Used Car Price Prediction

Faizan, Sonu, Yashwant, Sankalp

Problem Statement

- In 2022, the automotive manufacturing industry ranked 9th globally, with the U.S. automotive sector enjoying steady growth over the previous decade, marked by an annual sales increase of nearly 5%.
- The COVID-19 pandemic disrupted the industry, leading to a shift in consumer preferences towards private transportation and a surge in the second-hand car market due to economic challenges.
- To address this, a predictive model is being developed to estimate the selling price of used cars in the U.S., using factors such as manufacturer, model, mileage, and year.
- The model aims to provide accurate pricing predictions, improving transparency and efficiency in the used car market by reducing information asymmetry.

Data Characteristics

Independent variables:

- Manufacturer
- Model
- Prod. year
- Category
- Fuel type
- Mileage
- Cylinders
- Gear Box
- Airbags

} Strings/ Categorical

- Manufacturer
- Model
- Category
- Fuel type
- Gear box type
- Drive wheels

} Numerical

- Prod. year
- Mileage
- Cylinders
- Airbags

ID	Price	Levy	Manufacturer	Model	Prod. year	Category	Leather interior	Fuel type	Engine volume	Mileage	Cylinders	Gear box type	Drive wheels	Doors	Wheel	Color	Airbags
45654403	13328	1399	LEXUS	RX 450	2010	Jeep	Yes	Hybrid	3.5	186005	6	Automatic	4x4	04-May	Left wheel	Silver	12
44731507	16621	1018	CHEVROLET	Equinox	2011	Jeep	No	Petrol	3	192000	6	Tiptronic	4x4	04-May	Left wheel	Black	8
45774419	8467	-	HONDA	FIT	2006	Hatchback	No	Petrol	1.3	200000	4	Variator	Front	04-May	Right-hand drive	Black	2
45769185	3607	862	FORD	Escape	2011	Jeep	Yes	Hybrid	2.5	168966	4	Automatic	4x4	04-May	Left wheel	White	0
45809263	11726	446	HONDA	FIT	2014	Hatchback	Yes	Petrol	1.3	91901	4	Automatic	Front	04-May	Left wheel	Silver	4
45802912	39493	891	HYUNDAI	Santa FE	2016	Jeep	Yes	Diesel	2	160931	4	Automatic	Front	04-May	Left wheel	White	4

[snippet from dataset]

Model Construction

Decision Tree Regressor:

- Chosen for its interpretability and the fact that some independent variables are categorical. The regressor type was chosen because the dependent variable/feature variable is continuous.

Random Forest Regressor:

- Chosen to further improve and advance the Decision Tree Regressor. The regressor type was again picked because of the numerical nature of the dependent variable.

Evaluation Metrics

The following evaluation metrics were used:

- Means Squared Error:
- Root Mean Squared Error
- Mean Absolute Error
- R2 score

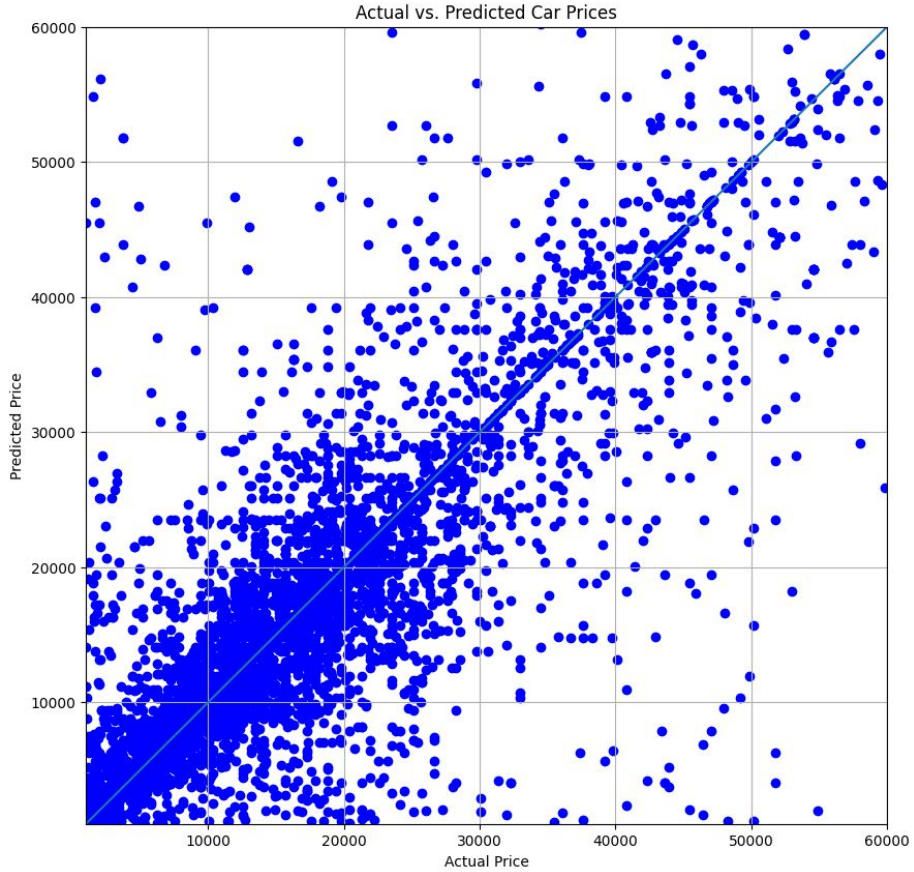
Decision Tree Regressor:

```
Mean Squared Error: 116412941.91
Root Mean Squared Error: 10789.48
Mean Absolute Error: 5233.85
R-squared (R2) Score: 0.60
```

