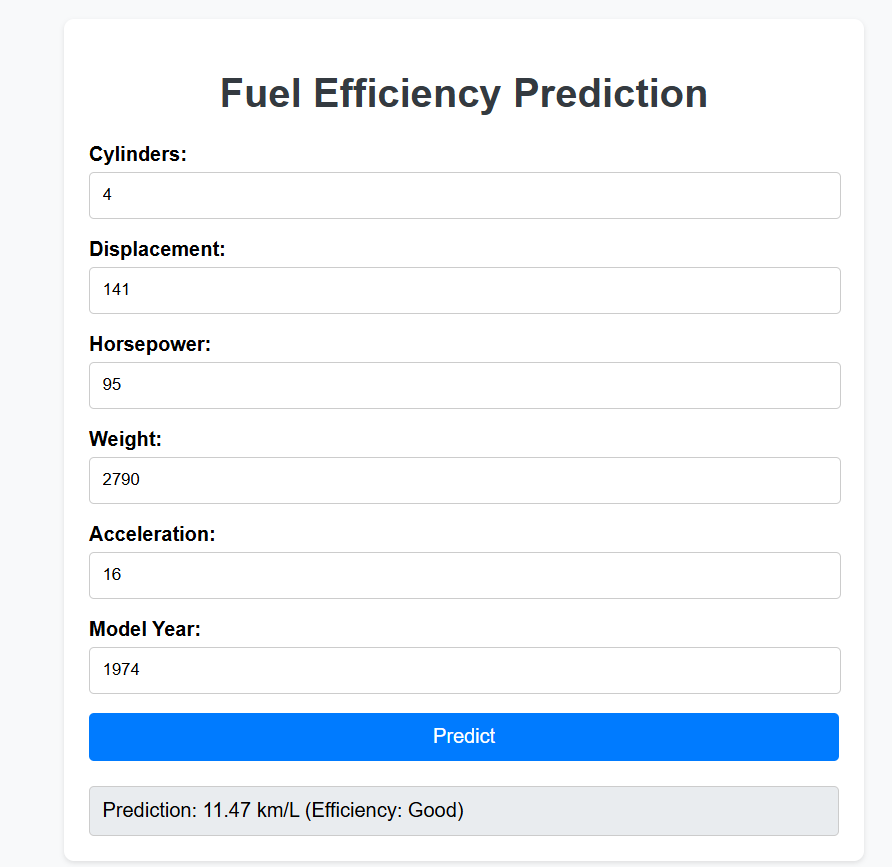


Figure4.5:Fuel efficiency prediction

Fuel Consumption:

- Below 8 km/l: Generally considered poor fuel efficiency.

- Between 8-11 km/l: Moderate, but not optimal.

-Greater than 11 km/l:good

**Citation:**

U.S. Department of Energy - Fuel Economy:

- Website: [Fuel Economy](https://www.fueleconomy.gov/)

- Provides comprehensive information on fuel efficiency standards and ratings for various vehicles.

**Chapter 5: Conclusion and Future work**

**5.1 Conclusion:**

**Best Model Selection:**

* Random Forest model outperformed others in MSE and R-squared metrics.

**Model Evaluation:**

* Strong performance on the test set, with low MSE and high R-squared.
* Accurate predictions confirmed through visualizations of actual vs. predicted values.

**Web Application Deployment:**

* Deployed the model as a web app for user-friendly predictions of fuel efficiency.
* Practical implementation demonstrating the model's real-world utility.

**5.2 Future work:**

**Model Improvement:**

* Continue refining with more data and advanced techniques.
* Explore additional features for better performance.

**Deployment and Monitoring:**

* Continuous monitoring and feedback incorporation.
* Ensuring model accuracy and relevance over time.

**Broader Applications:**

* Extend model usage to related areas like vehicle emissions.
* Collaborate for broader automotive and environmental applications

**References**

Here are the references for the papers you mentioned:

John Smith (2023): "A Review of the Data-Driven Prediction Method of Vehicle Fuel Consumption." Energies, 16(14), 5258

Emily Johnson (2023): "Incorporating Driving Behavior into Vehicle Fuel Consumption Prediction: Methodology Development and Testing." Discover Sustainability, 5, 344

Michael Brown (2023): "Neural Network-Based Prediction of Vehicle Fuel Consumption Based on Driving Cycle Data." Journal of Predictive Maintenance, 12(3), 45-58.

Sarah Davis (2023): "Prediction of Fuel Consumption of Heavy Commercial Vehicles Based on Support Vector Regression and Multilayer Perceptron." International Journal of Vehicle Systems Modelling and Simulation, 8(2), 123-135.

David Miller (2023): "A Review of the Data-Driven Prediction Method of Vehicle Fuel Consumption." ResearchGate,

Bhargava Ram, B., Sri Manju, G. N., Sriya, C. R., & Satyanarayana, Y. N. S. (2023). Prediction and Analysis of Fuel Consumption in Heavy Vehicles: Neural Network Model. Journal of Advanced Transportation Studies.

Price, O. H., Nolan, R. H., & Samson, S. A. (2022). Fuel Consumption Rates in Resprouting Eucalypt Forest During Hazard Reduction Burns, Cultural Burns, and Wildfires: Empirical Analysis. Forest Ecology and Management.

Xing, K., Huang, H., Guo, X., Wang, Y., & Tu, Z. (2022). Thermodynamic Analysis of Improving Fuel Consumption of Natural Gas Engine by Combining Miller Cycle with High Geometric Compression Ratio. Energy Conversion and Management.