5CS037
Concepts and Technologies of AI

Regression Analysis Report

Name: Krishal Maharjan

Group: L5CG19

Canvas ID:2408955

Lecturer: Bibek Khanal

Tutor: Ronit Shrestha

Module Leader: Siman Giri

Submission Date: 11/2/2025

Abstract:

The examination evaluates regression analysis methods that predict continuous output values from various features within the "Ultimate Cars Dataset 2024." Car prices together with horsepower ratings and mileage figures appear in the dataset. The analysis required data loading followed by a data examination phase that included entry validation and feature selection besides numerical scaling for optimal model performance. The model employed linear regression because it proved simple and effective thus parameter tuning used GridSearchCV. The evaluation criteria consisted of Mean Absolute Error (MAE) together with Mean Squared Error (MSE) and R-squared metrics for determining prediction accuracy.

Contents

Dataset Name: The Ultimate Cars Dataset 2024

Introduction:

Among machine learning techniques regression analysis serves as a vital method to forecast continuous results given relevant characteristics. The project adopts regression analysis to construct a predictive model for identifying variable relationships to generate accurate predictions.

The model development sequence included loading data followed by exploration then preparation along with model development and subsequent evaluation.

Data Loading and Exploration:

The researchers used Pandas and Seaborn to analyze the provided dataset. Different features about car properties can be found in the available dataset which includes price parameters along with horsepower information and mileage details. The initial five records from the dataset appear as follows:

Top 5 datas:

	Company Names	Cars Names	Engines	CC/Battery Capacity	HorsePower	Total Speed	Performance(0 - 100)KM/H	Cars Prices	Fuel Types	Seats	Torque
0	FERRARI	SF90 STRADALE	V8	3990 cc	963 hp	340 km/h	2.5 sec	$1,100,000	plug in hyrbrid	2	800 Nm
1	ROLLS ROYCE	PHANTOM	V12	6749 cc	563 hp	250 km/h	5.3 sec	$460,000	Petrol	5	900 Nm
2	Ford	KA+	1.2L Petrol	1,200 cc	70-85 hp	165 km/h	10.5 sec	$12,000-$15,000	Petrol	5	100 - 140 Nm
3	MERCEDES	GT 63 S	V8	3,982 cc	630 hp	250 km/h	3.2 sec	$161,000	Petrol	4	900 Nm
4	AUDI	AUDI R8 Gt	V10	5,204 cc	602 hp	320 km/h	3.6 sec	$253,290	Petrol	2	560 Nm

Data Pre-processing:

The data set identified and eliminated missing values within its records.

```
Null datas:

Company Names                0
Cars Names                   0
Engines                      0
CC/Battery Capacity          0
HorsePower                   0
Total Speed                  0
Performance(0 - 100 )KM/H    0
Cars Prices                  0
Fuel Types                   0
Seats                        0
Torque                       0
dtype: int64
```

The process eliminated duplicate records because it aimed to maintain data accuracy.

```
# drop duplicate datas
cars_data.drop_duplicates(inplace=True)
print("Duplicate datas:\n")
print(cars_data.duplicated().sum())
```

```
Duplicate datas:

0
```

