Forecasting Pesky SKUs for Auto Parts Retailer

Adithya Bharadwaj Umamahesh, Apoorva Singh, Jose Manuel San Martin Galindo, Saachi Lalwani,
Shubham Agarwal, Yang Wang
Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907
aumamahe@purdue.edu; singh749@purdue.edu; jsanmar@purdue.edu; lalwanis@purdue.edu,
agarw275@purdue.edu; yangwang@purdue.edu

ABSTRACT

The current issue faced by the client involved lost sales and increased holding costs for leftover inventory. Both issues have a direct impact on the economic profits of the firm and are thus of pressing importance to the company. We have used historical sales data in our project in order to better understand the patterns in sales which can then give us an idea of future sales.

Through this study, we have identified anomalous SKUs based on outlier detection and understanding the statistical significance of each input predictor. We have defined thresholds in sales per store amount to classify each SKU as "pesky", i.e., underperforming in some stores and overperforming in others, or not. Further, we have attempted to forecast the demand for these pesky SKUs in order to improve the inventory management and sales reporting of the firm. We explored and applied prediction models including linear, random forest and lasso regression. This will not only reduce holding costs and avoid lost sales, but also streamline the supply chain as it gives the client a better understanding of the parts that need to be supplied to each store.

Keywords: machine learning, predictive modeling, forecasting, demand prediction, pesky skus, outlier handling

INTRODUCTION

In the retail industry, understanding the movement of products, assortment planning, and demand forecasting play a key role in staying profitable. Companies invest quite heavily in understanding the customer demands based on a multitude of factors, including but not limited to individual demographics, geography, seasonal changes. “The better organizations become at forecasting, the greater their ability to make viable preparations for the future.” (Business Insider, 2011)

A stock keeping unit (SKU) is an identifier that is used to track each unique item in the inventory. The ability to forecast demand at an item level can help with effective inventory management. Sometimes, companies need to make production and procurement decisions at the beginning of a product’s lifecycle before any demand is realized. (Kurawarwala & Matsuo, 1996). SKU forecasting is a constantly evolving field with consistent growth and development. SKU prediction methodologies vary from simple trend forecasting to AI models. Shelf Engine is a company that uses an AI engine to predict consumer demand with high precision. It has helped multiple retailers in eliminating food waste by automating ordering for every SKU, every day, in every store based on the demand prediction. (Dhinakaran, 2022)

The auto part industry, commonly known as the automotive aftermarket industry, involves manufacturing, distribution, retail, and installation of light auto parts. It is a direct-to-consumer market that meets the needs of individual vehicle owners. As of 2018, the United States automotive aftermarket was worth USD 75.31 billion and was expected to continue to grow. (Grand View Research, 2019). The client in this study is an auto part retailer across North America. With over 4000 stores across the continent, they supply parts for motorcycles, cars, and trucks. This gives an insight into the wide assortment of inventory carried by them. The sales naturally depend upon the local terrain and seasons, which is what may lead to anomalous retail trends.

In this project, we have attempted to build an accurate forecast for a group of SKUs that have unusually low performance in certain stores as compared to the majority. For the purpose of this paper, we will refer to them as ‘pesky SKUs’. Identifying outlier SKUs can become critical to the demand forecasting of the retail outlets, as there are heavy costs associated with improper inventory management. If a product is not stocked adequately, the company will lose sales and customers may turn to competitors. However, if there is an oversupply of certain products, inventory holding costs apply. The forecast can help determine the right time to add or remove a SKU from the store’s portfolio to optimize profits. Our research aims to answer the following questions – How can we identify pesky SKUs? Are there signals in the input variables? What outlier handling techniques could be used? What is the best way to develop an accurate forecast for pesky SKUs?

There could be a myriad of reasons why an SKU has higher demand in certain stores as opposed to others, e.g., adverse weather conditions in some areas would cause higher demand for certain repair tools, smaller areas with low-end customer base would not have demand for parts related to luxury vehicles, new competitor stores that provide vehicle accessories could lead to reduced sales for such products in the area. We intend to include such factors in our forecasting model and provide predictions at an SKU-store level.

In the remainder of this paper, we will walk you through our process in the following manner: A literature review to familiarize you with the existing work being performed in this domain, a description of the dataset we have used, and methodologies proposed by us for pesky SKU forecasting. Further, we will discuss the various models we have developed and tested for this project. Finally, we will be discussing the outcome of our research, its implications for the client, and future scope for this topic.

