**Mileage prediction analysis**

The project submitted to the

SRM University – AP, Andhra Pradesh

For the partial fulfillment of the requirements to award the degree of

**Bachelor of Technology/Master of Technology**

In

**Computer Science and Engineering**

**School of Engineering and Sciences**

Submitted by

**Candidate Name**

**M Sree Annapurna AP21110010791**

**B.Jyothirmai AP21110010794**

**G.Madhurya AP21110010798**

**M.NanditaAP21110010829**

**M.Veni AP21110010835**

**A picture containing text

Description automatically generated**

Under the guidance of

**Krishna Siva Prasad**

**SRM University–AP**

**Neerukonda, Mangalagiri, Guntur**

**Andhra Pradesh – 522 240**

**[Month, Year]**

**11,2023**

# Certificate

Date: 27-Nov-23

This is to certify that the work present in this Project entitled “**Mileage prediction analysis**” has been carried out By M. Sree Annapurna (AP21110010791), B.Jyothirmai (AP21110010794), G.Madhurya (AP21110010798), M.Nandita (AP2111001829), and M.Veni (AP21110010835) (under my/our supervision. The work is genuine, original, and suitable for submission to the SRM University – AP for the award of Bachelor of Technology/Master of Technology in the **School of Engineering and Sciences**.

**Supervisor**

(Signature)

Prof. / Dr. [Name]

Designation,

Affiliation.

**Co-supervisor**

(Signature)

Prof. / Dr. [Name]

Designation,

Affiliation

**TABLE OF CONTENTS:**

**Contribution of each student**

**Novelty of aspects**

**Abstract**

**Introduction**

**Project background**

**Description of the project**

**Proposed solution using data mining technique**

**Model architecture**

**Experimentation details**

**Conclusion and Future Trends**

**Contribution of each student:**

|  |  |
| --- | --- |
| **student** | **Role** |
| **M sree Annapurna** | Contribution of Decision Tree, final conclusion, and Contribution of PPT and report. |
| **B jyothirmai** | Data preprocessing (data cleaning, data transformation) and helping with the PPT as well as report. |
| **G.madhurya** | Clustering, graphical plotting, and helping with the report. |
| **M.Nandita** | Data preprocessing (data discretization), naive Bayes, and helping for the report. |
| **M.Veni** | The Linear regression, (Test and Train of the dataset) and the final conclusion with the MPG result. Contribution of PPT and report. |

**Novelty of aspects:**

**1. Innovative Feature Engineering:**

- Developed novel features and applied creative transformations to enhance the predictive power of the models.

**2. Effective Data Cleaning and Transformation:**

- Addressed unique challenges in the dataset through advanced data cleaning methods, ensuring the quality of input data.

- Implemented effective data transformation techniques to extract meaningful patterns for mileage prediction.

**3. Sophisticated Discretization Strategies:**

- Utilized advanced discretization methods, enhancing the interpretability and accuracy of the predictive models.

**4. Diverse Model Selection and Integration:**

- Employed a diverse set of models, including decision tree, naive bayes, clustering, and linear regression, to capture various aspects of mileage prediction.

- Integrated model results in an innovative way, leveraging the strengths of each model for a more comprehensive understanding.

**5. Creative Visualization Approaches:**

- Developed and employed innovative visualizations to represent mileage predictions graphically, aiding in the interpretation of results.

**6. Unexpected Insights:**

- Uncovered unexpected insights from the analysis, providing a unique perspective on factors influencing auto mileage.

**7.** **Comparative Analysis with Existing Methods:**

- Conducted a thorough comparison with existing methods or benchmarks, highlighting the project's improvements and advantages over established approaches.

**Abstract:**

The automotive industry has benefited from innovative applications made possible by the swift progress of machine learning and data analysis algorithms. This project aims to create a reliable Auto Mileage Prediction system that uses machine learning methods to calculate a car’s fuel efficiency. The suggested model aims to help car owners, automakers, and legislators optimize fuel efficiency, lower carbon emissions, and make well-informed decisions about automotive technologies.

The study entails gathering a wide range of data on variables including acceleration, vehicle weight, model year, and car name. The best model for mileage prediction is determined by comparing a number of regression algorithms, such as ensemble methods, decision trees, and linear regression.

**Introduction:**

The goal of this research is to create a Mileage Prediction system that can accurately and consistently estimate a vehicle's fuel efficiency by applying analysis (like decision tree, naive Bayes, clustering, and data preprocessing). In addition to providing individual auto owners with insights into their fuel usage habits, mileage prediction technology gives automakers a useful tool for developing and refining energy-efficient automobiles. Furthermore, these systems can help legislators enact rules and incentives that will effectively encourage environmentally friendly transportation practices. A number of intricate factors, such as engine specifications, vehicle weight, aerodynamics, and driving conditions, interact to predict an automobile's mileage. The complexities of these relationships are often too intricate for traditional methods to fully capture. On the other hand, machine learning algorithms are highly proficient in managing extensive datasets and recognizing intricate patterns, which makes them highly appropriate for accurately predicting automobile mileage**.** This research investigates the use of Data mining and analysis methods to create a reliable Mileage Prediction model. A broad dataset covering a wide range of variables affecting fuel efficiency is gathered and pre-processed for the study. Our goal is to determine the best model for predicting auto mileage for various vehicle types by utilizing the most recent regression algorithms and performing an extensive comparative analysis. The research findings have noteworthy consequences for both the automotive sector and more extensive endeavors aimed at achieving sustainable transportation. With the information provided by the proposed MileagePrediction system, people may be better equipped to make decisions regarding their driving practices and vehicle usage. Furthermore, by using these insights, automakers can create vehicles that are more fuel-efficient, helping to ensure that the transportation industry has a more sustainable and environmentally friendly future.

