Car Price Prediction Model Project Report

Executive Summary:

The Car Price Prediction Model project aims to develop a machine-learning model capable of predicting car prices based on various features. The project utilizes multiple algorithms, including Random Forest Regressor, Linear Regression, Extra Trees Regressor, and CatBoost Regressor, to evaluate their performance in predicting car prices accurately.

Project Objectives:

Develop a machine learning model for predicting car prices.
Evaluate the performance of different regression algorithms.
Hyperparameter tuning using RandomizedSearchCV for the RandomForestRegressor.
Implement the CatBoostRegressor and assess its predictive accuracy.

Methodology:

Data Collection and Preprocessing:

- The dataset, containing car information, was loaded from the 'audi.csv' file.
- Exploratory Data Analysis (EDA) was conducted using both Pandas Profiling and manual methods to understand the data.

Feature Engineering:

- Selected features for prediction include 'model', 'year', 'mileage', 'fuelType', 'transmission', 'engineSize', etc.
- Label encoding and one-hot encoding were performed to convert categorical variables into numerical representations.

Model Development:

- Standardisation of features using StandardScaler.
- Splitting the dataset into training and testing sets.
- Created regression models using RandomForestRegressor, Linear Regression, Extra Trees Regressor, and CatBoost Regressor.

Hyperparameter Tuning:

- Utilized RandomizedSearchCV for hyperparameter tuning on the RandomForestRegressor model.

Model Evaluation:

- Evaluated the models using metrics such as R2 Score, Mean Absolute Error (MAE), and Mean Squared Error (MSE).

Data Analysis:

Dataset Overview:

- The dataset consists of [number of rows] rows and [number of columns] columns.
- Features include 'model', 'year', 'mileage', 'fuelType', 'transmission', 'engineSize', etc.

Exploratory Data Analysis:

- Conducted automated EDA using Pandas Profiling.
- Checked data types, null values, and basic statistics of the dataset.

Model Evaluation:

- Random Forest Regressor: R2 Score [R2_SCORE], MAE [MAE_RF], MSE [MSE_RF].
- Linear Regression: R2 Score [R2_SCORE_LR], MAE [MAE_LR], MSE [MSE_LR].
- Extra Trees Regressor: R2 Score [R2_SCORE_ET], MAE [MAE_ET], MSE [MSE_ET].
- CatBoost Regressor: R2 Score [R2_SCORE_CAT], MAE [MAE_CAT].

Results:

- Random Forest Regressor achieved an R2 Score of [R2_SCORE_RF] and Mean Absolute Error of [MAE_RF].
- Linear Regression achieved an R2 Score of [R2_SCORE_LR] and Mean Absolute Error of [MAE_LR].
- Extra Trees Regressor achieved an R2 Score of [R2_SCORE_ET] and Mean Absolute Error of [MAE_ET].
- CatBoost Regressor achieved an R2 Score of [R2_SCORE_CAT] and Mean Absolute Error of [MAE_CAT].

Recommendations:

Model Selection: Consider using the Random Forest Regressor or CatBoost Regressor for accurate car price predictions.

Feature Importance: Analyze feature importance to understand which features significantly impact the predictions.
Data Enhancement: Collect additional data to improve model generalization and enhance predictive accuracy.

Limitations:

- The quality and quantity of available data may influence the model's predictions.
- Predictive accuracy may vary based on the choice of regression algorithm.

Future Work:

- Explore ensemble methods to combine predictions from multiple models.
- Incorporate user feedback for continuous model improvement.
- Evaluate the model's performance on a larger and more diverse dataset.

Conclusion:

The Car Price Prediction Model project successfully develops and evaluates machine learning models for predicting car prices. By employing various regression algorithms and conducting hyperparameter tuning, the project provides insights into the performance of different models in the context of car price prediction.