

Yuwen Hong, Jingda Zhou, Matthew A. Lanham

Purdue University Krannert School of Management

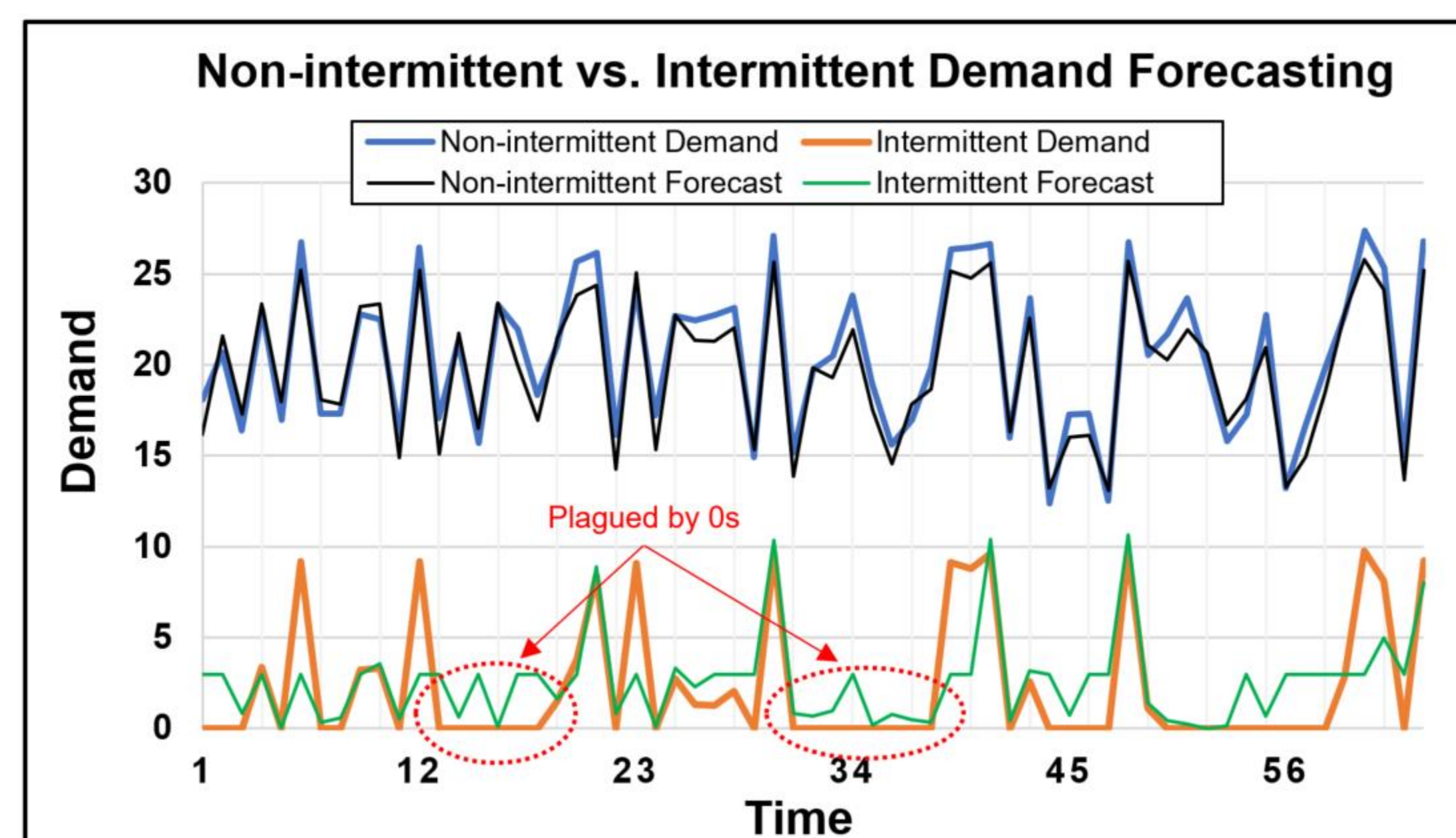
hong142@purdue.edu; zhou854@purdue.edu; lanhamm@purdue.edu

