

CUSTOMER RETENTION PROJECT

Submitted by:

LALIT JOSHI

**ACKNOWLEDGMENT**

**INTRODUCTION**

* Business Problem Framing

We were in a lockdown situation, occurred due to the pandemic caused by COVID-19 virus. The most effected industry due to this are: Travel and Tourism, Vehicle Manufacturing. As, we were in the lockdown situation, we are not allowed to move outside of our respective houses which causes an alarming situation. Vehicles industry was also hit badly with this. So, here we are trying to build a machine learning model that would pick the data from few used cars websites and predict the pricing of the vehicles.

* Conceptual Background of the Domain Problem

We can say that we have successfully outcome the situation, however, we are still getting cases. This model could help the user with the price predication of the vehicles they wish to purchase and also will help the vendors at what price the user can buy the vehicle.

* Review of Literature

In order to build this project, we have visited various websites, in order to find the attributes deciding the price of a used vehicle (in case the user wants to purchase a used car). Using this model, the user can check all the basic parameters and can decide wisely which vehicle is best suited for them and they wish to purchase for their day to day use. I have checked several articles and found that basic attributes for deciding the price of a used car depends upon the fuel type, manufacturing year, variant and ownership details.

* Motivation for the Problem Undertaken

I believe this could be a very helpful project for a person going to purchase a 4 wheeler. They can easily compare among different brands and also can decide as per their budget.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

We have used statistics and machine learning to build this model. Several mathematical statistics were used for the model like mean mode for null values, z-score or IQR for removing outliers and so on.

* Data Sources and their formats

We have scrapped several websites dealing in used cars few of them are: - carwale.com, cardekho.com, olx.in and etc. We have scrapped the basic information like company name, variant, transmission, city, owner type and price for these vehicles to build our dataset.

* Data Preprocessing Done

What were the steps followed for the cleaning of the data? What were the assumptions done and what were the next actions steps over that?

* Data Inputs- Logic- Output Relationships

Describe the relationship behind the data input, its format, the logic in between and the output. Describe how the input affects the output.

* State the set of assumptions (if any) related to the problem under consideration

Here, you can describe any presumptions taken by you.

* Hardware and Software Requirements and Tools Used

We have used python version 3.8.5 to write the code. Also, we have used intel icore i5 7th gen machine to run python and google chrome. We have used selenium web driver to automate the google chrome and extract the data. We have used several python libraries here to run and build this model like pandas, NumPy, matplotlib, seaborn, sk-learn etc.

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

After scrapping the e have checked the target variable that we need to predict. We have used matplotlib and seaborn libaray to plot the variables in the form of graph to check which variable is effecting the price of the vehicle most and least. On the basis of that and heatmap we proceed to drop or continue with the data.

* Testing of Identified Approaches (Algorithms)

As we need to predict the price of the vehicle which is a running float values, we can say that this is a regression problem. We have used 4 different regression algorithms to build the model and calculated the score using r-mean squared, mean squared and compare the algorithms and at last we chose the best algorithm with the highest score.

* Run and Evaluate selected models

Describe all the algorithms used along with the snapshot of their code and what were the results observed over different evaluation metrics.

* Key Metrics for success in solving problem under consideration

What were the key metrics used along with justification for using it? You may also include statistical metrics used if any.

* Visualizations

Mention all the plots made along with their pictures and what were the inferences and observations obtained from those. Describe them in detail.

If different platforms were used, mention that as well.

* Interpretation of the Results

Give a summary of what results were interpreted from the visualizations, preprocessing and modelling.

**CONCLUSION**

* Key Findings and Conclusions of the Study

We found that customer retention depends upon the best promotional offers and product information a website/app is offering. Also, other features that helps to retain a customer is usage of the app, how well connected the app is, how well the UI is developed.

* Learning Outcomes of the Study in respect of Data Science

List down your learnings obtained about the power of visualization, data cleaning and various algorithms used. You can describe which algorithm works best in which situation and what challenges you faced while working on this project and how did you overcome that.

* Limitations of this work and Scope for Future Work

What are the limitations of this solution provided, the future scope? What all steps/techniques can be followed to further extend this study and improve the results.

https://www.livemint.com/money/personal-finance/know-how-second-hand-cars-are-valued-to-get-a-good-price-11600274256773.html