Role of Machine Learning in Inventory Optimization using Time-series Forecasting

Archit Bansal

June 27, 2021

ROLE OF MACHINE LEARNING IN INVENTORY OPTIMIZATION USING TIME-SERIES FORECASTING

Archit Bansal

Indian Institute of Technology Delhi
New Delhi, India

Abstract. A key concern for manufacturers today is to maintain optimum inventory levels to drive business growth with better prediction of future sales. With rapid advancements in analytics and machine learning (ML), companies can now proactively examine master and transactional data in near real-time and use the insights derived to plug the gaps and revenue losses. ML algorithms and the models they are based on, excel at finding anomalies, patterns, and predictive insights in large data sets. Predictive analytics can anticipate any spikes or dips in demand and suggest which items should be replenished when along with quantity and location/store. I have developed an ML model to forecast sales demand to help optimize inventory and save significant cost due to high or short inventory caused by inaccurate demands. The model also looks at the return orders data to optimize the returns, thereby resulting in customer satisfaction and cost reduction.

1. **Introduction.** The supply chain is becoming complex by the day with increased globalization, and enterprises, especially manufacturers and retailers, are facing challenges to manage inventory effectively. Traditional inventory management techniques are not adequate to assist manufacturers with inventory optimization in today's global supply chain ecosystem. Even advanced inventory management packages and tools are found lacking in delivering significant outcomes and benefits on reduced inventory costs.

In the retail industry, inventory demand planning and sales forecasting is the key to being competitive and having happy customers without excess spending on inventory. Inventory demand forecasting determines sufficient just-in-time replenishment, timely buying decisions, effective supply chain decisions, and inventory spend optimization. Overprediction or underprediction of demand costs in terms of out-of-stock percentages, inventory aging, and logistics costs.

The non-predictive inventory planning, which is the spreadsheet-based manual forecasting or tool/package based approach is not suitable for staying ahead of the competition and improving the bottom line. Manual adjustments of forecasts in the traditional methods leave all inventory spend decisions to human judgment, which pushes towards generalized decisions rather than going with product, vendor, and market-specific indicators. Issues arise when decisions are mostly based on historical sales numbers without considering dynamically evolving business factors.

Key words and phrases. Supply chain, inventory optimization, machine learning, sale forecasting prediction, return order.

Supply chain risks can be circumvented if demand and opportunities based planning are plugged in the inventory management techniques. Demand-driven supply-chain ecosystem can help reduce inventory and operational costs, and enhance customer service, thus ultimately boosting profits. Uncertainties trigger a series of business and operational issues. Inventory planning sees many uncertainties, the root causes of which differ on a day-to-day basis and are difficult to forecast using pre-defined rules. For instance, customer demand fulfillment takes more than 25 days instead of the scheduled 15 days in a dairy equipment delivery. Typically, the delays are triggered due to the unavailability of a finished product in a particular store. In contrast, certain other low demand items are lying in the store for more than three months, thereby locking the value.

ML allows real-time analytics and action points, which can play a vital role in understanding the challenges related to inventory planning and predicting them well in advance based on the sales forecast. ML uses algorithm models that can process large volumes of data very quickly, something that is not possible through manual methods. Here, I have used a time-series ARIMA model, which makes use of vast historical data to equip machine learning algorithms to visualize, draw insights, and predict future needs. This approach can drive efficient order execution, reduce revenue losses from delayed and return orders, and improve overall customer satisfaction.

2. **Building Machine Learning Models: A Use Case.** There are myriad use cases to demonstrate the impact ML approaches bring in inventory optimization for manufacturers. The application of ML algorithms has been able to predict better and faster, thereby improving supply chain effectiveness. Here, the data of a leading American full-service dairy solutions provider was used. They provide barn design and consultation, and manufacturing and installation of the equipment put in it. To build ML models to predict the future demands accurately, the following steps were used. The sequence of steps is shown in Figure 1.

