CAR RENT PROJECT

Submitted by:

Abhay surma

ACKNOWLEDGMENT

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

https://www.geeksforgeeks.org/

https://github.com/

https://www.mckinsey.com/

https://www.counterpointresearch.com/

INTRODUCTION

- Problem statement

ABC is a car rental company based out of Bangalore. It rents cars for both in and out stations at affordable prices. The users can rent different types of cars like Sedans, Hatchbacks, SUVs and MUVs, Minivans and so on.

In recent times, the demand for cars is on the rise. As a result, the company would like to tackle the problem of supply and demand. The ultimate goal of the company is to strike the balance between the supply and demand inorder to meet the user expectations.
The company has collected the details of each rental. Based on the past data, the company would like to forecast the demand of car rentals on an hourly basis.

- ## Conceptual Background of the Domain Problem

  A good knowledge of after sales market of cars is necessary. What makes a car valuable will be key.

- ## Review of Literature

  Not a lot of research is available on car prices after covid-19 impact.

  https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/how-consumers-behavior-in-car-buying-and-mobility-changes-amid-covid-19

  https://www.counterpointresearch.com/weekly-updates-covid-19-impact-global-automotive-industry/

- ## Objective - The main objective of the problem is to develop the machine learning approach to forecast the demand of car rentals on an hourly basis.

Analytical Problem Framing

- ## Mathematical/ Analytical Modeling of the Problem

Root Mean Squared Error (RMSE)and Mean Absolute Error (MAE) are metrics used to evaluate a Regression Model. These metrics tell us how

accurate our predictions are and, what is the amount of deviation from the actual values.

Technically, RMSE is the **R**oot of the **M**ean of the **S**quare of **E**rrors and MAE is the **M**ean of **A**bsolute value of **E**rrors. Here, errors are the differences between the predicted values (values predicted by our regression model) and the actual values of a variable. They are calculated as follows :

$$RMSE = \sqrt{\frac{\sum (y_i - y_p)^2}{n}}$$

$$MAE = \frac{|(y_i - y_p)|}{n}$$

y_i = actual value
y_p = predicted value
n = number of observations/rows

On close inspection, you will see that both are average of errors.

Let's understand this with an example. Say, I want to predict the salary of a data scientist based on the number of years of experience. So, salary is my target variable (Y) and experience is the independent variable(X). I have some random data on X and Y and we will use Linear Regression to predict salary. Let's use pandas and scikit-learn for data loading and creating linear model.

Concordance index was also used. The concordance index or c-index is a metric to evaluate the predictions made by an algorithm. It is defined as the proportion of concordant pairs divided by the total number of possible evaluation pairs.

1) Data in CSV format
2) Data Pre-processing
3) EDA (Exploratory data analysis)

- Univariate analysis

a) Demand vs hours

b) Demand vs week day

AND so on.

Hardware and Software Requirements and Tools Used

Pandas, Seaborn, ploty and sickit libraries were used throughout the project.

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)

  Regression and co relation.

  In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables

- Testing of Identified Approaches (Algorithms)

  Linear regression

  Random forest regressor

  Gradient boosting regressor

  Xtreme gradient boosting regressor

- Run and Evaluate selected models

  Random forest regression: RMSE = 33.18
  And as per this data random forest was chosen as the best model; further hyper parameter tuning was performed.

- Key Metrics for success in solving problem under consideration

  RMSE VALUE