International Journal of Economic, Business, Accounting, Agriculture Management and Sharia Administration (IJEBAS)

International Journal of Economic, Business, Accounting, Agriculture Management and Sharia Administration (IJEBAS Journal) | ISSN (e): 2808-4713
Volumes 6 No. 3 (2026)

DEMAND PATTERN-SPECIFIC FORECASTING FOR SPARE-PART INVENTORY: ENSEMBLE MODEL EVIDENCE FROM INDONESIAN HEAVY EQUIPMENT DISTRIBUTION

Like Ati Handayani[1]*, Dermawan Wibisono[2]
[1,2]Institut Teknologi Bandung, Bandung

Received : 15 April 2026 Accepted : 11 May 2026
Revised : 25 April 2026 Published : 26 May 2026

Abstract

PT Cakra Harmoni Sentosa (HCS), the largest heavy equipment distributor in Indonesia, faces a supply chain performance problem at its highest-revenue plant, Plant Sungai Danau (SDU). With IDR 1.1 trillion in annual spare-part transactions, SDU records a customer service level of 75%, which is five percentage points below the 80% target, while Days of Inventory reaches 87 days against the 75-day benchmark. Both shortfalls originate from reliance on a 12-month Moving Average forecasting method applied uniformly across a portfolio where 90% of SKUs show intermittent or lumpy demand patterns. This paper evaluates 23 forecasting model variants per demand pattern through walk-forward validation. Methods tested range from classical statistical approaches including Croston, SBA, and TSB to machine learning models such as XGBoost and LightGBM, deep learning architectures including LSTM and N-Beats, and ensemble combinations. Performance is assessed using RMSE and MAE on non-zero periods alongside the Stock-Keeping-Oriented Prediction Error Cost metric with stockout-weighted parameters. Findings show that a demand-pattern-specific Ensemble of SES, SBA, LSTM, and N-Beats achieves the strongest performance for lumpy and intermittent SKUs, cutting RMSE by 43.9% and SPEC by 36.3% on lumpy items relative to the MA12 baseline.

Keywords: demand forecasting, ensemble model, intermittent demand, inventory optimization, spare part, SPEC.

INTRODUCTION

PT Cakra Harmoni Sentosa (HCS) is Indonesia's largest distributor of heavy equipment, with operations spreading the mining, construction, and agro-forestry sectors. Three revenue streams of the business came from: unit sales, spare parts, and after-sales service. Of these, spare parts carry the greatest operational sensitivity. In mining operations, a single excavator failure affecting ten dump trucks within one fleet has been estimated to cost USD 49,850 in lost production per incident. When customers run hundreds of such fleets, the financial exposure from parts unavailability is substantial, and purchasing decisions for spare-part stock become critical rather than routine.

Coal price movements add a further dimension to this problem. Prices reached record levels in 2022, driven by the energy supply shock that followed the Russia-Ukraine conflict, which pushed global buyers toward coal as an alternative fuel source. From 2023 the market found a lower equilibrium, and by August 2025 the Newcastle benchmark had dropped to USD 114.50 per tonne. When coal revenue per tonne compresses, mine operators tend to defer new equipment purchases and run existing machines longer. This extends maintenance cycles and shifts the commercial weight of HCS's business toward spare-part fulfilment rather than unit sales, which raises the cost of getting the spare-part forecast wrong.

Plant Sungai Danau (SDU) sits at the center of this challenge. It is the single highest-revenue plant in HCS's 53-plant network, recording IDR 1.1 trillion in annual spare-part transactions, roughly 10% of the company's total. Yet its customer service level (CSL) runs at 75%, five points below the 80% target, and its Days of Inventory (DOI) sits at 87 days against a 75-day benchmark. Holding more stock than planned while still missing service targets is the classic symptom of misallocated inventory: stock accumulates on items that are not being ordered while the items customers actually request run dry. The root cause is the forecasting method. SDU currently uses a 12-month Moving Average (MA12), which averages the 12 most recent months of demand with equal weight for every month. For items where demand is sporadic or trending, this produces forecasts that lag reality in one direction or another. Seventy percent of SKUs at SDU show intermittent demand patterns and 20% show lumpy patterns; together they account for nine in ten items in the portfolio. For these items, MA12 is the wrong tool by design. This paper asks whether a method matched to each demand pattern can close both the service level and the inventory gaps at the same time, and tests that question empirically across 23 forecasting variants and all four demand categories found in the SDU portfolio.

LITERATURE REVIEW

The Role of Forecasting in Inventory Management

In supply chain planning, the forecast is what connects anticipated customer demand to actual stock decisions. Waters (2003) frames it as a buffer: better forecasts let a company carry less safety stock for the same probability of meeting demand, because less uncertainty needs to be covered by buffer inventory. The quality of the forecast therefore shapes the trade-off between the cost of holding stock and the risk of running out of it. When demand estimates are off, replenishment orders are wrong in one direction or another, and the inventory system drifts away from its intended balance.

