An Approach to Predict Demand for A Grocer's Most Challenging Products: Meat and Seafood

Gopi Krishna Mashetty, Jiacen Liu, Hui Zeng, Prachi Priyam, Matthew A. Lanham

Purdue University, Krannert School of Management

gmashett@purdue.edu; liu2480@purdue.edu; zeng162@purdue.edu; ppriyam@purdue.edu; lanhamm@purdue.edu

ABSTRACT

This study analyzes the trend of demand for meat and seafood of a mid-west grocery chain and develops a comprehensive set of forecasting models with sophisticated workflow to improve the prediction accuracy. By using our approach to forecasting, the grocer can gain a more in-depth view of where and how accuracy can be improved. Moreover, we provide the grocer with ready-to-use models that can help them ensure and improve prediction accuracy. We show 1) how feature engineering and variable selection affect the accuracy of modeling, and 2) how our set of modeling properly predicts product demand accurately and helps the grocer make more strategic decisions to improve their future forecasting.

INTRODUCTION

Predicting the demand for seafood and meet accurately is challenging due to various factors including the perishable nature of the products and the drastic changes in the demand for seafood and meat combined with the change of customer behavior due to COVID-19. The difficulty of predicting seafood and meat demand is indicative of the grocery stores' challenges to cope up with the supply and demand chain. Therefore, the forecasting models need to be tuned so that the demand can be accurately captured by superstores. Using machine learning models helps to understand the various metrics and build the ecosystem of demand and supply. Moreover, the integration of high-performance computing can benefit the process by improving the running speed of various experiments.

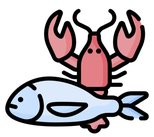


Fig 1a. Meat and Seafood

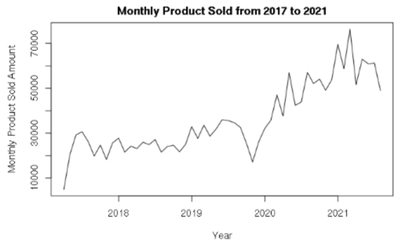


Fig 1b. Change of Trend from 2017 to 2021

Research Questions:

- What are the important features that affect meat and seafood products?
- How can the demand forecasts for meat and seafood products be improved?

LITERATURE REVIEW

Previous studies have examined the complex relationship between promotions, trends, and seasonality using time-series models to advanced LSTMs to address these research questions. Our methodology, ensemble of all such models, is used to identify important features and improve forecasting accuracy.

Study	Prophet	LASSO	SVM	ARIMA	LSTM	KNN	GBM	RF
(Ali, et al., 2009)			✓					
(Ma, et al., 2015)		✓						
(Elmasdotter, et al., 2018)				✓	✓			
(Odegua, 2020)						✓	✓	✓
(Zunic, et al., 2021)	✓							
Our Study	✓	✓	✓	✓	✓	✓	✓	✓

METHODOLOGY

BUSINESS AND DATA UNDERSTANDING

RAW DATA

- 25GB
- 165 Variables
- 15 Product Hierarchy Levels

EXPLORATORY DATA ANALYSIS

- Created data dictionary and built ER diagram.
- Explored relationship among variables.
- Analyzed distributions and patterns of the data.

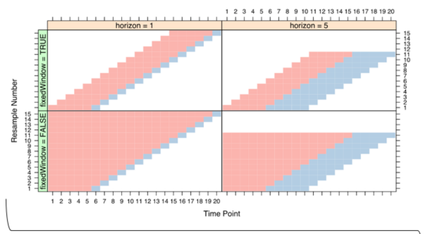
DATA CLEANING AND PREPARING

DATA PRE-PROCESSING & FEATURE ENGINEERING

- Filtered data based on sales and store open date.
- Dropped unnecessary variables.
- Removed outliers.
- Eliminated linear dependencies and multi-collinearities.
- One-hot encoding.
- Removed variables with low variation.
- Transformed variables.
- Standardized data.

