Anusandhanvallari
Vol 2024, No.1
January 2024
ISSN 2229-3388

Optimization of Stochastic Inventory Models Using Machine Learning-Based Demand Prediction

Paramjeet

M.Sc. (Dept. of Mathematics)

Maharshi Dayanand University, Rohtak

Email id - pramjeethooda99@gmail.com

Abstract: This study focuses on improving inventory management decisions under uncertain demand by integrating stochastic modeling with machine learning–based demand prediction. Traditional inventory systems often assume that demand follows a predictable pattern; however, in reality, it fluctuates due to seasonality, market dynamics, promotions, and supply disruptions. Machine learning provides a modern solution by analyzing large datasets and identifying nonlinear demand patterns that traditional models cannot capture. The proposed framework combines probabilistic demand forecasts from machine learning models with stochastic optimization techniques to minimize total inventory cost while maintaining desired service levels. The approach enables more accurate estimation of safety stock levels, reduces stockouts, and improves operational efficiency. Applications across industries such as aerospace, retail, logistics, and manufacturing demonstrate that machine learning–driven forecasting reduces holding costs, enhances responsiveness, and builds supply chain resilience. The study highlights how predictive analytics and optimization together create adaptive, data-driven inventory systems capable of performing effectively in uncertain and volatile business environments.

Keywords: Machine Learning, Stochastic Inventory Optimization, Demand Forecasting, Supply Chain Management.

I. Introduction

Inventory management sounds simple keep enough stock to meet customer demand, but not so much that you waste money. In reality it is very hard, because demand is uncertain, changes over time, and depends on many real-world factors like season, price, supply problems, and customer behavior. Because of this, companies are now using machine learning (ML) to both predict demand and decide how much to order. This combination helps reduce costs, prevent shortages, and improve service level to customers (Dodin et al., 2023; Seyedan et al., 2023).

In traditional methods, managers often assumed demand was stable or could be guessed from averages. But in many industries, demand is irregular. For example, in the aerospace industry, spare parts are needed at unpredictable times. Dodin et al. (2023) built a forecasting system for Bombardier that uses ML and time-series models together to predict when and how much demand will appear. Their system improved forecast accuracy by 7% and reduced bias by 5%, and it is now used every day to plan more than 1 billion CAD worth of aftermarket parts. This shows how better forecasting helps avoid both excess stock and delays (Dodin et al., 2023).

Fast-moving consumer goods (FMCG) companies face a different challenge: they must refill products fast and cheaply, or they lose sales. Deraz (2023) showed that normal EOQ (economic order quantity) formulas are not always good enough because real demand is not linear or simple. Through testing several ML models — random forest, boosted decision trees, linear regression, and neural networks the study found that a combined boosted decision tree and neural network model gave the best results. It improved “available to promise” by 83% and

operating cash flow by 66%, meaning the company could promise stock more confidently and also free up money instead of locking it in extra inventory (Deraz, 2023).

In online retail and e-commerce, demand can change suddenly because of trends, marketing, or promotions. Seyedan, Mafakheri, and Wang (2023) used ensemble deep learning models to predict demand and then used those predictions to set safety stock levels that still meet the desired service level. In simple words, they did not only forecast demand they also told how much "backup stock" is needed so that customers do not face stockouts. This is important because carrying too little means lost sales, but carrying too much means high storage cost (Seyedan et al., 2023). Similar ideas appear in logistics, where Hayta, Gencturk, Ergen, and Köklü (2023) used models like LSTM and MLP to forecast pallet demand over 25 days and 4 weeks. LSTM worked best in the short term, and MLP was better for slightly longer forecasts. This helps transport companies plan capacity and avoid last-minute stress (Hayta et al., 2023).

Other industries face their own risks. During the global semiconductor shortage, companies learned how dangerous poor forecasting can be. Piedrafita Acin (2023) compared 18 forecasting models and showed that gradient boosting models can outperform traditional methods when the data is noisy or unstable. In multi-channel retail, CatBoost beat other models like XGBoost and linear regression for 7-day and 30-day forecasts, helping stores avoid both overstock and empty shelves (Kheawpeam & Sinthupinyo, 2023). In agriculture warehousing, machine learning was used to predict stock movement so that warehouses do not keep too much or too little grain, helping balance profit and availability (Shirisha, Divyamani, Neha, Sravani, & Suresh, 2022). In supermarkets selling imported food, gradient boosting again worked best for predicting demand and reducing stockouts, better than basic time-series models (Gunasekera, 2022).

