Machine learning based Forecasting for Retail Supply Management

**A Mindtree thought paper**

**Background**

One of the important challenges the Retail industry currently faces is that of Overstocks. A study done by DynamicAction and IHL [1] in 2015 states that overstocks make up for $471.9 billion of the $1.75 trillion Ghost Economy that plagues the Retail Industry.

One of the more important strategic goals of any retailer is to hit the Aggregate Supply sweet spot with the most optimal Supply to Price ratio. At the center of this challenge lies the requirement of a strong Supply forecasting model that should be agile, scalable, and dependable. On top of that, there is a pressing need for the forecasting models to be malleable to dynamic forces like market changes, demand & demography shifts, and take into account local/global events that might affect a products ability to sell within a given time period. Having a Forecasting model that stands robust against Fluctuating demand is the need of the hour.

The best way to build such a forecasting model is to use the advancements made in the field of Machine Learning, apply it to existing Data Analytics that most retailers depend on and build a predictive application system that will help retailers have a scalable and agile forecasting model.

**CPFR, Reactive Models, and Forecast Accuracy**

CPFR (Collaborative Planning, Forecasting and Replenishment) is a GS1 US trademarked framework, which was created with the aim to help retail businesses enhance their supply chain integration and cooperation between trading partners. The Voluntary Interindustry Commerce Solutions (VICS) Association is the body that leads in defining the framework and guidelines for CPFR.

CPFR places high importance on the processes of Demand & Supply Management and Analysis. The central pillar of Demand & Supply management is Supply forecasting (which directly corelates to Sales Forecasting), whose accuracy is crucial in reducing Overstocking and prevention of Out-of-Stock scenarios.

Retailers do not depend on one forecasting model, instead using multiple models or a collection of interdependent models to help them provide insights into Sales, Supply, demand levels, competitors undercutting prices etc. Invariably almost all models used are Time Series Analysis techniques because of the stable nature of said analysis. In Supply forecasting, the most common practice is to perform varying forms of Reference class forecasting based on historical data, promotional buzz or consumer response and factoring in current economic factors (inflation, GDP growth, average median income in a given geo location etc.). This model relies heavily on historical data and therefore is more reactive. In today’s world where your average consumer is always online, with multiple inputs from various media and online services shaping their decision, a model which is primarily reactive leads to a wider error margin in forecasting, thereby leading to the Overstocking phenomenon.