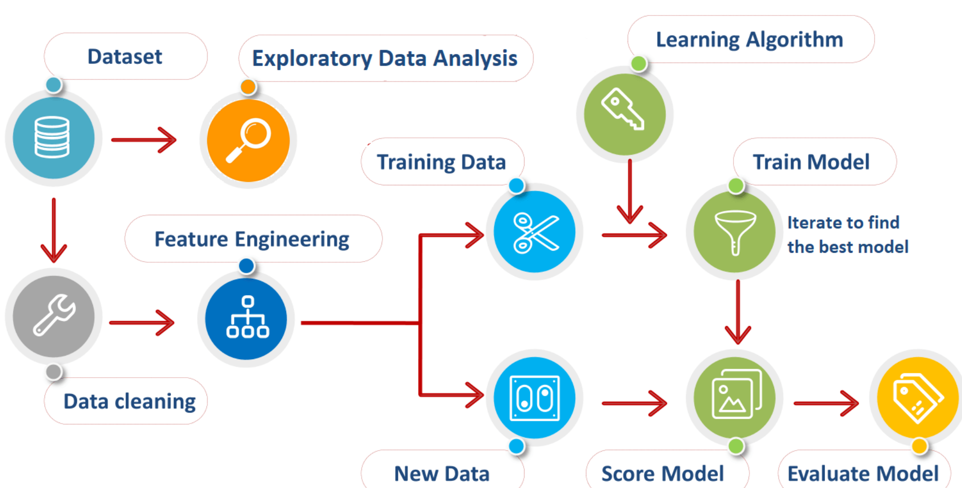

FIGURE 1. ML systems learn, and then infer results, from data

Item No	Transaction Date	Transaction Type	Store Code	Quantity	Customer Code	Country Code	Area Code	Unit of Measure Code	Item Category Code
3R03420	19-11-19 00:00	Sale	F06	1	0018291	USA	GA	EA	DAIRY
3R03435	19-11-19 00:00	Sale	A02	1	0018292	USA	MI	EA	DAIRY
3R05014	19-11-19 00:00	Sale	G07	15	0018293	CANADA	RP	EA	DAIRY
3R05025	19-11-19 00:00	Sale	F06	1	0018294	USA	VT	EA	DAIRY
3R05045	19-11-19 00:00	Sale	B01	20	0018295	USA	WA	EA	DAIRY
3R06120	20-11-19 00:00	Sale	A18	2	0018296	USA	SF	EA	DAIRY
3R01400	20-11-19 00:00	Sale	C11	1	0018297	USA	RP	EA	DAIRY
3R03030	20-11-19 00:00	Sale	C11	1	0018298	USA	MI	EA	DAIRY
3R03115	20-11-19 00:00	Return	F02	1	0018299	USA	GA	EA	DAIRY
3R03215	20-11-19 00:00	Sale	D15	1	0018300	USA	VT	EA	DAIRY
3R03265	20-11-19 00:00	Sale	A05	1	0018301	USA	WA	EA	DAIRY
1H00251	20-11-19 00:00	Sale	C09	60	0018302	USA	MI	EA	HARDWARE
1H00254	21-11-19 00:00	Sale	A18	60	0018303	USA	GA	EA	HARDWARE
1H07005	21-11-19 00:00	Sale	D12	15	0018304	USA	SF	EA	HARDWARE
1R00762	21-11-19 00:00	Sale	G07	360	0018305	CANADA	MI	LNFT	STEEL
1R01240	21-11-19 00:00	Sale	C02	122	0018306	USA	SF	EA	DAIRY
1R01260	22-11-19 00:00	Sale	C02	122	0018307	USA	VT	EA	DAIRY
1R01340	22-11-19 00:00	Sale	A02	200	0018308	USA	GA	EA	HARDWARE
3R03395	22-11-19 00:00	Sale	A18	120	0018309	USA	WA	FT	DAIRY
3K00143	22-11-19 00:00	Return	E13	23	0018310	USA	SF	EA	DAIRY
3F01625	22-11-19 00:00	Sale	B02	10	0018311	USA	WA	EA	DAIRY
1S30250	22-11-19 00:00	Sale	D07	200	0018312	USA	RP	LNFT	STEEL
3F00522	23-11-19 00:00	Sale	F16	18	0018313	USA	VT	EA	DAIRY
3F01181	23-11-19 00:00	Return	C11	48	0018314	USA	RP	EA	HARDWARE
3K00530	23-11-19 00:00	Return	A03	48	0018315	USA	RP	EA	HARDWARE

FIGURE 2. Snippet of the dataset (data changed due to confidentiality) to give a gist of the data requirements to build the model

2.1. **Data Collection.** Data for the past four years (Jan 2016 to Feb 2020) was collected and combined, and relevant information was sifted. Here are the key properties of the company:

- Total number of stores: 100+
- Total number of items: 9000+
- Planning cycle: monthly
- Average sales transactions: ∼2500 per month
- Average return transactions: ∼150 per month
- Average inter-store transfers: ∼1000 per month

Figure 2 shows the data set that was studied for this use case.