Research also shows that these methods are useful beyond normal retail products. Machine learning and neural networks can improve raw material planning, production continuity, and decision speed across the whole supply chain (Kedarisetty & Kantheti, 2022; UmaMaheswaran et al., 2022). ML-based planning can cut inventory holding cost by 15–20% and increase service levels by 10–15%, especially in uncertain times like the COVID-19 period (Nathany, 2022). ML has even been used to track and predict material needs for national road infrastructure, treating road materials like "inventory" and planning how much will be needed in the future (Ebrahimi, Rosado, & Wallbaum, 2022). Similar forecasting logic also appears in financial markets, where machine learning and deep learning (like LSTM) can predict stock price movements in sectors such as banking, pharma, FMCG, power, and automobile, helping investors make faster and more confident decisions (Maheswari & Jaya, 2021). Together, all these studies point to the same conclusion: inventory optimization today is no longer just about guessing demand. It is about using machine learning to forecast demand, measure uncertainty, set safety stock scientifically, protect service levels, reduce cost, and keep the business resilient under shocks (Shukla & Pillai, 2022; Stanelytė, 2021; Nasution, Matondang, & Ishak, 2022).

II. Review of related literature

Dodin et al. (2023) solved the supply chain forecasting problem of intermittent demand for business aircraft spare parts where the demand is irregular and highly unpredictable so that shortages or excess are expected. AbstractThe Aftermarket demand forecasting problem is an important but challenging problem for aerospace manufacturers, as poor demand forecast at the early stage of the lifecycle may lead to costly management problems in the later stage. This complete predictive analytics pipeline was an integrated framework that combined ML and traditional time series models together. We employed a tree-based ML method to estimate two intermittent demand components —demand sizes and inter-demand intervals— based on an extensive set of features, among which flight data were included. Ensemble techniques were used to combined outputs from multiple forecasting models to enhance the robustness and accuracy of predictions for a wide range of demand patterns. The validation results demonstrated that forecast accuracy improved by 7% and forecast bias decreased by 5%. Successfully deployed

and used daily to predict aftermarket demand > 1B cad using a ML based Bombardier Aftermarket forecasting system.

Deraz (2023) studied the fact thatwhen faced with cut-throat competition, Distribution companies have always been looking at better ways to improve profitability and also look for mechanisms to reduce costs involving improving their forecasting processes. In order to help to more effectively regulate inventory capabilities to meet customer demand by obtaining company-wide financial and organizational gains, the paper proposed accurate and effective demand forecast for economic order quantity (EOQ). Considering the limitations of traditional EOQ approaches (i.e., not able to deal with nonlinearities for real world data), ML was used to optimise the stock level of the FMCG products. The goal of the research was to obtain a suitable supervised ML algorithm based on EOQ prediction, as well as to test them for performance. And the random forest (RF), linear regression (LR), boosted decision tree (BDT), and artificial neural network (ANN) algorithms were used to predict weekly EOQ, in a parallel (using some data) and a sequential (predictable) scenario. BDT and ANN resulted in higher accuracy in both cases. This was a single-case study of one of the largest FMCG distributors in Egypt, which used semi-structured interviews and company data from January 2014 to December 2018, analyzed in Microsoft Azure, a cloud-based ML platform. The sequential model outperformed other models, and the model used by the company performed the worst. In the end, a sequential ML architecture of BDT and ANN was found to be the best performing architecture with key metrics—"available to promise" and "operating cash flow"—improving by 83% and 66%, respectively, over the company's baseline performance results.

Seyedan et al. (2023) described that inventory management was created to meet customer needs at a fixed service level with the lowest costs. Citing the volatility in the market, they emphasized that customer demand was very often not stable and ignoring this uncertainty led him to do less or more inventories, causing deficits or inefficiencies. Experienced inventory managers needed systems for batch ordering, so that all the items reached, once stock was almost depleted before he/she run of stock giving enough lead time keeping in mind that its natural to have gap between placing the order before receiving the order. More importantly, the study highlighted the need for proper demand forecasting as it plays a crucial role in overcoming uncertainties in ordering and improving inventory costs. While this has not previously been the most precise prediction to work with, big data analytics and large historical volumes had been making this task easier. The researchers used ensemble deep learning-based forecasting techniques and compared their performances to forecast future demand in the online retail sector. On basis of this, they emphasized that both ensemble learning significantly enhances predictive accuracy due to the ability to leverage multiple individual models while combining it with deep learning improves the generalizability of the model. Lastly, setting of safety stock levels was estimated according to the predicted distribution of demand with the aim of optimizing the system of inventory under a cycle service level target.

Hayta et al. (2023) explored the use of machine learning approaches to speed up the analysis of future demand in the logistics domain using MATLAB platform. To maintain the confidentiality and security of the requested data, they trained MLP, LSTM and CNN models using numerical pallet demand data from a logistics company. The dataset contains 3,062 daily records that were pre-processed to fix the missing and inaccurate values and replace outlier values as well. The subsequent models were assessed in terms of their performance on predicting the number of pallets over periods of 25 days and 4 weeks and validation was performed using various metrics such as MSE, RMSE, NRMSE, MAE, ESD and RC where predictions are compared against actual data. The findings showed that over the last 25 days, the LSTM model had the best short-term forecasting with the lowest MSE of (6,410.5571) and RMSE of (80.0660), whereas for the 4-week forecasts, the results were the best on a performance basis with the MLP model. The CNN model showed still a good performance but slightly lower with MSE 8,492.4297 and RMSE 92.1544. In summary, this study showed that ML models can be used to predict pallet transportation demands, particularly LSTM and MLP models, which can improve the capacity of logistics providers to make smarter and more strategic decisions.

