FUEL CONSUMPTION ANALYSIS

MINI PROJECT REPORT

Submitted by

Rohit M	**210701215**
Santhosh M	**210701233**

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

ANNA UNIVERSITY:: CHENNAI 600 025

APRIL 2024

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

BONAFIDE CERTIFICATE

Certified that this Report titled "**Fuel Consumption Analysis**" is the bonafide work of **"Rohit M(210701215) and Santhosh M(210701233)"** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Rakesh Kumar M,M.E.,Ph.D.
Assistant Professor,
Department of Computer Science and Engineering,
Rajalakshmi Engineering College,
Chennai – 602015

Submitted to Mini Project Viva-Voce Examination held on __________________

Internal Examiner **External Examiner**

ABSTRACT

Fuel Consumption analysis is crucial for optimizing vehicle performance, reducing operational costs, and minimizing environmental impact. This project analyzes fuel consumption and CO2 emissions data from various vehicle models using machine learning algorithms such as decision trees and random forests. By leveraging existing datasets that include variables such as vehicle type, driving patterns, and environmental conditions, the ML model can accurately forecast fuel consumption. Through exploratory data analysis and model building, insights into the factors influencing fuel consumption and emissions are gained. The study employs techniques like feature engineering, hyperparameter tuning, and evaluation metrics to optimize model performance. Results demonstrate the effectiveness of machine learning in predicting fuel consumption and CO2 emissions, contributing to environmental sustainability efforts in the automotive industry. Future research may focus on expanding the dataset and exploring additional predictive features for enhanced modeling accuracy.

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman **Mr. S.MEGANATHAN, B.E, F.I.E.**, our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN**, **Ph.D.,** for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. KUMAR, Ph.D.**, Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide, **Dr. RAKESH KUMAR M, M.E., Ph.D., Assistant Professor,**Department of Computer Science and Engineering for his encouragement and guiding us throughout the project to build our project. Rajalakshmi Engineering College for his valuable guidance throughout the course of the project.

Rohit M -210701215
Santhosh M-210701233

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

OC	Organized Crimes
SVM	Support Vector Machines
CMS	Crime Monitoring System
KNN	K Nearest Neighbours
FIR	First Information Report
JSON	Java Script Object Notation

CHAPTER 1
INTRODUCTION

1.1 GENERAL

The automotive industry focuses more on sustainability and efficiency, driven by concerns over environmental impact and the need for resource conservation. As a result, there is an increasing focus on understanding and optimizing fuel consumption in vehicles to reduce carbon emissions and operational costs while maintaining performance standards.In order to establish strategies to improve fuel efficiency and get insights into the elements driving it, fuel consumption analysis is essential.

1.2 OBJECTIVE

The main aim of this project is to to optimize model performance by using the effectiveness of machine learning in predicting fuel consumption and CO2 emissions, contributing to environmental sustainability efforts in the automotive industry by using techniques like feature engineering, hyperparameter tuning, and evaluation metrics.

1.3 EXISTING SYSTEM

The Consumption analysis method of the past depended on less complex machine learning algorithms because of limitations in processing power and data accessibility. For example, an outdated system had analyzed past fuel usage data using conventional statistical techniques like multiple linear regression or simple decision trees. It's possible that these systems' predictive power was constrained in comparison to more recent methods. Still, they could offer insightful information about variables like car attributes, road conditions, and maintenance records that affect fuel economy. Even though they were basic, these early systems showed that machine learning could be used in transportation management, which paved the way for more sophisticated Fuel Consumption analysis tools.

1.4 PROPOSED SYSTEM

This project suggests Fuel Consumption analysis based on machine learning (ML). The machine learning algorithm is able to predict fuel usage with high accuracy by utilizing pre-existing datasets that contain characteristics like driving habits, vehicle type, and environmental conditions. Furthermore, the model offers predictive and analytical capabilities in real-time, allowing for proactive efforts to improve fuel economy. Our results show that ML models can be a good substitute for conventional sensor-based techniques, offering accurate predictions at a lower cost and complexity.

CHAPTER 2

LITERATURE SURVEY

[1]The predictive ability of three AI expectation models to predict the fuel consumption of a long-distance public transportation vehicle was evaluated by the authors. Several essential characteristics, including as load, motor RPM, and traffic, are not included in the selected dataset, even though they directly affect fuel consumption. In fact, they demonstrated that the RF model could predict fuel use even more accurately while accumulating data patterns in the absence of such fundamental components. One example of such a model is the identification of gasoline theft via comparing the actual mileage of the car with the estimated value based on several parameters such as distance, area, elevation, speed, and day of the week.

