Predictive Analytics in the Food Supply Chain

iFoodDS®

Affordable, high-quality fresh products in grocery stores and restaurants have become increasingly important to consumers. A 2022 report released by FMI, the Food Industry Association, stated that the number one factor produce consumers consider when shopping at grocery stores is price, on par with appearance and quality[1]. In another report, 68% of diners surveyed ranked food quality as the most important factor when deciding where to dine[2]. For brands throughout the fresh produce supply chain, growing, delivering, and selling affordable, high-quality perishables is critical to being a leader in the industry.

Inspectors commonly complete inspections on paper, and these processes are costly, inefficient, and inconsistent. Inconsistent inspection processes between inspectors and distribution centers lead to an unpredictable consumer experience in stores and restaurants, impacting sales and customer loyalty.

Additionally, paper inspection processes don't allow tracking the quality of perishables provided by suppliers over time, resulting in blind spots and limited visibility into the historical performance of suppliers during different seasons, weather cycles, and more.

The iFoodDS Quality Management Solution has allowed companies to evolve, making it easy to capture quality metrics during various inspection workflows digitally. Inspection data is automatically stored in our secure cloud-based platform and used to analyze the performance of the food supply chain, enabling buyers to optimize sourcing decisions to ensure they are procuring the most affordable, highest-quality product during any given season.

But instead of using quality inspection data to only report on the historical performance of suppliers and commodities, imagine if you could use that same data to predict the quality of a product before it arrives.

This white paper describes how iFoodDS engineers use machine learning and artificial intelligence to answer this question.

We outline the data preparation process, feature engineering, model development, and model evaluation to show how brands could predict the quality of incoming produce based on historical inspection results. Specifically, we will predict the unsatisfactory rates of produce at a distribution center based on previous orders received from suppliers.

Data Preparation

The iFoodDS engineering team isolated inspection data to one customer. The data focused on the receiving workflow activity for the fruits and vegetable category from 2018 to 2021. When a distribution center receives a product, one or more quality checks occur. Inspection results are classified into several different reasons at the receiving dock of the DC with unsatisfactory rates varying over time and seasonality. Common reasons for an unsatisfactory inspection are due to disease, physical damage, packaging issues, blemishes, and maturity.

In addition to only focusing on a single customer, the engineering team focused on suppliers, with the total number of product suppliers limited to those with more than ten inspections and those with more than one year of transaction history. Engineers further edited this data to focus on a single product item, strawberries.

The team categorized the time series into:

- year
- month
- bi-weekly subsections

This categorization enabled us to create ‘Ddata buckets’ using discrete trailing bi-weekly periods.

Feature Engineering

Before developing a model, the engineering team chose features to ensure the model would encompass a wide array of scenarios. The grain of the data focused on produce inspections (audits) received in DCs after inspectors had accepted the product for processing and before delivery to the end customer. All audits took place at the DC receiving dock.

We used variations of yearly, monthly, and bi-weekly time series ranges for dimensions. Additionally, we used regional and product specificity for categorical characteristics.

The engineering team used the count of audits and a derived unsatisfactory rate for measures. Any audit with a rating of ‘Accept’ was deemed satisfactory. In contrast, audits labeled ‘Reject,’ ‘Accept with Issues,’ or ‘Missing’ were considered unsatisfactory for this evaluation. The unsatisfactory rate is defined as the number of unsatisfactory audit counts divided by total audit counts.

Model Development

iFoodDS evaluated several machine learning models, including:

- SARIMA
- TBATS
- Neural Network Autoregression
- Random Forest Regression (RFR)

After careful consideration, the iFoodDS engineering team used a combination of TBATS and RFR for this analysis. Several produce types were evaluated, with strawberries being the most reliable.

The TBATS and RFR hybrid model uses a set of 1,000 decision trees and set the random state to 2,000. The team used 80% of the data to train the model, with the remaining 20% allocated for testing. iFoodDS set a forecast horizon at 26 periods or one year. As a TBATS model lends itself well to seasonality, a seasonal period of 26 or one year was also used.

The baseline uses a Naive forecast. In this case, it represents the observed inspection results from the last period as the forecast for the next period without considering any predictions or factor adjustments. The Naive forecast was developed over time, as opposed to year over year, as this proved less accurate and had a significantly worse MAPE score.