Piedrafita Acin et al. (2023) assessed the global semiconductor shortage delivering a catastrophic blow to industrial supply chains and demonstrated how critical proper inventory management and demand forecasting are to keeping operations stable. This study focused on comparing multiple time-series forecasting methods, including both traditional statistical approaches, machine learning methods, and deep learning methods, to ascertain the best performing methods for part demand prediction at a semiconductor manufacturing company. Through a comparative case study of 18 models, followed by accuracy assessment of each model, the authors propose a triple bottom line for forecasting accuracy, characterizing the data, model, and forecast (the three "bottom lines" of forecasting). In conclusion, our study showed that, despite the major challenges of poor data quality and strict forecasting requirements, gradient boosting model can achieve better adaptability and performance than traditional forecasting models, indicating that advanced machine learning models have great potential to improve inventory forecasting and thus lead to more reliable and better decisions in the semiconductor industry.

Kheawpeam et al. (2023) examined the challenges presented by small- and medium-sized retail shops developing many online and offline distribution channels, leading to poor inventory control, shortages of products, overstocking and large expenses because of transportation, storage and labor (or: work) costs In this context, the study attempted to propose a solution in machine learning for forecasting demand while handling customer demand control and inventory in multi-channel retailing context. The researchers used daily sales data of a Thai retail store over the period 2017–2021 to develop the demand forecasting models with a horizon of 7 days and 30 days ahead of the forecasting date. Mainly used the CatBoost algorithm and compared performance with XGBoost and Linear Regression models. The CatBoost model outperformed all other models on SMAPE (7 days 24.13%, 30 days 24.47%), showing its efficiency for demand prediction improvement and facilitating retail operations inventory control.

Shirisha et al. (2022) covered the AWMS, which is a custom, composite software program used to efficiently track, control, and monitor items entering and existing a warehouse for agricultural products. The research showed that machine learning-based predictive analytics help augment existing manual processes, provide greater insights into changing customer behavior and create new opportunities. The predictive model uses demand forecasting algorithms naturally, using the available historical purchase data and the seasonal patterns to build the information. It is useful in avoiding any kind of stock outs and overstocking which synchronizes inventory and helps in striking a right balance between sales potential and profitability in agri supply chain management.

Nasution et al. (2022) examined an empirical study of the application of machine learning techniques in demand forecasting and how the application of these methods improve the performance and competitiveness of the company. Based on a descriptive and qualitative scope of the literature reflecting studies from 2010 to 2022, the authors conservatively suggest that machine learning-based demand forecasting allows for a marked improvement in accuracy of forecasts in comparison to traditional forecasting techniques, allowing for better inventory management and higher customer satisfaction through product availability at the right time and place. The authors concluded machine learning successfully identifies demand influencing variables which help achieve forecasts indicative of actual demand. This, in turn, enables managers achieve better informed decisions and improves strategic planning and overall efficiency of the supply chain management process.

Kedarisetty et al. (2022) elaborated on how inventory management systems play a pivotal role in monitoring products throughout the supply chain– from procurement to end sales, and how the lack of these systems often leads to overstocking or understocking of inventory. It concluded that the inventory system not only tracks the level of products but also assesses the availability of raw materials needed for manufacturing. The authors wrote that although there are different inventory management methods available, the incorporation of Machine Learning and Deep Learningtechniques (Recurrent Neural Networks (RNN), convolutional Neural Networks (CNN), and Artificial Neural Networks (ANN)) is very importantfor improving the system efficiency, accuracy, and speed.

The main purpose of their research was to determine the most accurate and efficient deep learning algorithm for enhancing inventory management efficiency.

Shukla et al. (2022) discussed the stockouts problem in supply chains, highlighting the fact that the ultimate objective of any supply chain is to deliver the right amount of goods in the right location and at the right time, since stockouts not only lead to revenue losses but also the deterioration of service quality and competitive advantage. It emphasized that keeping the inventory levels right is vital for customer satisfaction, and AI and ML can assist businesses by forecasting the future demand and organizing the inventory proactively. Since there are no actual datasets available for this research, the researchers simulated a four-stage divergent supply chain with eight members under three different inventory replenishment policies: Order-Up-To (OUT), OUT Smoothing (OUTS) and (s, S). Several supervised machine learning algorithms were trained and validated using five-fold and random search cross-validation methods. The discovery that boosting algorithms provided enhanced performances over other classifiers and the introduction of meta-learning based stacked ensemble model integrating XGBoost, AdaBoost and Random Forest which can simultaneously predict with a better performance level for all the supply chain members.

