Predicting Stock Outs for Buy-Online Pickup-in-Store Orders

Sushree S. Patra, Shantam D. Mogali, Sachin U. Arakeri, Pranav Saboo, Zaid Ahmed, Matthew A. Lanham
Purdue University, Krannert School of Management
patras@purdue.edu; smogali@purdue.edu; sarakeri@purdue.edu; psaboo@purdue.edu; ahmed152@purdue.edu; lanhamm@purdue.edu

ABSTRACT

This work provides insights on how to leverage 'Buy-Online Pickup-in-Store' data to understand customer preferences, demand patterns and when products go out-of-stock (OOS) to improve replenishment decisions for grocery chains. The solution uses different product categories based on the grocer's business segments, and then specific predictive models are implemented to predict stockouts. This work is novel in how OOS data from brick-and-click is utilized to advise the grocery stores on the timely replenishment of stock to reduce overall lost sales. This study aims to evaluate and compare multiple classification algorithms for predicting OOS at a store-product level on highly skewed data.

INTRODUCTION

In the retail industry, on-shelf availability of products is critical for the company's profitability, customer retention, and customer satisfaction. We have defined 'OOS rate' as the ratio of out-of-stock transactions to total online transactions. The typical industry average is 8.3%, and the client, a national supermarket chain, has an average OOS of 4.6%. However, the OOS rate varied considerably across business segments requiring a custom product solution for **3400+ products across 240+ stores.**

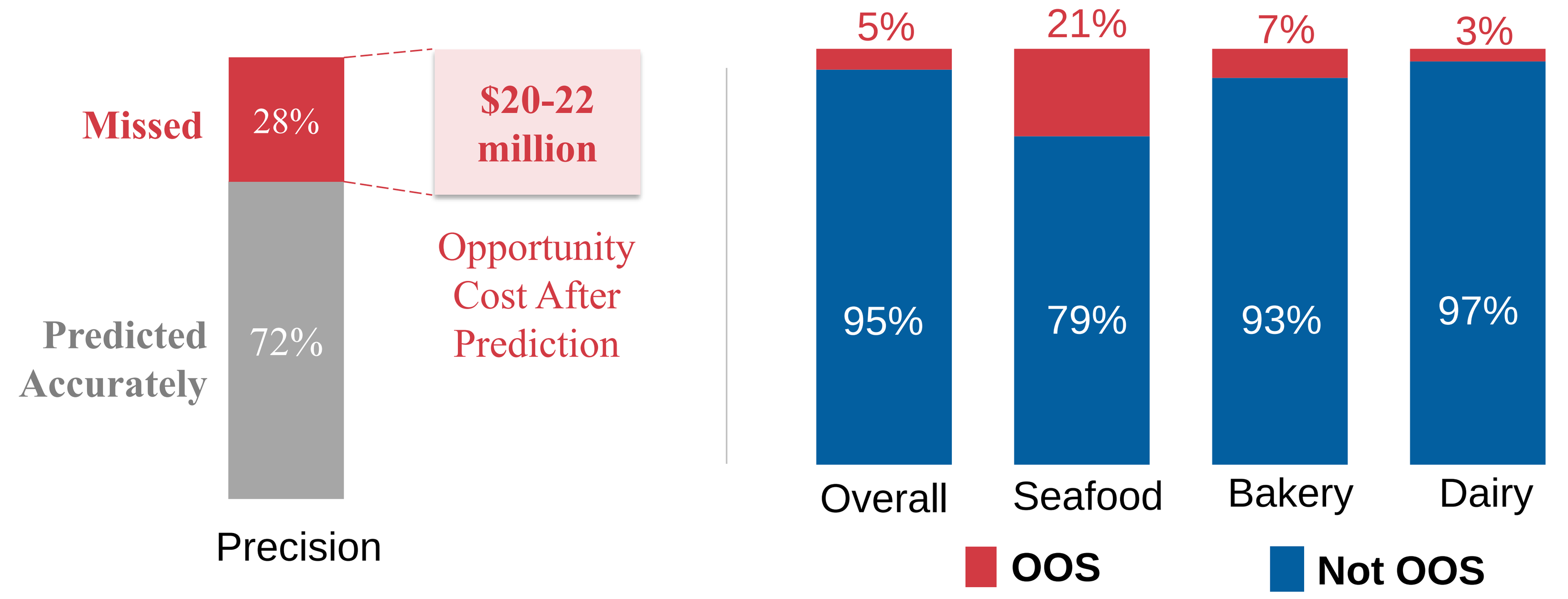

Fig 1. Opportunity Cost & Variance in OOS Rates

The primary research questions of this study:

- What are the most important drivers for identifying OOS rate at product-store level?
- What is the business impact and stock availability across Business Units?