Abstract

The study aims to generalize the predictive accuracy of various machine learning approaches, along with the widely used Croston's method for time-series forecasting. Using multiple multi-period time-series we see if there is a method that tends to capture intermittent demand better than others. The motivation of this study is that demand forecasting is an important component of business planning, but it is also a challenge, especially in the case for intermittent demand. If companies cannot capture the demand accurately, companies either risk losing sales and customers when items are stocked out, or are burdened with excessive inventories. Therefore, being able to identify even small improvements in prediction accuracy of intermittent demand can translate into significant savings (Aris A. Syntetos, Zied Babai, & Gardner, 2015). In collaboration with a supply chain consulting company, we investigated over 160 different intermittent time-series to identify what works the best.

Introduction

Because of its irregularity and unpredictable zero values, intermittent demand is typically much more challenging to forecast than non-intermittent demand. As shown in the figure below, rarely will your forecast be zero when demand is zero. Sometimes practitioners will truncate forecasts for such products where if predicted demand is less than some threshold τ , the forecast is set to zero. However, we ignore such heuristics in our study and focus on solely on the performance of each model.



Our research question are as follows:

- How well do popular machine learning approaches perform at predicting intermittent demand?
- How do these machine learning approaches compare to the popular Croston's method of time-series forecasting?
- Can combining models via meta-modeling (what we call two-stage modeling) improve capturing the intermittent demand signal?
- Can one overall model be developed that can capture multiple different intermittent time-series and how would it perform compared to the others?

Methodology

Data Sources

The dataset used was provided by an undisclosed industrial partner. It contains 160 time-series of intermittent demand for unknown products. These time-series are observed either in daily or weekly frequency. There are only three features: series number, time, and value. Each series represents the demand of a distinct item.

Feature Engineering

Five features were created from the original dataset. These features are divided into two components to capture the unique characters of the intermittent time series problem. The components and their corresponding features are:

- **Time lags:** lag1, lag2, lag3
- **Intermittent demand:** non-zero interval, cumulative zeros

The three "lag"s are just the demand values directly lagged up to three periods, no matter if it was zero or there was indeed demand. The "Non-zero interval" is the time interval between the previous two non-zero values. This feature identifies the length of time during when demand occurs without interruption. The "cumulative zeros" is the number of successive zero values until lag one. It shows the length of time during which no demand occurs.

Sequential Data Partitioning

Each series were performed at the individual level to capture the unique profile of each item. We used sequential data partitioning to split each series into training and testing sets, which comprised 75-25% splits of the total number of observations respectively. The training set was used to train models, and the models were then used to predict outcomes on the testing set.

Data Pre-Processing

All sets were normalized using Min-Max Scaling to ensure all series were on the same scale. Training and testing sets were pre-processed separately, because forecasts are made on a rolling basis. The mode predict D_t (demand at time t) based on the inputs from previous forecasts (lags) were easily set with R `caret`.

Methodology (Approach) Selection

Neural networks (NN) are robust in dealing with noisy data and flexible in terms of model parameters and data assumptions. With multiple nodes trying various combination of weights assigned to each connection, NN can learn around uninformative observations, which indicates great potential to find out relationships within intermittent time series data without other extra information.

Gradient Boosting Machines (GBM), as a forward learning ensemble method, is robust to random features. By building regression trees on all the features in a fully distributed way, we expect GBM to capture some features of the unstable intermittent demand.

Random Forests (RF), similar to GBM, is based on decision trees. The difference is that GBM reduces prediction error by focusing on bias reduction (via boosting weak learners), while RF focuses on reducing error by focusing on variance reduction (via bagging or bootstrap aggregation).

Meta-modeling (a.k.a. two-stage modeling in our project), is suggested by some researchers to have better performance than using single base learners in isolation. Particularly, more information could be gathered via models of different focuses.

