Project Report

Of

Artificial Intelligence (INT 404)

On

USED CARS PRICE ESTIMATOR

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

By

Sr. No.	NAME	REG. NO.	ROLL NO.
1.	Astha	11815369	37
2.	Devashish Jandgid	11813424	72
3.	Ramandeep Kaur	11803348	57
4.	Mohit	11814437	70

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

LOVELY PROFESSIONAL UNIVERSITY

Phagwara,Punjab

Table Of Contents

INTRODUCTION

According to the data obtained from the National Transport Authority (2014), there has been an increase of 254% in the number of cars from 2003 (68, 524) to 2014 (173,954). We can thus infer that the sale of second-hand imported (reconditioned) cars and second-hand used cars has eventually increase given that new cars represent only a very small percentage of the total number of cars sold each year. Most individuals in Mauritius who buy new cars also want to know about the resale value of their cars after some years so that they can sell it in the used car market.

Price prediction of second-hand cars depends on numerous factors. The most important ones are manufacturing year, make, model, mileage, horsepower and country of origin. Some other factors are type and amount of fuel per usage, the type of braking system, its acceleration, the interior style, its physical state, volume of cylinders (measured in cubic centimeters), size of the car, number of doors, weight of the car, consumer reviews, paint colour and type, transmission type, whether it is a sports car, sound system, cosmic wheels,
power steering, air conditioner, GPS navigator, safety index etc. In the Mauritian context, there are some special factors that are also usually considered such as who were the previous owners and whether the car has had

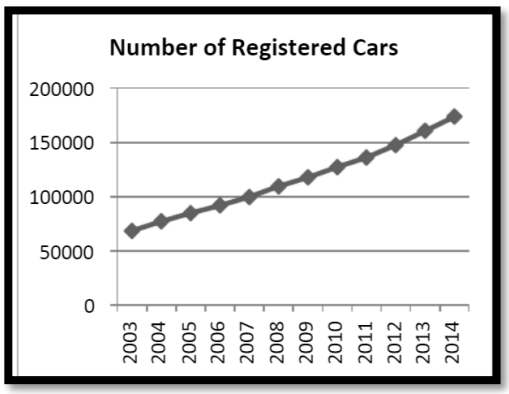


Figure 1Number of registered cars from 2003-2014

Thus, predicting the price of second-hand cars is a very laudable enterprise. In this PROJECT REPORT, we ill assess the methodologies used to accurately predict the price of second- hand cars. The results have also been compared with other methods like linear regression and support vector regression.

PURPOSE

The prices of new cars in the industry is fixed by the manufacturer with some additional costs incurred by the Government in the form of taxes. So, customers buying a new car can be assured of the money they invest to be worthy. But due to the increased price of new cars and the incapability of customers to buy new cars due to the lack of funds, used cars sales are on a global increase.

There is a need for a used car price prediction system to effectively determine the worthiness of the car using a variety of features.

Even though there are websites that offers this service, their prediction method may not be the best. Besides, different models and systems may contribute on predicting power for a used car's actual market value.

It is important to know their actual market value while both buying and selling.

INDIVIDUALS ISSUE

There are lots of individuals who are interested in the used car market at some points in their life because they wanted to sell their car or buy a used car. In this process, it's a big corner to pay too much or sell less then it's market value.

A car comparison and second hand car value Estimation System. Now a days a lot of classified websites allow people to post their used cars at their own price. A buyer at this time does not know the real value of the car and may get fooled. So our system requires the seller to enter car details such as car manufacturing year, model, engine condition and many other factors. Based on these values our algorithm calculates the most accurate value for the car. These factors need to be verified at the time of buying the car. Thus our system allows users to get an Estimation of the Price Of Second hand or Used Cars.

