Forecasting Pesky SKUs for Auto Parts Retailer

PURDUE UNIVERSITY | Krannert School of Management

Adithya Bharadwaj Umamahesh, Jose San Martin Galindo, Apoorva Singh, Saachi Lalwani, Shubham Agarwal, Yang Wang
Purdue University, Krannert School of Management
aumamahe@purdue.edu; jsanmar@purdue.edu; singh749@purdue.edu; lalwanis@purdue.edu; agarw275@purdue.edu; yangwang@purdue.edu

ABSTRACT

The client is facing issues in the form of lost sales and increased holding costs of leftover inventory, both rooted in the fact that SKU sales are difficult to predict. These issues have a direct impact on the economic profits of the firm and are thus of pressing importance to the company. Our first step was to use past sales data to identify anomalous SKUs based on outlier detection techniques. Tukey's boxplot method proved most effective to flag an SKU as "pesky" based on the number of sales per store. Further, we used the past year's sales to perform forecasting. Supervised learning algorithms such as multiple linear regression, lasso regression, random forest etc. were used in this process. We have been able to successfully identify and predict these pesky SKUs' behavior through our forecasting techniques, leading to improved inventory management and sales reporting of the firm.

INTRODUCTION

In the retail industry, understanding the movement of products, assortment planning, and demand forecasting plays a key role in staying profitable. Companies invest quite heavily in understanding customer demands based on a multitude of factors, including but not limited to individual demographics, geography and seasonal changes.

As of 2018, the United States automotive aftermarket was worth USD 75.31 billion and expected to continue growing. In this project, we have attempted to build an accurate forecast for a group of SKUs that have unusually high performance in certain stores as compared to the majority.

Research questions:

- How to define, identify and forecast pesky SKUs?
- What kind of splits or outlier handling techniques can be leveraged using the signals from input variables?
- What is the best way to develop an accurate forecast for pesky SKUs?

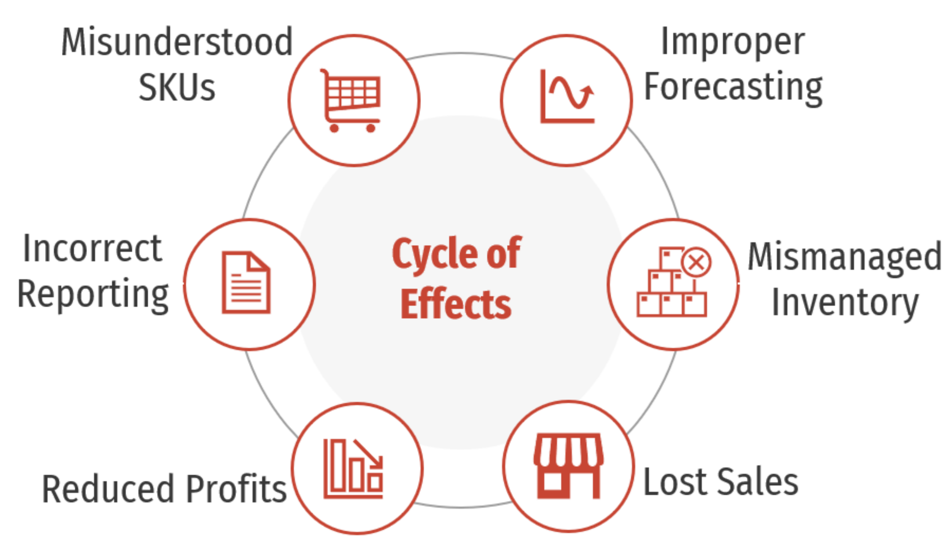

Fig 1. Business Problem

LITERATURE REVIEW

Author	Motivation	Model
Cheriyan et al. (2018)	Integration of decision analysis and predictions through data mining techniques	Gradient Boost Algorithm
Kumar & Patel (2010)	Using clustering to improve sales forecasts in retail merchandising	Clustering Methods
Li & Lim (2018)	Intermittent demand forecasting for a retailer: self-improvement procedure for Croston based methods	Hierarchical forecasting using seasonal exponential smoothing
T.M. Williams (1984)	Classification strategy for handling SKUs with varying demands	Variance partition (to split demand into groups)
Gutierrez et al. (2008)	Forecasting lumpy demand	Multi layered perceptron Neural Networks

