

# Machine Learning Engineer Nanodegree

Using Supervised learning we need to predict if a customer can purchase a vehicle looking at its features and decides if spending some amount on a vehicle is meaningful or not.

**Sushma Ponakala**

28<sup>th</sup> June 2018

**Proposal:**

## **Domain Background:**

In today's world we are spending so much of our expenses on purchasing motor bikes and cars. Based on their brands some of them are so expensive. So, when we are affording too much on the vehicles, it is necessary to check and examine all the features of a vehicle. We should check our safety as well. I took a dataset that represents the features of cars. So we should check if a car with specific features and specifications is worth to buy and as well as safe or not and it really matters. Supervised learning deals with two important concepts called Regression and Classification. These concepts help for this prediction.

<https://archive.ics.uci.edu/ml/machine-learning-databases/car/>

The above link describes the data set I picked for making predictions about purchasing a car.

## **Problem statement:**

A dataset with the features of the car is taken and based on the features we have to predict whether purchasing a car is safe as well as worth enough or not.

Features :

overall price

buying price

price of the maintenance

technical characteristics

comfort

number of doors

capacity in terms of persons to carry

the size of luggage boot

estimated safety of the car

Dataset and Inputs:

Sales: Level of Sales

Maintenance: Level of maintenance

Doors: Number of doors for the car

Persons: Capacity of the car (Number of persons in a car)

lug\_boot: Size of the luggage boot

Safety: Safety level of customers

Target: This is our target variable

Overall 6 attributes are considered and dataset description can be viewed here:

<https://archive.ics.uci.edu/ml/machine-learning-databases/car/car.names>

### **Solution Statement**

When a dataset is given, we make predictions about the car i.e. whether a customer purchasing a particular car is quite safe or not. We make predictions if it is worth enough spending such huge amount on car so that customer saves his money from purchasing a less worthy car for huge amounts.

Benchmark model:

Initially we will use logistic regression. There F beta score and accuracy will be taken as reference and we will use other models to get better results for accuracy and F beta score.

### **Evaluation Metrics**

In this we will use the Accuracy, F beta score and others like recall, precision.

We define them as :

$$\text{accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{precision} = TP / (TP + FP)$$

$$\text{recall} = TP / (TP + FN)$$

$$f \text{ beta score} = (1 + \beta^2) * (\text{precision} * \text{recall}) / (\beta^2 * \text{precision} + \text{recall})$$

$$\beta = 0.5$$

TP-total positives and TN -Total negatives

### **Project Design**

I will encode the categorical data into numerical data.

After that I will plot a map of all the features and will come to know how

the features are correlated with each other. I will then split the data into training and testing sets. 80% will be used for training and 20% will be used for testing. Classification model is evaluated on the basis of accuracy, precision, and recall. As mentioned Logistic regression is used as a benchmark model. FBeta score of benchmark model is the reference. I will then test the accuracy of the training and the testing sets and check if the accuracy of training data as well as testing data is almost the same (high).