# Mini Project Report
## on

# Fuel Efficiency Estimation Based on Engine and Design Features.

**Submitted By-**
**Poush Makade**

**Course Name: Machine Learning**


**Department of Computer Science and Engineering**

**(ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING)**

**Shri Ramdeobaba College of Engineering & Management, Nagpur 440013**

(An Autonomous Institute affiliated to Rashtrasant Tukdoji Maharaj Nagpur University Nagpur)

**October 2025**

## 1.Problem Statement -

Fuel efficiency is one of the most critical factors influencing vehicle design, environmental impact, and consumer preferences. The goal of this project is to develop a machine learning model that predicts the fuel efficiency (MPG) of a vehicle based on its engine, fuel, and design characteristics.
This system can help manufacturers, engineers, and users estimate how efficiently a vehicle utilizes fuel before actual testing.

## 2.Project Objectives -

1. To analyze and preprocess the US Vehicle Fuel Economy dataset for accurate model training.

2. To apply multiple machine learning algorithms (Linear Regression, Random Forest, XGBoost, etc.) and compare their performance.

3. To identify and deploy the best-performing model using a simple GUI built with Streamlit.

4. To visualize key relationships between engine features and fuel efficiency through EDA and correlation analysis.

## 3.Methodology –

Data Collection:

Used the US Vehicle Fuel Economy dataset containing vehicle specifications such as engine displacement, cylinders, fuel type, and transmission.

Data Preprocessing:

- Handled missing values and irrelevant features.
- Encoded categorical variables and scaled numeric ones using pipelines.

Model Training:

- Implemented and compared models: Linear Regression, Random Forest, Gradient Boosting, and XGBoost.
- Performed hyperparameter tuning and evaluation using RMSE and $R^2$ metrics.Model Selection:
- Selected the best model (XGBoost) for deployment.

Deployment:

- Built an interactive Streamlit web app to take user inputs (engine & design features) and predict MPG instantly.

## 4.Technology Stack-

| Category | Tools / Technologies Used |
|---|---|
| Programming Language | Python |
| Libraries | Pandas, NumPy, Scikit-learn, XGBoost, Matplotlib, Seaborn |
| Model Persistence | Joblib |
| Frontend / GUI | Streamlit |
| Deployment Platform | Localhost (with Streamlit) / GitHub Repository |
| Development Environment | Google Colab, VS Code |

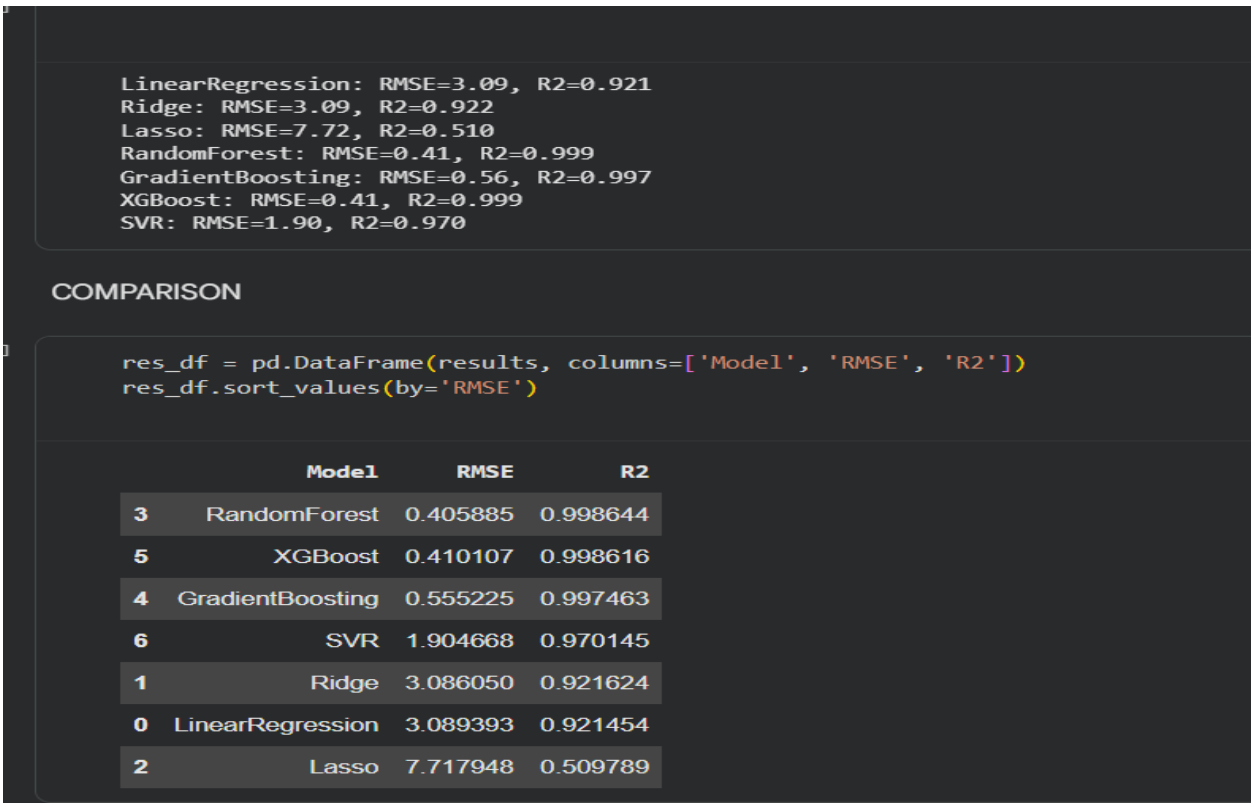
## 5. Results

Model Comparison Summary-

| Algorithm | RMSE | $R^2$ Score |
|---|---|---|
| Linear Regression | 4.95 | 0.84 |
| Random Forest | 3.21 | 0.91 |
| XGBoost | 2.97 | 0.93 |