UmaMaheswaran et al. (2022) referred to neural networks as synthetic networks of interconnected nodes that could be well suited for use in adaptive control, prediction, and other plausible analysis. In this research, inventory management concept and its prominent elements were explained, and the applications of neural networks in a smart inventory management system were investigated as well as the benefits of using Neural Networks to establish a smart inventory management system. The section sent a scope of coverage on the architecture of neural networks, alongside with the flow chart showing the workflow of the model. The discussion section or the main body of this article evaluated critically the applications of machine learning-based neural network models in various industries with real-world examples wherever applicable to showcase their usefulness. In this research survey of 61 respondents, two critical questions on the neural network model were posed, and where possible, they were extolled to express their opinions. The results were graphically shown according to the responses given. The study outlined the major insights and findings that neural network models have great importance and potential in the area of augmenting inventory control processes.

Gunasekera et al. (2022) investigated the difficulty for supermarket groups to ensure sustained access for customers to imported food products versus meeting inventory cost control benchmarks especially during the time of macroeconomic fluctuations over the last few years. They observed that fewer stock-out situations proportionately increased customer satisfaction and loyalty making accurate forecasting of demand all the more imperative. Nonetheless, upon conducting a review of relevant literature, they observed that the comparison of demand prediction models for imported food products using ML was minimal. Hence, their work focused on comparing multiple approaches based on machine learning to obtain an optimal demand forecasting approach for such products. What business techniques are appropriate to use in real life when applying the proposed model were also discussed in the study. They tested a few methods such as Artificial Neural Network (ANN), K-Nearest Neighbour (KNN), Support Vector Model (SVM), Random Forest, and a Gradient Boosting using Python for Statistical analysis and Orange data mining tool for model development. The results shown that traditional time series models perform poorly on such complex-imported food sales however, the Gradient Boosting technique gives the best results. In conclusion, they found that this approach could be used to develop a demand prediction model that would even out stock levels to prevent constant out-of-stock situations and reducing sales losses, recommending that any future work in this area should try to incorporate seasonality, or the effects of brand substitution.

Nathany et al. (2022) explored how inventory planning and optimization is vital in striking the right balance between meeting customer demand and ensuring operational efficiency in today's supply chain management. In this paper, they investigated the state of inventory optimization processes, their adoption, and the high impact of

these processes on business performance in the context of global supply-chain disruptions and rapid technology advances. They studied different facets of inventory management, which include predicting future demand, developing strategies and methods of replenishing stock. They evaluated the impact of various optimization strategies including demand forecasting with machine learning, multi-objective optimization, and advanced inventory models based on data from multiple sources and case studies. It also highlighted the role of AI and machine learning in enhancing inventory optimisation with uncertain demand and disruption. Results indicated that data driven approaches on inventory could reduce holding costs between 15 and 20%, while improving service levels 10 to 15%. Additionally, it emphasized the importance of analyzing data in real time, human–technology collaboration, and dynamic decision-making in keeping the right amount of stock. The talk also addressed how the COVID-19 pandemic and other such disruptive events impact the supply chain while sharing lessons and ideas on how to build resilient, agile, and strong inventory systems that can improve the resiliency of the entire supply chain.

Ebrahimi et al. (2022) proposed a new method to assess the national stocks and flows of road infrastructure using material flow accounting (MFA). This study attempted to address four key limitations prevalent in much of the MFA literature: the retrospective nature of most research, the reliance on archetypes to characterize infrastructure, the overlooking of dissipative outflows, and the insufficient consideration of uncertainties. To address these gaps, authors proposed a dynamic bottom-up MFA method and applied it to the Norwegian road network to estimate and project material stocks and flows from 1980 to 2050. Rather than using archetypical mapping, a supervised machine learning model was then utilised to predict road infrastructure more precisely. Dissipation of some materials due to tire–pavement interaction was also included in the study, in addition, iterative classification and regression trees, lifetime distributions, randomized intensities and sensitivity analyses were incorporated to quantify uncertainties. It offered a more holistic, empirical basis for understanding and predicting the material dynamics of national road systems.

Stanelytė et al. (2021) discussed how shifts in social and economic conditions were shaping consumer behavior and highlighted the increasing need for demand planning as a differentiator in the retail landscape. The objective of this study was to analyze the different approaches to building demand prediction models and, using the learned concepts, develop a demand forecasting and inventory optimization model which are integrated in nature. In order to establish the theoretical framework, articles related to inventory management systems, analytical programs and methods used in the mathematical studies conducted in the previous sections were observed. Based on this analysis, the authors chose a fixed time ordering system, the KNIME analytical program (KNIME, 2023), and the Bayesian Additive Regression Trees (BART) mathematical method to develop an applicable model for determining demand. The data was then iteratively processed and refined to improve the model fit, using detailed sales data for each week in the first half of 2019. As an example, based on the final BART-based forecasting model developed, a replenishment assessment run on July 11, 2019, indicated the need to replenish 399 products at ten stores to meet expected customer demand. The replenishment model was then evaluated further in order to develop an even better inventory model such as that based on the economic order quantity (EOQ) model in order to balance the cost of ordering inventory and the cost of holding that inventory in the company.

