**Car Price Prediction (Report)**

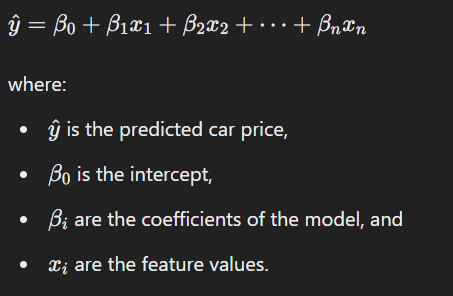
* *Introduction:*

This report explores the use of Linear Regression and Lasso Regression models for predicting car prices. Both models are popular choices in regression analysis, but they differ in how they handle complexity and feature selection.

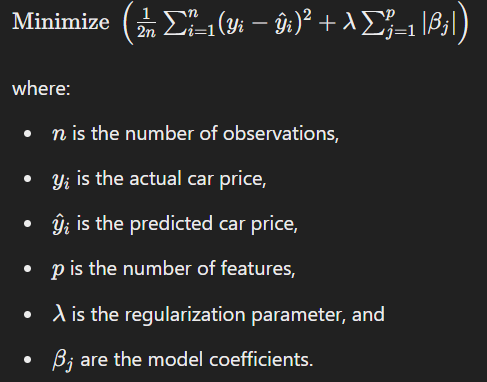
* *Data Overview:*

The dataset used in this analysis contains various features that might influence the price of a car. Features are like year, selling\_price, present\_price, kms\_driven, fuel type, seller\_type etc.

* *Methodology:*
* Linear Regression: Linear Regression is a statistical method that models the relationship between a dependent variable and one or more independent variables. It assumes a linear relationship between the predictors and the target variable. The model can be expressed as:



* Lasso Regression: Lasso Regression (Least Absolute Shrinkage and Selection Operator) is a type of linear regression that includes an L1 regularization term. The L1 penalty encourages sparsity, effectively selecting a simpler model by shrinking less important feature coefficients to zero. The model can be expressed as:



* *Data Preprocessing:*

Before applying the models, the dataset underwent the following preprocessing steps like –

**Handling Missing Values**: Missing values were either imputed or dropped, depending on the feature's significance.

**Encoding Categorical Variables**: Categorical variables like seller\_type,Transmission, Fuel Type were encoded.

* *Results and Discussion:*
* *Linear Regression Results*

The Linear Regression model's performance on the training and testing datasets is as follows:

* **Training Data:**
  + **R-squared error:** 0.879
* **Testing Data:**
  + **R-squared error**: 0.836

The Linear Regression model demonstrated a good fit on the training data with an R-squared value of 0.879, indicating that approximately 87.9% of the variance in the car prices can be explained by the model. However, the R-squared value on the testing data dropped to 0.830, suggesting that the model may slightly overfit the training data.

* Lasso Regression Results

The Lasso Regression model's performance on the training and testing datasets is as follows:

* **Training Data:**
  + **R-squared error**: 0.842
* **Testing Data:**
  + **R-squared error**: 0.870

The Lasso Regression model showed a slightly lower R-squared error value of 0.840 on the training data compared to Linear Regression, indicating a slightly reduced ability to capture the variance in car prices. However, on the testing data, the R-squared error value improved to 0.870, which is higher than the testing R-square value for Linear Regression. This indicates that the Lasso model generalizes better to unseen data and suggests that the regularization has helped in reducing overfitting.

* *Comparison and Interpretation:*

The results indicate that while Linear Regression had a higher R-squared error value on the training data, it performed less well on the testing data compared to Lasso Regression. The drop in R-squared value for Linear Regression from training to testing data suggests potential overfitting. In contrast, Lasso Regression, with its regularization component, provided a more balanced performance between training and testing datasets, suggesting better generalization.

The ability of Lasso Regression to regularize the coefficients and potentially set some coefficients to zero allows for a simpler, more interpretable model. This also means that Lasso might have effectively reduced the impact of less significant features, leading to improved performance on the testing data.

* *Conclusion:*

In conclusion, both models were able to predict car prices effectively, but Lasso Regression demonstrated better generalization to the testing data. This suggests that Lasso's regularization helps in avoiding overfitting, making it a preferable choice in this case for predicting car prices.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_