LITERATURE REVIEW

Most of the studies done to predict out of stock events apply either time series or classification methods.

Study	ARIMA	ANN	RF	LMT	Ensemble
Dimitris Papakiriakopoulos, 2011			✓	✓	✓
M.W.T. Gemmink, 2017	✓				
Enzo Morosini Frazzon, 2019		✓			
Bart L MacCarthy, 2019	Used a generalized store wave picking model				
Our Study, 2020			✓	✓	✓

After evaluating various modeling approaches, we zeroed down our approach to build ensemble models using RF, LR and AdaBoost.

METHODOLOGY

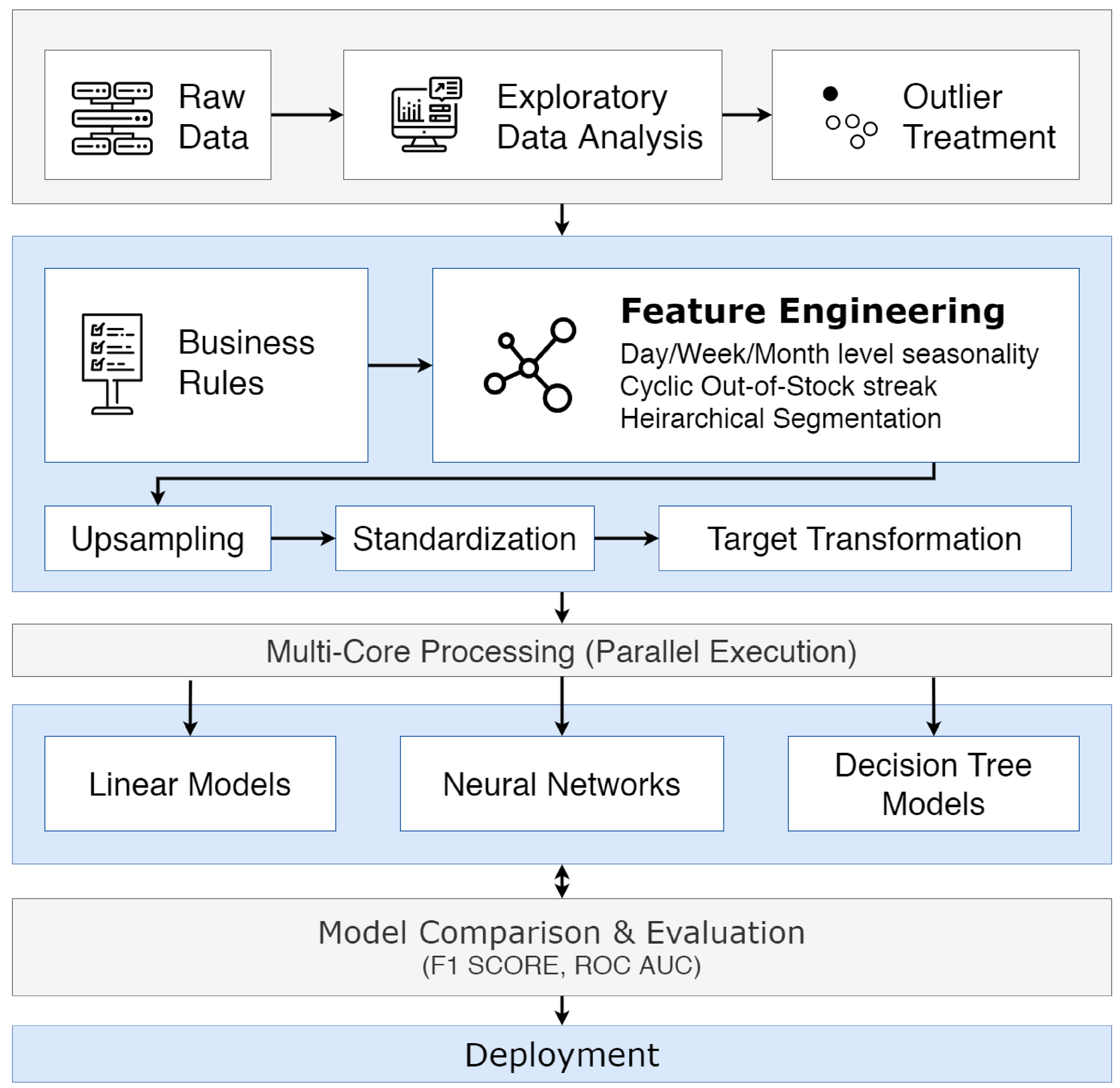

Fig 2. Study Design

FEATURE ENGINEERING

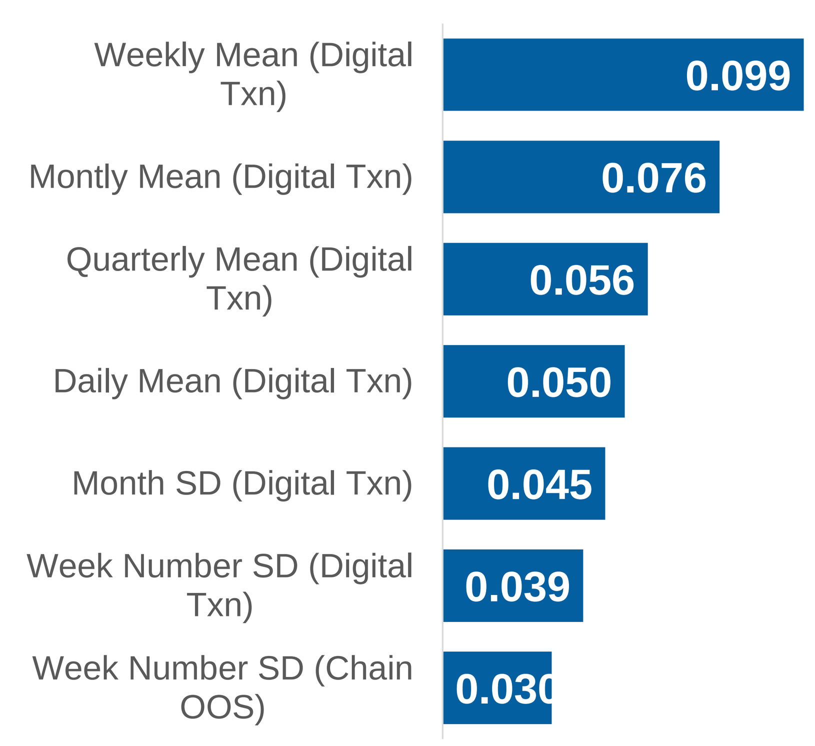

Fig 3. Important Features

To predict if a product will go out of stock, certain exogenous variables could not be used such as unit, sales, etc. for the day to be predicted due to their unavailability on the day itself.

The features were synthesized to incorporate the cyclic nature of demand and on-shelf availability.

16 new features were engineered to boost model performance. Expected number of digital transaction on the prediction day contributes most to model to classify OOS at store product level.

RESULTS

Utilizing multiple classification algorithms, we concluded that RF, LR and Adaboost performed relatively well over others. Empirical results show, RF can handle the high dimensionality of the data and capture the pattern from noise in the presence of highly imbalanced data.