LITERATURE REVIEW

A stock keeping unit is a unique item for sale, purchase, or tracking in inventory, such as a product or service, and any features connected with the item type that distinguish it from other item types. In inventory management and control, SKUs are the bricks that make up the inventory structure. Any good inventory forecasting software solution will be based on a similar level of examination of prior sales data. They commonly forecast at the SKU level because SKU-level data allows you to learn more about your purchases and the people who made those purchases.

One of the key questions we are addressing is whether the methodology to forecast demand depends on the nature of the demand. A major issue in many previous studies was to predict intermittent demand for slow-moving items as this is short term in nature. Moving averages or exponential smoothing methods have been widely used for such cases with intermittent demands. Researchers have also tried to use regression which identifies the relationship between factors affecting demand. For long term demand forecasting, sophisticated mathematical models have been developed that use regression techniques to estimate coefficients and sequentially apply nonlinear relationships.

Initial papers on this topic assumed that demand is completely known. However, research, conducted by Arrow, Harris & Marschak, 1951 relaxed this assumption and mentioned that only the mean and the standard deviation of the demand is known. They proposed three approaches which can be used to predict the demand accurately. The first approach is to choose the most important subset of predictors, the second approach was to build prediction models on the summaries of predictor variables and the third approach talked about backward subset selection to arrive at the best predictor variables. John Kohavi, &Pfleger, 1994 proposed the use of best sub selection for the choice of predictor variables. However, when the numbers of predictors are large, this method was not very efficient, and the use of heuristic optimization models have been widely advocated like stepwise regression, forward and backward feature selection algorithms.

Tibshirani (1996) used a method called LASSO which is a regression technique that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the using the simple regression techniques and then adjust for any additional resulting statistical model. Many studies have added the target product's promotional characteristics into their forecasting models to improve SKU sales forecasting in the context of promotions. (Cooper, Baron, Levy, Swisher, & Gogos, 1999; Huang et al., 2014). Research try and predict the baseline case scenario promotional sales that might add to the total sales.

Since we have a variety of SKU demand patterns, some studies show that simple statistical models like linear regression are unlikely to produce accurate results. Research suggests the use of nonlinear models which is an improvement to the linear regression models. (Kneib, 2013) proposed quantile regression, where specific quantiles of the response variable (demand) are linked to covariates.

One of the premium papers written on this topic was written by Croston (1972). Croston’s model was based on the assumption that when stock is replenished in the system, it will always certainly be a function of demand occurred in the most recent interval. Forecasting based on this model results in overestimating the mean demand and underestimating the variance. Croston further went on to revise the model using exponential smoothing scheme for updating the expected demand and expected time gap in active periods. Another research by Snyder (2002) identified some inconsistencies in the original paper of Croston and used a time dependent Bernoulli process along with exponential smoothing.

Hierarchical forecasting, another significant forecasting framework, is frequently employed in the field of business forecasting and consists of two approaches: top-down (TD) and bottom-up (BU). (Li & Lim, 2018). However, this is not a very widely used forecasting method for demand determination. There are two important factors that should be considered while forecasting demand for SKU's. First are the demand properties that should be captured while forecasting and second is the appropriate time period. To address these points, a study was conducted by Williams (1984) who created a classification strategy for handling different stock keeping units. Williams used method of demand pattern categorization based on the idea of variance partition. The researched published developed a 5-quadrant model which helped in identification of most appropriate method of forecasting and inventory control method. Two other research studies proposed a multi layered perceptron (MLP) model for forecasting this (Gutierrez et al., 2008, Mukhopadhyay et al., 2012). The model factored in the demand for the immediately preceding period and last two non-zero demand transactions at the end of the immediately preceding period as inputs.

According to a recent research paper published in the Journal of Retailing, "Optimizing assortments and portfolios is essential to decrease failure rates of individual SKUs. ML approaches can evolve to complementary support tools for such management problems" (Farris et al., 2021). The research also revealed that the distribution of SKUs is influenced by several factors, the most important being store sizes, store category specialization, brand line length, parent brand overall performance, and sales consistency. The methodologies and results presented aided CPG marketers (suppliers and retailers) in determining whether SKUs are under-performing, performing well or over-performing, as well as the factors that may be contributing to that performance. The model that was used was weighted random forests, which accurately predicted 83% of under- and over-performing SKUs in the velocity model.

