# 1. Introduction

Regression can be defined as a statistical technique that allows us to study and identify how one or more features relate to one another. Regression analysis is aided by a technique that adapts to assist you to discover which aspects of your data are essential, as well as which aspects are irrelevant.

There are thousands of used cars that are being sold every year in the UK. The used cars are usually posted on platforms like Gumtree, Facebook marketplace, and other platforms. When a new person decides to sell his used vehicle, the biggest confusion occurs on pricing. Unknowingly people ask for overpricing for their vehicles or value under-price their vehicles in most cases. To avoid this confusion based on all the ads that were posted we can build a regression model that automatically predicts the price of the vehicle based on certain features. This would be helpful for the new sellers in the market and also will help buyers by avoiding them paying overprice for a vehicle.

For the regression analysis, I will be using multiple techniques of regression models that are available in Machine Learning and Deep Learning. Techniques such as Linear Regression, Lasso Regression, Ridge Regression, Support Vector Regression, Decision Tree Regression, Recurrent Neural Networks, etc., are the models that we will be using as part of this project.

I will be using Python language and Jupyter IDE for the whole project. To successfully completing the project I will be using Python libraries such as NumPy, Pandas, Scikit Learn, etc.

## 1.1. Research Questions

* How does traditional classification algorithms such as SVM and decision trees perform for regression tasks compared to the regular regression algorithms such as linear regression?
* Regression models in Machine Learning and Deep Learning which of these models are capable of accurately predicting the prices of the vehicles?

## 1.2. Objectives

* Clean the data by removing the outliers and handling the missing values.
* Make interpretations from the data visualization during the data analysis.
* Use feature engineering to extract new useful features from the data available.
* Use Regression Analysis to accurately predict the prices of the cars to help the customers correctly price their vehicles on selling platforms.
* Evaluate the regular regression models against the deep learning model and the traditional classification models for regression task.

## 1.3. Dataset

The data is readily available on the Kaggle platform for the use of research purposes. I will be using this data to perform the regression analysis on the data. This data was scraped from the listings of the used cars. There are different cars from different manufacturers are listed in the dataset. In total there are close to 100k data points in the data.

This data will be further divided into train, validation, and test sets in the project for evaluation and building robust models.

# 3. Ethical, Social, and Legal Issues

## 3.1. Ethical Issues

Ethical issues are those issues that let us understand what is fundamentally right and wrong. The ethical issues that are possible in this case of the project are as follows:

Machine Ethics & Algorithmic Biases:

This issue under the ethical OS toolkit helps us understand right and wrong while building the machine learning algorithms. In our project, we use Machine Learning for training the regression models and the model might behave biased due to the presence of outliers in the data. This needs to be checked and the outliers need to be removed from the data to avoid such issues of bias in the models.

## 3.2. Legal Issues

Legal issues are those issues that let us understand what is legally right and wrong. The legal issues that are possible in this case of the project are as follows:

The predictions that are being passed from the algorithms to the users should be accurate and should not have a drastic difference in the original value of the cars being sold and the predicted value. Any misclassification of prices could misguide the people to lose their money or face loss. This could be a reason for the legal issues. To avoid this issue the algorithms should be properly trained on the data and should cross-check the performance before production and the predictions should be monitored while the algorithm is in production.

## 3.3. Social Issues

Social issues are the issues that harm society. The social issues that are possible as an outcome of this project are as follows:

Issues related to unemployment as a result of automation can be a possible social issue that might occur as a result of this project.