**MINI PROJECT REPORT**

**Car Resale Value Prediction using Regression**

| **NAME** | **ROLL NO** |
| --- | --- |
| Shail Pancholi | 33253 |
| Natanshi Sharma | 33251 |
| Gargi Meshram | 33247 |

**Abstract:**

The market for used cars is a changing and complex system influenced by many factors, which makes it difficult for buyers and sellers to determine the fair value of vehicles. In this project we aim to solve this problem by using regression analysis to predict the prices of used cars.

To start we gather a dataset that includes important details about different car models. This dataset covers attributes like fuel type, seller type and transmission. We use data preprocessing techniques to handle categorical variables and prepare the data for analysis. This thorough preparation process is crucial in building predictive models.

Next we divide the dataset into training and testing sets so that we can train our models and evaluate their performance. We explore two regression models. Linear Regression and Lasso Regression. To provide both buyers and sellers with a way to confidently estimate car prices. This will enhance transparency and efficiency in the car market.

To support our analysis we create visualizations such as scatter plots that clearly show the relationship between actual prices and predicted prices. We use the R error metric for evaluation, which demonstrates that Lasso Regression performs better than Linear Regression in terms of predictive accuracy.

This project provides a tool for individuals involved in the second hand car industry giving them the knowledge necessary to make well informed choices regarding the pricing of previously owned vehicles.

**Objectives:**

1. Develop a Robust Prediction Model: The primary objective of this project is to create a reliable predictive model for estimating used car prices. By using regression techniques, we aim to provide a tool that can accurately predict the market value of pre-owned vehicles.
2. Enhance Data Handling Skills: As an aspect of the project our main emphasis lies in data preprocessing. This involves tasks such as encoding variables and effectively managing missing values. The objective here is to showcase effective data preparation techniques that are crucial for building accurate machine learning models.
3. Model Comparison: We aim to compare the performance of two regression models, Linear Regression and Lasso Regression, in predicting used car prices. This objective allows us to determine which model offers superior accuracy in the context of the used car market.
4. Empower Buyers and Sellers: Our final goal is to empower participants in the used car market, be it buyers or sellers, with a tool that enhances transparency and efficiency. We intend to provide a means for users to confidently estimate the fair value of pre-owned vehicles, ultimately contributing to informed decision-making.

**Software Requirements:**

1. **Python 3.11:** The project is implemented in Python, making Python 3.11 an essential requirement. Although the version can vary, it's recommended to have the latest version to access the latest features and security updates.
2. **IDE:** While you can use any text editor, it's advisable to have an Integrated Development Environment like Jupyter Notebook, Visual Studio Code. They streamline code writing, execution, and visualization, enhancing productivity.
3. **Python Libraries:**

a) Pandas: This library is indispensable for data manipulation, analysis, and preprocessing tasks.

b) Matplotlib: To create static, animated, or interactive visualizations in Python.

c) Seaborn: An additional data visualization library that complements Matplotlib with a higher-level interface.

d) Scikit-Learn: Essential for machine learning tasks, as it provides access to a wide range of machine learning models. In this project, Linear Regression and Lasso Regression from Scikit-Learn are used.

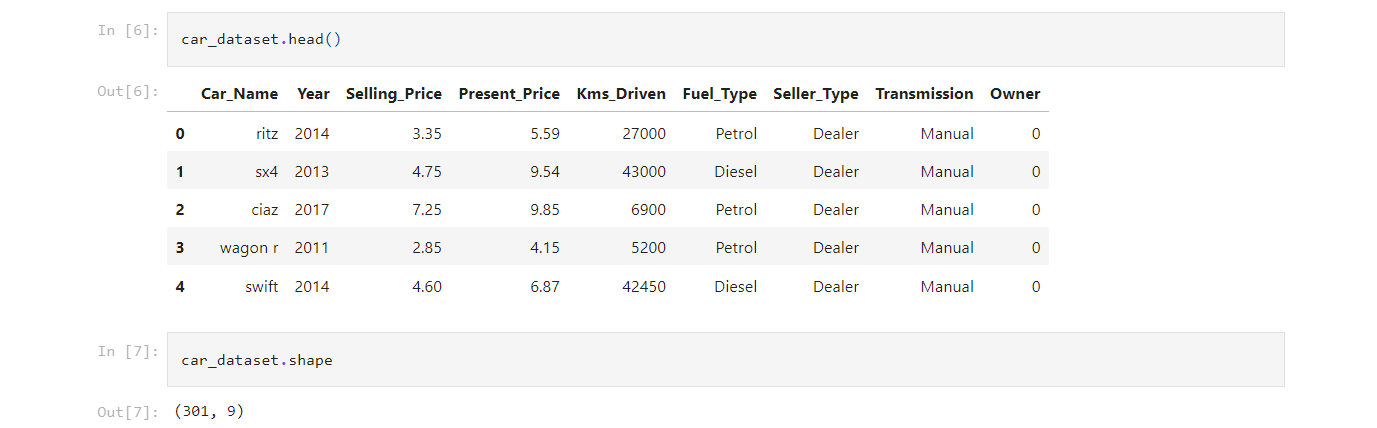
Ensure you have these libraries installed. Use pip, the Python package manager, to install any missing libraries.