METHODOLOGIES USED

Importing Dependencies :

The code has a few dependencies. We have loaded the dependencies using the following code snippet:

```
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.neural_network import MLPRegressor
from sklearn.model_selection import train_test_split
from sklearn import preprocessing
from sklearn.preprocessing import StandardScaler
import numpy as np
```

Creating the dataset :

Before we can do any machine learning, we need data. We collected the data using a online used car selling platform from different parts of the World. Different Companies of the Cars,Their Color, Mileage, Odometer and on Many Other Aspects the data have been taken into Consideration.

Cleaning the dataset

There are a few issues with the dataset. First of all, not all variables are of ordered types. For example, there is no logical way of ordering the color. Therefore, we will convert these columns to binary columns which simply says per color whether the car is that color or not. We have used the following script for transforming categorical variables to binary variables:

```
# Convert the color column to one binary column for each color

df_colors = df['Color'].str.get_dummies().add_prefix('Color: ')
# Convert the type column to one binary column for each type
df_type = df['Type'].apply(str).str.get_dummies().add_prefix('Type: ')


# Add all dummy columns
df = pd.concat([df, df_colors, df_type], axis=1)
# And drop all categorical columns
                    df = df.drop(['Brand', 'Type', 'Color'], axis=1)
```

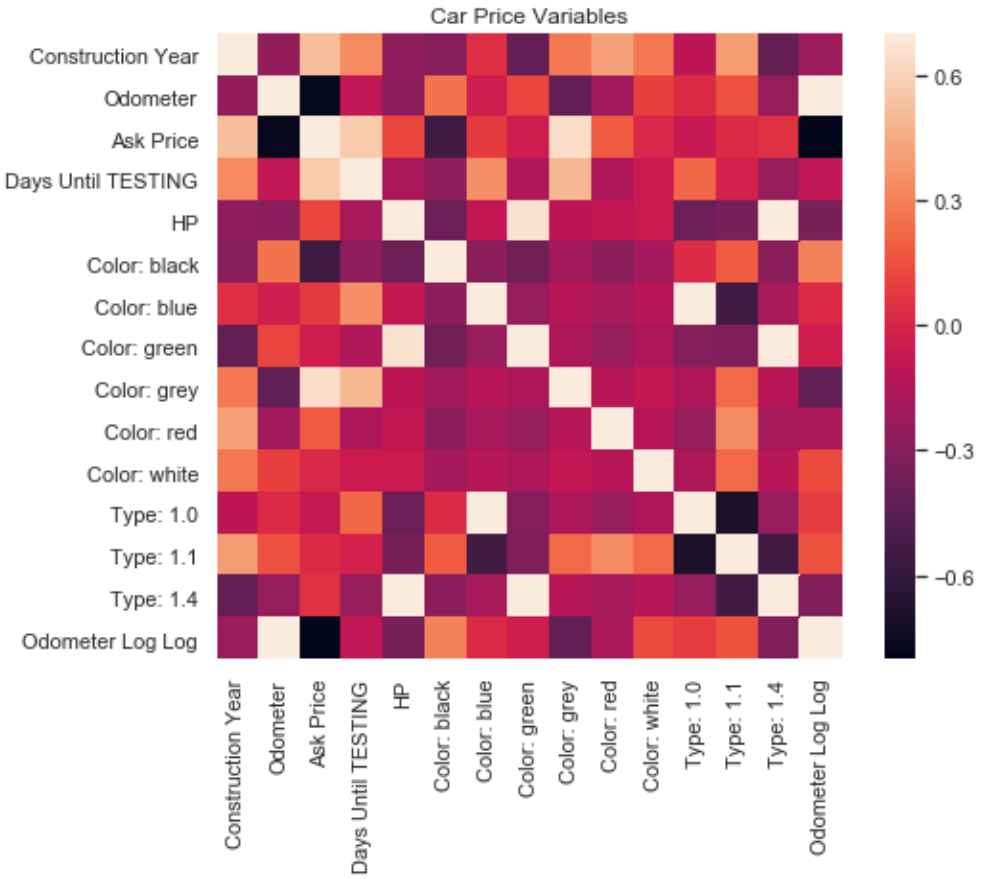
The Following Output Has Been Received :