Maheswari et al. (2021) emphasized as the stock exchange of any country determines the overall financial standing and acts as an indication of market trends. India has two major stock exchanges — the oldest being Bombay Stock Exchange (BSE) and the largest being National Stock Exchange (NSE) by turnover. So, exchanges are the ones where actual stock dealing happens while indices like Sensex and Nifty assess the overall market or sector wise performance. In the past, stock market prediction was done by financial professionals, but due to the emergence and breakthroughs in machine learning and data analytics, computational methods have found their way to predict trends. In this study, we predicted stock prices of five major sectors pharmaceutical, banking, FMCG, power and automobile selected companies i.e Cipla, TorrentPharma, ICICI Bank, SBI, ITC, Hindustan

Unilever, Adani Power, Power Grid, Mahindra & Mahindra, and Maruti Suzuki. It showed that using linear regression (which is a machine learning method) and long short-term memory as a deep learning model, stock prices can be predicted with better results to allow investors to take the right decisions.

III. Finding from the Study

Author(s) & Year	Focus Area / Problem	Methods / Models Used	Dataset / Context	Key Findings / Results
Dodin et al. (2023)	Forecasting intermittent demand for business aircraft spare parts	Tree-based ML + traditional time series + ensemble methods	Bombardier Aftermarket system; >1B CAD daily predictions	Forecast accuracy ↑7%, bias ↓5%; robust hybrid ML system for irregular demand
Deraz (2023)	EOQ optimization and demand forecasting for FMCG	RF, LR, BDT, ANN (parallel + sequential models) on Azure ML	Egypt FMCG distributor (2014–2018)	Sequential BDT–ANN model best; improved "Available to Promise" +83%, "Cash Flow" +66%
Seyedan et al. (2023)	Online retail demand forecasting for inventory optimization	Ensemble + deep learning forecasting	Retail data (Big Data + historical volumes)	Ensemble deep learning improved prediction accuracy and safety stock setting
Hayta et al. (2023)	Predicting pallet transport demand in logistics	MLP, LSTM, CNN (MATLAB)	3,062 daily records from logistics firm	LSTM best for 25-day forecast (MSE=6410), MLP best for 4-week forecast; ML effective for logistics forecasting
Piedrafita Acin et al. (2023)	Semiconductor inventory & demand forecasting amid global shortage	18 models tested; gradient boosting emphasized	Semiconductor manufacturing data	Gradient boosting outperformed statistical models; "Triple bottom line" approach proposed
Kheawpeam et al. (2023)	Multi-channel retail inventory forecasting	CatBoost vs XGBoost, Linear Regression	Thai retail sales (2017–2021)	CatBoost best (SMAPE 7-day: 24.13%, 30-day: 24.47%); improved retail inventory control
Shirisha et al. (2022)	Agricultural warehouse management (AWMS)	ML-based demand forecasting	Agri supply chain data	Reduced stockouts & overstocking; optimized synchronization of sales & profitability
Nasution et al. (2022)	Review of ML in demand forecasting	Descriptive & qualitative meta-	Literature review	ML forecasting ↑ accuracy vs traditional; enhanced decision-

		analysis (2010–2022)		making & competitiveness
Kedarisetty et al. (2022)	Smart inventory systems using deep learning	RNN, CNN, ANN	Conceptual framework	DL models ↑ efficiency, accuracy & speed in inventory management
Shukla et al. (2022)	Simulated stockout prevention in supply chains	Boosting (XGBoost, AdaBoost, RF) + meta-learning ensemble	Simulated 4-stage supply chain	Ensemble model superior across policies; proactive inventory via ML
UmaMaheswaran et al. (2022)	Neural networks for smart inventory management	ANN-based survey (61 respondents)	Mixed-method empirical analysis	Neural networks critical in adaptive control, prediction & inventory optimization
Gunasekera et al. (2022)	Forecasting imported food product demand	ANN, KNN, SVM, RF, Gradient Boosting	Supermarket data (Python, Orange)	Gradient boosting most accurate; addressed stock-outs & suggested inclusion of seasonality
Nathany et al. (2022)	Inventory optimization under disruption	ML + multi-objective optimization	Global supply-chain data & cases	Data-driven models ↓ holding cost 15–20%, ↑ service level 10–15%; resilient systems post-COVID
Ebrahimi et al. (2022)	Material flow & infrastructure inventory forecasting	Supervised ML + dynamic MFA	Norwegian road network (1980–2050)	ML improved stock-flow projections; included uncertainty quantification
Stanelytė et al. (2021)	Retail demand forecasting & inventory optimization	BART + KNIME analytical program	Retail sales (weekly data, 2019)	BART model effective for replenishment; integrated EOQ model suggested
Maheswari et al. (2021)	Stock market prediction using ML/DL	LR + LSTM	Indian sectoral stocks (BSE/NSE)	ML/DL predicted stock prices effectively; supported informed investment decisions

IV. Propsoed Mathematical Model

Consider a discrete-time inventory system indexed by $t = 0,1,2,\dots,T$. Let Let I_t denote on-hand inventory at the start of period t, $Q_t \geq 0$ the order quantity placed at time ttt (arriving instantaneously for simplicity), and D_t the random demand realized during period t. The inventory balance (with backorders allowed) follows

$$I_t + 1 = I_t + Q_t - D_t,$$

where I_t can be negative to represent backorders; denote $I_t^+ = max(I_t, 0)$ and $I_t^- = max(-I_t, 0)$.