Table 1: Papers referenced in the literature review

Authors	Motivation	Algorithm Used
J.D. Croston (1972)	Forecasts in stock control systems: overcoming inappropriate stock levels due to intermittent demands	Exponential Smoothing
Snyder et al. (2002)	Demand forecasting for items with intermittent demand	Time dependent Bernoulli process
Li & Lim (2018)	Intermittent demand forecasting for a retailer: self-improvement procedure for Croston based methods	Greedy 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
John Kohavi, &Pfleger	Variable selection for demand prediction	Best sub selection
Tibshirani (1996)	Variable selection and standardisation	LASSO
Kneib, 2013	Overcome the challenges of simple linear regression	Quantile regression

DATA

In this project, we have used data from our client to perform data modeling and forecasting. As we are working with a client in the retail sector, our data revolves around sales along with store and SKU features.

The data is primarily divided into prediction inputs (pi) and demand prediction (dp). Prediction inputs include various data around SKUs as described in Table 1 below. These prediction inputs will be used as inputs to train our model. Demand prediction contains the data generated by the client’s current forecasting system. This data will be used to determine accuracy of the current model and measure the performance of our new forecasting model.

Table 2: Data used for modeling (pi)

Column	Description
sku_number	unique number used to internally track an inventory
store_number	store id
merchandise_group_desc	product category/ bpg
qty_sold	quantity of products sold
sum_py_qty_sold_on_hand	sum of product stocked in the store past year
sum_cy_qty_sold_on_hand	sum of product stocked in the store current year
sum_cy_qty_sold_transfer	sum of product transfered from other stores current year
sum_py_qty_sold_transfer	sum of product transfered from other stores past year
lookup_cnt	lookup count (looked up product for customer but did not sell; demand quantity)
lookup_cnt	lookup count
failure_sales	related to vio
ts_forecast	time series forecast
lost_qty	lost sales current year
ss_sales	specific sales type
sales_signal	sales indicator for the year (maybe similar to trend)
unit_sales	unit sales
projected_growth_pct	projected growth percent
other_unit_pls_lost_sales	subsection of lost sales
other_unit_pls_lost_sales	subsection of lost sales for current year
weighted_lookup_cnt	weighted lookup count
cy_periods_in_stock	current year periods in stock (13 periods, 4 weeks each)
sales_cost	sales cost
pop_est	population estimate
lifecycle	useful life of part
sku_existence	sku existence current year
age	age demographics
pop_density	population density
other_unit_pls_lost_sales_py	subsection lost sales past year
total_vio	total vehicales in operations (based on store area)
median_household_income	household income
other_gross_sales	gross sales
lifecycle_pre_peak_post	pre if new sku, peak if prime selling years, post for older cars/past prime
adjusted_lifecycle_cy	adjusted useful life
sold_since_maxi	sold since added to store
part_type	part type
sku_store_pdq	sku store (maybe a 0/1) (Is 0 or 12)
unadjusted_total_vio	unadjusted total vehicles in operation
sku_existence_py	sku existence vehicles in operation
application_count	sku-specific types of vehicles it would fit (Some do not have a count if it goes in everything)
sku_store_pdq_cy	sku store pdq current year

establishments	census data, registered businesses in area
road_quality_index	road quality index
filter_reason	filter reason (superceded, discontinued, etc.)
avg_cluster_unit_sales	avg unit sales for cluster
avg_cluster_lost_sales	average lost sales per cluster
adjusted_avg_cluster_sales	adjusted average cluster sales
avg_cluster_total_sales	avg total sales for cluster
bpg	base product group
pct_white	percent of population that is white
pct_college	percent of the population that is in college
pct_blue_collar	percent that is blue collar
pct_of_lifecycle_remaining	Percent of lifecycle remaining

Table 3: Data used for measuring performance (dp)

store_number	store id
sku_number	sku id
prediction	Sales Qty prediction from model