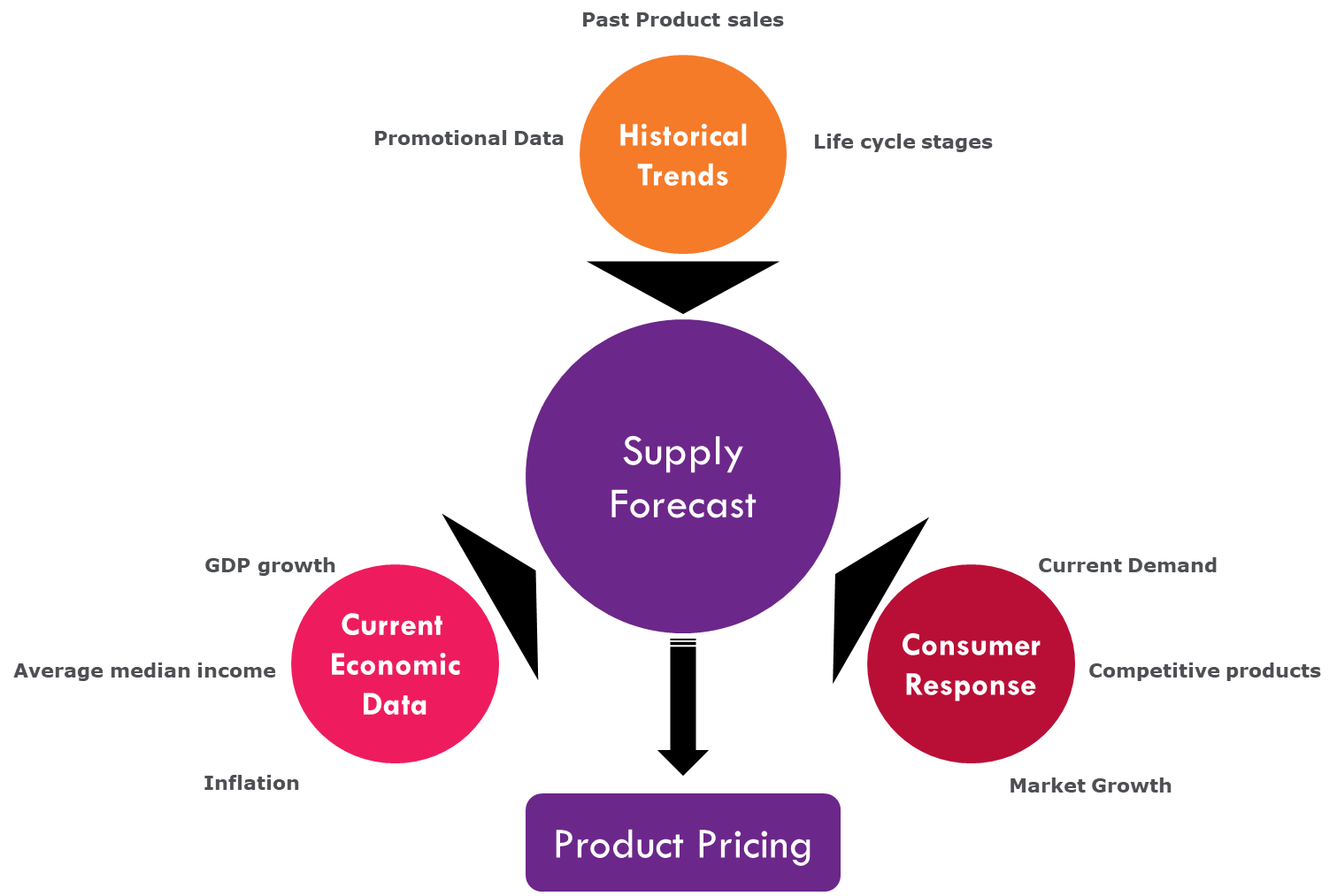


Figure 1: Standard Supply Forecast model

For pricing, the current trend is to use the above model, link it to existing product inventory and provide a dynamic pricing model for the products. All leading retailers use this dynamic pricing based on supply/demand parameters, which only serves to show how important supply forecast models are.

Getting the right supply forecast is tricky but it becomes even more so when dynamic pricing is involved, as Overstocking leads to price reduction which in turn reduces profit margins. Usually, a competing brand suffering from Overstocking invariably affects the pricing of similar brands involved across the market leading to a price slump in the curve which in turn affects the data for the next year supply forecasting.

**AI or Where does Machine Learning come into picture?**

Artificial Intelligence broadly unifies multiple specializations that enable automated machines to perform cognitive functions and independently evolve heuristic capabilities. To enable AI within any industry, engineers use Machine learning or Deep learning algorithms to enable programs to perform independent tasks.

Machine learning (ML) is the ability for computer programs to self-learn from experience without any explicit programming for said learning. The term was coined by Arthur Samuel, a researcher computer gaming and a pioneer in Artificial intelligence, in 1959 while working at IBM [2]. Great strides are being made in the world of Artificial intelligence with Google, Microsoft, IBM and Facebook leading the charge. Google’s Alphabet company Deepmind and IBM’s Watson are among the most famous examples which showcase how Machine Learning is helping overcome challenges within the real world.

With programs that have the ability to constantly learn from various events and adapt to latest data sets and provide predictive analysis, Machine Learning is the right tool to boost data analytics that can solve challenging problems for the Retail industry, like Overstocking. Data Analytics has had a boost in speed and versatility in the last decade with the use of Big Data. The standard rule of thumb in Statistics has always been – the more data you have for your sample, the better your inferences would be. One of the most obvious examples is Online Marketing. As mentioned before about the always online consumer, Big Data analytics has made it easier than ever before for retailers to have localized demographical targeting data to use in their promotional content.