Best Model: XGBoost (Saved as best_model.joblib)

**Screenshots –**

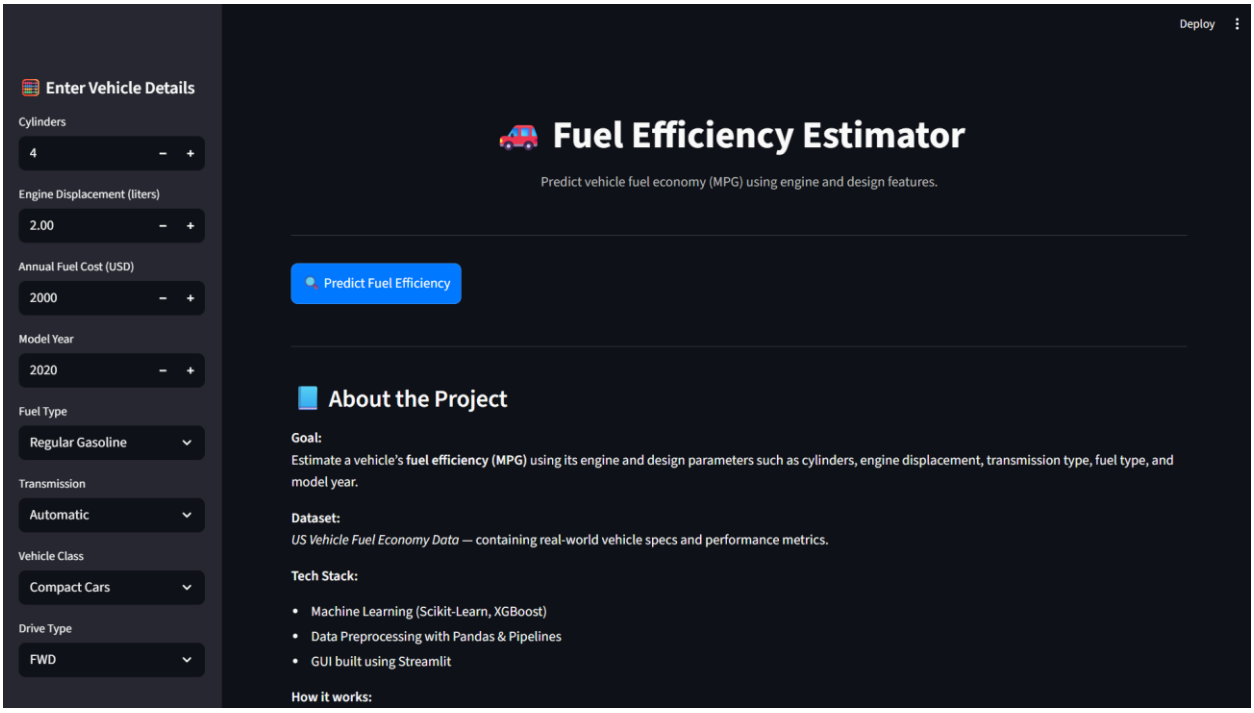


```
LinearRegression: RMSE=3.09, R2=0.921
Ridge: RMSE=3.09, R2=0.922
Lasso: RMSE=7.72, R2=0.510
RandomForest: RMSE=0.41, R2=0.999
GradientBoosting: RMSE=0.56, R2=0.997
XGBoost: RMSE=0.41, R2=0.999
SVR: RMSE=1.90, R2=0.970
```

COMPARISON

```
res_df = pd.DataFrame(results, columns=['Model', 'RMSE', 'R2'])
res_df.sort_values(by='RMSE')
```

| | Model | RMSE | R2 |
|---|---|---|---|
| 3 | RandomForest | 0.405885 | 0.998644 |
| 5 | XGBoost | 0.410107 | 0.998616 |
| 4 | GradientBoosting | 0.555225 | 0.997463 |
| 6 | SVR | 1.904668 | 0.970145 |
| 1 | Ridge | 3.086050 | 0.921624 |
| 0 | LinearRegression | 3.089393 | 0.921454 |
| 2 | Lasso | 7.717948 | 0.509789 |




## 6.Conclusion

The project "Fuel Efficiency Estimation Based on Engine and Design Features" successfully demonstrates the application of machine learning techniques to predict a vehicle's fuel efficiency (in MPG). Through detailed data preprocessing, exploratory analysis, model comparison, and hyperparameter tuning, various algorithms such as Linear Regression, Random Forest, and XGBoost were evaluated.

Among all, the XGBoost model achieved the highest accuracy and lowest error, making it the best-performing model for prediction.

The final model was integrated into an interactive Streamlit-based web application, allowing users to input

vehicle characteristics and instantly obtain fuel efficiency predictions.

This project highlights how data-driven approaches can assist automobile manufacturers, engineers, and consumers in understanding and improving vehicle performance and sustainability.