[2]Manufacturers, regulators, and consumers are interested in fuel consumption models for automobiles.They are necessary throughout every stage of the vehicle lifetime. In this study, we mainly focus on predicting the average fuel consumption for large vehicles during the maintenance and operating phase. Methods that are often ineffective in creating fuel consumption models fall into three basic categories: models grounded on physics and developed from a thorough comprehension of the underlying system. These models use rigorous mathematical equations to describe the dynamics of the vehicle's components at each given stage.

[3]The topic of minimizing fuel use in a microgrid with several power sources is explored. The situation of a big system and its conventional economic dispatch issue differs significantly from the optimization of a small power system. The existence of a local heat demand, which gives the optimization issue an additional dimension, is one of the most significant changes. Two reciprocating gas engines, a combined heat and power plant, a solar array, and a wind generator make up the microgrid that is the subject of this research.

Jagannathan et al. [4] proposed an ensemble of deep learning methods for the detection and classification of moving vehicles. The model initially ap- plied image processing techniques to enhance input image quality, followed by the use of steerable pyramid transform (SPT) and Weber local descriptor (WLD) for feature ex- traction. When simulated on two datasets (MIO-TCD and BIT vehicle dataset) of RGB images, the proposed model achieved impressive accuracies of 99.28% and 99.13%.

[5]An essential step in the design and operation of ground vehicles is the evaluation of fuel economy. The primary construction parameters of the vehicle may be decided upon at the design stage and measures to minimize fuel consumption may be implemented based on this evaluation, which is often carried out using mathematical modeling and simulation. The development of analytical models that enable precise prediction of vehicle consumption seems particularly desirable, as one of the primary aims of vehicle design is to minimize fuel consumption given anticipated operating circumstances. Inverse simulation, a well-known method for evaluating fuel consumption, uses power transferring functional blocks with preset efficiency to model the driving cycle-to-tank chain.

[6]The main trade-offs between the aforementioned methodologies are accuracy and relevance value in accordance with the requirements of the intended application. This study projects a model that might be easily created for individual important vehicles within a very large fleet.A fleet manager will optimize the route design for all cars backed by each unique vehicle's predicted fuel consumption, ensuring that the route assignments are aligned to reduce the fleet's total fuel consumption, assuming that all of the vehicles in the fleet are the proper models. These types of fleets are used in a variety of industries, including public transit, construction, garbage collection, and goods shipping along the road.

[7]The developers developed a model that accurately estimated a car's MPG based on a few vehicle-related parameters, and they came to the conclusion that the code worked by obtaining an RMSE score of 1.97 instead of the more notable 3.26 as the underlying RMSE score. As we had the option to code, we discovered a consistent increase in the value from 0.82 to 0.91, demonstrating that the model is significantly more solid when used. Additionally, it is useful for our competitor's future MPG evaluations for upcoming vehicles, allowing them to predict even the R2 score.

[8]The primary consumers of gasoline are automobiles. This has an adverse effect on the environment and significantly increases greenhouse gas emissions. Transportation was responsible for over 20% of global carbon dioxide emissions in 2000. Concerned with the long-term effects of carbon dioxide emissions are international environmental standards and specifications groups. The expected shortages in the production of petroleum products in the near future are contributing to the environmental effect and are driving forces behind the development of automobiles that are more fuel-efficient.

[9]The fuel market affects the income distribution of countries in a direct or indirect way, which has an impact on a number of areas such as the stock market, cost of living, education, and vital commodities. In return, a variety of factors that also affect everyday life for the average person influence fuel prices. Therefore, it is evident that forecasting fuel price patterns is important for the benefit of drivers as well as for economists in every country who need to foresee the economic trends resulting from these price variations and be ready for everything.

[10]The fuel of the automobile like cars mainly states the distance which is traveled by a car and particular fuel consumption is consumed by car or any vehicle. Fuel usage in cars and other vehicles contribute significantly to air pollution, and it differs greatly between countries. We are witnessing the fuel prices and customers being more particular about the features. To have a vehicle which is desirable and even more efficient, improvement of fuel has been carried out. Primarily, fuel prices differ throughout nations and also rely on the type of vehicle those nations utilize, as well as how frequently the customers use their cars.