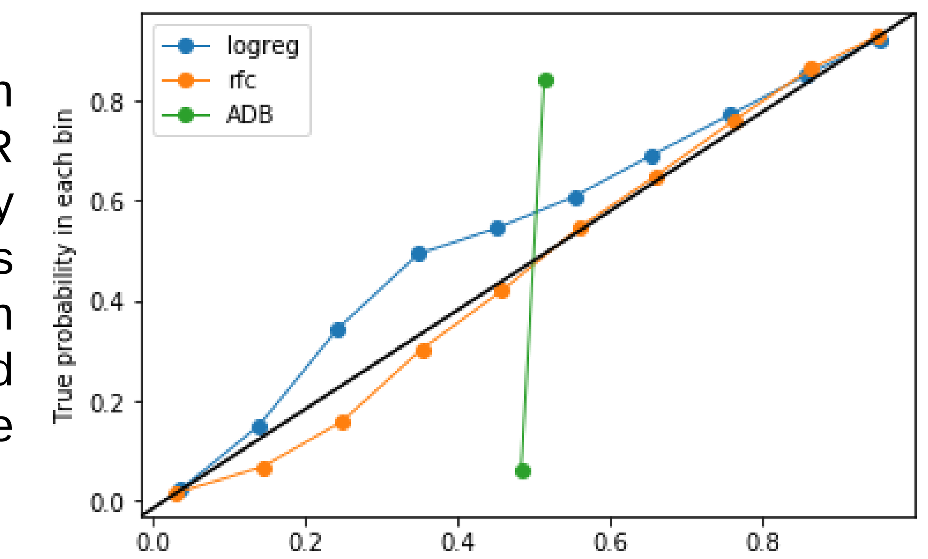

Fig 4. Probability Calibration Plot

BU	Base F1	Improvement
Grocery DSD	0.03	58%
Consumables	0.14	56%
Al. Beverages	0.14	54%
Dairy	0.07	50%
Packaged Meat	0.17	43%
Produce	0.11	40%
Seafood	0.86	5%
Meat	0.87	2%
Deli	0.9	1%

Fig 5. Improvement with Features

The empirical results show an improvement of 40% to 58% in F1 Score for products with high OOS Rate.

It also depicts marginal increment (1% to 3%) in F1 Score for products that are already being predicted with accuracy as high as 90%.

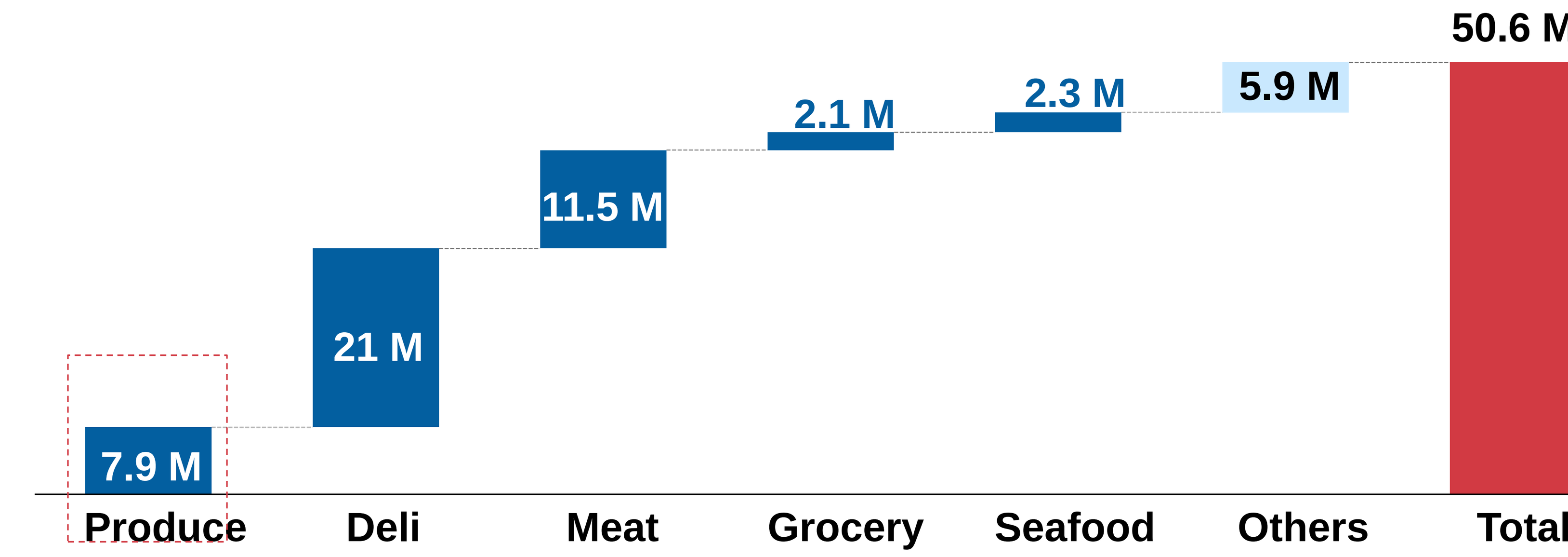

Fig 6. Total Net Savings ($) based on predicted OOS instances

CONCLUSIONS

Engineered features showed significant improvement on model performance such as weekly mean, monthly mean, quarterly mean, monthly standard deviation and weekly standard deviation for digital transactions proved to have the highest importance in identifying OOS rate at product-store level.

We incorporated the costs of incorrect classifications to compare models based on business measures and found RF to be the best performer. Based on Fig 6, the estimated potential savings from our predictions was ~**$50M per year**

ACKNOWLEDGEMENTS

We thank our industry partner for their trust and support. We also thank Professor Matthew Lanham for constant guidance on this project.