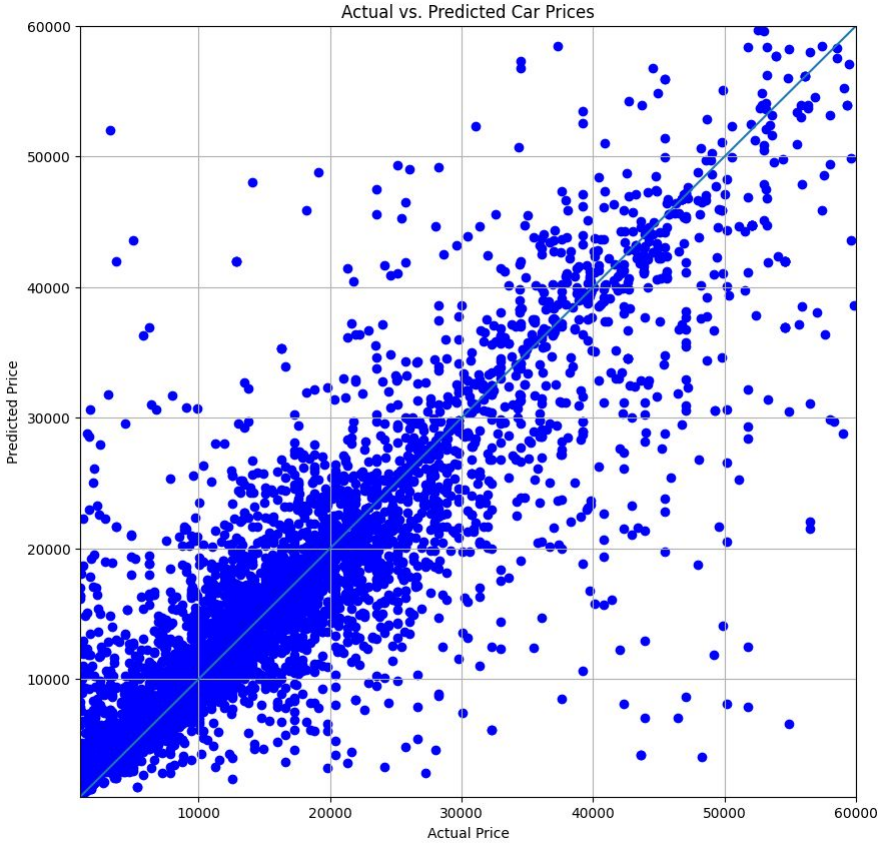
Random Forest Regressor:

```
Mean Squared Error: 85668865.85
Root Mean Squared Error: 9255.75
Mean Absolute Error: 4484.61
R-squared (R2) Score: 0.70
```

Decision Tree Regressor



Random Forest Regressor



Individual Contribution

Identifying dataset: Sankalp

Processing dataset: Yashwanth

Coding: Faizan, Sonu, Yashwanth

Report/Slides: Sankalp, Faizan, Sonu, Yashwanth