METHODOLOGY

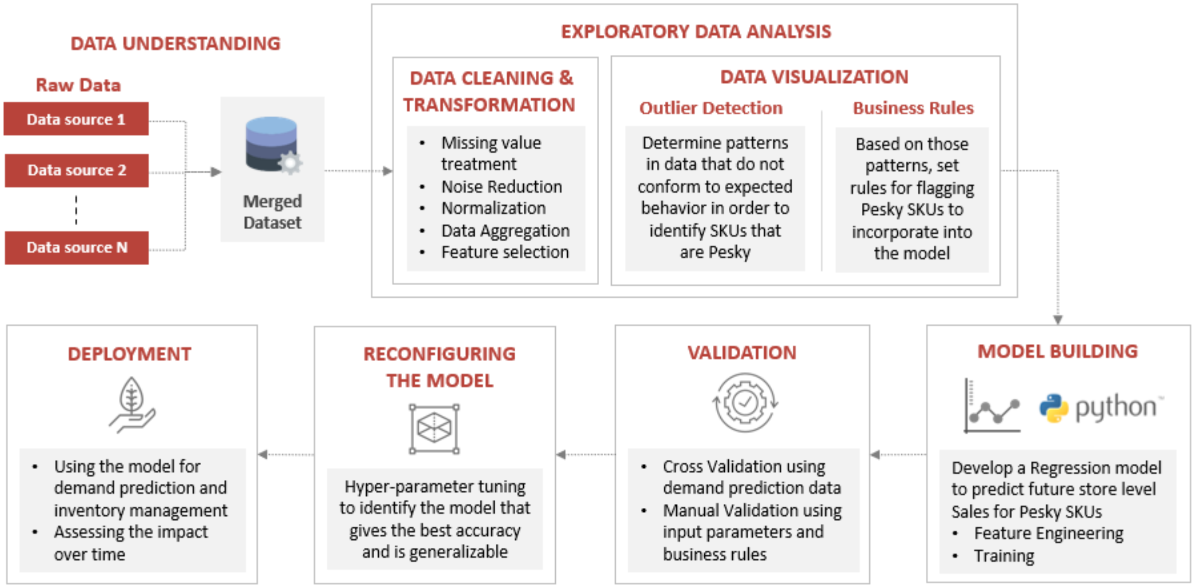
DATA UNDERSTANDING
Raw Data
Data source 1
Data source 2
Data source N
Merged Dataset
EXPLORATORY DATA ANALYSIS
DATA CLEANING & TRANSFORMATION
• Missing value treatment
• Noise Reduction
• Normalization
• Data Aggregation
• Feature selection
DATA VISUALIZATION
Outlier Detection
Determine patterns in data that do not conform to expected behavior in order to identify SKUs that are Pesky
Business Rules
Based on those patterns, set rules for flagging Pesky SKUs to incorporate into the model
MODEL BUILDING
python
Develop a Regression model to predict future store level Sales for Pesky SKUs
• Feature Engineering
• Training
VALIDATION
• Cross Validation using demand prediction data
• Manual Validation using input parameters and business rules
RECONFIGURING THE MODEL
Hyper-parameter tuning to identify the model that gives the best accuracy and is generalizable
DEPLOYMENT
• Using the model for demand prediction and inventory management
• Assessing the impact over time

Fig 1: Methodology used

The sales data that we have at hand is for the current year (with the CY tag) and previous year (with the PY tag). We used the previous year sales as predictors and current year data as target to train the model. We performed EDA on the data to select the best predictors for the model. Following methods were used for the same -

1. Backward Elimination: To eliminate the variables which are not significant thereby tuning the model to achieve a better accuracy.
2. Feature Selection and Importance: We used inbuilt feature selection libraries to pick the right features to run the model by analyzing the feature importance plot.
3. Correlation: We measured the correlation between the variables to see how they interact with each other and the target using the correlation matrix.

A correlation matrix can be used to summarize data, as an input to a more sophisticated study, or as a diagnostic tool for further analysis. The choice of correlation statistic, coding of the variables, management of missing data, and presentation are all key factors when producing a correlation matrix. We intend to pick out the relevant features by analyzing the matrix.

Before building the models, we split the data into training (60%), validation (20%) and test (20%) sets with a set random state. Next, we built the following models using the selected features. The model hyperparameters were then tuned to decrease the Root Mean Square Error (RMSE) for the validation set. Once a low error was achieved, the RMSE was calculated on the test set to make sure the model is generalized. The model with the lowest validation and test RMSE was selected as the best model.

