

OPPORTUNITIES AND LIMITATIONS OF MACHINE LEARNING ALGORITHMS FOR DEMAND FORECASTING IN INVENTORY MANAGEMENT

Dilfuzaxon Abdullayeva
New Uzbekistan University

Abstract: Accurate demand forecasting is essential for inventory control, yet increasingly volatile markets and intermittent product demand limit the reliability of conventional forecasting methods. This conference thesis critically examines statistical, ensemble machine learning, and deep learning approaches, focusing on their predictive capabilities, methodological weaknesses, and implications for inventory decisions. The analysis identifies data quality, information leakage, model overfitting, inappropriate evaluation metrics, and the gap between forecast accuracy and inventory costs as major challenges. A reproducible, time-aware comparison framework is proposed to support future empirical research. The contribution is methodological and analytical; no unperformed experiments or unsupported numerical improvements are reported.

Keywords: demand forecasting; inventory management; machine learning; intermittent demand; gradient boosting; LSTM; forecasting accuracy

INTRODUCTION

Demand forecasting is a fundamental component of inventory management because purchasing, replenishment, storage, and customer-service decisions depend on expectations about future product demand. An overestimated forecast can increase holding costs, working-capital requirements, and the risk of obsolescence, whereas an underestimated forecast may cause stockouts, lost sales, and lower service levels. These consequences are particularly significant when demand is unstable, product life cycles are short, and replenishment lead times are uncertain. Therefore, a forecasting model should not be judged solely by its ability to reproduce historical sales patterns; its usefulness must also be established in operational decision-making.

Artificial intelligence and machine learning (ML) are increasingly examined as alternatives to classical forecasting methods. Their main potential advantage is the ability to model nonlinear relationships between demand and factors such as prices, promotions, calendar effects, and previous sales. However, the frequent assumption that a more complex algorithm necessarily produces a more accurate or economically useful forecast is not supported in every context. This thesis aims to critically assess the opportunities and limitations of major forecasting approaches and to formulate a transparent framework for their comparison in inventory management.

COMPARATIVE AND CRITICAL ANALYSIS

Classical approaches, including seasonal naïve forecasts, moving averages, and exponential smoothing, remain important benchmarks. They are relatively interpretable, computationally inexpensive, and often appropriate when only a limited history is available. Nevertheless, their restricted flexibility can be problematic in markets where promotions, abrupt structural changes, and interactions among explanatory variables strongly influence demand. Intermittent demand creates an additional challenge: many periods contain no observed sales, while nonzero orders may vary considerably in size. Croston [1] addressed this problem by estimating demand size and the interval between positive-demand occurrences separately. Subsequent research demonstrated that the original procedure may produce biased estimates in some settings [2]. Thus, even specialized statistical methods cannot be regarded as universally reliable.

Ensemble ML algorithms, particularly Random Forest and gradient boosting, can integrate lagged demand, rolling statistics, product characteristics, pricing information, and calendar variables. Their adaptability is valuable for structured retail data containing diverse and interacting predictors. The M5 forecasting competition illustrated that advanced machine-learning-based submissions can perform strongly in a large-scale hierarchical retail setting [3]. However, the results of a particular competition should not be generalized automatically to smaller businesses, poorly documented inventory systems, or products with very sparse observations. Tree-based methods do not understand temporal order without carefully engineered predictors, and their performance may deteriorate when future conditions differ from the training period.

Long short-term memory (LSTM) networks and related deep learning models are designed to represent sequential dependencies. Despite this theoretical advantage, they may require substantial training data, computational resources, and extensive parameter tuning. For short or highly intermittent product series, a complex network can learn noise rather than a transferable demand pattern. Moreover, predictions concentrated near zero may reduce an average error measure while failing to detect the rare demand peaks that are most important for preventing stockouts. Limited interpretability further complicates the justification of replenishment decisions. Therefore, deep learning should be treated as a candidate approach, not as the default solution.

