Data Translation

Introduction: The original data set that we acquired from Kaggle consists of categorical values. Let’s consider a column buying which consist of the values “vhigh”, “vlow”, “med”, “high”, “low” in it. So while considering these categorical values and do task on the data. But we would like to the user to enter details of the car through the dropdown list provided in the frontend. So the basic ideas is to have meaningful relation with the user entered input values and the data set. As of now there is relationship between user and proposed system that we are going to derive with the dataset.

New Data Set: First thing we need when establishing the relationship with the categorical values of original data set and with user is that we need to have a new dataset with required information not in categorical. The car.csv dataset consists of 479 different Models and 35 different make with each models MSRP.

Make is the name of the company that manufactured.

Model is the manufactured company car name that is unique.

MSRP is Manufacturer’s Suggested Retail Price.

And the dataset has some other columns which is also important and may be used later. With required data gathered the next step is to establish the relationship.

How to calculate rate of depreciation and what is depreciation?

Depreciation is the reduced value of an asset with the passage of time in particular to wear and tear. The rate of depreciation tells about the car or an asset value after some years and how much its values is reduced when the asset was actually brought. For example lets consider a car. So the car was brought in 2018. After the purchase itself the car price is reduced 20% of its original value. According to standard depreciation rate system that generally used in market.

For calculating the value of a car after n years, we are using a formula that will calculate the value after some research over present market trend and some general scenarios.

A = P \* (1 – R/100)n

A = value of car after n years

P = Original Value of car

R= Rate of depreciation per annum,

n= No of year the car is used.

So with the above formula the value of the car after using n years can be calculated.

What does the code do?

For now the primary result is to obtain the categorical value of the entered value. For now we made it as command line input values. Make, Model, MSRP and year of car originally brought.

With the input values the target MSRP is found from the dataset and used for calculation of amount of car after n years(A).

The rate of depreciation is considered as 20% for the first year and 10% for ever year. Once the value ‘A’ is obtained. It is used to compare with the user’s value.

Here the value is compared with the user entered amount he wants to sell. The output will be the categorical value of the percentage increase or decrease of the original value(A). And the output will be [veryhigh, high, med, low, verylow], which is categorical.

Conclusion: Finally there is some meaning full relation that is established between with the users input and dataset. It is obtained with the above method.