Spare-part demand in the heavy equipment sector sharpens this challenge in two specific ways. Most SKUs record zero sales in the majority of months, so the demand signal is thin and irregular. At the same time, external drivers such as commodity prices and equipment operating hours affect consumption in ways that no method based purely on historical sales data can pick up. Both of these features push beyond what conventional moving average approaches can handle and call for methods that treat each demand pattern on its own terms.

Demand Pattern Classification

Syntetos, Boylan, and Croston (2005) put forward the framework that now serves as the standard starting point for spare-part demand analysis. They proposed sorting items on two axes: Average Demand Interval (ADI), which captures how often demand actually occurs, and the squared Coefficient of Variation (CV2), which captures how much the demand quantity varies from one occurrence to the next. Setting cut-offs at ADI = 1.37 and CV2 = 0.49 yields four groups. Items with low ADI and low CV2 are smooth, meaning demand comes in regularly and in fairly stable amounts. High ADI with low CV2 gives intermittent demand, where orders are infrequent but roughly consistent in size when they do arrive. Low ADI with high CV2 produces erratic demand, frequent but unpredictably sized. High ADI with high CV2 is lumpy demand, the hardest case to manage: orders arrive rarely and in quantities that jump around considerably. The practical value of this taxonomy is that it ties the choice of forecasting method to the statistical properties of each item rather than leaving that choice to intuition. At Plant SDU, the classification places 70% of active SKUs in the intermittent group and 20% in the lumpy group. These two categories together cover 90% of the portfolio, which immediately flags MA12 as a poor fit for the great majority of items the plant stocks.

Forecasting Methods for Intermittent and Lumpy Demand

Several statistical methods have been developed specifically for the intermittent demand problem. Croston's method (1972) separates the forecast into two components: one for how often demand will occur and another for how large it will be when it does, updating each component independently. The Syntetos-Boylan Approximation (SBA) later showed that Croston's original estimator carries a systematic upward bias and offered a correction factor that produces better-calibrated estimates. The Teunter-Syntetos-Babai (TSB) method built on this by allowing the demand probability component to decay over time, which helps when an item has effectively been discontinued or temporarily inactive. The ADIDA approach takes a different path: it aggregates demand to a coarser time period before forecasting and then disaggregates the result, reducing the proportion of zero-demand periods in the training data (Nikolopoulos et al., 2011).

Machine learning algorithms such as Random Forest, XGBoost, LightGBM, and CatBoost have been tested on spare-part data with uneven results. Their performance depends on having enough non-zero observations to train on; for items where demand occurs only a fraction of the time, the training signal is too thin for tree-based methods to find reliable patterns, and they frequently underperform even simple statistical baselines (Katyare et al., 2024). Deep learning models, particularly Long Short-Term Memory (LSTM) networks, work differently. They learn across sequences and across many series at once, which means they can pick up the conditional structure of sporadic demand patterns that would be invisible to a single-series model. Kourentzes (2013) found that neural networks outperformed Croston-family methods on intermittent series in part because of this cross-series learning. Neural Basis Expansion Analysis for Time Series (N-Beats) extends this further, decomposing the demand signal into trend and periodic components without needing stationarity to hold.

Combining methods in ensembles has consistently outperformed any single approach in benchmark comparisons. Khashei and Bijari (2011) demonstrated that a hybrid of ARIMA and neural networks beat either method used alone. The logic is straightforward: different methods fail in different circumstances, and averaging their outputs reduces the risk that any one failure mode drives the forecast badly off track. This study tests an ensemble specifically constructed for the intermittent and lumpy context, combining SES, SBA, LSTM, and N-Beats with weights optimised separately for each demand pattern.

Evaluation Metrics for Intermittent Demand

MAPE cannot be computed when actual demand is zero, which is the dominant condition for intermittent items, so it is not a usable metric for this context. Prestwich et al. (2014) addressed this by proposing modified versions of standard accuracy measures that compare forecasts to the mean of the underlying demand process rather than to the observed zero, giving stable rankings even when most observations are null. Kiefer et al. (2021) took a different angle with the Stock-Keeping-Oriented Prediction Error Cost (SPEC) metric. Rather than measuring how close the forecast is to the actual value, SPEC measures the cumulative inventory cost that results from the forecast being wrong, with separate weights for the cost of under-forecasting (stockout exposure) and over-forecasting (holding cost). In this study, a1 = 0.75 is applied to stockout errors and a2 = 0.25 to overstock errors, which reflects the business reality at HCS: a service level shortfall costs more in lost revenue and customer attrition than an equivalent amount of excess stock costs in holding charges.