METHODOLOGY

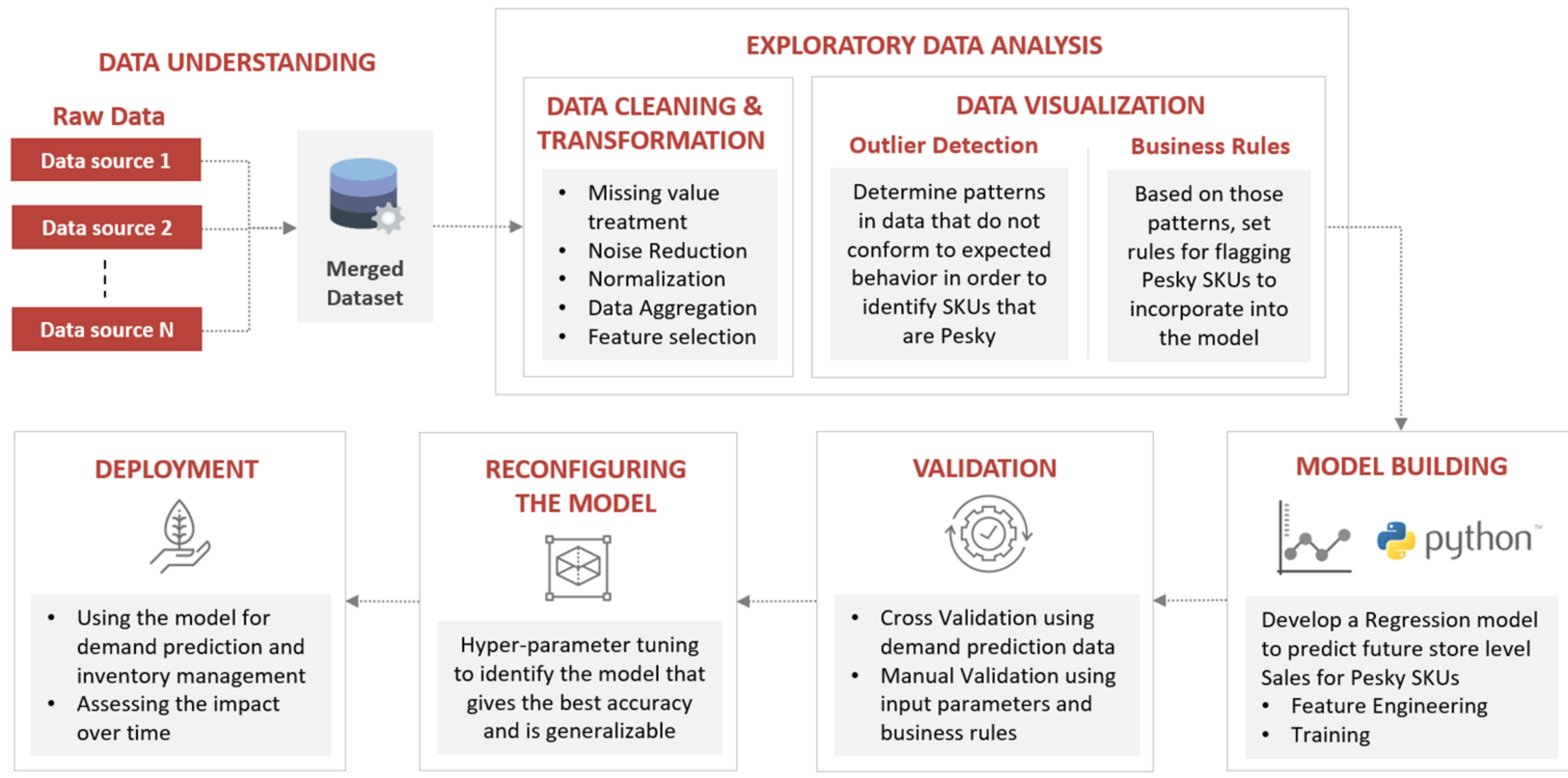

Fig 2. Study Design

The data consists of sales metrics, some SKU features and some store/area level features - all aggregated for a year at a Store-SKU granularity. We used the boxplot method on SKU-Store level sales to identify Pesky SKUs. Data for the identified SKUs was then used as input for the regression model. Separate generalized models were finalized using feature selection and engineering, and hyper-parameter tuning for different categories of SKUs.

STATISTICAL RESULTS

As a first set of results, we identified the Pesky SKUs from our client's portfolio. About 30% of items were flagged using a boxplot extended to the log-IQ method. These were items with unusually high sales in a few stores as compared to the others.