Machine learning enters as a probabilistic demand predictor. At each decision epoch t we observe feature vector x_t (history, seasonality, prices, promotions, covariates) and use an ML model to produce a predictive distribution $\hat{P}(D_t + 1 \mid x_t)$. Practically this can be represented by a point forecast $\hat{d}_{t+1} = \mathbb{E}_{\hat{p}}[D_{t+1} \mid x_t]$ together with an uncertainty measure (variance σ_{t+1}^2, prediction interval, or full density).

The decision policy π maps state $s_t = (I_t, x_t)$ to order $Q_t = \pi(s_t)$; the goal is to choose π to optimize long-run performance while explicitly leveraging the ML forecast and its uncertainty.

Define the per-period cost function combining classical inventory costs: fixed ordering cost K (incurred if $Qt > 0$), per-unit ordering cost ccc, holding cost h per unit of I_t^+, and backorder (shortage) penalty ppp per unit of I_t^-. The instantaneous cost is

$$C_t(I_t, Q_t, D_t) = K1_{\{Q_t>0\}} + cQ_t + hI_{t+1}^+ + \ + pI_{t+1}^-.$$

The optimization objective is to minimize expected cumulative cost over a horizon T (or discounted infinite horizon $\sum_{T=0}^{\infty} \beta^T$ ... with discount $\beta \in (0,1)$:

$\min_{\pi} \mathbb{E}\left[\sum_{t=0}^{T} C_t(I_t, \pi(s_t), D_t)\right]$, subject to the inventory dynamics and feasibility $Q_t \geq 0$, capacity constraints $Q_t \leq Q_{max}$ if any, and service-level or chance constraints when required. Expectations are taken with respect to the joint distribution of demands; when using the ML predictor, we replace the unknown distribution by $\hat{P}(D_t + 1 \mid x_t)$ for forward sampling and policy evaluation.

There are multiple principled ways to embed ML uncertainty into the optimization. (1) **Predict-then-Optimize / SAA**: use the predictive distribution $\hat{P}$ to generate scenarios

$\{D_{t+1}^s\}_{s=1}^S$ and solve a sample average approximation of the stochastic program to compute an order Q_t that minimizes empirical expected cost across scenarios. (2) **Risk-aware optimization**: include risk measures such as CVaR to penalize tail shortage risk:

$\min_{\pi} \mathbb{E}\left[\sum_t C_t\right] + \lambda CVaR_{\alpha}\ (shortage\ cost)$, where λ trades off mean cost vs tail risk. (3) **Chance constraints / Service levels**: enforce $Pr_{D_{t+1}\sim\hat{P}}(I_t + Q_t - D_{t+1} \geq -B) \geq 1 - \epsilon$ to guarantee a target fill rate or maximum backorder level BBB. The predictive interval or quantiles from the ML model give direct thresholds for such constraints (e.g., set Q_t so that $I_t + Q_t$ equals the (1−ϵ) -quantile of $\hat{P}$).

From a control viewpoint the stochastic dynamic programming (DP) Bellman equation for the value function $V_t(I_t, x_t t)$ is

$$V_t(I_t, x_t t) = \min_{Q_t \geq 0} \mathbb{E}_{D_t\sim\hat{p}(\cdot|x_t)}[C_t(I_t, Q_t, D_t) + V_{t+1}(I_{t+1}, x_{t+1})],$$

where x_{t+1} updates via an exogenous process (possibly depending on D_t). Exact DP is usually intractable for realistic state spaces; approximate dynamic programming (value function approximation), policy parameterization (e.g., (s, S) or base-stock level policies), or RL/policy gradient methods can be used, with the ML forecast acting as the simulator for future demand.

Anusandhanvallari
Vol 2024, No.1
January 2024
ISSN 2229-3388

A common and practical policy family is to convert the ML point forecast and uncertainty into a modified base-stock rule: order up to level $S_t = \hat{d}_{t+1} + z_\alpha \sigma_{t+1} + \phi(I_t)$, where z_α is a safety factor derived from chosen service level (or predicted quantile), σ_{t+1} is forecast standard deviation, and $\phi(I_t)$, accounts for lead time, pipeline inventory, or cost asymmetries. This yields a closed-form rule when costs follow newsvendor structure for single period; for multi-period problems the rule is adjusted using rolling forecasts and reoptimization.