Business Intelligence technologies that perform large scale analysis of structured and/or unstructured data have been the back bone of the Big Data revolution and all forecasting for major retailers are done using one of many Big Data / BI technologies that are available in the market. The ability to perform large scale data analysis on Terabytes and Petabytes of data have made it easier for retailers to streamline their supply chain and pricing models. But there exists an incredible opportunity to improve upon already existing applications and provide tangible cost savings in reducing the Ghost Economy most retailers face, with the use of Artificial Intelligence and Machine Learning.

**Predictive Analysis using ML based programs**

Mindtree is working on integrating Machine learning and Data Analytics into AI modules by training multiple simulated predictive programs, with Supervised learning (Active Learning) algorithms. The modules work on data available from Retailers, applying Time Series regression analysis to come up with supply parameters based on actual historical inputs. Using Machine learning modules enables Mindtree to run the following steps to come up with programs that have the capability of forecasting with significant accuracy –

* Each Simulation is measured against historical supply parameters to compare accuracy with the output supply data, for a specific sample year.
* The modules perform regression analysis and provide the output as year end supply projections for past data with factors leading to sales becoming the Independent variables.
* The simulations that are the closest to the year end supply data for any given year are filtered and then run against similar simulations, but with an intentional Bias introduced in some, for another sample year.
* The bias are data that have been identified as risk factors, Product lifecycle changes, competitions price tactics etc. for the past year.
* The simulations with the introduced bias that are closest to the supply data for the second sample year are again selected.

The above process is repeated iteratively multiple times with varying sample data.

