

Machine Learning Engineer Nanodegree

Capstone Proposal
AYIDH ALQAHTANI
JANAURAY 11th, 2019

Domain Background

[Elo](#), one of the largest payment brands in Brazil, has built partnerships with merchants in order to offer promotions or discounts to cardholders. But do these promotions work for either the consumer or the merchant? Do customers enjoy their experience? Do merchants see repeat business? Personalization is key.[1]

House price prediction paper was written by An Nguyen , this paper utilizes both the hedonic pricing model (Linear Regression) and various machine learning algorithms, such as Random Forest (RF) and Support Vector Regression (SVR), to predict house prices. [2]

Problem Statement:

The Elo wants to predict the customer loyalty score to reduce expenses on unwanted advertisement campaign which costs a lot of money. This problem is a supervised learning problem and since it's predicting the score (number) , so it is a regression problem. The input data are card_id ,merchant_Id , city_id , Category and other features. The output is going to be the target which is a customer's loyalty score.

Datasets and Inputs

I am going to use the provided three data set in this problem including the historical_transactions.csv and new_merchant_transactions.csv files which contain information about each card's transactions. historical_transactions.csv contains up to 3 months' worth of transactions for every card And merchants.csv files which contains aggregate information for each merchant_id represented in the data set beside other supporting files like train and test files which is provided by the company. The structure of data sets are categorical , the outcome variable is going to be number which is customer's score.

Solution Statement

I am going to do data exploration and preparation for the provided data sets and construct our model to predict the loyalty score the customer which is going to be a supervised model and since we are predicting numbers, so it's going to be a regression model. Our methodology is to do data preprocessing and split the data set into training data, validation, and testing data sets. Further, choose the regression model algorithm like support vector machine, lasso regression and others. In addition, apply the model on our training data set, until it achieves good score in accuracy and other metrics and then test the model on the testing data. And finally, validate the regression model.

Benchmark Model

We would like to use the Root Mean Squared Error (RMSE) **to evaluate our model results. The benchmark model is Stochastic Gradient Descent (SGD)**

<https://scikit-learn.org/stable/modules/sgd.html#regression>

we can compare our model to the benchmark model and we can know how our model performs by doing that.

Evaluation Metrics

The evaluation metrics for this project is going to be Root Mean Squared Error (RMSE)

Project Design

The project will have different phases including data preparation, exploration, building the model and then evaluating the model.

References:

[1] <https://www.kaggle.com/c/elo-merchant-category-recommendation>

[2] <https://pdfs.semanticscholar.org/782d/3fdf15f5ff99d5fb6acafb61ed8e1c60fab8.pdf>