**Project background:**

In this project, we focus on predicting auto mileage using machine learning techniques. Our approach involves comprehensive data preprocessing, application of decision tree and naive Bayes algorithms, and clustering analysis. We present the top 10 and least 10 MPG dataset values graphically for insights. Key findings and potential applications are discussed, providing a valuable contribution to the field. Auto mileage prediction holds paramount significance for fuel efficiency and environmental impact. This project aims to address this critical concern through machine learning. Leveraging a Kaggle dataset, we apply decision tree and naive Bayes algorithms, coupled with clustering analysis, to gain valuable insights. The visualization of the top and bottom 10 MPG values offers a practical perspective on the predictive capabilities. This paper outlines the methodology, dataset overview, and the significance of our findings in advancing auto mileage prediction.

**Description of the project:**

The goal of the auto mileage prediction problem is to precisely estimate a vehicle's fuel efficiency, which is typically expressed in miles per gallon (MPG) given a variety of input parameters. In light of the automotive industry's continuous efforts to improve sustainability, lower carbon emissions, and optimize fuel consumption, this issue is critical. The Mileage Prediction problem is important for manufacturers and policymakers, as well as for individual car owners who want to understand and maximize their fuel consumption. Accurate forecasting helps car owners plan their fuel budgets more effectively and encourages them to drive sustainably. Producing vehicles that are more fuel-efficient, satisfying consumer demands for sustainability, and adhering to stricter environmental regulations can all help manufacturers. With the aid of these predictive models, policymakers can create rules and incentives that effectively promote the development and use of energy-efficient automobiles. In order to solve the Mileage Prediction problem, a broad dataset covering a variety of cars and driving conditions must be gathered. To pre-process this data and choose and change pertinent variables for model training, feature engineering is essential. After that, data mining and analysis algorithms—which include ensemble methods, decision trees, and regression models like linear regression—are used to identify patterns in the data and produce precise predictions

**Proposed solution using data mining technique:**

**Data collection and preprocessing:** Compile a varied dataset containing data on a variety of cars from different manufacturers, models, and years. Clean the dataset by handling missing values, outliers, and ensuring data consistency

**Performing data analysis:** Determine possible trends and outliers by visualizing the correlations between the variables. Examine the ways in which various characteristics affect the differences in car mileage. Analyze different regression models, including clustering, naive Bayes algorithms, decision trees, and linear regression. Utilize cross-validation methods to evaluate model performance and choose the best algorithm**.**

**Model architecture:**

**Input Layer:**

Number of input features - Purpose: Take input features such as acceleration, vehicle weight, horse speed, etc.

**Feature Scaling:**

To guarantee that every feature contributes equally to the model, standardize or normalize numerical features**.**

**Model training:**

Split the dataset into training and validation sets.

**Interpretability Analysis:**

Conduct feature importance analysis to understand the contribution of each input feature to the predictions.

**Model Deployment:**

Deploy the trained model for making real-time predictions. - Implement user-friendly interfaces for users to input vehicle details and receive mileage predictions

**Continuous Monitoring and Updating:**

Set up mechanisms to monitor the model's performance over time. - Update the model as new data becomes available to ensure its accuracy and relevance

**Experimentation details:**

1. The dataset we used in our analysis:

[https://www.kaggle.com/datasets/uciml/autompg-dataset/](https://www.kaggle.com/datasets/uciml/autompg-dataset/data)

2. We have imported the necessary libraries

3. Import the dataset

4. We have implemented data preprocessing techniques: data cleaning, data transformation, data discretization

5. Then we implemented the decision tree technique, naive biased, clustering, Linear Regression

We included all the above-mentioned techniques in our program.

**Dataset overview table:**

|  |  |
| --- | --- |
| **Feature** | **description** |
| Mpg | Miles per gallon |
| Cylinders | Number of cylinders in an engine |
| Horsepower | Engine horsepower |
| Weight | Vehicle weight |
| Displacement | the total volume of all the cylinders in an internal combustion engine |
| Acceleration | the rate at which a vehicle can change its velocity |
| Model year | manufacturing year of a vehicle |
| Origin | Origin refers to the country or region where a vehicle is manufactured, |
| Car name | Name of the car |

**Conclusion and future trends:**

**Feature Importance:**

A vehicle's mileage is largely dependent on factors including engine specs, weight, aerodynamics, acceleration, model year, and place of origin. The significance of each feature was determined by the analysis, which also shed light on the variables that most influence fuel efficiency.

**Temporal Trends:**

One significant variable that showed the impact of changes in regulations and technology over time was the model year. Predicting how newer vehicles may show improved fuel efficiency as a result of changing manufacturing standards and technologies requires an understanding of temporal trends

**Driving Behaviour:**

Factors like acceleration were identified as significant contributors to fuel efficiency, emphasizing the importance of considering driving behaviour in mileage predictions. Aggressive or conservative acceleration patterns can affect how efficiently a vehicle uses fuel

**Model Performance:**

The effectiveness of several machine learning models, such as decision trees, and linear regression, in predicting mileage was assessed.

**Continuous Improvement:**

The analysis highlights how dynamic the automotive sector is and stresses the necessity of ongoing observation and updating the predictive model in response to new data. Preserving the model's accuracy over time requires making necessary adjustments for new technologies and modifications in driving behaviours.

Essentially, the analysis of Mileage Prediction provides a basis for developing a system that is both efficient and flexible. The analysis's conclusions help manufacturers, policymakers, and individual car owners make well-informed decisions that promote improvements in environmentally friendly and fuel-efficient transportation methods. With the information from this analysis, the predictive model is ready to help optimize fuel consumption and make a positive impact on a future where the automotive industry is more environmentally and energy-conscious

**References:**

1. Kaggle
2. Online platforms
3. Classroom lectures