MODELS

Following models were built and compared based on prediction accuracy:

1. Linear Regression: Since our aim here is prediction, linear regression would be used to fit a predictive model to a collection of observed response and explanatory variable values. If new values of the explanatory variables are obtained without an associated response value after creating such a model, the fitted model may be used to predict the response. We aim to predict the current year’s sales value using the different sets of previous year’s sales data that we have and use the existing current year’s sales to measure the model’s accuracy.
2. Random Forest: It is an ensemble learning approach for regression that works by building numerous decision trees during training. It works effectively on huge databases and can handle thousands of input variables. In the process, we can single out the list of significant variables that tend to impact the target effectively and tune the parameters accordingly. Random choice forests also adjust for the tendency of decision trees to overfit their training set.

1. Lasso Regression: LASSO (Least Absolute Shrinkage and Selection Operator) is a method that performs both variable selection and regularization to enhance the prediction accuracy and interpretability of the resulting statistical model. This is a good option for prediction since it improves the quality of predictions by shrinking regression coefficients.

To measure the accuracy of the models described, the following KPIs (key performance indicators) would be used to measure the accuracy of the models:

1. Root Mean Squared Error: It is the square root of Mean Squared error. It measures the standard deviation of residuals.
2. R-Squared: It represents the proportion of the variance in the dependent variable which is explained by the linear regression model. It is a scale-free score, meaning that regardless matter how little or huge the values are, R square will be less than one.

We will also simultaneously try to identify significant variables that contribute to the SKUs being ‘pesky’ to tweak our model in such a way that yields the best result. Eventually, we intend to derive insights from the variables to reason out why the inconsistencies are caused in the first place and if there are ways and methodologies to mitigate them.

RESULTS

As a first set of results, we identified the Pesky SKUs from our client's portfolio. About 30% of items were flagged using Tukey’s Boxplot extended to the log-IQ method. These were items with unusually high sales in a few stores as compared to the others. After a detailed EDA stage, we built different Supervised Learning regression models. Below are the summarized statistical results from the various models tried:

Table 4: Summary of results

Method	**Hyperparameters**	**R-Squared**	**Validation RMSE**	**Test RMSE**
Multiple Linear Regression	-	86.3	5.54	5.62
Lasso	Lambda $\lambda = 1$		5.08	5.37
Lasso	Lambda $\lambda = 0.011572$		4.72	4.84
Random Forest	No. of trees = 15 Max depth = 7 Max features = 5	90.2	4.62	4.75
Random Forest	No. of trees = 25 Max depth = 9 Max features = 7	90.6	4.61	4.64
Random Forest	No. of trees = 20 Max depth = None Max features = Auto	90.7	4.64	4.59

Random Forest provided the best predictions for the majority of SKUs in the holdout dataset. We were able to prevent overfitting by making sure both the validation and test RMSE are small.

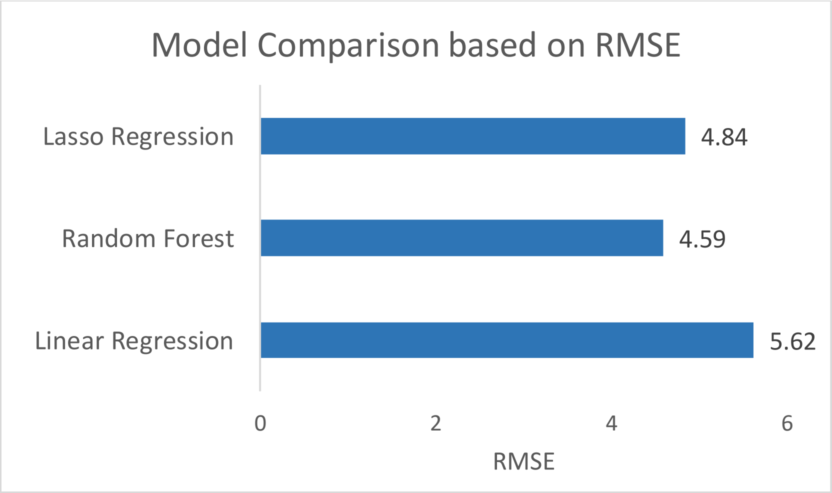

Fig 2: RMSE comparison for different algorithms

Scatterplots are useful to visualize how the model performs by plotting the actual values in our data against the values predicted by the model. The scatter plots below display the actual values along the X-axis and predicted values along the Y-axis.

