Journal of Advanced Zoology

ISSN: 0253-7214
*Volume **44** Issue **S-6 Year 2023** Page **1096:1101***

Smart Cars – Predicting Future Customers

K.Ajith Raj[1], G.Joshna Chowdary[2], K.Ruchitha[3],G.Sumanth[4]

[1,2,3,4]C&IT Department, REVA University.
Email: ajithkulkarni@hotmail.com[1] , joshnachowdary2000@gmail.com[2] , ruchitha911@gmail.com[3] , sumanthgannamani2000@gmail.com[4]

**Corresponding author's E-mail: ajithkulkarni@hotmail.com*

Article History

Received: 06 June 2023
Revised: 05 Sept 2023
Accepted: 30 Nov 2023

Abstract

This document gives Smart cars which are also known as autonomous cars or self-driving cars which moves with little or no human input. They are efficient and avoid accidents so many car manufacturing companies have started working on these autonomous cars and for this process they are investing huge money in R&D for best and efficient cars. But the main issue is as they are investing more money they are expecting to get huge profits, so we came up with a solution. In this paper we will set up a machine learning model which will predict the future sales or customers buying these car sand this will increase the income for the companies. So as to achieve this predictive model, we will make use of frame works and libraries of python.

Keywords: *Machine learning, Python libraries, Self-driving cars, Regression, NumPy, Pandas.*

1. Introduction

This document is a template. Our lives in the 21st century is mostly dependent on the technologies with the digital transformation is through the internet of everything, the biggest challenge faced by the organizations now a days is ability to use the real time insights to solve the most critical problems that give best possible output for the problem and best value to the organization. The development of new innovative technologies as Machine learning, Analytics, Edge computing and Artificial intelligence helped in developing the new solution for old problems such as instead of the human involved driving we can now use machine learning and artificial intelligence to make a car drive automatically all we need is the connection for the car. The autonomous cars or the self-driven cars or Robo cars are the next generation of the cars that can travel from one destination to another without any human intervention and able to observe its environment and move safely. The car must be able to follow the traffic rules and watch for any obstacles or pedestrians on the road and move safely where the person needs to go. With these new features there will be less road accidents and the driver or the traveler can do any other work or enjoy the trip without having to worry about the other vehicles and pedestrians. The autonomous cars need the below features to run without any interruption.

Cameras - These cars need cameras to observe and identify the traffic signs, signals and the pedestrians.

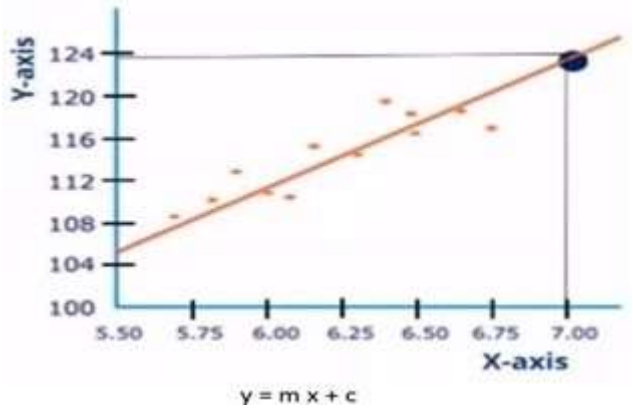
GPS – the GPS is needed to guide through the specific routes and reach the destination.

Control units – used to control and monitor operations performed by the vehicle.

Sensors – To get and process the ongoing information on the road collected from different sensors.

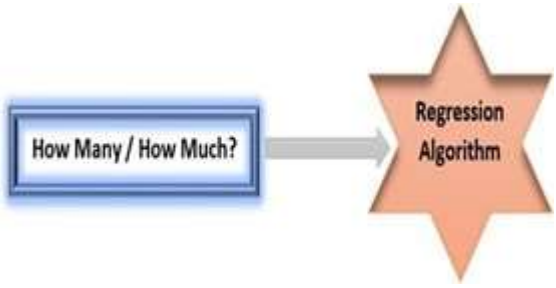
Regression

Regression is one of the popular and important technique used for creating models and analyse data. It is an algorithm which uses predictive modelling technique called supervised learning which helps in identifying the connection between a dependent variable called "target" and independent variable referred as "predictor".



