DAT 530 Advanced Statistical Methods

Project Review: Week 2

Reviewer: Brandon Hosley

September 5, 2020

Project Title:

Consumer Behavior Insights for Instacart

Author(s):

Sunny Lee, Josefa Sullivan, Michael Emmert and Jordan Runge

Source:

NYC Data Science Academy.

The Problem the Author(s) is Trying to Solve in the Project:

To examine online grocer customer's settled and exploratory purchasing habits. The Authors then use the patterns that they discover to build and test a product recommender.

Machine Learning (ML) Algorithm(s) used:

- XGBoost

- K-means followed by Hierarchical Clustering

- Associative Rule Mining

A Brief Description of One of the ML Algorithms used:

XGBoost (eXtreme Gradient Boost) is a type of decision tree modeling algorithm that uses gradient boosting on classification regression trees. It is also optimized for parallelization and use of cache and other hardware resources. It uses both Lasso and Ridge regularization to prevent over-fitting.

Metrics Used to Evaluate the ML Algorithms:

The Author's build and deploy a recommender, but fail to provide any metrics by which to test the efficacy of the recommendations. From the screen captures provided, it is doubtful that the recommendations actually made it to a test group. It may not have been feasible to track which items ultimately end up purchased after recommendation.