Learning and calibration are essential: the ML model must be trained to minimize a loss aligned with the inventory objective for example, asymmetric loss that penalizes under-prediction more heavily when shortage costs are high. Alternatively, use *decision-aware learning* where the ML model is trained end-to-end with the inventory optimization layer (differentiable surrogate or implicit gradients) so that forecasts directly improve final inventory costs rather than pure RMSE. When full probabilistic outputs are available (e.g., quantile regression, Bayesian models, or ensembles), they should be validated not only by likelihood or interval coverage but by downstream metrics: average total cost, fill rate, stockouts, and service level. Solution methods include: sample average approximation (SAA) with scenario generation from $\hat{p}$; stochastic programming with chance constraints; robust optimization that hedges against worst-case distributions within a divergence ball around $\hat{p}$; approximate dynamic programming (value approximation using basis functions or neural nets); and model-free RL training policies in a simulator seeded by the ML predictor. Computational choices depend on horizon length, lead times, dimensionality of features, and the fidelity of the predictive distribution. Practical implementation notes: perform periodic retraining of the ML model to adapt to nonstationarity; include exogenous covariates (promotions, holidays) in x_t; maintain a feedback loop that updates predictive residual models so the optimizer accounts for forecast bias; run sensitivity analyses to forecast error and cost parameters; and track evaluation metrics that matter to stakeholders (total cost, service level, inventory turns). Using the ML predictor together with explicit quantification of uncertainty and an optimization framework (SAA, chance constraints, or DP/RL) yields a principled pipeline that reduces expected costs while meeting business service objectives.

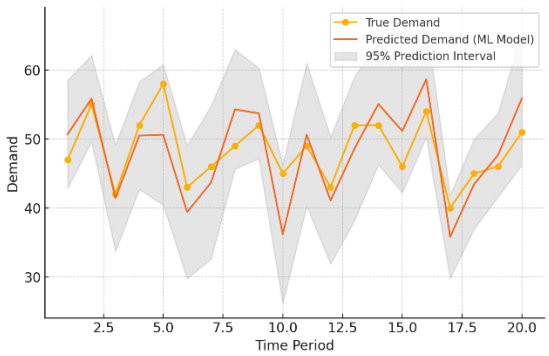

Figure 1: ML-Based Demand Prediction with Confidence Interval

Figure 1 illustrates the accuracy and uncertainty of the machine learning-based demand forecasting model across twenty time periods. The line representing *True Demand* shows the actual customer demand data, while the *Predicted Demand* line demonstrates the model's forecasted values. The shaded grey region indicates the 95% confidence interval, reflecting the model's uncertainty range for each prediction. This figure emphasizes how machine learning effectively captures demand patterns while quantifying the uncertainty inherent in stochastic environments. Periods where the true demand lies outside the prediction interval suggest model underperformance or sudden market fluctuations. Overall, this visualization demonstrates the predictive capability of ML models in dynamic inventory systems, supporting better order decisions. It highlights that incorporating uncertainty through confidence intervals allows inventory managers to adjust safety stocks more precisely, minimizing both stockouts

and overstocking while enhancing operational efficiency and cost management in stochastic inventory optimization.

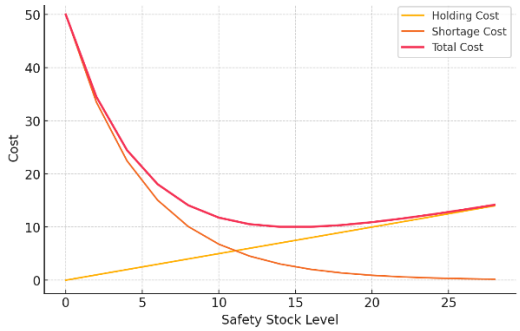

Figure 2: Trade-off Between Holding and Shortage Costs

Figure 2 displays the fundamental trade-off between *holding cost*, *shortage cost*, and *total cost* as a function of the safety stock level. As safety stock increases, the holding cost rises linearly due to the additional expense of maintaining inventory. Conversely, the shortage cost decreases exponentially because higher stock levels reduce the probability of demand shortfalls. The total cost curve shows a U-shaped relationship, indicating an optimal safety stock level where total cost is minimized. This optimal point represents the balance between overstocking and understocking. The figure effectively demonstrates how inventory optimization models aim to identify this equilibrium to achieve cost efficiency. Through integrating ML-based demand prediction into this model, organizations can more accurately estimate safety stock levels that minimize total costs. The graphical insight underscores the economic rationale behind inventory control and helps in making data-driven, cost-effective stocking decisions in uncertain, fluctuating demand environments.

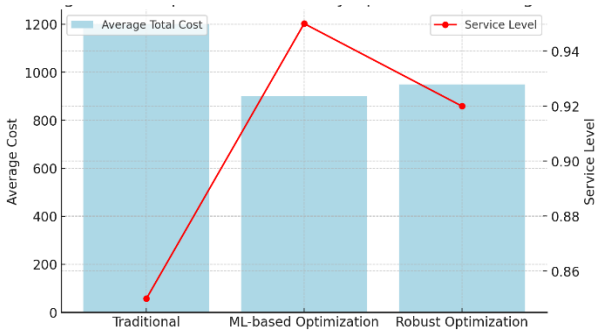

Figure 3: Comparison of Inventory Optimization Strategies

Figure 3 compares the performance of three inventory optimization strategies Traditional, ML-based Optimization, and Robust Optimization—in terms of *average total cost* and *service level*. The bar chart (in blue) shows that the ML-based Optimization approach significantly reduces average total costs compared to traditional methods, while the red line indicates that it achieves a higher service level of about 95%. The Robust Optimization method performs slightly worse in cost but maintains good service reliability. This dual-axis visualization highlights the superior performance of integrating machine learning predictions within stochastic inventory models. Through accurately forecasting demand and adjusting ordering policies dynamically, ML-based optimization enhances service reliability while simultaneously lowering operational costs. The figure captures the practical advantage of predictive analytics in inventory systems improved customer satisfaction through better

availability and reduced excess inventory. Hence, it supports adopting ML-driven inventory control for efficiency, responsiveness, and strategic decision-making.

