

Retail Analysis with Walmart Data

(Data Science with Python: Project4)

Writeup

We have a sales prediction topic for one of the leading retail stores in the United States, Walmart. The data we have are available for the weekly sales of 45 Walmart stores, for the period between 2010-02-05 and 2012-11-01. Sales are impacted by several factors such as unforeseen requests and stock shortages for example.

We have the task to produce some statistics and produce an ML model that will incorporate factors such as economic conditions, including CPI, unemployment index, etc.

Firstly, our work consisted in producing some results by answering some questions about statistics sales and produce the sales prediction model. Several algorithms will be tested by studying several hypotheses.

We were particularly, interested in studying stores with maximum sales and standard deviation. We also, our study covered Q3'2012 sales by looking for the best performing stores. Furthermore, we have made the comparison with the previous quarterly the same year and the same quarterly of last year.

Particular attention was given to identify monthly and semesterly sales, as well as holidays that had performed compared to the average non-holiday sales.

Overall, several figures have been given for better illustration.

As for the sales prediction model, we tested the following algorithms LinearRegression, DecisionTreeRegressor, Ridge, GradientBoostingRegressor, RandomForestRegressor and XGBRegressor.

The best performers were the last 3 with scores exceeding 93%. These algorithms remain efficient even without scaling, which is not the case for the first two LinearRegression and Ridge, DecisionTreeRegressor remains stable.

Finally, we used the GridSearchCV technique, and we were able to find the parameters to improve the score of the model from XGBRegressor by going from 94.55% to more than 95.57%.