MODEL BUILDING AND TUNING

CROSS VALIDATION



MODEL CREATION

- Multiple Linear Regression
- Support Vector Machine
- Random Forest
- H2O AutoML
- Optimal reconciliation with Exponential, ARIMA, SARIMA and Prophet models
- XGBoost
- Advanced Neural Networks (RNN, LSTM)

MODEL EVALUATION AND COMPARISON

1. Identify candidate models.
2. Compare performance of models using different key performance measures.
3. Select model with the best accuracy score.

ERROR METRICS USED

Weighted Absolute Percentage Error (WAPE)
Root Mean Square Error (RMSE)
Mean Absolute Percentage Error (MAPE)
Symmetric Mean Absolute Percentage (sMAPE)

Fig 2. Study Design

STATISTICAL RESULTS

Because the product level sales can be high, low, or intermittent, WAPE was the main model metric we used in our evaluation. WAPE weights errors based on their significance. Based on this study, we observed that the XGBoost model produced the best results. Moreover, we plan to examine deep neural networks like LSTM and RNN to further improve the model's performance.

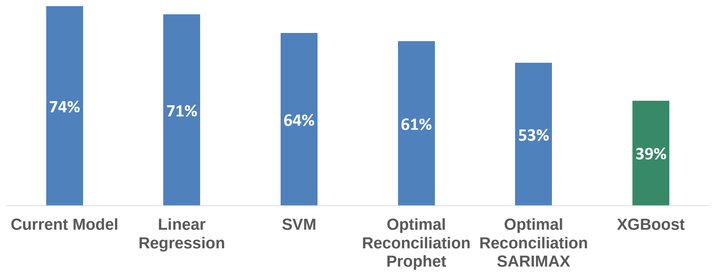


Fig 3. Summary of best model results (WAPE Metric)

IMPORTANT FEATURES

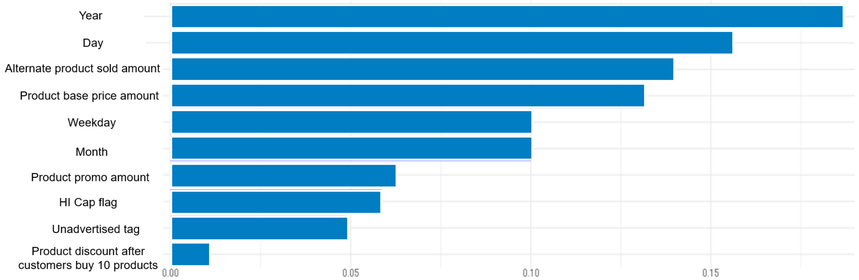


Fig 4. Most Important Features

PURDUE UNIVERSITY'S BELL CLUSTER

With large amount of data provided by the client we used the high-performance Bell Cluster at Purdue University to perform data processing and to conduct thousands of forecasting experiments. By parallelizing tasks, we were able to process the data quickly, enhancing our team's productivity by 50%.

EXPECTED BUSINESS IMPACT

With the new model, we reduced the error metric by ~35%. The substantial decrease in error metrics results in better accuracy of the model, which would allow clients to make better decisions. As a result, for the products in meat and seafood segments, clients will be benefitted with below:

- Efficient inventory management
- Efficient demand and supply channels
- Increase in sales of products by ~10%
- Reduction in cost resulting in profitability by ~5%
- Better resource management leading to enhanced productivity

Fig 5. Business Impact

CONCLUSIONS

- We have developed a predictive modeling workflow that will allow thousands of forecasting experiments to be performed on meat and seafood products.
- By using XGBoost model, we have improved the error metric by ~35%
- Promotions and seasonality are the significant features that affect the sales prediction and model accuracy.
- Using HPC capabilities we have designed a way to process massive datasets, capture and share results, and summarize key findings of areas where product forecast performance can be improved.
- Through cooperation with the grocery store's Analytics team, we have analyzed performance across various metrics and therefore provided new capabilities.

ACKNOWLEDGEMENTS

We would like to thank Professor Matthew Lanham for his guidance on this project.