CAR PRICE PREDICTION PROJECT

PREPARED BY :

ANKEET SAMBYAL

ACKNOWLEDGEMENT

Firstly, I would like to thank Fliprobo technologies for giving me the opportunity to work on this project .I would thank Data trained for providing me the

information related to this project And a big thanks to family and my child who always support me for my work.

INTRODUCTION

PROBLEM FRAMING

In this project, we have to make car price valuation model by using new machine learning models from new data. Because with the change in market due to covid-19 impact, our client is facing problems with their previous car price valuation machine learning models.

CONCEPTUAL BACKGROUND OF DOMAIN PROBLEM

Firstly, we will prepare our own dataset using web scrapping.

After that we will check whether the project is a regression type or a classification.

We will also check whether our dataset is balanced or imbalanced. If it is imbalanced one, we will apply sampling techniques to balance the data set.

Then we will do model building to check its accuracy.

Our main motto is to build a model with good accuracy and for that we will also go for hyper parameter tuning.

REVIEW OF LETERATURE

I am summarizing my research on the topic.

I have created my own dataset using web scraping and imported important libraries for my project.

I have created the data frame.

I have analyzed my data by checking it shape, number of columns, presence of null values if any and checking the data types.

Then I have some done data cleaning steps, e.g Checking the value counts of the target variable, dropping some irrelevant columns from the dataset, checking correlation between the dependant and independent variables using heatmap, visualizing data using distribution plots, detecting and removing skewness in my data if any, outliers detection using box plots and removing them, balancing dataset using random over's sampler method , splitting the data into independent and dependant variables and finally scaling the data.

Then I have used 5 regressor models, out of which XGBregressor is giving a good accuracy score of 98% after hyperparameter tuning .

ANALYTHICAL PROBLEM FRAMING

MATHAMETICAL/ANALYTHICAL MODELING OF PROBLEM

If you look at data science, we are actually using mathematical models to model (and hopefully through the model to explain some of the things that we have seen) business circumstances, environment etc and through these model, we can get more insights such as the outcomes of our decision undertaken, what should we do next shall we do it to improve the odds. So mathematical models are important, selecting the right one to answer the business question can bring tremendous value to the organization.

Here I am using XGBregressor with accuracy 98% after hyper parameter tuning.

DATA SOURCES AND THEIR FORMAT

Data Source: The read csv function of the panda's library is used to read the content of a csv file into the python environment as pandas data frame.

The function can read the files from the OS by using proper path to the file.

Data Description: Pandas describe () is used to view some basic statistical details like percentile, mean, std etc. of a data frame or a series of numeric values.

DATA PREPROCESSING DONE

I have checked for null values and there are some null values and outliers are present.

I have used label encoder the object type columns in the dataset.

THANKYOU