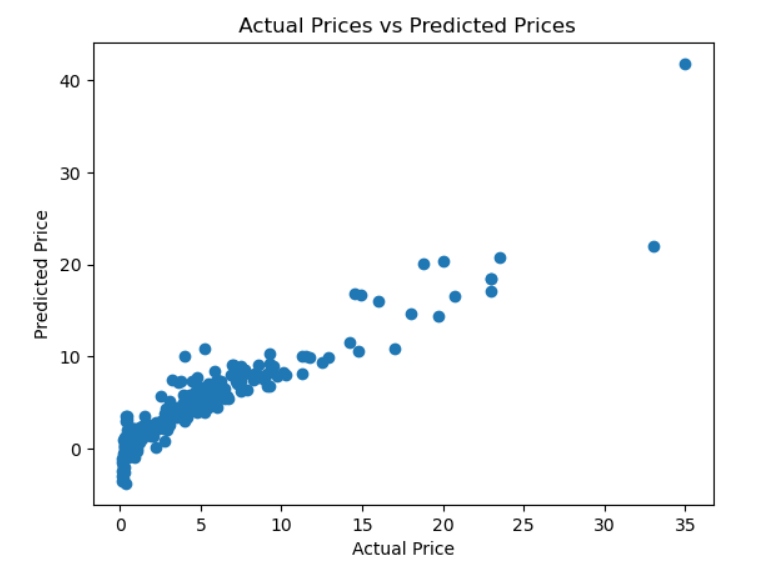
1. **Dataset:** Download the "car data.csv" dataset containing information about used cars. Ensure it's in the project directory or specify the correct file path in the code.
2. **Operating System:** This project is platform-independent and can be executed on Windows, macOS, or Linux.

**System Architecture:**

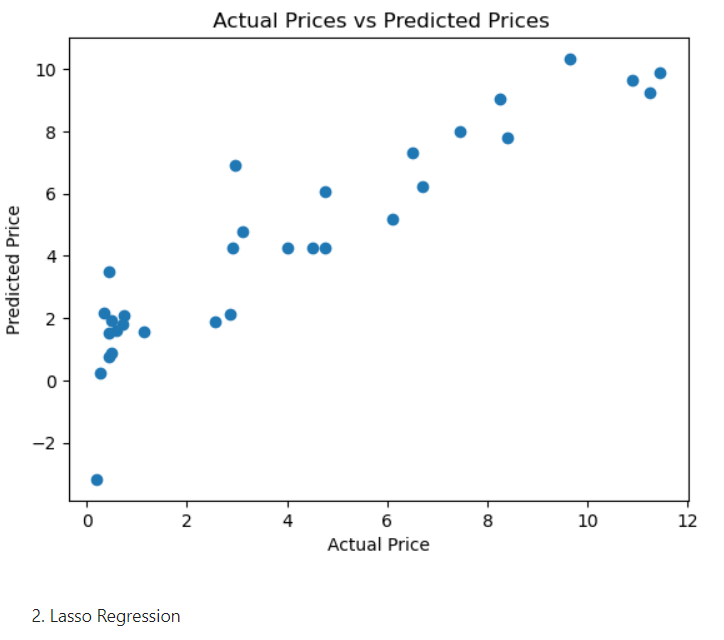
1. Data Source:
2. The project relies on a dataset containing information about used cars.
3. The dataset can be in CSV format and should include columns for car attributes like year, kilometers driven, fuel type, seller type, and more.
4. Data Preprocessing:
5. Data preprocessing involves cleaning, transforming, and encoding the dataset.
6. Common tasks include handling missing data, converting categorical data into numerical format, and normalizing features.
7. Machine Learning Models:
8. The project uses two machine learning models: Linear Regression and Lasso Regression.
9. Linear Regression is used to establish a baseline model for predicting car prices.
10. Lasso Regression helps to improve prediction accuracy by adding L1 regularization.
11. Training and Testing:
12. The dataset is split into training and testing subsets (typically 90% for training and 10% for testing) using tools like Scikit-Learn's train\_test\_split.
13. Model Training:
14. Both Linear and Lasso Regression models are trained on the training dataset.
15. Model Evaluation:
16. The R-squared error metric is used to evaluate model performance on both the training and testing datasets.
17. Matplotlib and Seaborn are used to create scatter plots for visualizing the predicted prices against actual prices.
18. Results and Visualization:
19. The project presents the evaluation results, including the R-squared error score and scatter plots to visualize the model's predictions.
20. These results help assess the model's accuracy in predicting used car prices.
21. User Output:
22. The system can provide output to users or stakeholders.
23. Users can use this information to estimate the selling price of a used car based on the input attributes.
24. Software and Tools:
25. The project relies on Python 3.x, Jupyter Notebook or Anaconda, and various libraries like Pandas, Matplotlib, Seaborn, and Scikit-Learn.
26. Operating System:
27. The project is designed to run on multiple platforms, including Windows, macOS, and Linux.

**Screenshots**

****

****

1. Linear Regression

****

**References**

1. <https://github.com/pancholi-s/car-price-estimation>