2.2. **Exploratory Data Analysis.** After analyzing the data, it was discovered that forecast information was diffused across multiple stores and regions, which added to data complexity. While some novel sales trends happened at one store, other kinds of spikes occurred at another store. Those novelties arise for various reasons and ask for modeling all the stores into a single model to capture the overall dynamics of the business. All items were divided into different groups based on their total and monthly balance of share of the sales vector. These were also reviewed for the number of returns. Figures 3-6 show some examples of the results of the data analysis.

2.3. **Data Cleaning and Feature Engineering.** Irrelevant and error-prone data was cleaned by filling-out/removing missing values, deleting duplicate data, removing columns with low variance, and handling outliers having a z-score greater than 3 or 4. Performed feature engineering by improving the existing features and brainstorming/creating new features.

2.4. **Model Training and Scoring.** The dataset was divided into training data (Jan 2016 to Feb 2019) and test data (Mar 2019 to Feb 2020). ML algorithms like linear regression, random forest, support vector machine, gradient boosting, and time-series were then applied to train the transactional dataset and help in forecasting, and their accuracy was compared. The statistical time-series method called ARIMA performed better for demand planning in most cases.

The method of differencing was applied to the time-series data to make it stationary. The sales of items sold by the manufacturer were forecasted for the coming months, and the R2 score was calculated. R2 was chosen as it gives scale-free results, which is advantageous in cases where items have varied sales trends. The ARIMA model worked quite well with an average R2 score between 0.7 and 0.8, and some items scoring as high as 0.9.

2.5. **Hyperparameter Tuning.** With the help of autocorrelation and partial autocorrelation plots, p (number of autoregressive terms), d (number of nonseasonal

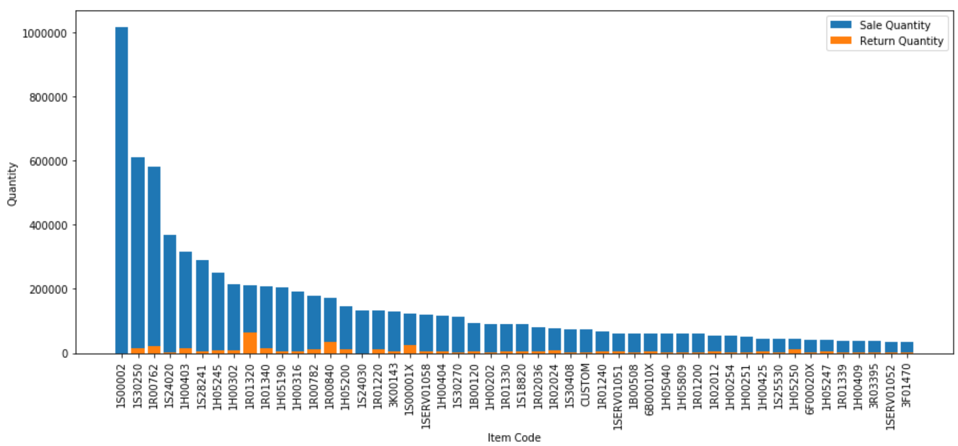

FIGURE 3. Top 50 selling items (in descending order), with return quantity for each item – to get a glimpse of the fast-moving items

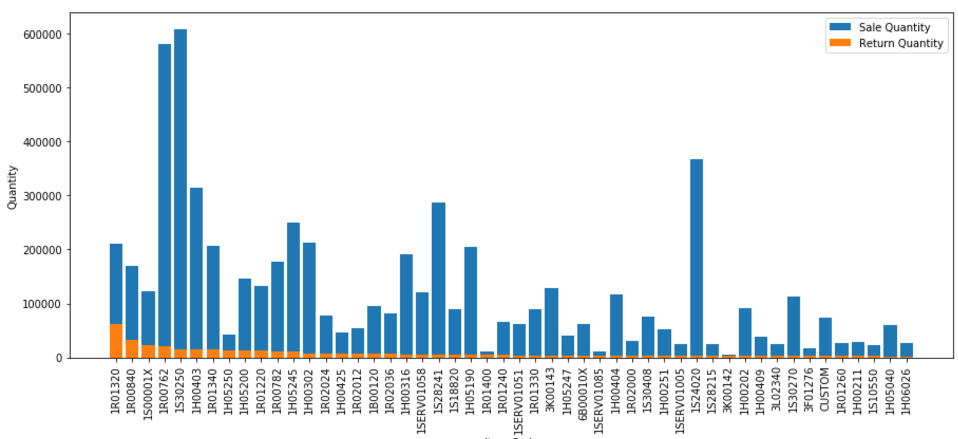

FIGURE 4. Top 50 returned items (in descending order), along with sales – to formulate strategies to minimize returns thereby saving costs