$y = m x + c$

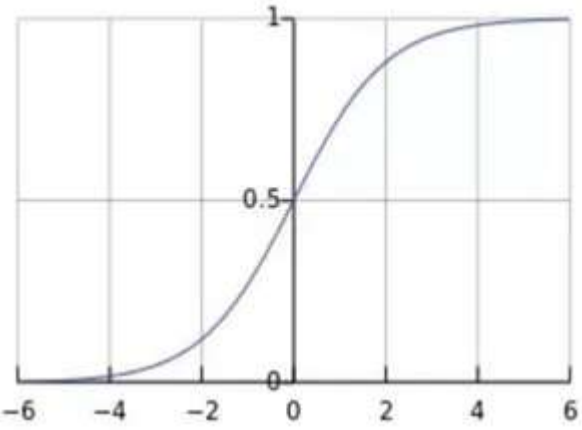
Regression algorithms are used to find the numerical value for the problem needed by giving the consisting values of the target variable and of a explanatory variable to get a mathematical value as an output, we go for the regression for these kind of problems



Using the regression algorithm, we fit a curve to the data points in a way that the difference between the points is minimized. In this example we are going to predict the number of customers are interested in buying the providing model of the autonomous car based on the features given by the company and comparing the previous history of the customers who bought the same featured cars in the past.

Logistic regression

Logistic regression is an algorithm used for classification which comes under the category of supervised algorithm and to find probabilities of event of success or failure. It is used when the dependent variable is of a binary nature like 0 or 1, true or false, yes or no types. It helps to classify data into discrete classes by studying the relationships of a certain labelled data set it learns a linear relationship from the given data set and then introduces a nonlinear relationship in the form of a sigmoid function.



$y = \text{logistic}\ (c + x1*w1 + x2*w2 + x3*w3 + \ldots\ldots + xn*wn)$

$y = 1 / 1 + e\ [- (c + x1*w1 + x2*w2 + x3*w3 + \ldots\ldots + xn*wn)]$

The final result of this method will be the sigmoid curve, which is recognised as the S-curve where the horizontal x-axis consists the value of independent variable and the vertical y-axis consists value of the dependent variable.

The sigmoid function equation is:

$$S(x) = \frac{1}{1+e^{-x}}$$

The sigmoid function in the logistic regression converts any value between the -∞ (-infinity) and +∞ (+infinity) to the distinct values which we need in the algorithm.

The uninterrupted nature of the input variable can provide more flexibility in the logistic regression and therefore provide the more appropriate responses.

For example: spam detection is a problem in binary classification in which we are given a mail with some information in it to verify whether the email given is a spam or not. If the email given is spam we mark it as a 1, if it is not a spam email we mark it as a 0. In order to apply logistic regression to spam detection problem we need to extract the following details from the email which are as follows:

- The author of the email
- The number typing mistakes in the email
- The presence of the words like “offer”, “prize”, etc.

The resulting feature vector is then used to train a logistic classifier which emits a score of in the range between 0 and 1. The resulting score is more than the number 0.5, we label the email as a spam. If the score is less than 0.5, we label it as not spam or we don’t label it as a spam email. Using the insights generated by the output of the logistic regression the users can streamline their emails and accomplish less wastage of time and not fall for the tactics played by the spammers.

Literature survey

In 2019, K.Madhuvathi, Nallakruppan kailasanathan, Senthil kumar N.C, Siva Somayaji [1] proposed an approach for “Cars sales prediction using machine learning algorithms “ using analytic hierarchy process, although it is strongly built it has few disadvantages like it may be not able to fix the careful numerical qualities to examination judgements and it can just settle direct models. They also used linear regression and Random forest to get the results. The main disadvantage of linear regression is it assumes there is a straight line relationship between dependent and independent variables which is wrong most of the times. The disadvantage of random forest models are not at all interpretable and for very large data sets , the size of trees can take up a lot of memory.

In 2015,Keshav Bimbraw[2] proposed an approach for “Autonomous cars, past, present and future”. This paper discussed about the history, advancements, developments and detailed future of fully smart cars.

In 2020,Todd litman[3] proposed an approach for “Autonomous vehicle implementation prediction”. This paper clearly explained the how the autonomous cars are going to take the technology to next level and transportation professionals will be having an major role to play in the smart cars production.

Objectives

The main objective of our project is to predict future sales/customers that may buy self driving cars using logistic

regression. To improve the efficiency of sales based on prediction and to collect the datasets of the previous sales information.

2. Materials And Methods

At first, we have to predict the customers who are willing to buy the smart cars. For this we need to get entrance to the customer dataset and their previous sales info. And with the help of logistic regression we also have to find the factors that influence the customers to buy smart cars. Here, we are going to use jupyter notebook to create a predictive model and access the open source libraries such as Pandas, Numpy, Seaborn, Matplotlib and so on.

