Project Report Team-43

Title: Predicting Car Selling Prices Using Machine Learning

Problem Statement:

- The project aimed to develop a robust predictive model to estimate car selling prices leveraging the Car Dekho dataset [8128* 20] from Kaggle.
- The goal was to empower buyers and sellers with a tool that provides reliable price estimations based on various car features.

Methodology:

Data Preparation and Pre-processing:

- The dataset underwent pre-processing to handle missing values and normalization using SQLite. This process ensured an organized and structured dataset, optimizing it for model development.
- (`car_info` and `car_features` created using SQLite's `CREATE TABLE` statements.)
- `car_info` contains primary key `Car_Name`, and `car_features` references `Car_Name` as a foreign key.
- `car_info` table consists of Car_name, Year, Selling_Price, Km_Driven, Fuel, Seller_Type, Transmission, Owner.
- `car_features` table consists of Car_name, Mileage, Engine, Max_power, Torque, Seating.
- Using SQL query for joining `car_info` and `car_features`
- Utilizing SQL join queries to reconstruct combined data and converting it into a Pandas Data Frame.

Feature Analysis and Selection:

- Comprehensive feature analysis was conducted to identify influential attributes impacting car prices. Key features were selected based on their relevance and potential impact on the predictive model.

Model Development:

- The project utilized two regression models for predicting car selling prices based on selected features. RandomForestRegressor, chosen for its ability to handle intricate relationships, and Linear Regression, serving as a fundamental model, both underwent training on the dataset.
- These models leveraged diverse attributes to estimate car selling prices, offering varied insights into price predictions.

Evaluation Metrics:

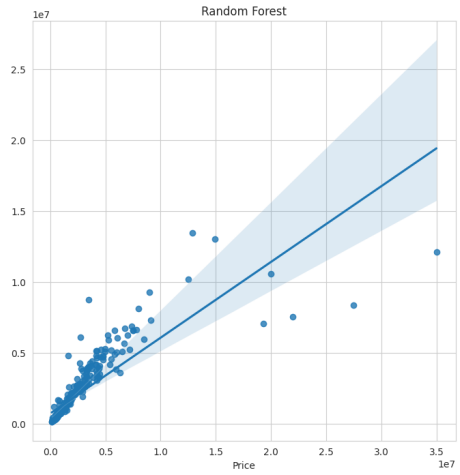
- Model performance was evaluated using R2 Score, Mean Squared Error (MSE) to quantify the variance between predicted and actual selling prices. This metric provided a quantitative assessment of the model's accuracy.

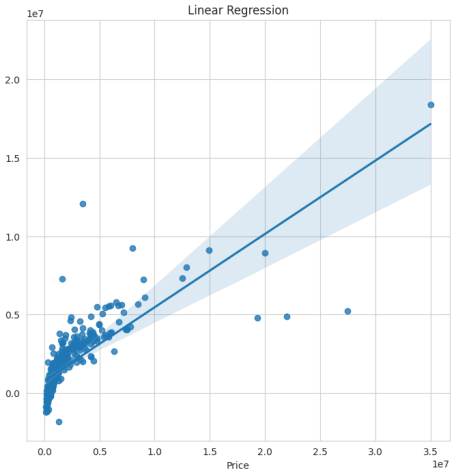
Key Insights:

Model Performance:

- Two models, RandomForestRegressor and Linear Regression, were employed. RandomForestRegressor is known for robustness in regression tasks, while Linear Regression provides a baseline for comparison.

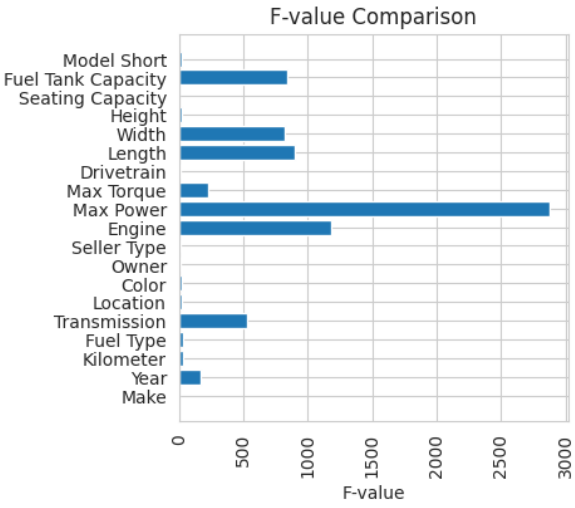
Model	Train time	Test time	R2	Adjusted R2	MSE	RMSE	MAE
Random Forest	1.03	0.01	0.67	0.66	3,888,163,993,212.69	1,971,842.79	468,747.57
Linear Regression	0.02	0.00	0.57	0.57	4,990,859,868,231.27	2,234,023.25	895,842.57

