# *IEOR E4999 Internship Report*

*(Zi Ying) Sheila Teo, Summer Data Scientist at Nestlé USA*

## Executive Summary

This summer, I interned as a Summer Data Scientist (Digital Analytics) at Nestlé USA. I worked at Nestlé's Supply Chain Analytics team, which utilizes Data Science to forecast demand for hundreds of thousands of Nestlé's products, as well as to improve its supply chain efficiencies. As the largest food & beverage company in the world, the opportunity to work at Nestlé's supply chain domain has been exhilaratingly fulfilling, and the project I embarked on brought about dollar savings of $3.4 million for the firm. I am thrilled to be able to share my experience with IEOR faculty through this report.

## Project Introduction & Motivation

I was engaged in an end-to-end Machine Learning project throughout the course of my internship. The problem the Supply Chain Analytics team was facing was that of less-than-ideal accuracy rates when forecasting demand for Nestlé's products. Prior to the start of my internship, the root cause was identified to likely be a situation of "garbage in, garbage out"- to put it colloquially- in which the problem lay with the features (variables) the team selected for usage as inputs to their models.

The precise problem the team faced with its features were that they were collecting data from many avenues, and thus had access to many relevant features that could be used in forecasting product demand. To provide further context, examples of such features relevant in forecasting demand include past demand, promotional data, holiday data, and competitor data. With this large amount of features available to be used for modelling, the team built its demand forecasting models interchangeably in either one of the two ways below:

1. Utilize all the data available to them as features for the model
2. Cherry-pick which data to use as features through intuition and domain expertise

These led to the 2 respective scenarios:

1. The model was built with hundreds of features, which made it difficult to explain the model to stakeholders. This was crucial at Nestlé since the Supply Chain Analytics team would pitch their demand forecasting models to other Demand Planning teams, who would then have the final say of whether or not they bought into the model's forecasts. It was thus key for the Supply Chain Analytics team to produce models that could be understood and interpreted by non-technical Demand Planners in order for them to trust its forecasts.
2. The model was built with a subset of the hundreds of original features, however it is unclear if this subset is an ideal one, and if model accuracy could be improved further if a different subset of features were selected instead. There was room for improvement to base feature selection on more technical methods instead of intuition.

## Project Description & My Contributions

The above outlined situation thereby motivated my internship project, in which I executed the following:

**(a)** Designed an end-to-end framework for feature selection through multiple Data Science techniques. Through this framework, the team would be able to select a subset of features from the hundreds of original ones, thereby firstly increasing model explainability to stakeholders. Using these technical methods for feature selection instead of human intuition, this subset of selected features would also more certainly be the "ideal" subset of features to use, thus alleviating the aforementioned situation of "garbage in, garbage out".

In designing this feature selection framework, I researched extensively on best practices and tested out multiple approaches before settling on the final approach as follows:

- Obtain the importance of each feature using Permutation Feature Importance method with a Random Forest model. In layman terms, this is best explained through the following scenario: Say we have 2 features and we want to forecast the value of a target variable using them. We thus build a Random Forest model with both features and record the accuracy of the model. Now, in order to determine the importance of Feature 1, we randomly permute (shuffle) its values such that we break the relationship between that feature and the target. With this relationship now broken, we then use the same Random Forest model to predict the target. The importance of Feature 1 is thereby the fall in accuracy of the model from before and after it was permuted, since this measures how significant Feature 1 was in helping the model to predict the target. In order to later determine the importance of Feature 2, we repeat the above. This allows us to thereby obtain the feature importance of all features within our dataset.
- Determine the optimal subset of features to use via Recursive Feature Elimination (RFE) with cross-validation. RFE is a recursive process in that it progressively considers smaller and smaller subsets of the original features. It starts with using all the original features, builds a Random Forest model using them, obtains feature importance scores for all the features as well as a model accuracy score on the validation set of cross-validation. Next, it discards the feature identified to be the least important, and repeats the above. This repetition continues until the size of the feature set has been incrementally reduced by 1 until it becomes 0. Following which, we then determine the optimal feature subset to be used as the one that produces the highest model accuracy score on the validation set.

The above explains the technical details behind my devised framework for feature selection. Next, following from this brainstorming process, I proceeded to build the code for running this feature selection tool.

**(b)** Built code on Microsoft Azure Databricks using Apache Spark (PySpark) to bring the above designed framework to life. The above designed framework thus transforms into a concrete feature selection toolkit that can be used by the team in future work.

**(c)** Back-tested this designed feature selection tool by running it on various different time periods to determine if it improves model forecasting accuracy.

**(d)** Explore different levels in which to conduct feature selection. For example, feature selection could be conducted at the "business" level, wherein the term "business" at Nestlé refers to a category of products that Nestlé sells, examples being the Baking business, the Beverage business, and so on. Under this method, one would conduct feature selection for an entire business of products, and utilize this same selected features in the various models for forecasting demand for all products under that business.

On the other hand, feature selection could also be conducted at more granular levels, such as at a product level instead. Within any one business, there are hundreds of products. For example, two different products within Nestlé's Baking business are cookies and brownies. As such, conducting feature selection at the product level instead would mean selecting features to use for forecasting demand for each unique product, and using different features in modelling between different products even within the same business.

To determine the optimal level at which feature selection should be conducted, I further built code for the feature selection process at these various levels, and back-tested from there to obtain this optimal level that resulted in the highest forecasting accuracy.

**(e)** Constructed data visualization dashboard in Microsoft Power BI to convey results of utilizing new feature selection tool. I specifically designed this dashboard to be interactive and visually impactful, allowing users to toggle the dashboard to view products or time periods of particular interest. I further ensured that the dashboard was intuitive and easy to read, allowing users to obtain a summary of the results at a quick glance- a key feature of any successful data visualization. A screenshot of this dashboard is included in Appendix A.

