Prediction of costumer’s purchase black Friday dataset

Sviatoslav Fedoriv

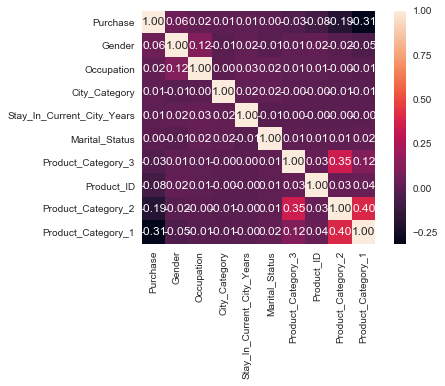
Introduction

Imagine that you are supermarket. You want to maximize your profits on black Friday. That why you want to understand the customer purchase behavior (specifically, purchase amount) against various products of different categories. That way it is good to build a model to predict the purchase amount of customer against various products which will help to create personalized offer for customers against different products.

Methods

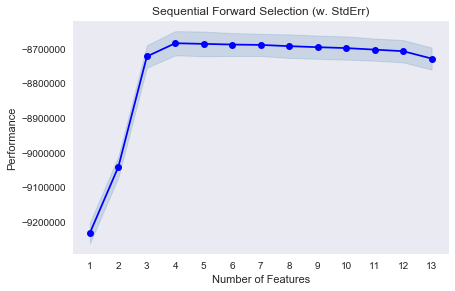
Before looking at the data it is important to understand how does the company expect to use and benefit from this model? This first brainstorming helps to determine how to frame the problem, what algorithms to select and measure the performance of each one. So first it is the best to make correlation analysis for better understanding the model we should build. When I understood that purchase depend only on kinds of product that were cohoused by costumer, I decided first to predict the category of product by classifier and from it I will predict purchase by regression.

First, let’s have a look on correlation matrix.



As you can see from it only product categories and city category have the influence on purchase. So I decided to predict these fitches and from them I will predict purchase.

Let’s find the optimal model for regression. It is DecisionTreeRegressor() first I tried to train on all fitches, than to prove my theory I use SequentialFeatureSelector.



These 4 fitches are 'City\_Category', 'Product\_Category\_1', 'Product\_Category\_2', 'Product\_Category\_3'.

In classification I did it all in similar manner, the model that I cohouse is called RandomForestClassifier.

But first I reduced 'Product\_Category\_1', 'Product\_Category\_2', 'Product\_Category\_3' by PCA to 1 column and use it like label. After that we can predict these label and chose equivalent 'Product\_Category\_1', 'Product\_Category\_2', 'Product\_Category\_3' + 'City\_Category' to predict purchase. As we can see, we predict categories and then money it can be useful for us.

Conclusion

In this work the process of data preparation was studied for purchase prediction. The ML algorithm that perform the best was Decision Tree Model with RMSE = 2920. The next step will be looking at Hyperparameter Tuning and Ensembling.