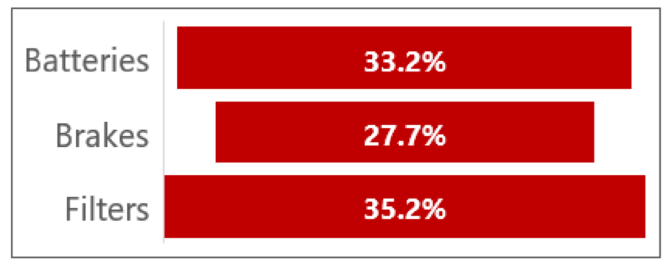

Fig 3. Percentage of Pesky SKUs by Product Type

After the EDA stage, we built different regression models like Multiple Linear Regression, Lasso Regression, and Random Forest, using 60% training data, fine-tuned the results on validation set (20%) and assessed our models on the holdout set (20%). The models were compared based on Root Mean Square Error. Random Forest provided the best predictions for majority of SKUs.

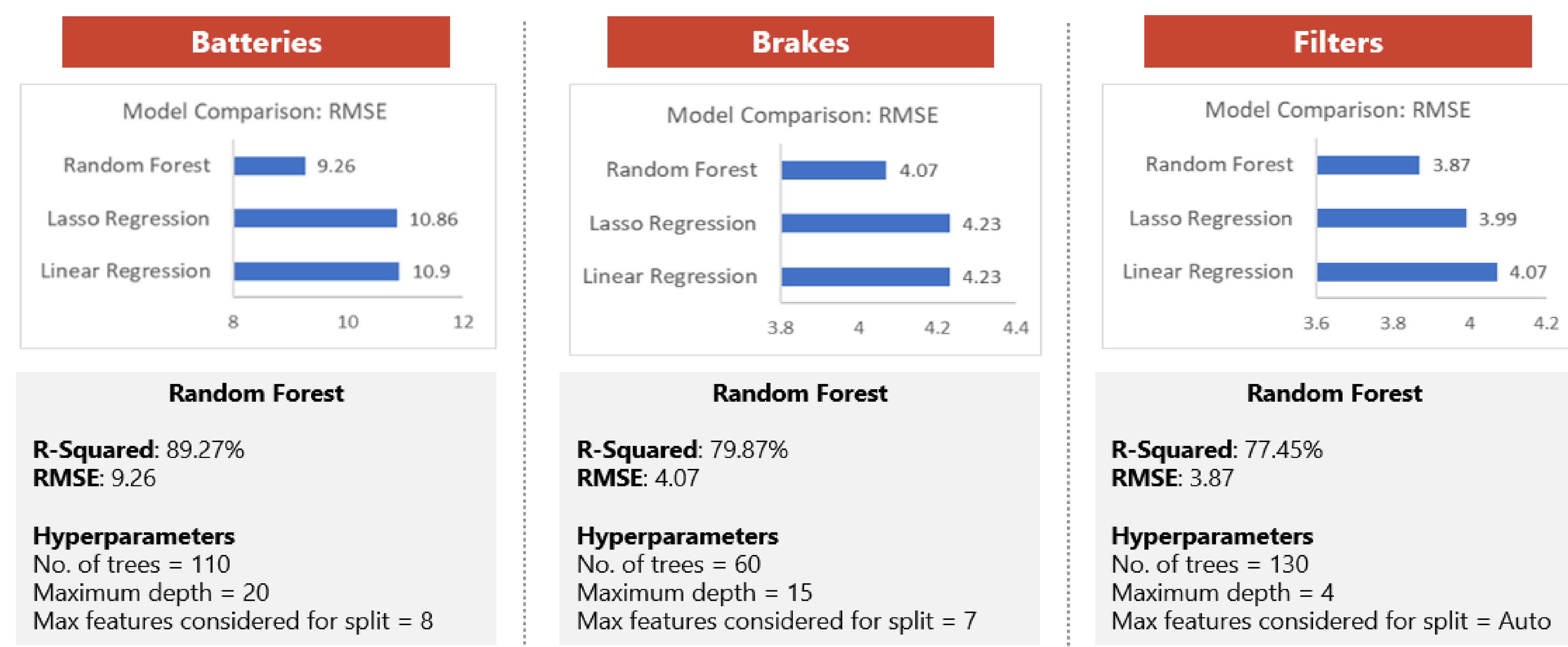

Fig 4. Results and Model Comparison

BUSINESS IMPACT

The outlier detection methodology will identify SKUs with high variability in sales across stores. These items are the main cause of inefficient stock management and lost sales, since it is difficult to predict their demand. Building a sales forecasting model for pesky SKUs can result in the following business benefits:

1. **Reduced holding costs** - Inventory in stores and warehouses will be streamlined according to the predicted demand and thus cost of holding items in stock will be reduced
2. **Reduced lost sales** - Customers will not need to wait for the products they require as the forecasting will ensure a more accurate supply of product is provided in each store based on predicted demand
3. **Improved employee efficiency** - As the client stores have customer facing retail, the employees will be able to provide more efficient customer service and will not be overwhelmed by improper inventory

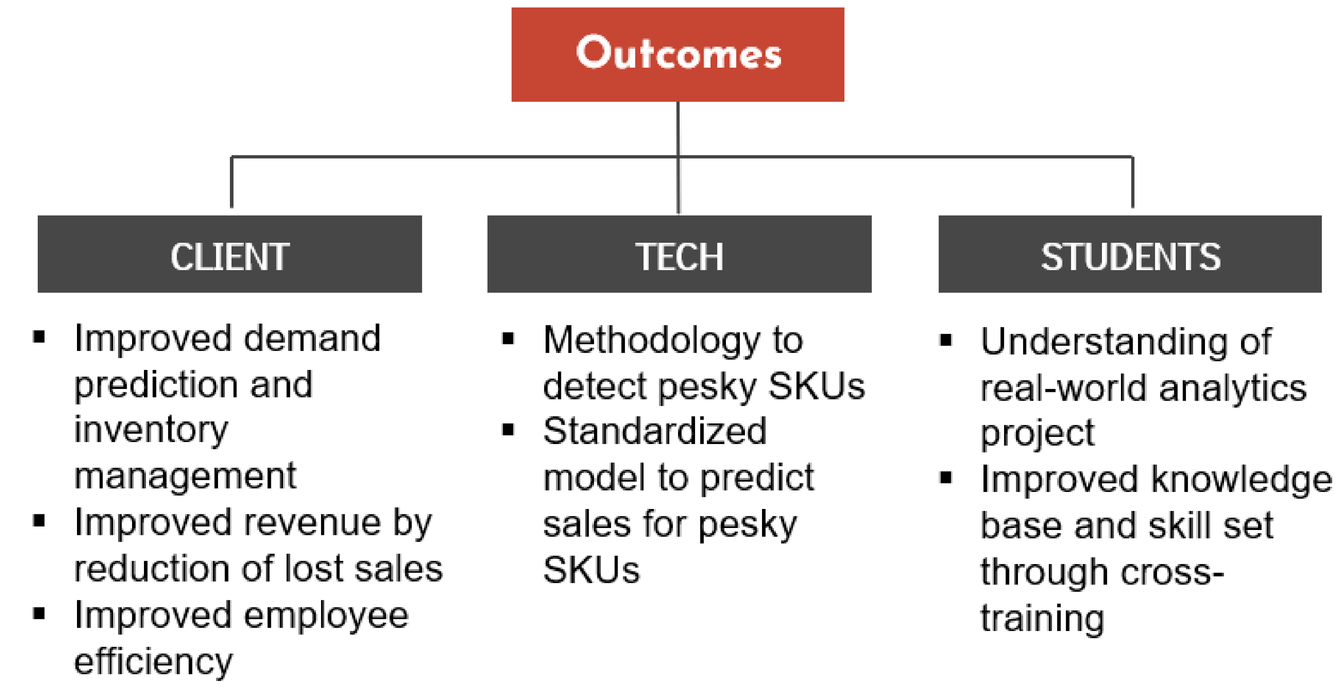

Fig 5. Project Outcomes

CONCLUSIONS

- Pesky SKUs are the items that have high variability in sales across different stores. Tukey's boxplot extended to the log-IQ method showed more accurate results identifying them in comparison to the z-score method. Thresholds were customized for each data type.
- A surprisingly large number of the SKUs with sales were identified as pesky when checked at a store level.
- The best model built for all three product types was based on Random Forest Regression, on comparing RMSE for the holdout set with other algorithms. Hyper-parameter tuning was also performed accordingly.

ACKNOWLEDGEMENTS

We would like to thank Professor Yang Wang, Professor Matthew A. Lanham and our client for their guidance and support throughout this project.