[11]Random Forest is a tree-based ensemble with each tree depending on a collection of random variables. It can be used for either a categorical response variable, referred to in as “classification”, or a continuous response, referred to as “regression”. Similarly, the predictor variables can be either categorical or continuous.

CHAPTER 3
SYSTEM DESIGN

3.1 DEVELOPMENT ENVIRONMENT

3.1.1 HARDWARE SPECIFICATIONS

This project uses minimal hardware but in order to run the project efficiently without any lack of user experience, the following specifications are recommended

Table 3.1.1 Hardware Specifications

PROCESSOR	Intel Core i5
RAM	4GB or above (DDR4 RAM)
GPU	Intel Integrated Graphics
HARD DISK	6GB
PROCESSOR FREQUENCY	1.5 GHz or above

3.1.2 SOFTWARE SPECIFICATIONS

The software specifications in order to execute the project has been listed down in the below table. The requirements in terms of the software that needs to be pre-installed and the languages needed to develop the project has been listed out below.

Table 3.1.2 Software Specifications

FRONT END	Streamlit,Pickle
BACK END	Python, Django
SOFTWARES USED	Visual Studio, Jupyter Notebook

3.2 SYSTEM DESIGN

3.2.1 ARCHITECTURE DIAGRAM

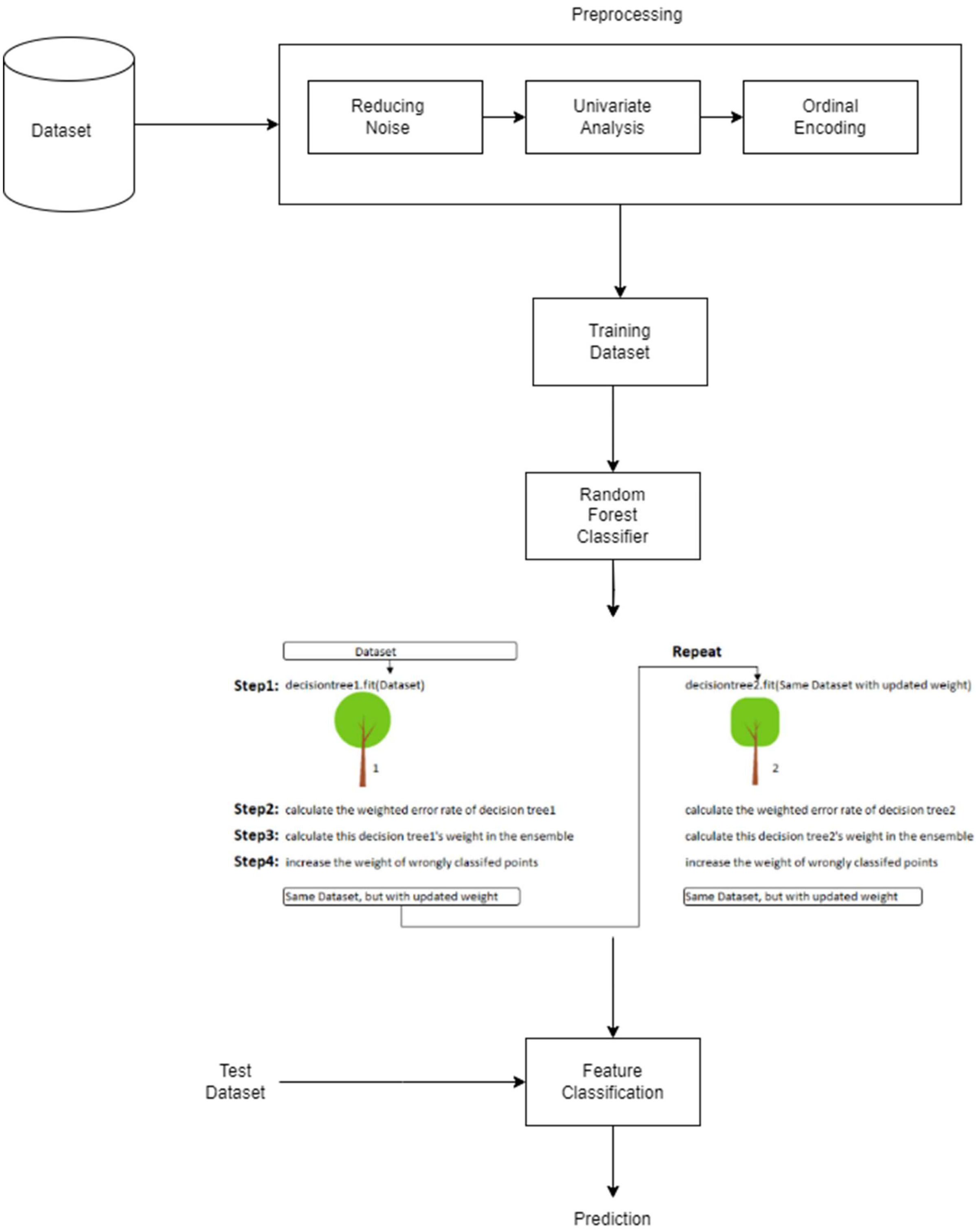

Fig 3.2.1 Architecture Diagram

PRE-PROCESSING:

The first step is to remove any extraneous columns from the dataset in order to make it simpler and need less computing power. The mode value of the column is then used to fill in any missing data in crucial fields, including Fuel Type. By preserving the most common category, this imputation technique preserves the dataset's integrity. It's also crucial to reduce noise in category columns, such as Transmission; this entails standardizing the inputs to a common format and fixing any irregularities to guarantee data homogeneity.

Univariate analysis is then performed to understand the distribution of individual variables, identifying outliers and data patterns that may influence the model. This analysis helps in making decisions about further data transformations or feature engineering. Following this, bivariate analysis examines the relationship between two variables, particularly how independent variables interact with the target variable, Fuel Consumption. This step is crucial for uncovering dependencies and correlations that inform feature selection and model building.

Finally, ordinal encoding is applied to categorical variables that have a meaningful order but are not numerical, such as Transmission type or vehicle size categories. This technique converts these categories into numerical values, facilitating their use in ML algorithms. By systematically following these preprocessing steps, the dataset becomes clean, consistent, and ready for effective model training and prediction of Fuel Consumption.

TRAINING SET:

When training a random forest classifier, the dataset is divided into multiple subsets. The classifier constructs numerous decision trees, each trained on a different subset of the data. By considering various combinations and splits of the features like number of cylinders, transmission mode, fuel type, engine size,CO2 rating, each tree makes independent predictions. The random forest then aggregates these predictions to form a final, consensus output. This ensemble approach enhances the model's robustness and accuracy. Ultimately, the training dataset with these diverse and informative features enables the random forest classifier to learn intricate patterns associated with fuel efficiency. The model leverages the interplay between engine characteristics, transmission type, fuel properties, and emissions to predict fuel consumption accurately.The quality of each potential split is evaluated using a metric like Gini impurity or entropy. These metrics measure the homogeneity of the target variable (fuel efficiency) within the resulting subsets. A good split results in subsets that are more homogeneous with respect to fuel efficiency, meaning the vehicles in each subset have similar fuel efficiency ratings.

CHAPTER 4
PROJECT DESCRIPTION

4.1 MODULE DESCRIPTION

4.1.1 DATA PRE-PROCESSING:

This module mainly consists of the preliminary step of collecting values of data from various vehicles and conversion of cleaning the data by removing unnecessary columns in the dataset. Next, missing values in essential columns, such as Fuel Type, are addressed by replacing them with the mode value of the column. Noise reduction in categorical columns, like Transmission involves standardizing the entries to a consistent format and correcting any anomalies.

4.1.2 TRAINING SET:

The data containing number of cylinders, transmission mode, fuel type, engine size,CO2 rating are fed to a machine learning algorithm for classification of data.This is a very important step as this step involves dividing the data into two subsets based on a condition related to one of the features.

4.1.3 TRAINING MODEL:

The system is then trained with a model which is Random Forest algorithms.The algorithm selects the split that results in the highest increase in homogeneity (i.e., the best split) among the randomly chosen features. This process is repeated recursively, creating new nodes and further splitting the data within each subset.

4.1.4 PREDICTION:

Once all the decision trees are built, the random forest aggregates their predictions. For a regression task like predicting fuel efficiency, this aggregation typically involves averaging the predictions from all the trees. Consequently, the random forest classifier effectively predicts fuel efficiency by leveraging the complex relationships between the number of cylinders, transmission mode, fuel type, engine size, and CO2 rating.

CHAPTER 5
IMPLEMENTATION AND RESULTS

5.1 IMPLEMENTATION

5.1.1 RANDOM FOREST:

The initial dataset comprises various features such as the number of cylinders, transmission mode, fuel type, engine size, and CO2 rating. This dataset is typically divided into training and testing subsets, where the training subset is used to build the model, and the testing subset evaluates its performance.

