Fittlyf's Data Science Internship Entrance Test

Part - 2

How many **Red coloured top Honda SUV cars** do you think will be sold in India in **2022**? Questions to consider:

1. What factors do you think will impact sales? You may google about guesstimate questions to answer this question better.

The important factors are mileage, condition, fuel type, and colour for any car. Even natural calamities and financial crisis may also impact the sales.

If it is a second-hand car, how many kilometres was car driven, seller is an owner or dealer, and age of the car will impact the car sales.

1. Assume you have all the data you need, what statistical methodology or algorithm will you use to make this sales forecast? Please give a brief explanation of why you choose this model.

First of all, the car price is not majorly depending upon time. It depends upon various factors like, I mentioned above. So, I will be not stick on time series model. For this sales forecast, I will use supervised machine learning regression technique. I will not select a specific model, Instead I will build different models with different regression algorithms like logistic regression, decision tree regressor, random forest regressor, etc. Based on each model performance, I will select the best for further process.

1. How would you evaluate your model or determine its accuracy?

For evaluation purpose, I will use test data. I will consider the r2 score and root mean squared error metrices for validating my model performance. How r2 score is nearer to 1 and how RMSE is minimal will tell how good our model performance is.

Part - 3

7. My latest data science project was Back-order prediction

To giving a solution to the problems due to lack of Inventory management, this project was selected for my capstone project. The objective of this project is to predict whether the products of the ecommerce website will be backordered or not based on some characteristics of that product by using the past data. In this problem, I started with exploratory data analysis and based on the insights of the EDA, the data was pre-processed. The data is highly skewed in every numeric feature. so, the mean imputation is not a correct thing here. So that I imputed the missing values with that features median value. All the categorical features are in binary values Yes or No. In this I replaced those Yes and No values with 0 and 1 manually. The value count of the minority class should be atheist 25 % or greater. This is called the Balanced data. But the value count of minority class in the target variable is only 0.72 %. So, I balanced the data by using smote technique. After all these steps, my approach was building a predictive model with different supervised machine learning classification algorithms such as Logistic regression, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier and Ada Boost Classifier. Out of these, got a best result in Decision Tree Classifier model with 98% of accuracy, 98% of F1 score and 99% of AUC score. With that Decision Tree Classifier model, the web app was created by using flask framework. Then it was deployed in one of the PaaS (Platform as a Service) cloud platform called Heroku.

To see the demo, click the link <https://backorder--prediction.herokuapp.com/>