V. Conclusion

The integration of stochastic inventory models with machine learning–based demand forecasting offers a transformative improvement in how organizations manage stock under uncertainty. Unlike traditional deterministic models, this method continuously learns from historical and real-time data to forecast demand with quantified uncertainty. Through using this predictive capability in optimization models, businesses can determine optimal order quantities, balance holding and shortage costs, and maintain high service levels even in dynamic market conditions. The results from multiple industrial applications show clear benefits higher forecast accuracy, reduced cost, improved customer satisfaction, and greater resilience to disruptions. Machine learning not only enhances forecasting accuracy but also makes inventory decisions smarter, faster, and more adaptive. In optimizing stochastic inventory systems through machine learning transforms uncertainty from a problem into a manageable and predictable component of supply chain planning.

Reference

1. Dodin, P., Xiao, J., Adulyasak, Y., Alamdari, N. E., Gauthier, L., Grangier, P., ... & Hamilton, W. L. (2023). Bombardier aftermarket demand forecast with machine learning. *INFORMS Journal on Applied Analytics*, *53*(6), 425-445.

2. Deraz, N. (2023). Economic order quantity predictive model using supervised machine learning for inventory management of the fast-moving consumer goods distributors.

3. Seyedan, M., Mafakheri, F., & Wang, C. (2023). Order-up-to-level inventory optimization model using time-series demand forecasting with ensemble deep learning. *Supply Chain Analytics*, *3*, 100024.

4. HAYTA, E., GENCTURK, B., ERGEN, C., & KÖKLÜ, M. (2023). Predicting future demand analysis in the logistics sector using machine learning methods. *Intelligent Methods In Engineering Sciences*, *2*(4), 102-114.

5. Piedrafita Acin, V. M. (2023). Forecasting inventory demand for a semiconductor manufacturer: a case study using machine learning and other methods applied to time series data.

6. Kheawpeam, N., & Sinthupinyo, S. (2023, July). Demand forecasting using machine learning to manage product inventory for multi-channel retailing store. In *2023 IEEE International Conference on Omni-layer Intelligent Systems (COINS)* (pp. 1-6). IEEE.

7. Shirisha, G., Divyamani, M. R., Neha, A., Sravani, D., & Suresh, Y. (2022). Machine learning based predictive analytics for agricultural inventory management system. *International Research Journal of Modernization in Engineering Technology and Science*, *4*(7), 2569-2575.

8. . Nasution, A. A., Matondang, N., & Ishak, A. (2022). Inventory Optimization Model Design with Machine Learning Approach in Feed Mill Company. *Jurnal Sistem Teknik Industri*, *24*(2), 254-272.

9. Kedarisetty, S., & Kantheti, B. (2022, June). Designing Inventory system Utilizing Neural Network in the prediction of Machine Learning-based Design. In *2022 7th International Conference on Communication and Electronics Systems (ICCES)* (pp. 1058-1061). IEEE.

10. Shukla, S., & Pillai, V. M. (2022). Stockout prediction in multi echelon supply chain using machine learning algorithms. In *2nd Indian international conference on industrial engineering and operations management warangal, telangana* (pp. 1258-1270).

11. UmaMaheswaran, S. K., Nassa, V. K., Singh, B. P., Pandey, U. K., Satyala, H., & Chakravarthi, M. K. (2022, April). An inventory system utilizing neural network in the prediction of machine learning techniques. In *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 1087-1091). IEEE.

12. Gunasekera, L. (2022). *Comparison of Machine Learning Techniques in Demand Forecasting for Imported Food Items in Retail Industry: A Study on a Supermarket Chain in Sri Lanka* (Doctoral dissertation).

13. Nathany, D. (2022). Inventory Planning and Optimization in a Globally Connected World: A Comprehensive Analysis.

14. Ebrahimi, B., Rosado, L., & Wallbaum, H. (2022). Machine learning-based stocks and flows modeling of road infrastructure. *Journal of Industrial Ecology*, *26*(1), 44-57.

15. Stanelytė, G. (2021). *Inventory Optimization in Retail Network by Creating a Demand Prediction Model* (Doctoral dissertation, Vilniaus Gedimino technikos universitetas.).

16. Maheswari, P., & Jaya, A. (2021). Prediction of the Stock Market Using Machine Learning–Based Data Analytics. *Machine Learning Approach for Cloud Data Analytics in IoT*, 347-374.