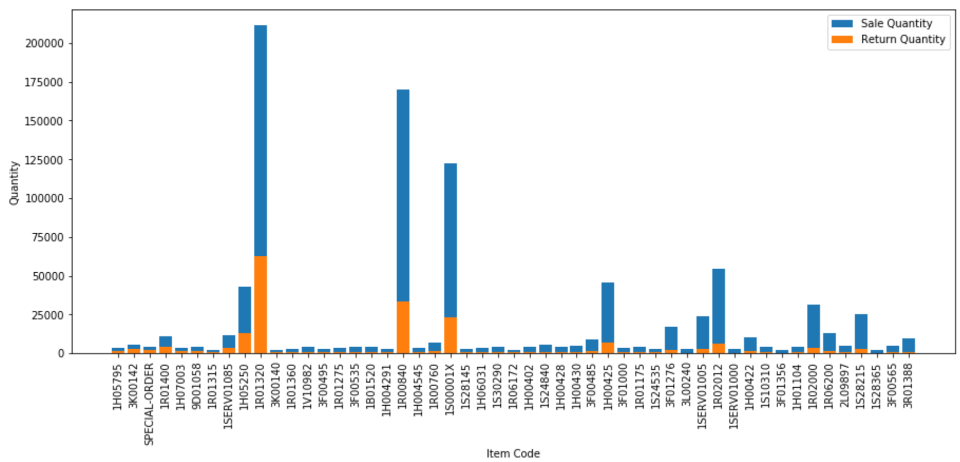

FIGURE 5. Items with a high return percentage (≥10% of sale) – to look for reasons to circumvent in future, thereby increasing customer satisfaction

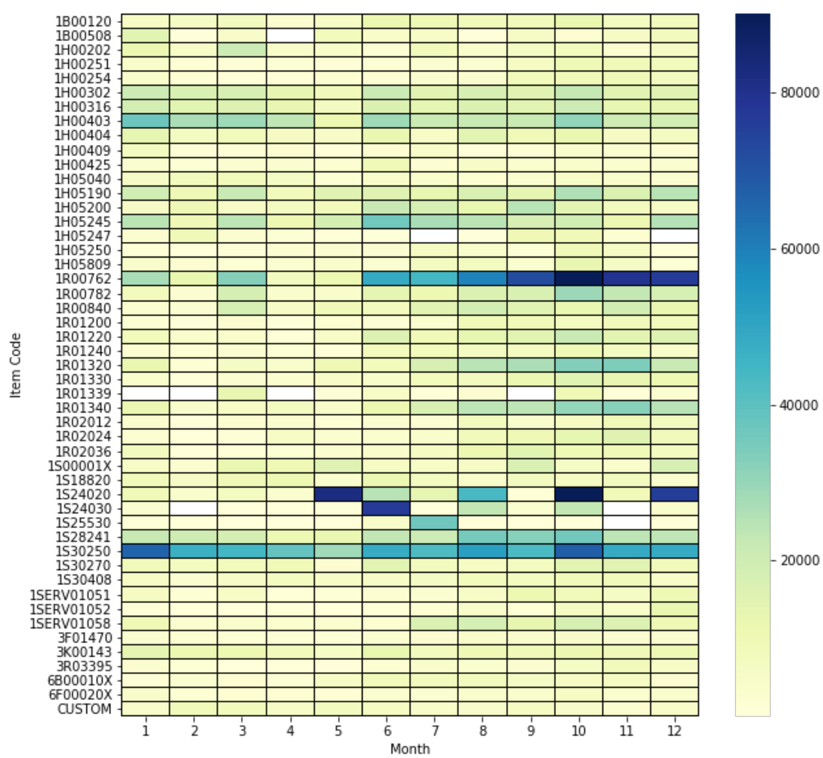

FIGURE 6. Monthly trend of top-selling items – to predict monthly sales thereby planning optimum inventory to execute orders efficiently