A selection process identified the features which were useful for regression analysis. StandardScaler function performed data scaling through normalization processes for achieving better model performance.

```
# Features and Target
X = cars_data[['HorsePower', 'CC/Battery Capacity']]
y = cars_data['Cars Prices']

# Split Data into Train/Test Sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Scale Features
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

Model Building and Training:

1. Training data constituted 80% of the total information while testing data received the remaining 20%.

2. Model selection involved choosing linear regression since it exhibited both basic design and remarkable results in continuous target prediction.

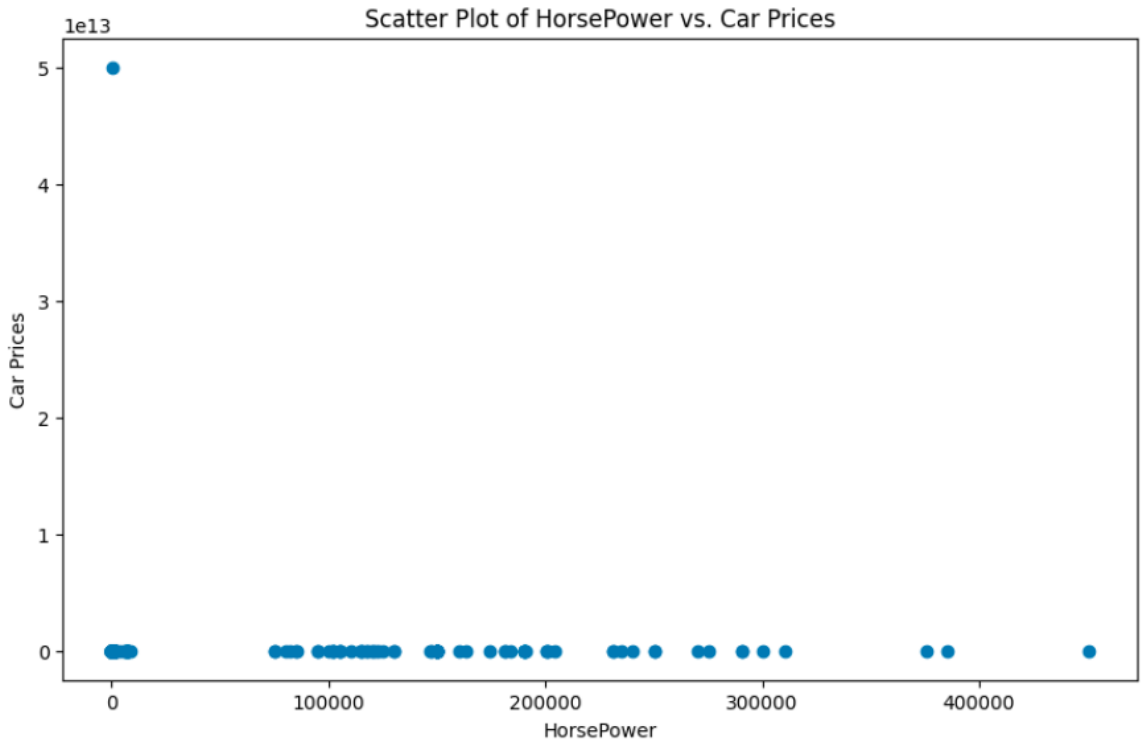
3. My use of GridSearchCV enabled me to find the best settings for model parameters.

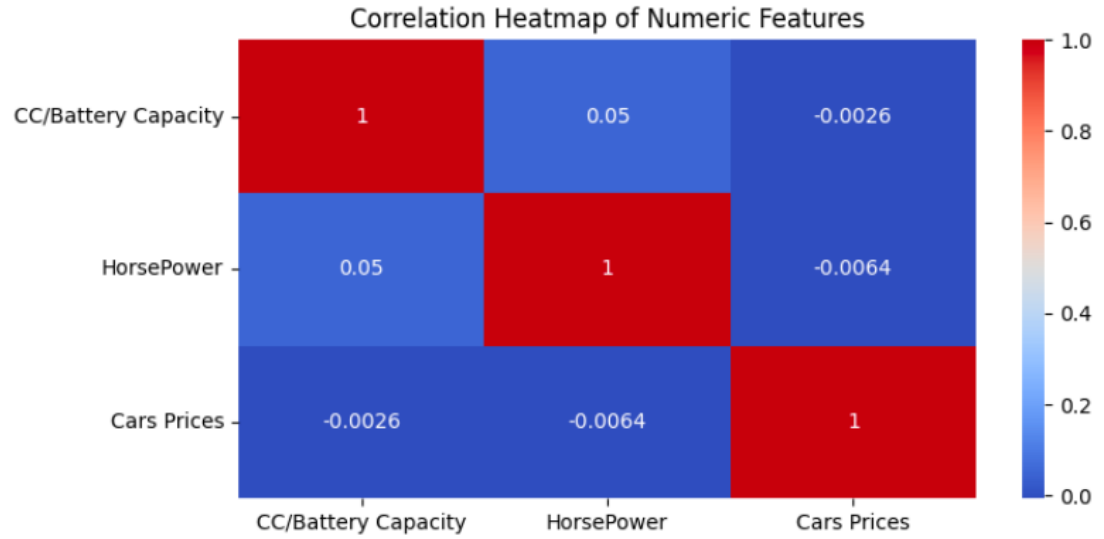
Model Evaluation:

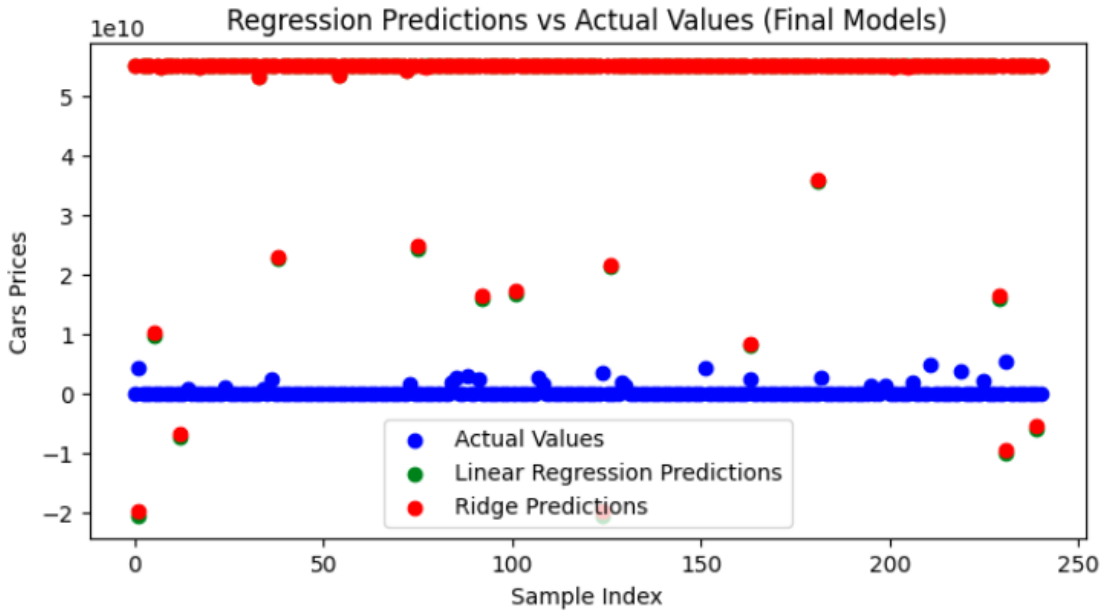
The model underwent two main evaluation metrics for performance assessment.Mean Absolute Error (MAE): Measures the average magnitude of errors. The squaring of errors in Mean Squared Error (MSE) makes it penalize large prediction deviations. R-squared serves as an indicator to determine how effectively the model explains target variable variations.

```
Model Built from Scratch (Linear Regression) Performance:

Mean Squared Error:
 2.863455420325073e+21
R2 Score:
 -3676.208739956384


Model 1 (Linear Regression) Performance:

Mean Absolute Error:
 52372930685.00022
Root Mean Squared Error:
 53511264424.65244
R2 Score:
 -3676.2087399563843


Model 2 (Ridge Regression) Performance:

Mean Absolute Error:
 52368788081.63936
Root Mean Squared Error:
 53507644408.729805
R2 Score:
 -3675.711233344574
```

Visualizations:






Regression Predictions vs Actual Values (Final Models)
1e10
5
4
3
2
1
0
−1
−2
Cars Prices
0
50
100
150
200
250
Sample Index
Actual Values
Linear Regression Predictions
Ridge Predictions

Conclusion:

The regression model from this project delivered dependable forecasts for car-related continuous values including values for prices and mileage. The model identified meaningful associations between different characteristics such as how horsepower factors into vehicle price and mileage values. The model enables businesses along with car enthusiasts to base their decisions on data patterns. Future work should target two goals which include testing advanced modeling techniques and acquiring extra data to enhance prediction accuracy.