The random forest algorithm begins by creating multiple decision trees. For each tree, a random sample of the training data is drawn with replacement, a process known as bootstrapping. This ensures that each tree is trained on a slightly different subset of data, introducing variability and reducing overfitting.

At each node of a decision tree, a random subset of features is selected from the full set of features. This randomness helps in making each tree unique and further reduces the risk of overfitting, as different trees will likely consider different features at each split.

The program assesses many split options for each feature subset that has been picked. Splitting data entails splitting it into two parts according to a feature-related criteria. A split may divide the data between cars with less than six cylinders and those with six or more cylinders, for instance, if the feature is the number of cylinders.

A measure like entropy or Gini impurity is used to assess each possible split's quality. Within the generated subgroups, these measures quantify the homogeneity of the objective variable (fuel efficiency). A well-executed split produces more homogenous subsets in terms of fuel efficiency, i.e., each subset's automobiles have comparable fuel efficiency ratings.

Among the randomly selected characteristics, the algorithm chooses the split that produces the greatest gain in homogeneity, or the optimal split. Recursively, this procedure is repeated, producing new nodes and further dividing the data inside each subset. Until a halting condition is satisfied, like as reaching a maximum depth, having a minimum number of samples per leaf, or attaining a particular degree of homogeneity, the tree keeps growing by splitting nodes. A subset of the data with a projected value for the goal variable (fuel efficiency) is represented by each leaf node in the tree.

The random forest collects the predictions made by each decision tree once it has been constructed. In the context of regression tasks such as fuel efficiency prediction, this aggregation usually consists of taking the average of all the trees' predictions.

5.2 OUTPUT SCREENSHOTS

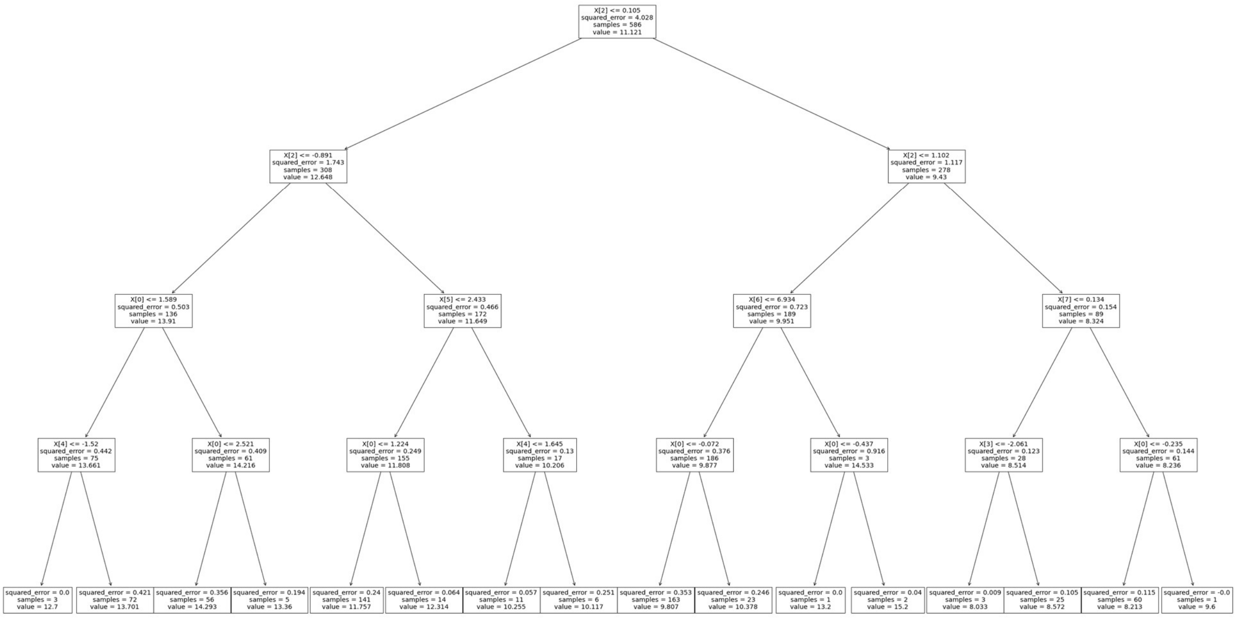

Fig 5.2.1 Classification tree

The graph shows how the data are splitted based on various conditions. The output is given based on the Random Forest’s tree weight and splitting is based on that weight. The Random Forest gives accuracy of 0.88.

