**SMA 390:CAT 2 PROJECT**

**Car Price Prediction Model**

**Group Members**

1. Kiunga Vivian Sifa- I163/4448/2021
2. Awuor Brenda Ochieng – I163/7220/2021

Project Link:

**1.Introduction**

**Problem Statement:** The process of determining accurate and reliable car prices in the automotive market is often challenging. Buyers and sellers face difficulties in predicting the fair value of a vehicle, considering various factors such as model, year, mileage, condition, make and market trends. Existing methods may lack precision, leading to inaccurate approximations and potential financial loss.

**Objective of the Project:** The goal of this project aims to address the problem of the lack of a robust car price prediction simulation. It aims to create an efficient and effective model that will be able to predict the price of used cars using Linear Regression algorithm.

**2.DATASET**

**Dataset description:** The dataset used to train and test the model is comprised of historical data on used automobile sales.  
The features include:  
**Mileage** is the total distance travelled by the vehicle.   
**Age**: The car's age in years.   
**Condition**: variable that represents the car's general condition (great, good, fair, or poor).   
**Model**: The make and model of the automobile.

**Data pre-processing:** 1. Handling missing values: Missing values are estimated using appropriate methods such as mean, median, or mode.  
2. Encoding categorical variables: Categorical variables, such as 'Condition' and 'Model', are encoded using one-hot encoding.   
3.Feature scaling: Continuous features such as miles and age are scaled to provide consistency in feature magnitudes.

**Model Building:** **Linear regression** was chosen as the predictive modelling technique since it is simple and easy to understand. The dataset is divided into training and testing sets with scikit-learn routines such as train\_test\_split.   
The model is trained on a subset of the dataset and validated with techniques such as k-fold cross-validation to verify generalizability.

**Linear Regression Estimator**

Learning Algorithm

Training Set

Predictive Model

Testing Set

Predicted Results

**3.IMPLEMENTATION**

**Pseudo Code**

**Step 1:** Import the required packages

**Step 2:** Download the dataset and link it to google colab

**Step 3:** Read Dataset

**Step 4:** Data cleaning and processing

**Step 5:** Saving the cleaned data set

**Step 6:** Start training the Machine Learning Model

**Step 7:** Split features and target X and Y respectively

**Step 8:** Split the data into training data and testing data

**Step 9:** Train the model with training data and Testing data

**Step 10:** Applying Linear Regression to the model

**Step 11:** Fit the Linear Regression model

**Step 12:** Applying Lasso Regression to the model

**Step 13**: Fit Lasso Regression model

**Step 14:** Reading the model and dataset and making the prediction using python

**4.PREDICTION**

**Results and Interpretations:**   
The trained model gives information about the relationship between automotive attributes and price reductions. The coefficients of the linear regression model represent the impact of each feature on price reduction. The interpretation of coefficients aids in determining which features contribute most significantly to price reduction.

**5.CONCLUSION**

The linear regression-based car price reduction model offers a framework for estimating used car price reductions. Further modification and optimization of the model may result in higher accuracy and predictive capacity.   
The model can be applied in real-world scenarios to help vehicle retailers and customers make informed judgments about pricing and acquiring used cars.