Construction Year	Odometer	Ask Price	Days Until MOT	HP	Color: black	Color: blue	Color: green	Color: grey	Color: red	Color: white	Type: 1.0	Type: 1.1	Type: 1.4
2002	166879	999	138	60	0	1	0	0	0	0	1	0	0
1998	234484	999	346	60	0	1	0	0	0	0	1	0	0
1997	219752	500	-5	60	1	0	0	0	0	0	0	1	0
2001	223692	750	-87	60	0	0	0	0	1	0	0	1	0
2002	120275	1650	356	59	0	0	0	1	0	0	0	1	0
2003	131358	1399	266	60	0	0	0	0	1	0	0	1	0
1999	304277	799	173	57	0	0	1	0	0	0	0	1	0
1998	93685	1300	0	75	0	0	1	0	0	0	0	0	1
2002	225935	950	113	60	0	0	0	0	0	1	0	1	0
1997	252319	650	133	75	0	0	1	0	0	0	0	0	1
1998	220000	700	82	50	1	0	0	0	0	0	1	0	0
1997	212000	700	75	60	1	0	0	0	0	0	0	1	0
2003	255134	799	197	60	1	0	0	0	0	0	0	1	0

Relations between the variables

It is quite simple using Pandas to find out correlations between our variables:

The Ask Price has a strong positive correlation with the construction year which makes a lot of sense. Newer cars are worth more money. It also has a positive correlation with the days until the TESTING. This also makes a lot of sense! If you just did a TESTING, the buyer is more certain that nothing is wrong with your car. The person is definitely willing to pay more if you just did a TESTING. There is a strong negative correlation with the odometer. If you drove a lot with your car, it is more likely that the car will get more and more issues. Also interesting: the color grey is also correlated. But that is because of underlying (latent) variables.



The Ask Price function :

How can we approximate the ask price given the variables?

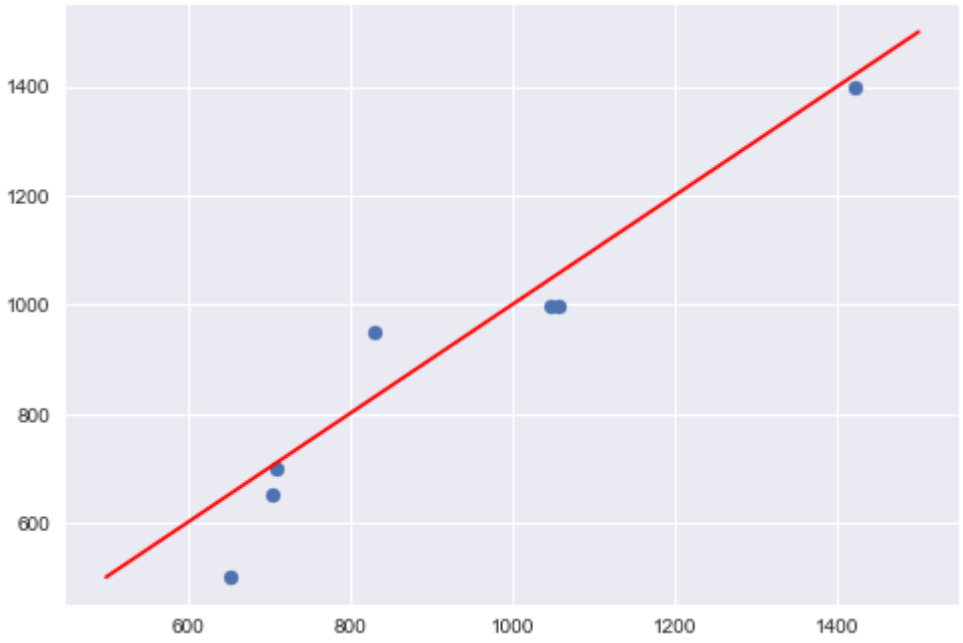
Using a neural network ofcourse! We have used a **Multi-Layered Perceptron to estimate the price**. Also the variables are not normalized.