**(f)** Presented to Nestlé leadership on business impacts of harnessing this new feature selection tool in demand forecasting. Importantly, I designed impactful slides that visually broke down technical machine learning concepts for a non-technical audience, as well as synthesized the impact of my internship project in terms of business-focused metrics more relevant to leadership priorities, such as dollar and labor-hour savings. These presentation slides are included in Appendix B.

The above (a) to (f) were the agreed-upon aspects of my project from the first week. Throughout the course of working on my project, however, I self-initiated a sixth aspect, upon realizing that there was scope for furthering my project and creating even greater value for the team. Below outlines the added issues that the team faced in which I brainstormed a solution for which I could step in and add further value to.

The team had been using traditional Statistics-based models such as ARIMA, ARIMAX, and other regression-based models for time-series demand forecasting. However, they were keen to branch out and adopt more modern machine learning-based models that were commonly used for time-series forecasting work as well, such as ensemble-based models or deep learning neural networks. I had learned this interest upon reaching out and speaking to various team members on my own accord. Upon learning this, I proposed to the team that I could extend my project by not only working on the feature selection, which would have been one aspect of any typical machine learning workflow, but instead also create a new forecasting model for the team that had not yet been used by the team prior. As such, the scope of my project widened to include also the following:

**(g)** Develop code on Microsoft Azure Databricks using Apache Spark for deploying new machine learning model to forecast demand, in which I tested out the usage of multiple types of models, including Random Forests, Gradient Boosted Trees, and Regularized Regression, before settling on a Random Forest model.

One interesting thing about this model development is that I designed a custom scoring metric in order to choose optimal hyperparameters for the model, instead of simply utilizing the traditional metrics of R-squared or Mean Squared Error for a regression task. My custom error metric instead took into account the exact measure that the team was concerned with- a figure termed Demand Planning Accuracy (DPA). As such, instead of optimizing on these aforementioned traditional metrics in tuning the model, I optimized on this DPA metric that was of most significance to the task at hand of demand forecasting, so as to directly influence the yardstick that the team was concerned with. This was a concept that I was introduced to in a machine learning course at Columbia, and I was elated to put it to practice in this modelling aspect of my internship.

Another exciting aspect of this is feature engineering. Upon moving away from traditional Statistics-based time-series models such as ARIMA, to more modern machine learning-based models such as Random Forests, it became crucial to engineer features that would help such models learn better by converting such a time-series problem into a machine learning one. As such, I created new features from original ones that would help these ML models uncover patterns and relationships within the data better. For example, I created lag variables of past demand and consumption information, converted the mean and standard deviation of such features into new features to be explicitly fed into the model, and engineered new features that provided the model with useful time information such week number of year, which was a high-cardinality categorical variable that then required encoding. I thus further explored different types of categorical variable encodings such as Hash Encoding in order to determine one that was optimal for my specific use-case. All these proved to be fruitful work in the model-building process.

**(h)** Back-test this new machine learning model against the current Statistics-based models used by the team by running this it on various different time periods to determine if this new model improves forecasting accuracy.

Lastly, as a wrap-up of (a) to (h) above, I further took the initiative to hand over all of my work to the team as seamlessly as possible by not only sharing my codes with them, but also creating a separate documentation of my codes wherein I explained the motivation behind each code chunk written and rationale for organizing the code in a certain manner, so that anyone reading these codes in the future can easily interpret the rationale and thought process behind it and pick it up quickly. This was instrumental as the team would work off of my codes after the internship concluded to integrate my feature selection tool into their current modelling workflow, as well as to adopt my new machine learning models designed.

### Making An Impact at Nestlé

From the various back-testing processes carried out, it was determined that my new feature selection tool as well as machine learning model developed were able to increase Demand Planning Accuracy (DPA) of the demand forecasts across 6 categories of Nestlé products, as follows:

1. Baking products: 12.8% increase in DPA
2. Beverage products: 7.4% increase in DPA
3. Infant Nutrition products: 18.3% increase in DPA
4. Global Brands products: 16.7% increase in DPA
5. Prepared Foods products: 4% increase in DPA

These translate to the following conservative estimates of dollar savings for Nestlé, calculated based on inventory costs saved from more accurate demand forecasting:

1. Baking products: $2 million in savings
2. Beverage products: $351K in savings
3. Infant Nutrition products: 565K in savings
4. Global Brands products: 400K in savings
5. Prepared Foods products: 87K in savings

Overall, the summation of the above leads to a total figure of $3.4 million dollar savings.

After my internship, the Supply Chain Analytics team intends to incorporate my various developed tools & methods into their current modelling workflow in order to bring these figures to fruition.

I have been incredibly fortunate to have been able to bring about the above business impacts with the guidance and support of other Data Scientists and Supply Chain Analysts on the team, and I am endlessly grateful for their mentorship.

### Conclusion

In summary, my work as a Summer Data Scientist (Digital Analytics) at Nestlé consisted of designing and developing a feature selection tool, a new machine learning model, and integrating this developed tool at different levels into the model. Thereafter, I proved that these new approaches taken would indeed improve forecasting accuracy through rigorous back-testing. My journey at Nestlé concluded with a detailed handover of my developed tools and models to the team to ensure seamless integration of my work with the team’s current modelling workflow.

My internship at Nestlé has been an immensely fruitful experience for me. With my newfound knowledge in time-series forecasting for supply chain, I am inspired and excited to work on future big data problems in similar business domains in my future career!