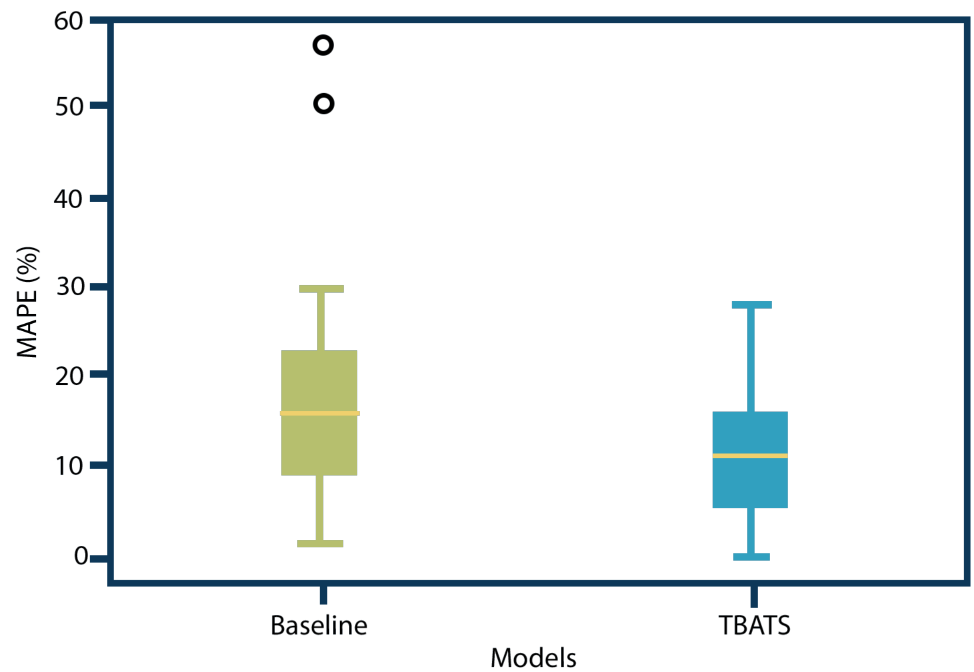
Model Evaluation

To evaluate the results, iFoodDS compared the Naive baseline to the more advanced TBATS and RFR hybrid predictive model.

We evaluated the performance metrics based on the Symmetric Mean Absolute Percentage Error (SMAPE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Scaled Error (MASE).

The iFoodDS engineering team ran both models against historical data having a 33.5% unsatisfactory inspection rate; the baseline prediction was 28.5%. In comparison, the TBATS and RFR hybrid model predicted an unsatisfactory rate of 27.7%. While both predictions were lower for the first period, the MAPE for the baseline is 18.11% and 11.26% for the TBATS and RFR hybrid model for the entire 26 periods predicted. This means that, on average, the hybrid model could predict the unsatisfactory rate for the following two-week period within a smaller variance than the baseline.

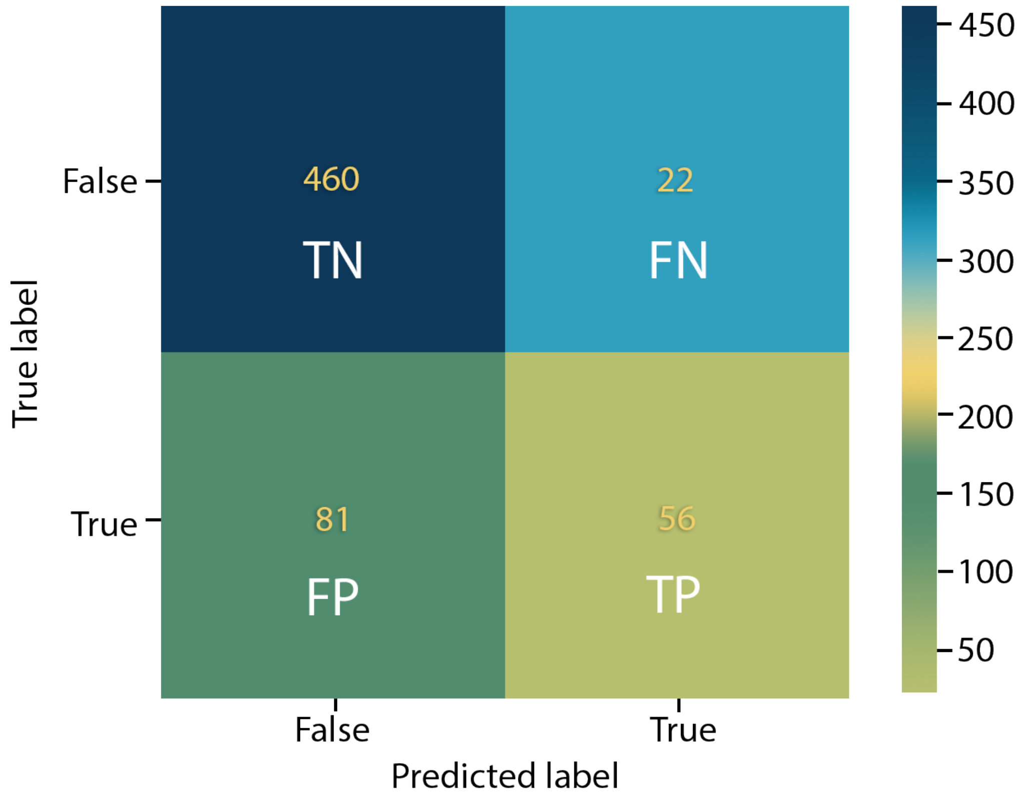
A box plot shows the marginal improvement of the TBATS & RFR hybrid model to the baseline prediction in MAPE and interquartile range (IQR). The IQR of the baseline is 0.131, and the hybrid is 0.108. Note the outliers returned by the baseline as well.



