Capstone Project Proposal
Car Price Prediction Using Linear Regression

Business Understanding

An automobile company aspires to create a linear regression model that uses machine learning to estimate used car prices from a given dataset. My goal is to develop a linear regression model that uses the dataset to predict car prices.

Data Understanding

This project's dataset was obtained from Kaggle and been directly downloaded from there. Kaggle is the website which provides the opportunity to find and publish dataset. This platform is helpful for data scientist to solve data science challenges. The most significant part is that most of the dataset are easily to accessed and be able to download.

Data Preparation

My goal is to understand the dataset properly then make the strategy to solve problems. My first approach will be to import data and then I will create some useful libraries for this dataset for example matplotlib for data visualization. Linear model to create linear regression model for the dataset. Linear Regression will provide preprocessing data, reducing dimensionality, implementing regression, classifying, and clustering. Train_test_split to split data into train and test sets. Before starting training processes, I would do pre-process techniques on the data set. Pre-processing mean cleaning the data set which means if some unwanted columns are present in the data set and that will create no impact on your model you can simply drop those columns from your data set. Finding missing values prior to build model is significant most of the machine learning models provide an error if you pass NaN value into it. To eliminate those errors and get data accuracy it’s better to find any null values prior in the dataset.

Modeling

In this dataset selling price is the dependent variable on other features such as present price, fuel type, transmission and other. Selling price will be my targeted variable whereas all other dependent variable plays important in the dataset to determine how the selling price directly affects. Linear Regression model suits to the problem because regression helps predict a continuous quantity, classification predicts discrete class labels. To be able to predict the values of selling price based on the other dependent variables.

Evaluation

I will be using MAE and MSE metrices for the success of my Capstone project.

Tools/Methodologies

I will be using Random Forest methodologies as it will be good for supervised learning and can be helpful for predicting the used car prices.