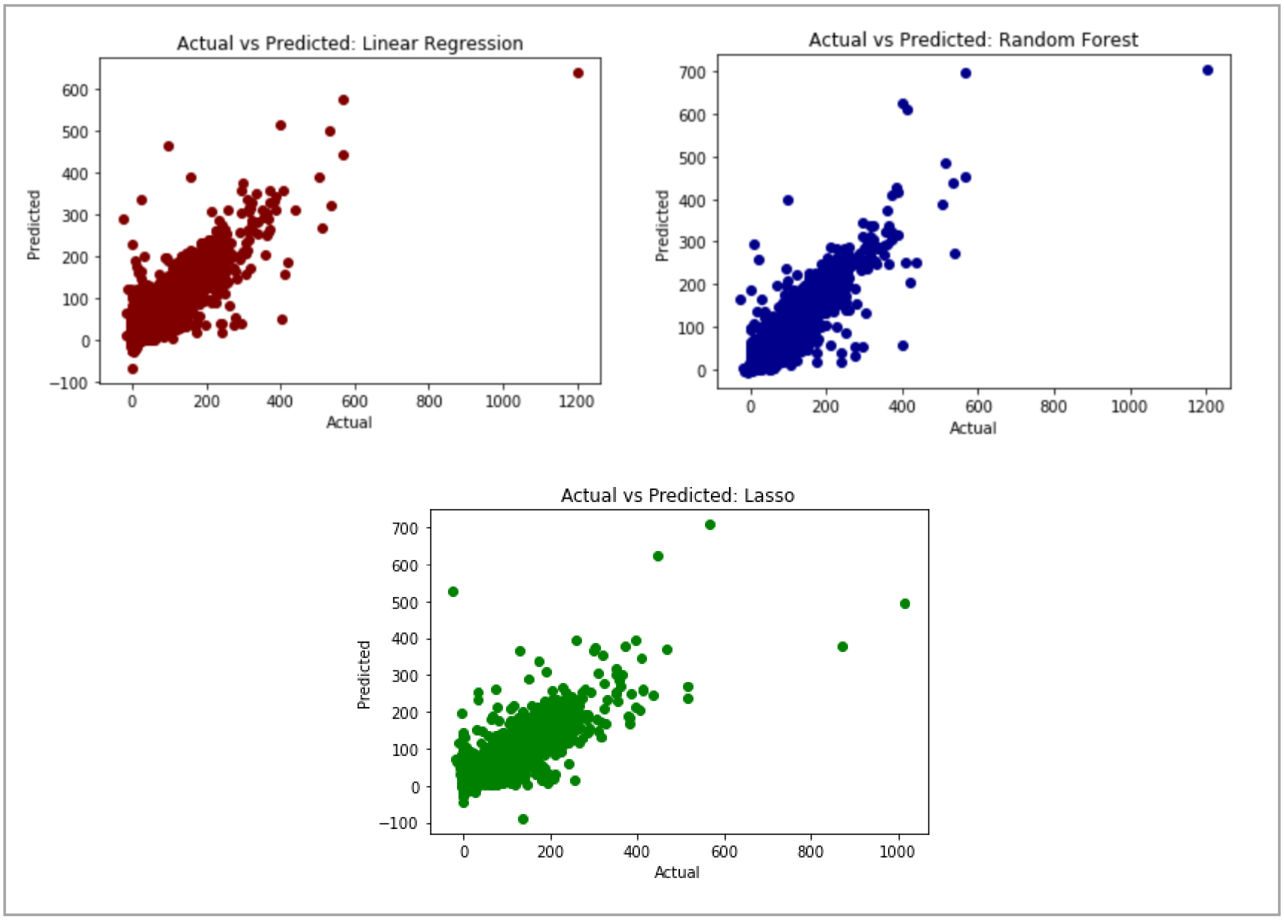

Fig 3: Scatterplot of actual vs predicted sales quantity for different algorithms

CONCLUSIONS

The research as well as the work implemented in this paper will be vital in helping our client have a fundamental understanding regarding how anomalous SKUs should be identified and treated in the modeling process. Furthermore, the precision of our forecasting will be used by them to effectively manage their inventories correctly for each store and avoid lost sales due to unavailability or high wait times. Eventually, our client will benefit from higher profits and better sales reporting, which is crucial in getting data-driven insights thereby improving their business.