Normalization is important for a neural network. If We will not normalize the data, then certain variables will have a different scale than other variables.

Take for example the construction year variable. The construction year will be way lower than the traveled distance. An example construction year is 1989 and an example traveled distance is 300.000 km. This is approximately 1000 x more than the construction year! Therefore, we will normalize such that all numbers are centered around zero (and have a standard deviation of approximately 1). Luckily, **scikit-learn (sklearn)** can take care of that!

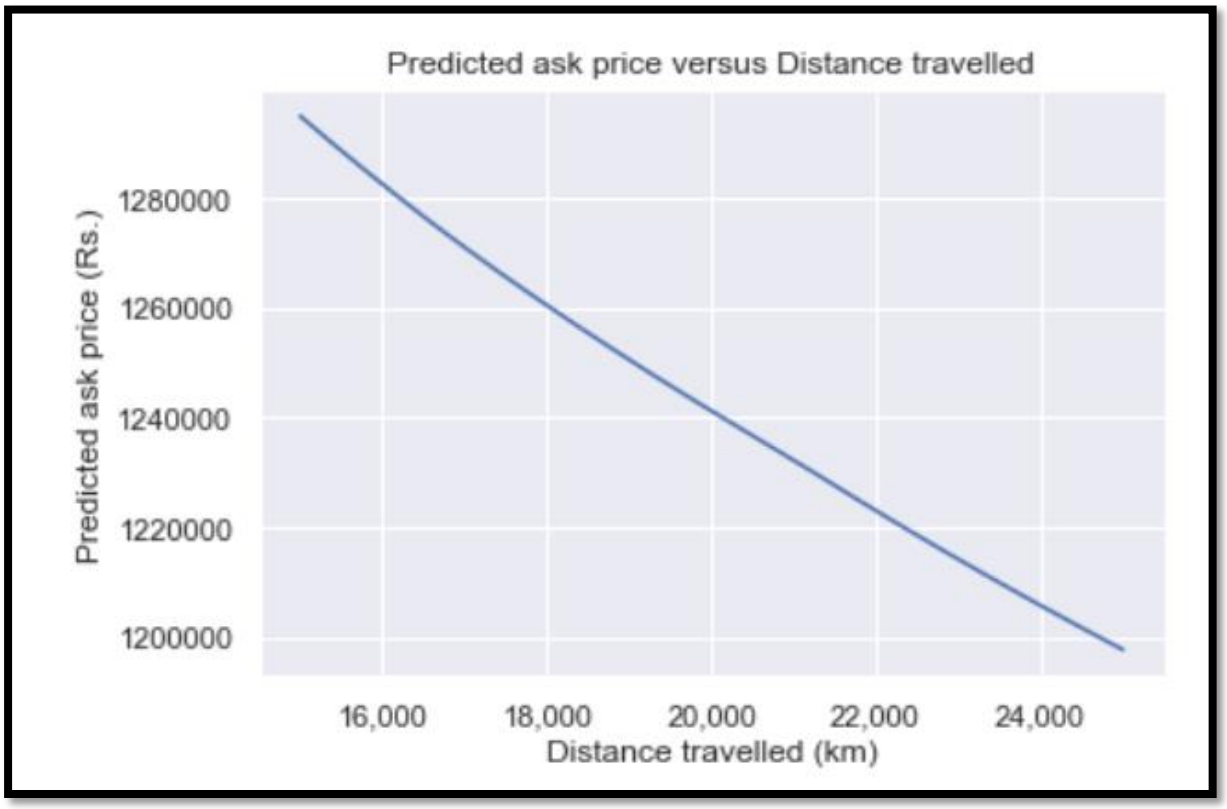
Now Price can be Predicted :

The red line shows the perfect predictions. If all blue dots would lie on this line, then the predictions would be perfect. But this is not bad! We are able to predict the ask price quite good.