Model Comparison / Statistical Performance Measures

The statistical measures adopted here were Mean Absolute Error (MAE) and Mean Absolute Scaled Error (MASE). MAE is easy to interpret and understand, and it treats all errors equally. However, you cannot use this measure to compare across time-series because it is scale dependent. MASE allows comparison across series, because it is scale independent and has been shown to be a great measure to assess intermittent demand (Hyndman, 2006).

Study Design/ Workflow

Figure 1. Overall Flow

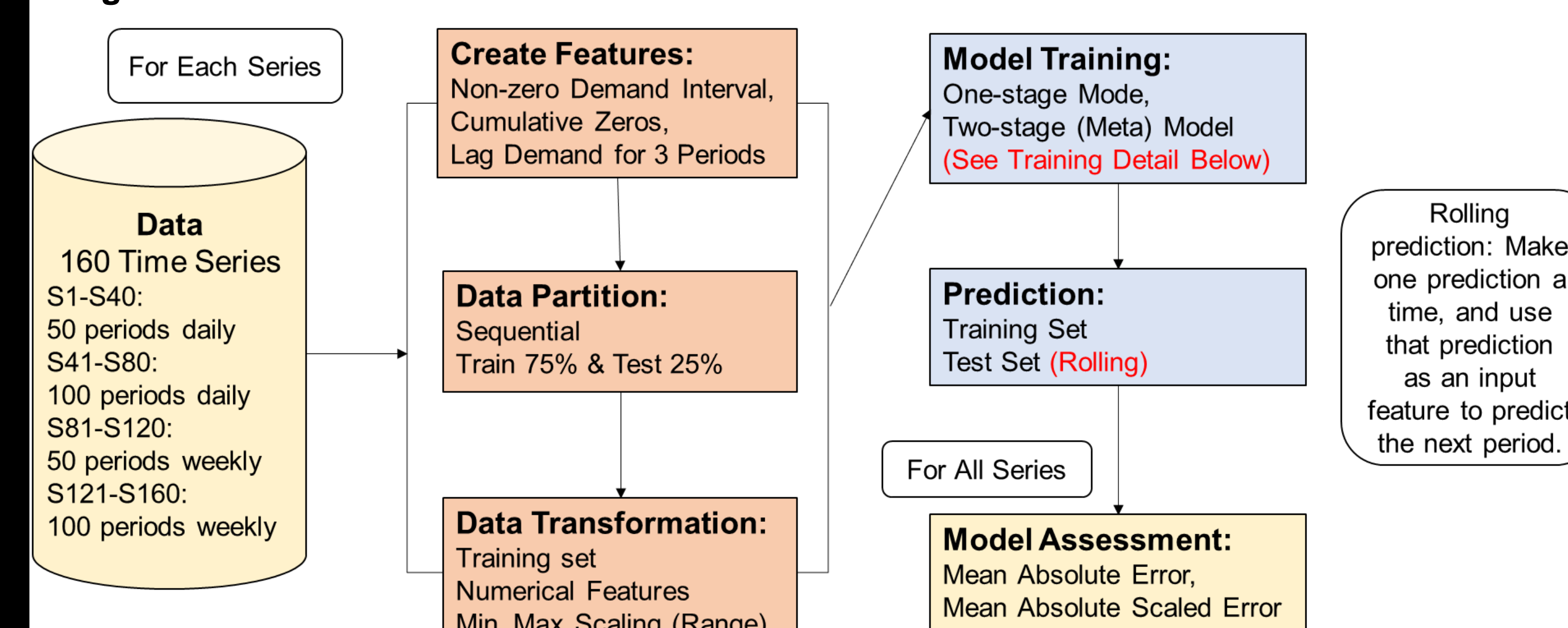
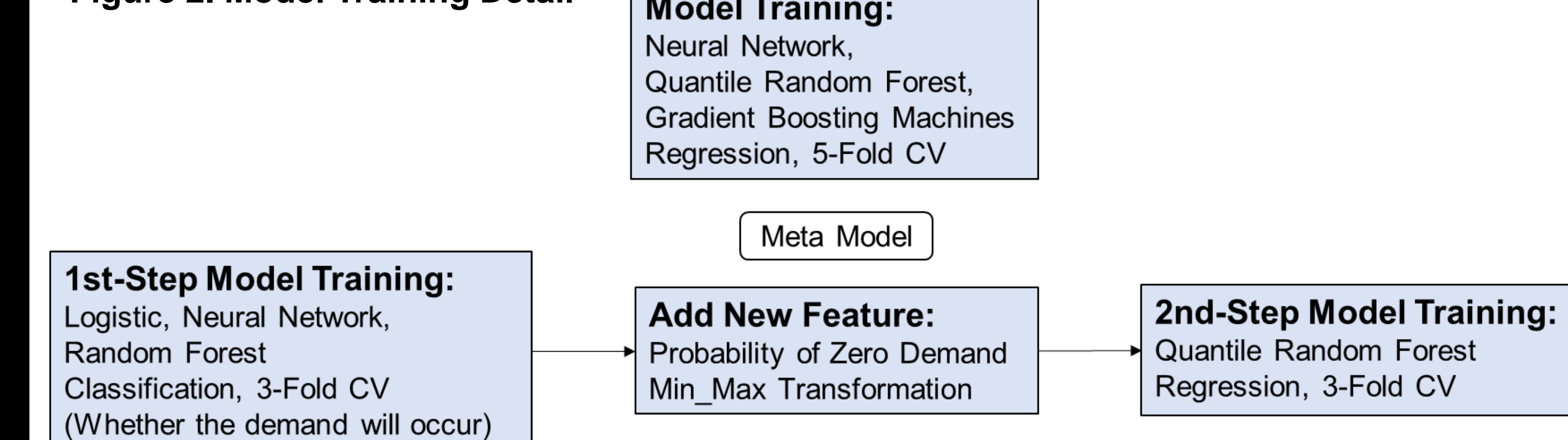