Another limitation concerns the distinction between recorded sales and actual demand. If a product is unavailable, zero sales may reflect a stockout rather than the absence of customer interest. Treating these observations as genuine zero demand can introduce systematic downward bias. Missing transactions, changing product codes, inconsistent promotional records, and delayed data updates can create similar distortions. In addition, information leakage occurs when features contain values that were unavailable at the forecast origin. This can make a model appear successful during validation but unreliable after deployment. These data-related weaknesses may have a greater practical impact than the choice between two sophisticated algorithms.

EVALUATION AND INVENTORY IMPLICATIONS

A defensible evaluation procedure must preserve chronological order. Product histories should be divided into training, validation, and untouched test periods, while rolling-origin evaluation should be used to reproduce repeated forecasting decisions. Feature construction, scaling, and hyperparameter selection must use only information available at each origin. Comparable forecasting horizons and identical test periods are essential for fair comparisons. Tashman [4] emphasized that out-of-sample evaluation design materially affects the reliability and reproducibility of forecasting studies.

Forecast error should be reported using more than one metric. Mean absolute error (MAE) is readily interpretable, while root mean squared error (RMSE) penalizes large errors more heavily. Mean absolute percentage error (MAPE) is undefined when actual demand equals zero and can be unstable for small values; this is a serious limitation for intermittent-demand series. Mean absolute scaled error (MASE) offers a useful alternative for comparisons across products, provided its scaling denominator is valid [5]. Where inventory decisions rely on uncertainty, interval or quantile forecasts should also be examined, not only point predictions.

Most importantly, superior predictive accuracy does not necessarily imply lower inventory costs. Forecasting errors interact with replenishment lead time, ordering policy, shortage penalties, and storage costs. A model that performs marginally better on MAE might still produce excessive safety stocks or frequent shortages. Research on feature-based combinations for intermittent demand considers both forecasting performance and inventory consequences [6]. Future

comparisons should therefore connect each forecasting model to the same replenishment policy and evaluate stockout frequency, fill rate, average inventory, and total relevant cost.

For a reproducible empirical extension, products can first be grouped by demand intermittency, variability, and seasonality. Seasonal naïve and Croston-type forecasts can then be compared with Random Forest, gradient boosting, and LSTM under identical rolling-origin splits. After measuring predictive errors and uncertainty calibration, forecasts should be translated into reorder points and safety-stock decisions. A hybrid or segmented approach may be justified when different model families perform better for different product groups, but this hypothesis must be tested rather than assumed.

CONCLUSION

Machine learning offers meaningful opportunities for demand forecasting, particularly when reliable explanatory variables and sufficient historical observations are available. However, its practical use is constrained by sparse and biased sales records, methodological leakage, overfitting, limited interpretability, and computational costs. The principal conclusion of this thesis is that algorithmic complexity is not an adequate substitute for rigorous validation and inventory-oriented assessment. Future work should examine whether segmented or hybrid forecasting methods improve both out-of-sample performance and replenishment outcomes under uncertain demand. The present study is a critical literature-based analysis and methodological proposal; it does not claim experimental superiority for any algorithm.

References

1. Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, 23(3), 289–303. <https://doi.org/10.1057/jors.1972.50>
2. Syntetos, A. A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting*, 21(2), 303–314. <https://doi.org/10.1016/j.ijforecast.2004.10.001>
3. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4), 1346–1364. <https://doi.org/10.1016/j.ijforecast.2021.11.013>
4. Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting*, 16(4), 437–450. [https://doi.org/10.1016/S0169-2070\(00\)00065-0](https://doi.org/10.1016/S0169-2070(00)00065-0)
5. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
6. Li, L., Kang, Y., Petropoulos, F., & Li, F. (2023). Feature-based intermittent demand forecast combinations: Accuracy and inventory implications. *International Journal of Production Research*, 61(22), 7557–7572. <https://doi.org/10.1080/00207543.2022.2153941>