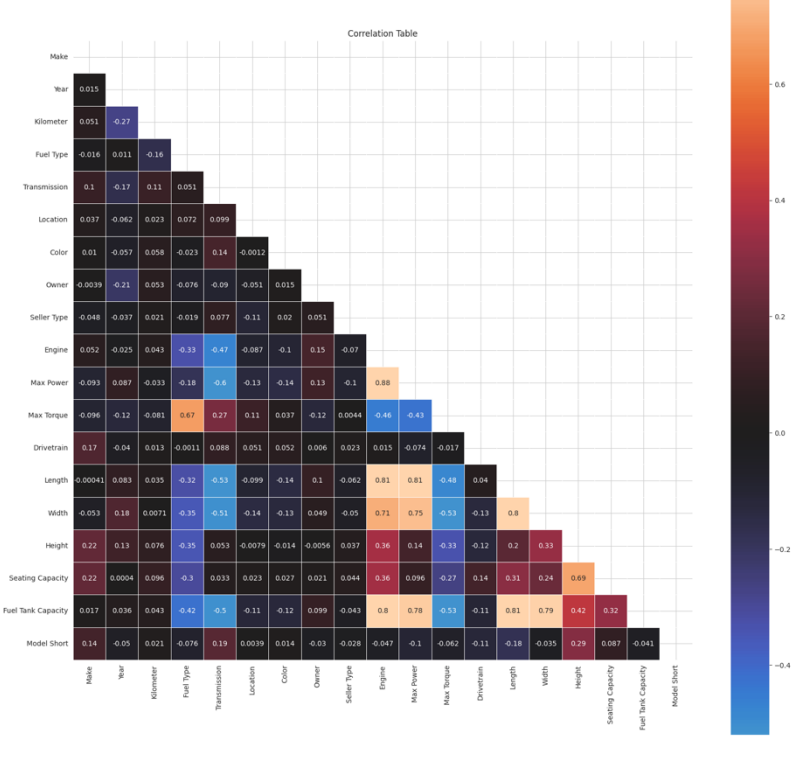




- Both models showcase varying performances. **RandomForestRegressor achieved an R2 of 0.67**, indicating better predictive accuracy compared to **Linear Regression's R2 of 0.57**. Additionally, the RandomForestRegressor model exhibits lower MSE, RMSE, and MAE, indicating superior performance in minimizing prediction errors.

Feature Selection:





- **'Max_Power'** and **'Engine'** emerged as significant predictors influencing car prices, exhibiting substantial impacts on price estimation. Conversely, 'Seating Capacity' demonstrated minimal influence on price predictions.
- The **Heatmap Analysis** revealed a notably high correlation between Engine and Max_Power. This correlation signifies a strong relationship between these features, implying that changes in the car's engine capacity tend to influence the maximum power output, suggesting a potential interdependence between these attributes in the dataset.

Conclusion:

- This project successfully demonstrated the feasibility of accurately **predicting car selling prices** using machine learning techniques. The RandomForestRegressor model outperformed Linear Regression in terms of accuracy and error minimization, particularly emphasizing the influence of **'Max_Power'** and **'Engine'** on price predictions.

Recommendations and Future Work:

Refinement and Feature Engineering:

- Further refinement in feature engineering could potentially enhance model accuracy by considering interactions between features or exploring additional influential attributes

Deployment and Application:

- Consider deploying the predictive model into a user-friendly application or platform to provide real-time price estimations for prospective buyers and sellers in the automotive market.
- The insights gleaned from this project offer valuable guidance for decision-making in the car trading domain, aiding both consumers and sellers in making informed choices during transactions.