Predicting inventory demand

Elliot Blackstone, Amirhossein Tavakoli, S. C. Park (mentor)

Github - https://github.com/ElliotBlackstone/EWinter25_Product_Inventory

Overview – We travel back in time to participate in the Kaggle competition "Grupo Bimbo Inventory Demand". Grupo Bimbo is a large Mexican company that sells baked goods worldwide. In this project, we predict the demand for perishable baked goods in central Mexico. Each week, clients receive products arriving along various routes originating from numerous sales depots. Any unsold goods at the end of the week are bought back at half price by Grupo Bimbo (i.e. returns). For each product purchased by each client, the target variable is adjusted demand, which is equal to the number of sales minus the number of returns. The goal is to predict the future adjusted demand for many clients based on their past purchases.

Stakeholders – Companies which sell perishable goods on regular time intervals

Key performance indicators – Accuracy of predicted demand vs true demand

Methods and Results – We began by making a simple baseline model based on aggregate statistics of overall client demand and product demand. An important caveat of this project is that new clients and new products can appear in our testing data, so our simple model handled this issue with aggregate stats. The demand for client/product combinations, which are present in the training and test data, is predicted by an autoregressive AR(1) model. This simple baseline model achieved a Kaggle score of 0.579.

The baseline model demonstrates that the aggregate stats of clients and products are reasonable predictors of demand. Next, we trained models using various features of XGBoost and LightGBM. Our best-performing model was a grid search cross-validated XGBoost model using 3 weeks of lagged demand, client and product IDs (as categorical variables), client mean/median/min/max and product mean/median. This model achieved a Kaggle score of 0.491.

Conclusion – With the proper selection of features, we created a model that accurately predicts demand. For reference, the top score on the Kaggle leaderboard is 0.442.

Future work – There are many ways our results can be improved. Computing performance was a major issue when training models using XGBoost. With a more powerful machine or GPU acceleration, we would have access to more parameters within our XGBoost model. Another place for improvement could be imputing missing values in our training set.