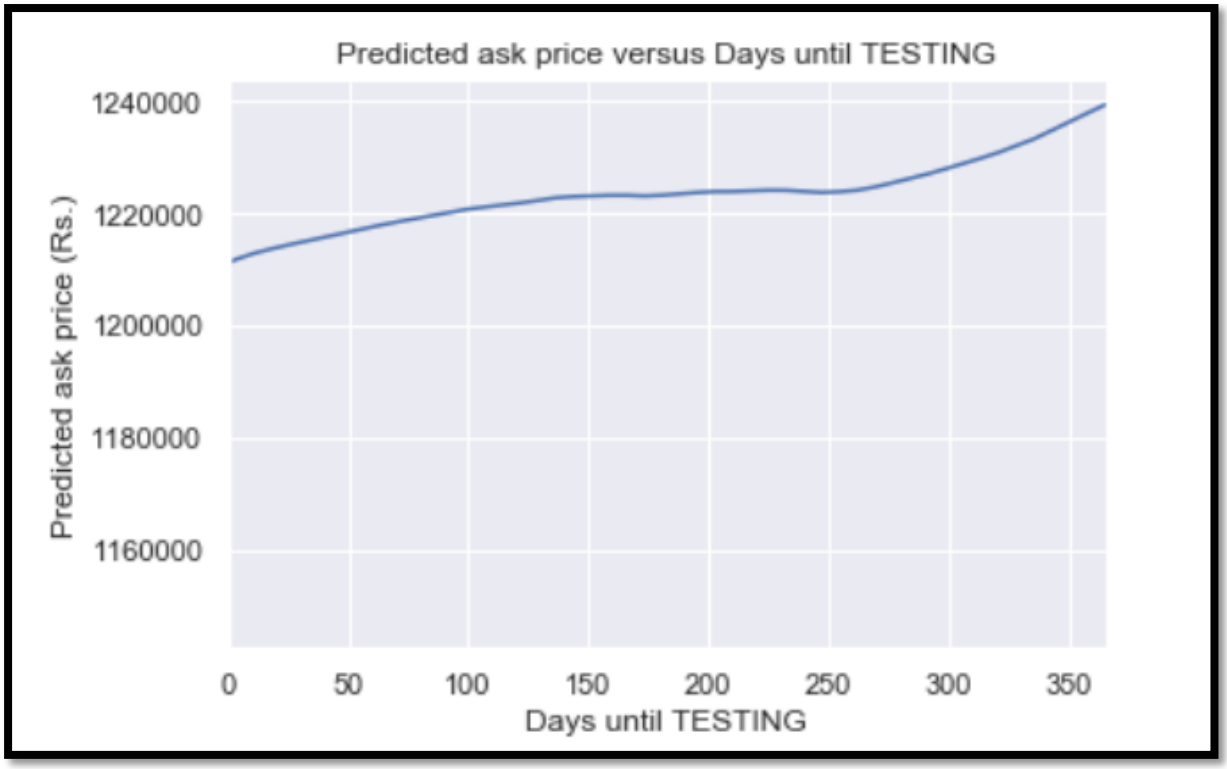
Distance traveled (odometer) versus ask price

I think this relation is general for most cars. Please report if you find a different relation for you car! As you can see, the more you travel with a car, the less the ask price becomes. The difference in the price is almost Rs.2000k.



Annual (TESTING) versus ask price

If the TESTING costs less than 1100k-700k Indian Rupee, then it is definitely worth it to do a check-up! The predicted ask price is approximately 700k Indian Rupee, when the TESTING is soon. there are 365 days until TESTING left when the TESTING just happened. When there are 365 days until TESTING, the predicted ask price is approximately 1100k Indian Rupee. So definitely do a TESTING if you want to sell your car and for this car, the TESTING costs less than Rs.400k



What is the value of my Car?

If My car has the following properties:

- Construction year: 1998.
- Traveled distance: 220.000 km.
- Days until TESTING: 150.

The output was 828.86k Indian Rupee.

----------THE END----------