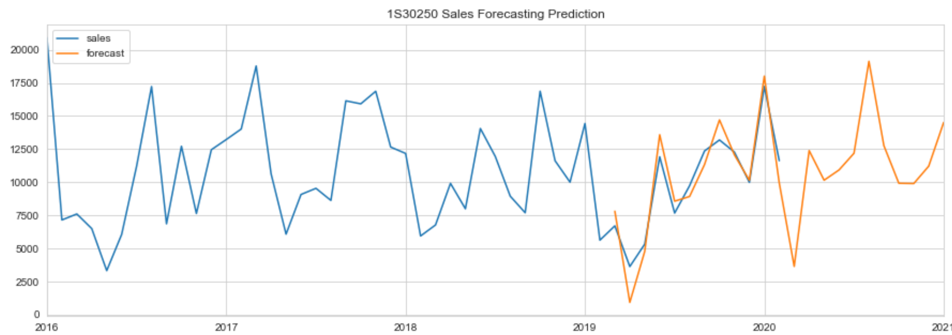

FIGURE 7. Sales forecasting prediction for item – 1

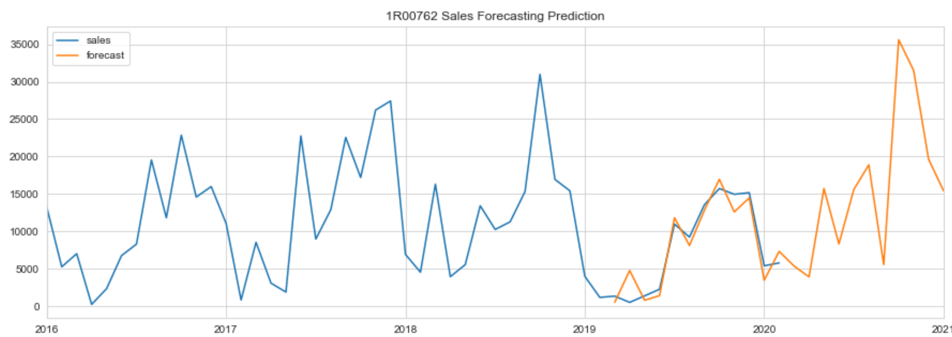

FIGURE 8. Sales forecasting prediction for item – 2

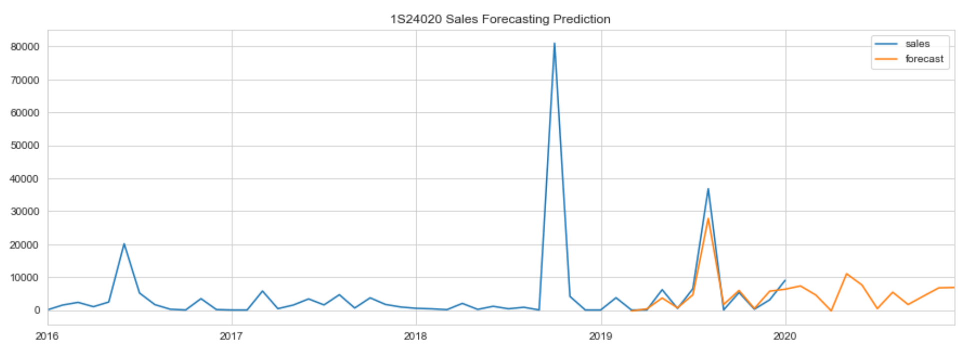

FIGURE 9. Sales forecasting prediction for item – 3

differences needed for stationarity), and q (number of lagged forecast errors in the prediction equation) were determined to fine-tune the model. Figures 7-9 show the forecasting predictions for the three most popular items.

3. **Conclusion.** Inventory and demand forecasting is a challenge because the sales and demand curves are highly unpredictable with unexpected spikes and other artifacts. The ARIMA model used is a very basic time-series method, yet we get a

good average R2 score for more than 60% of the items. Businesses can rely on innovative next-generation ML technologies to help optimize inventory, serve customer demands on time, save operational costs, and derive desired business outcomes, thus providing better customer experience in this digital age of ever-changing customer needs.

REFERENCES

[1] L. Aburto and R. Weber, Improved supply chain management based on hybrid demand forecasts, *Applied Soft Computing*, **7(1)** (2007), 136–144.

[2] B. Xing, W. Gao, K. Battle, T. Marwala and F. V. Nelwamondo, *Artificial Intelligence In Reverse Supply Chain Management: The State Of Art*. Available from: https://pdfs.semanticscholar.org/60fc/df407a55224a7d7e56ab443afc4b409990cd.pdf.

[3] T. Šustrová, *A Suitable Artificial Intelligence Model for Inventory Level Optimization*. Available from: https://dspace.vutbr.cz/handle/11012/63168.

[4] D. Faggella, *What Is Machine Learning? - An Informed Definition.*, TechEmergence, 2017. Available from: www.techemergence.com/what-is-machine-learning/.

[5] Wen-Jing Gao, et al, Computational Intelligence in Used Products Retrieval and Reproduction, *Research Methods*, 1188–1230.

[6] B. Lantz, Machine Learning with R, Birmingham, UK: Packt Publishing Ltd.

E-mail address: architbansal28@gmail.com