Figure 2. Model Training Detail



Model Building

All models were trained using 3-fold cross-validation. Considering the measures used in this research, all regression models were optimized on MAE.

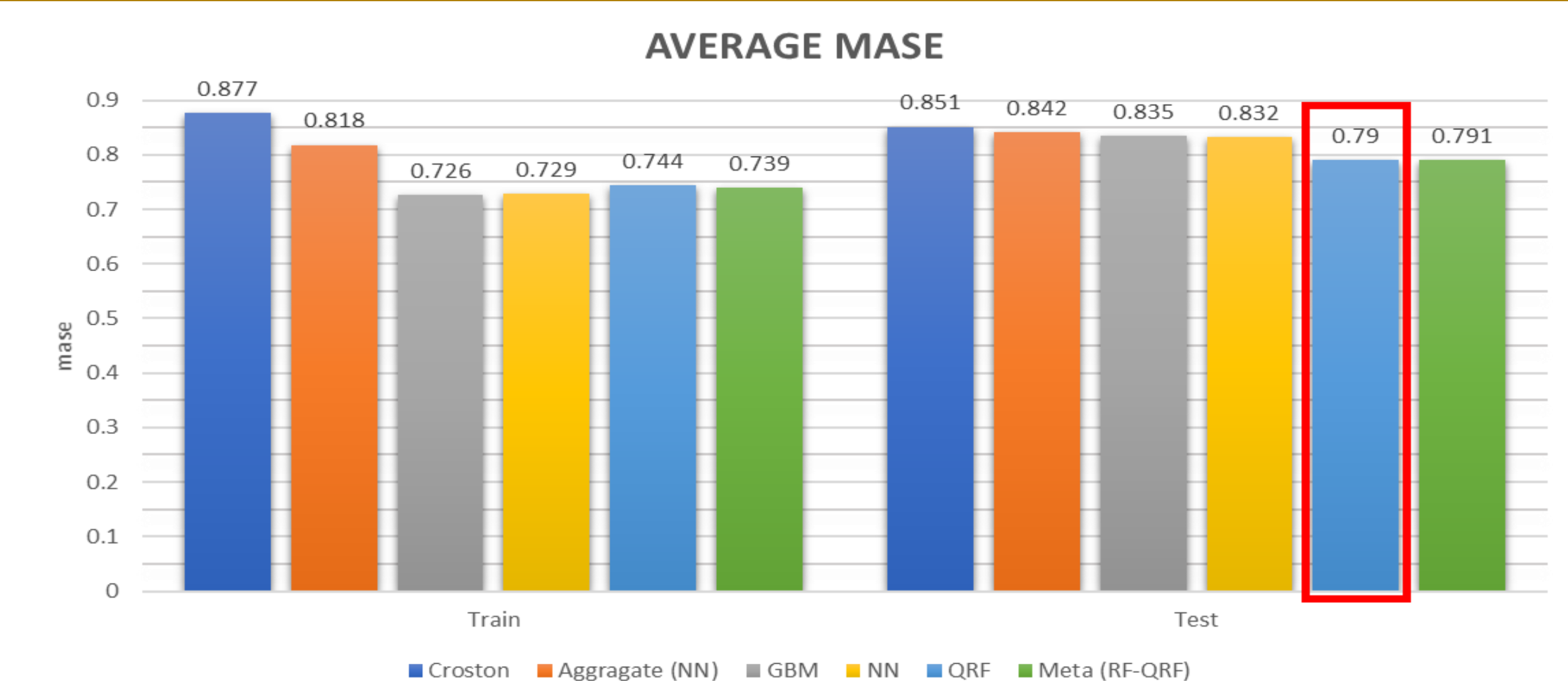
The formula below is a general one used in all single models, as well as the 1st stage model. The 2nd stage of the meta-model, the probability of whether the non-zero demand will occur was added as another variable.