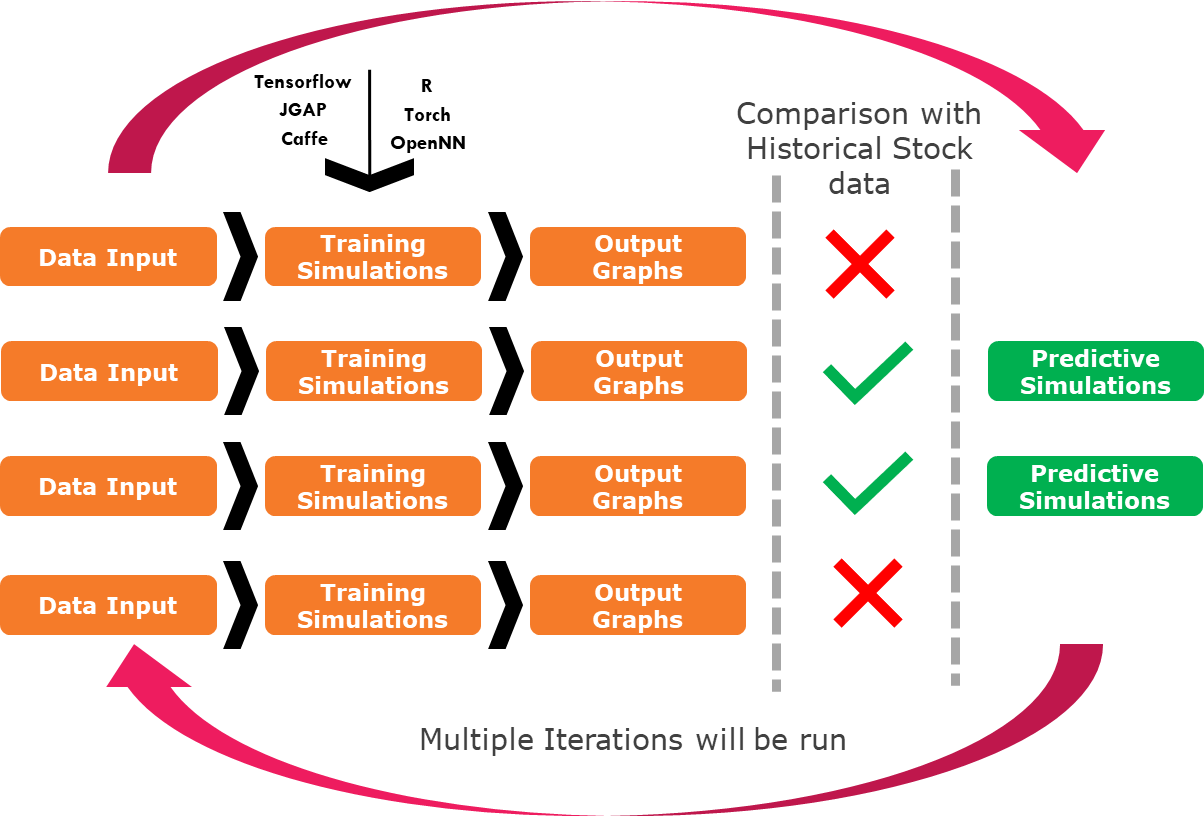


Figure 2: Training Simulations running with historical data, using Supervised learning

With Mindtree’s predictive modules handling multiple sample data to run regression, the following advantages can be realised -

**Multiple iterations advance predictive capabilities**: Running Regression analysis over multiple iterations of Machine learning programs that are simulating historical supply forecasts help narrow simulations that come closest to historical supply data.

**Bias leads to modules that learn for changing Data**: Adding Bias to the simulations helps the modules learn from changes in data. This results in the modules being capable of handling sporadic parameter changes, close to Local/Global events affecting sales.

**Dependable and Scalable forecasting modules**: Predictive Simulations that have “Survived” through multiple iterations of regression analysis, provide dependable and scalable modules that lead to Supply forecasting for a varied set of parameters. Each simulation that “survived” is provided with specific parameters, in relation with other simulations and the forecasting model is solidified.

**Better analysis for the overstocking problem**: The simulations that are tweaked to prioritise those learning algorithms that provide elasticity curves that favour accurate supply data, lead to providing more accurate analysis to the overstocking problems for any given product.

**Iterative modules need smaller Sample data**: As the modules are refined over multiple iterations, the need for a larger sample data reduces, thereby reducing the time taken for the simulations to run. This reduces dependency on compute requirements, quicker analysis and faster forecasting.

The predictive modules will be constantly evolving, are able to factor changes in data inputs at various levels, and will be more agile in forecasting. The predictive modules are built on self learning algorithms using Machine learning libraries provided by frameworks such as Google Tensorflow, JGAP (Java Genetic Algorithm Programing), Caffe, PyTorch etc. Based on requirements from Retailers, the Simulation modules can be tweaked to give the required level of supply parameters, providing a safe margin between overstocking and out-of-stock scenarios.

