

# Customer review rating prediction

By Simantini Patil

## Machine Learning:

---

In this section I applied Machine learning models to the dataset to predict customer review rating score.

## Hyperparameter tuning:

---

After feature creation, it was observed that different features in the data set had values in different ranges. For example, in 'delayed' column the value ranged from (-189.0, 146.0) and for product\_value the value ranged from 2 - 13444. That means some column were more weighted compared to other.

Differences in variable ranges could potentially affect negatively to the performance of an algorithm so I used scikit-learn's Normalizer to normalize the data.

After normalizing the data between [0,1] I took the log values of the data. This introduced a few -inf and NaN values which I later on replaced with 0.

The data was still widely distributed so I set a lower and upper bound on feature values using clip function.

Also, the categorical columns were encoded using Labelencoder.

Dataset was then split into test and train datasets which was later using in modeling.

## Modeling:

---

To predict the review rating score, the base model was Random forest regressor. This model gave an accuracy score of 0.26 with a 3 fold grid search cross validation.

Other models like Linear regression and gradient Boost regressor gave the less score than Random forest.

## Conclusion:

Unfortunately, unlike NLP, it is by no means easy to determine the state of the art model for review prediction. To predict review rating on the basis of product and delivery poses a data challenge.

Purchase decision processes are composed of several variables that influence customer's choice for certain products and many factors influence review rating such as

- Incorrect descriptions of the product
- Poor client service
- Poor response to information request
- Lack of communication
- Rude or Uninformed customer care staff
- Needs not accurately defined
- Promises not carried out
- Repeated complaints from the same customer

To conclude, customer satisfaction can not be predicted solely based on objective facts without taking the vast amount of subjective sensor and service process data into account.