

Title: Regression Analysis for Car Price Prediction Using Machine Learning

1. Introduction: The project focuses on using regression analysis and machine learning techniques to build a model capable of accurately predicting the prices of cars based on various features and characteristics. The dataset used for this project was obtained from Kaggle, a popular platform for data science and machine learning competitions.

2. Data Collection and Preprocessing: The first step in the project involved collecting data from the Kaggle website. The dataset contains information about different cars, including features such as make, model, year, mileage, horsepower, fuel type, and more. The dataset was imported into the project using the Pandas library, a popular Python data manipulation library.

3. Exploratory Data Analysis (EDA): After importing the data, the next step was to explore and understand the dataset. EDA was performed to gain insights into the data distribution, relationships between variables, and the presence of any missing values. Null values, if present, were handled by removing them to ensure the dataset's integrity and completeness.

4. Outlier Detection and Removal: To enhance the model's performance, it was essential to handle outliers effectively. The boxplot method was used to identify and subsequently remove outliers from the dataset. Outliers can have a significant impact on the model's learning process, leading to inaccurate predictions. By eliminating them, the model

becomes more robust and capable of capturing meaningful patterns in the data.

5. Data Encoding and Standard Scaling: Machine learning models require numeric input, and therefore, categorical features need to be encoded. In this project, encoding techniques were applied to convert categorical variables into numerical values. This step ensures that the model can process all features effectively.

Additionally, standard scaling was applied to the numerical features. Standard scaling helps to normalize the features by transforming them into a standard distribution, ensuring that no particular feature dominates the model's learning process.

6. Model Selection: Several regression models were considered for this project, including:

Linear Regression: A basic regression model that assumes a linear relationship between the features and the target variable.

K Neighbors Regressor: A non-parametric regression model that predicts the target value based on the average of the k-nearest neighbors.

Decision Tree Regressor: A tree-based model that recursively splits the data into subsets to make predictions.

Support Vector Machine (SVM) Regressor: A powerful model that finds the hyperplane that best separates the data points, optimizing for the maximum margin.

7. Model Evaluation and Selection: Each of the above models was trained and evaluated on the dataset. Evaluation metrics such as Mean Squared Error (MSE) or Root Mean Squared

Error (RMSE) were used to assess the models' performance. By comparing the evaluation results, the model that demonstrated the best performance in terms of accuracy and generalization was selected.

8. Hyperparameter Tuning: After selecting the Decision Tree Regressor as the best model, hyperparameter tuning was performed to further optimize its performance.

Hyperparameters are configuration settings that are not learned during the training process. Techniques such as Grid Search or Random Search were used to search through different combinations of hyperparameters, aiming to find the best set of values that yield the highest model performance.

9. Conclusion: In conclusion, the project successfully implemented regression analysis and machine learning techniques to build a model capable of predicting car prices based on various features. Through thorough data preprocessing, outlier removal, encoding, scaling, and model selection, the Decision Tree Regressor was identified as the best performing model. With hyperparameter tuning, the model's performance was further optimized, providing accurate car price predictions. The developed model could have valuable real-world applications in the automotive industry, assisting both sellers and buyers in estimating car prices based on their features.