Further Development

In addition to regression models, classification models can predict the inspection status of incoming ordered cases that still need to be audited. The current order data provides the feature of the model with the inspection status as the target. This assumes no split audits where part of the order is accepted and the remaining is unsatisfactory.

By focusing on a single product category, such as strawberries, we can use customer order data to train a CatBoost Classification model. In this example, we divided order data into four sections, dating from 2018 through March 15, 2023. The model trains with three sections and tests its results on the fourth. Each section rotates until all data are trained. On average, a model tested four times returned an accuracy score of 76%. This trained model was then used to predict the results of the audited produce from March 16th, 2023.



The graph classifies the results of the model into four categories, True Negative (TN), False Negative (FN), False Positive (FP), and True Positive (TP). These categories allow us to evaluate the model's accuracy, which resulted in a 72% precision rate.

A True Positive (TP) indicates that the model predicted the record would have an unsatisfactory inspection result, and the record's inspection result was unsatisfactory. Similarly, a False Positive (FP) indicates that the model predicted a satisfactory inspection result while the record's actual inspection result was unsatisfactory.

The model predicted that 78 orders would be considered "unsatisfactory" upon arrival considering the negative outcomes. When compared to the actual inspection data, 56 orders were accurately predicted (TP), but 22 were accepted and, therefore, an incorrect prediction (FN).

Conclusion

This analysis shows that we can predict when a brand will receive a product with an unsatisfactory status by using a combination of TBATS and Random Forest Regression models.

The results outperform the baseline models, but many opportunities exist for further exploration and refinement. We aim to refine the results to the supplier grain and increase the prediction accuracy. Future iterations will expand the scope and leverage different independent variables, such as travel distance, inspections at shipping points, and weather-related data points.

Looking Ahead

iFoodDS engineers are working to integrate predictive analytics into our Quality Insights solution. Here is just one example:

The iFoodDS Quality Management Solution allows users to flag items requiring special attention from inspectors manually. Our predictive model will use an algorithm to provide a predictive score that Quality Inspectors can use to evaluate suppliers and products as they are received. This will maximize the effectiveness of Quality Control Teams by alerting them when there is a high probability of unsatisfactory products the distribution center is receiving that day.

Endless Possibilities

Leveraging the iFoodDS applications and analytics with a cloud-based machine-learning platform offers endless possibilities.

The most valuable asset in the inspection process is the data. At iFoodDS, we have moved beyond just data collection. We deliver insights that help optimize sourcing decisions and move toward the power of artificial intelligence and machine learning to take quality and freshness to the next level. Predictive Analytics from iFoodDS will give you the tools to foster better supplier partnerships, source the highest quality products, and optimize inspection resources leading to increased consumer satisfaction and sales.

References

1 https://theproducenews.com/headlines/consumers-rank-price-par-appearance-and-ripeness-fresh-produce-purchases

2 https://www.restaurantbusinessonline.com/consumer-trends/dine-customers-rank-food-quality-top-priority-when-choosing-restaurant

iFoodDS®