$$Demand \sim nzInterval + zeroCumulative + lag_1 + lag_2 + lag_3$$

Three sets of single stage models were trained using Neural Network (NN), Quartile Random Forest (QRF), and Gradient Boosting Machines (GBM), respectively. We also tried another NN model using aggregated training sets of the 160 series.

The 1st stage classification models adopted Logit, NN, and RF, respectively. The output of 1st stage model was then feed into the 2nd stage meta-model forecast, where QRF and NN were used.

Results



A paired t-test showed that QRF generates significantly lower MASE than the traditional Croston's Method, indicating that this model achieves higher predictive accuracy. Moreover, the predictions on test datasets performed reasonably well without obvious overfitting issues. As show in the figure above, QRF yields the most accurate predictions among the one-step model. The MAE table shows similar results. Again, MAE cannot be compared across series, but we used the same 160 series for model training and testing, so we used the average to gain a basic idea of the overall performance of each model.

Model	Croston	Aggregate (NN)	GBM	NN	QRF	Meta (RF-QRF)
Train	33.51	30.45	26.31	28.11	29.03	28.88
Test	31.85	30.91	31.93	31.62	30.66	30.68

The most accurate meta-model, which used RF in the first step and QRF in the second step, did not outperform the one-step QRF as expected. All the other meta-models performed worse than the corresponding one-step models.

The aggregated NN (our one overall model) showed similar results as the series-level NN with time series ID included as a feature. Companies carrying large numbers of SKU with intermittent demand may want to adopt this approach to simplify their model training.

Conclusions

A small increase in predictive accuracy can help firms save substantial amount of inventory costs while maintaining acceptable service levels. As the results show, machine learning techniques, such as Quartile Random Forest, can improve predictive accuracy for intermittent demand forecasts. We consider the limitation of our models to be fitting small number of inputs into a data-hungry model. Future analysts should explore more input features related to intermittent demand prediction. There are possibilities some models will perform badly in terms of statistical measures, but perform well in achieving business measure. In this case, the decision maker would need further information about the costs associated with a low service level, to leverage between statistical and business measures.

Acknowledgements

We thank our Professor Matthew Lanham for constant guidance and our industry partner for giving us the opportunity to work on this project. The Purdue BIAC partially funded this project.