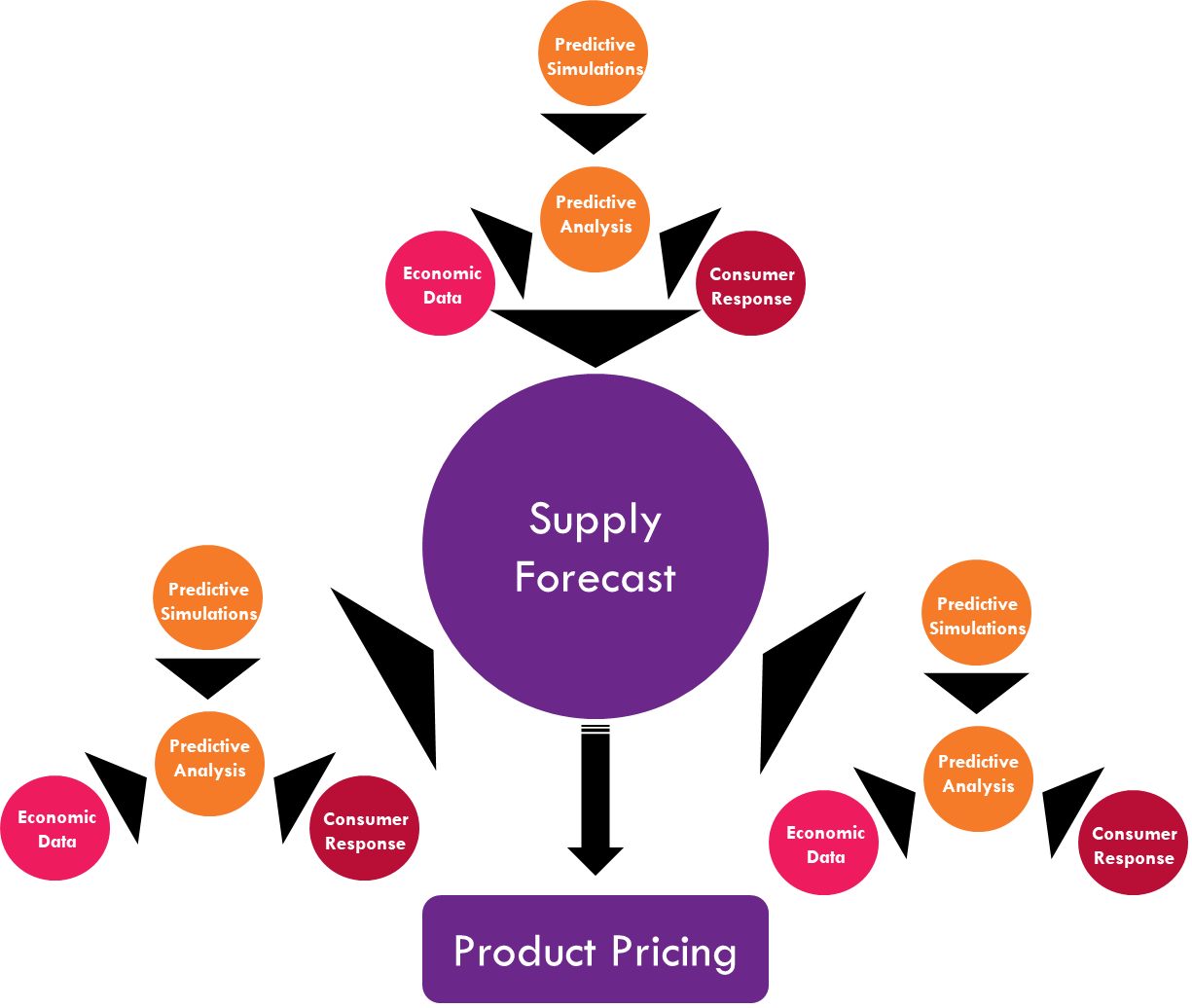


Figure 3: Predictive Analysis Simulations working together to provide better Supply Forecast

Given that the Self learning and evolving Simulations accommodate not only historical data but also localized parameters such as shifts in economic situations, Price changes from competitors, and so on, their Predictive Forecasting capabilities are far more accurate than reactive forecasting models. Automated analysis and shorter run times lead to having Forecasting models that are not only more accurate but also adaptable to rapid changes.

**Conclusion**

Forecasting models can gain a major boost when the data analytics behind them are augmented with Artificial Intelligence. A Machine learning based Forecasting model for Supply Management can reduce Overstock burdens for many retail clients while making sure Out-of-Stock scenarios are avoided at the same time by being malleable enough to market changes, demand & demography shifts, and consider local and global events that might affect a products ability to sell within a given time period. This results in an agile, scalable and dependable forecasting model, which can evolve faster based on the self-learning algorithms used.

Mindtree, has been working on the latest Machine Learning platforms and has a deep understanding of the vast benefits this revolution will bring and has the required value driven expertise in the AI space. Mindtree currently is servicing various businesses in employing this framework and is an enabler in their AI success stories.

**References**

[1]http://engage.dynamicaction.com/WS-2015-05-IHL-Retailers-Ghost-Economy-AR\_LP.html

[2] Samuel, Arthur (1959). [*"Some Studies in Machine Learning Using the Game of Checkers"*](https://doi.org/10.1147/rd.33.0210). IBM Journal of Research and Development.

[3] Google Deepmind’s AI beats human Go player - <https://deepmind.com/research/alphago/>