CHAPTER 6
CONCLUSION AND FUTURE ENHANCEMENTS

6.1 CONCLUSION

In this project, we aimed to develop a fuel efficiency prediction system using a random forest classifier through meticulous data preprocessing steps including dropping unnecessary columns, handling missing values, reducing noise, and encoding categorical variables, we prepared the dataset for model training.

The utilization of features such as the number of cylinders, transmission mode, fuel type, engine size, and CO2 rating allowed the random forest classifier to effectively capture the complexities influencing fuel consumption. By leveraging the ensemble of decision trees and the randomness in feature selection, the model exhibited robustness and accuracy in predicting fuel efficiency across diverse vehicle types and driving conditions.

Our findings demonstrate that machine learning-based approaches, particularly the random forest algorithm, offer a cost-effective and scalable solution for fuel efficiency analysis. By leveraging existing datasets and advanced data analytics techniques, we can provide valuable insights into optimizing vehicle performance, reducing operational costs, and minimizing environmental impact.

6.2 FUTURE ENHANCEMENTS

Incorporating more advanced feature engineering techniques can help extract additional insights from the data. For example, deriving new features based on domain knowledge or employing dimensionality reduction methods to capture underlying patterns more effectively could enhance the model's predictive power.

Furthermore, continuous model monitoring and evaluation are essential to identify and address any performance degradation over time. Implementing robust validation techniques and periodic retraining of the model with updated data can ensure that it maintains its accuracy and relevance in evolving real-world scenarios.

REFERENCES

[1] 1. Aceves, S.M. , Flowers, D.L. , Martinez-Frias, J. , Smith, J.R. , Dibble, R. , Au, M. , Girard, J. (2001). ‘HCCI combustion: analysis and experiments’. SAE International, paper # 2001-01-2077S.

[2] B. Lee, L. Quinones, and J. Sanchez, “Development of greenhouse gas emissions model for 2014-2017 heavy-and mediumduty vehicle compliance,” SAE Technical Paper, Tech. Rep., 2011.

[3] C. A. Hernandez-Aramburo, T. C. Green and N. Mugniot, "Fuel consumption minimization of a microgrid," in IEEE Transactions on Industry Applications, vol. 41, no. 3, pp. 673-681, May-June 2005, doi: 10.1109/TIA.2005.847277.

[4] Jagannathan, P., Rajkumar, S., Frnda, J., Divakarachari, P.B., Sub- ramani, P.: Moving vehicle detection and classification using gaus- sian mixture model and ensemble deep learning technique. Wire- less Communications and Mobile Computing 2021 (2021). DOI 10.1155/2021/5590894

[5] Froberg, A.; Nielsen, L. Efficient drive cycle simulation. IEEE Trans. Veh. Technol. 2008, 57, 1442–1453.

[6] F. Perrotta, T. Parry, and L. C. Neves, “Application of machine learning for fuel consumption modelling of trucks,” in Big Data (Big Data), 2017 IEEE International Conference on. IEEE, 2017, pp. 3810–3815.

[7] Varun Shirbhayye, Deepesh Kurmi, Siddharth Dyavanapalli, Agraharam Sai Hari Prasad, Nidhi Lal, “An Accurate Prediction of MPG (Miles Per Gallon) using Linear Regression Model of Machine Learning”.

[8] Elalem, A., EL-Bourawi, M.S. Reduction of Automobile Carbon Dioxide Emissions. Int J Mater Form 3 (Suppl 1), 663–666 (2010). https://doi.org/10.1007/s12289-010-0857

[9] Ensemble Machine Learning: Methods and Applications (pp.157-176)Chapter: 5 Publisher: SpringerEditors: Cha Zhang, Yunqian Ma.

[10] L. Shalini, S. Naveen and U. M. Ashwinkumar, "Prediction of Automobile MPG using Optimization Techniques," 2021 IEEE Madras Section Conference (MASCON), Chennai, India, 2021, pp. 1-6, doi: 10.1109/MASCON51689.2021.9563597.

[11] Harsh H. Patel, Purvi Prajapati, “Study and Analysis of Decision Tree Based Classification Algorithms,” International Journal of Computer Sciences and Engineering, Vol.6, Issue.10,pp.74-78,2018.https://doi.org/10.26438/ijcse/v6i10.747