Sno	Customer Number	Customer Name	Location	Age	Gender	Salary	High End Model	Category
1	100023456	John	Los Angles	45	Male	120000	1	B
2	100023786	Maria	Irvine	34	Female	100000	0	B
3	100023894	Sneha	Los Angles	32	Female	145000	1	B
4	100046578	Harry	Santa Ana	27	Male	132000	1	L
5	100090876	Thomas	Irvine	26	Male	125000	1	B
6	100034501	Jerry	Sandiego	36	Male	160000	1	B
7	100009876	Kate	Bakersfield	41	Female	90000	0	B
8	100034976	William	Beverly Hills	20	Male	225000	1	B
9	100020986	Hudson	Sandiego	25	Male	155000	1	L
10	100012785	Mike	Carmel Valley	33	Male	135000	0	L
11	100028734	Angelo	Escondido	28	Male	124000	0	B
12	100000678	Daniel	Bakersfield	30	Male	110000	0	L
13	100023567	Nick	Irvine	19	Male	150000	1	L
14	100028765	Pavan	Los Angles	39	Male	180000	0	B
15	100098790	Charles	Long Beach	24	Male	130000	1	B

There are six steps in this process:

Importing dataset into python environment

By using jupyter notebook we need to bring in the dataset into python environment with the help of Pandas, Numpy, Seaborn, Matplotlib libraries.

Loading the Historical Sales Data of Previous Customers

```
import seaborn as sb
import math
import pandas  as pds
import numpy   as npy
import matplotlib.pyplot as mplp
%matpalotlib inline
```

```
autocar_data= pd.read_csv("Autonomous_Cars.csv")
```

Analysing the data

We need to divide the dataset and examine its relationships by asking the few questions.

1) Group of customers with high-end car, buy or lease.
2) What is the location which have more high-end car purchases?
3) What is the category of customers with certain age who have high salary?

After accessing and analyzing the loaded data we are going to create different graphs, distribution and correlation graphs how one variable affects another.

This is done using seaborn library

```
sb.countplot(x="leased", data=autocar_data)
```

```
sb.countplot(x="leased", hue="Gender", data=autocar_data)
```

```
autocar_data["Age"].plot.hist()
```

Data Cleaning

Here we are going to check for null values and we are going to delete the whole column.

```
autocar_data.isnull()
```

This code tells us if there are any null values in the dataset resulting in Boolean values.

```
autocar_data.drop("Sno", axis=1, inplace=True)
```

By using this code the column can be removed or the value can be altered.

Building the model using training dataset

Here, We are going to split the dataset into two parts: training dataset and testing dataset.

```
from sklearn.cross_validation import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(x, y test_size=0.30, random_state=0)
```

Using this we are going to split the dataset. Now using "customer age" and "salary" (two different columns and independent variable) and by using this we find whether the “High end model” variable is purchased or not.

For Independent variable

```
X = autocar_data.iloc[:,[3,5]].values
```

For Dependent variable

```
Y = autocar_data.iloc[:,6].values
```

After this step we are going to apply the logistic regression

```
classifier=LogisticRegression(random_state=0)
classifier.fit(X_train, Y_train)
```

Testing Predictive Model

By applying the below code and the testing dataset and we are going to predict the model.

```
y_pred=classifier.predict(X_test)
```

Accuracy of the Model

Here, We are going to evaluate the accuracy of the predictive model by using “accuracy_score” function.

```
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)
```

applications

1. It will be useful for car manufacturing companies as we are predicting the sales of the cars
2. It will increase the overall profits of the companies
3. This also helps the users to check the popularity of the car and decide which car is best for them

3. Conclusion

Our proposed system is designed to predict the future sales/customers of smart cars .This research paper explains the correctness of model that helps to forecast future sales or customers who are interested in buying these autonomous cars. This prediction will help the companies to grow economically.

Acknowledgment

We would like to take this opportunity to express our gratitude to our Project Guide, Prof. Shashikala, REVA University, for continuously supporting and guiding us in our every endeavor as well for taking a keen and active interest in the progress of every phase of our Project. We would also like to extend our sincere thanks to Dr. Sunil kumar S Manvi, Director of C&It dept., REVA University, for providing us with all the support.

References:

[1] https://www.researchgate.net/publication/332072545_C ar_Sales_Prediction_Using_Machine_Learning_ Algorithttpsmns

[2]https://www.researchgate.net/publication/283757446_Autonomous_Cars_Past_Present_and_Future_A_Review_of_the_Developments_in_the_Last_Century_the_Present_Scenario_and_the_Expected_Future_of_Autonomous_Vehicle_Technology

[3] https://www.machinedesign.com/mechanical-motion-21122520/predictionmarket-for-artificial-intelligence-in-cars-will-grow-1200-in-next-six-years

[4] https://link.springer.com/article/10.1007/s40534- 016-0117-3

[5] https://www.machinedesign.com/mechanical- motion-systems/article/21838234/how-ai-ispaving- the-way-for-autonomous-cars

[6] “Logistic Regression.” Wikipedia, Wikipedia, 2019. Web.10 April 2019. <Logistic Regression.>