Identifying a pesky SKU was made using the Tukey's box plot statistical method and considering the existence of possible and probable outliers in the sales data distribution per SKU overall the stores. The best way to develop an accurate forecast for pesky SKUs is stated in the methodology used in this paper and the models with significantly good performances have been described in our results section.

There were signals in the input variables that led us to model and forecast the pesky SKUs correctly. Essentially, these features were obtained in our Exploratory Data Analysis work and the many plots we created during that stage.

Assumptions were made regarding the definition of pesky SKUs and how to treat them in the modeling stage. However, we believe that more investigation is required about a standardized way to define "pesky" that can help measure and compare performance with competitors.

Business Impact

The outlier detection methodology will identify SKUs with high variability in sales across stores. These items are the leading 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 the 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. **Better vendor and supplier relationships**: This will help the company to determine which products sell and at what volume. Companies can use this knowledge to leverage better vendor and supplier contracts
4. **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
5. **Increased productivity and profits**: Improved inventory management helps to save time for employees which can be utilized in other activities. This also leads to better inventory planning which results in higher inventory turnover, resulting in higher profits.

Another beneficial takeaway from this project has been the immense knowledge sharing – between the students and the clients, between the students and the professor and more importantly, we were able to connect with other students working on the program to create a healthy learning environment impacting the entire group significantly.

REFERENCES

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Journal of the Operational Research Society*, *23*(3). https://doi.org/10.1057/jors.1972.50

Dhinakaran, A. (2022, January 27). Shelf engine's CEO on disruptive innovation without disruptive adoption and the AI-driven future of grocery retail. Forbes. https://www.forbes.com/sites/aparnadhinakaran/2022/01/27/shelf-engines-ceo-on-disruptive-innovation-without-disruptive-adoption-and-the-ai-driven-future-of-grocery-retail

Grand View Research (2019, October). U.S. automotive aftermarket size, share & trends analysis report by replacement part, by distribution channel (retailers, wholesalers), by service channel, by certification, and segments forecasts, 2019 – 2025. https://www.grandviewresearch.com/industry-analysis/us-automotive-aftermarket

Gutierrez, R. S., Solis, A. O., & Mukhopadhyay, S. (2008). Lumpy demand forecasting using neural networks. *International Journal of Production Economics*, *111*(2). https://doi.org/10.1016/j.ijpe.2007.01.007

Hirche, M., Farris, P. W., Greenacre, L., Quan, Y., & Wei, S. (2021). Predicting under and overperforming skus within the distribution–market share relationship. *Journal of Retailing*, 97(4). https://doi.org/10.1016/j.jretai.2021.04.002

Insider Studios & Salesforce. (2021, November 11). How to drive predictable revenue with more accurate sales forecasting. Business Insider. https://www.businessinsider.com/sc/how-to-improve-sales-forecasting-2021-11

Kurawarwala, A. A., & Matsuo, H. (1996). Forecasting and inventory management of short life-cycle products. *Operations Research*, *44*(1). https://doi.org/10.1287/opre.44.1.131

Li, C., & Lim, A. (2018). A greedy aggregation–decomposition method for intermittent demand forecasting in fashion retailing. *European Journal of Operational Research*, 269(3). https://doi.org/10.1016/j.ejor.2018.02.029

Mukhopadhyay, S., Solis, A. O., & Gutierrez, R. S. (2012). The accuracy of non-traditional versus traditional methods of forecasting lumpy demand. Journal of Forecasting, 31(8). https://doi.org/10.1002/for.1242

Snyder, R. (2002). Forecasting sales of slow and fast moving inventories. *European Journal of Operational Research*, 140(3). https://doi.org/10.1016/S0377-2217(01)00231-4

Williams, T. M. (1984). Stock control with sporadic and slow-moving demand. *Journal of the Operational Research Society*, 35 (10). https://doi.org/10.1057/jors.1984.185

Hohberg, M., Peter, P., & Kneib, T. (2018). Generalized additive models for location, scale and shape for program evaluation: A guide to practice. *Working Paper*. https://arxiv.org/pdf/1806.09386.pdf.

Hastie, T. J., & Tibshirani, R. J. (1986). Generalized Additive models. *Statistical Science, 1*(3), 297–318. https://pdodds.w3.uvm.edu/files/papers/others/1986/hastie1986a.pdf