Customer Service Level and Inventory Policy

Customer service level at HCS is measured as the share of order lines filled from stock held in the serving warehouse at the time the order arrives. Corsten and Gruen (2004) found in a large retail study that stockouts caused 31% of customers to switch to a competing supplier and another 26% to buy a substitute product. Those figures likely understate the switching cost in a B2B spare-part context, where equipment downtime adds urgency and the switching decision may affect the entire service contract, not just one transaction. Christopher (2016) frames inventory policy around a two-by-two matrix of volume and profit contribution. Items that are both high-volume and high-margin deserve the most investment in availability. Items that are low-volume and low-margin are candidates for rationalisation or very lean stocking. Applying this lens to the SDU portfolio, SKUs with consistently high zero-demand frequencies are natural candidates for centralised or on-demand sourcing rather than local safety stock, which is where the demand classification analysis connects directly to stocking policy decisions.

METHOD

Research Design

The study takes a quantitative and comparative form. The performance gap at Plant SDU is expressed in numbers that can be directly measured and compared across forecasting methods, which makes an empirical model-evaluation design the natural choice. Four sequential phases structure the work: classifying each SKU by demand pattern, building the feature set and preparing the data, running the model comparisons through walk-forward validation, and computing the error metrics that rank each method.

Data Collection

Six data sources were brought together into one analytical table covering January 2021 through July 2025, giving 55 consecutive months of history for Plant SDU. The sources are: spare-part demand transactions from the ERP system, recorded per SKU per month across more than 700,000 active items; heavy equipment active population data logged by on-unit hour-meter sensors, with a unit counted as active when its average monthly operating hours exceed four; coal price data scraped monthly from TradingEconomics; weather variables including rainfall, soil temperature, and evaporation rate drawn from the NASA MERRA-2 reanalysis database; public holiday flags covering Ramadan and Eid periods; and per-SKU stock position and lead-time statistics extracted from the ERP warehouse module. The combined dataset reaches 231 features per SKU-month record.

Demand Classification

ADI was calculated as the mean number of months between non-zero demand events for each SKU, and CV2 was calculated from the coefficient of variation of non-zero demand quantities. The Syntetos-Boylan-Croston (2005) thresholds of ADI = 1.37 and CV2 = 0.49 were then applied to assign each SKU to one of the four demand categories. At Plant SDU this produced: intermittent 70%, lumpy 20%, erratic 2%, and smooth 8%.

Forecasting Models Evaluated

Twenty-three model variants were run against each demand pattern. Statistical methods in the test included MA12, SES, Holt-Winters, ARIMA with automatic order selection via AIC minimisation, Croston, SBA, TSB, and ADIDA. Machine learning models included Random Forest, XGBoost, LightGBM, CatBoost, and three MLP variants each tuned against a different objective function (RMSE, MAPE, and SPEC). Deep learning models included LSTM without exogenous inputs, LSTM with 17 external features, and N-Beats. A hybrid ARIMA-MLP residual model was also tested. A Mix Model that picks the best single method per SKU from in-sample error rounded out the individual approaches. Finally, two ensemble configurations were evaluated, the main one being SES + SBA + LSTM + N-Beats with weights optimized per demand pattern.

Walk-Forward Validation

All models were assessed through walk-forward validation on a held-out test set. Each monthly forecast was produced using only data that would have been available up to that point in time, so no future information was allowed to influence any forecast in the evaluation window. MA12 served as the baseline throughout. All percentage-difference figures compare each model's metric value against this baseline: a negative number means the alternative method improved on MA12.

Error Measurement Metrics

Four metrics were computed for each model within each demand pattern. RMSE and MAE were restricted to non-zero demand periods to prevent the statistics from being dominated by the many zero months. SPEC was computed following Kiefer et al. (2021) with a1 = 0.75 for stockout errors and a2 = 0.25 for overstock errors. The formula accumulates demand-supply mismatches across periods rather than averaging point errors, which captures the inventory cost consequence of a wrong forecast in a way that single-period metrics cannot. A demand-weighted accuracy measure was also included to assess how often each method forecasts in the correct direction. Final model rankings were based on average rank across all four metrics. The SPEC formula is expressed as: SPEC(a1,a2) = (1/n) multiplied by the double summation of the maximum of zero and the minimum of the stockout cost and overstock cost terms, each scaled by the time distance (t - i + 1), where n is the series length, y(t) is actual demand at period t, and f(t) is the forecast for that period.

RESULTS AND DISCUSSION

Demand Pattern Composition and Data Characteristics

The demand classification produced the distribution shown in Table 1. Intermittent and lumpy SKUs together account for 90% of the portfolio, which confirms that MA12 is the wrong method for the bulk of what Plant SDU stocks. Demand quantity distributions are right-skewed across all patterns, with the median well below the mean, which is the standard Pareto shape for spare-part inventories. For intermittent and lumpy SKUs, zero-demand months make up between 40% and 75% of all observations. This level of sparsity makes MAPE unusable and pushes the analysis toward SPEC as the primary yardstick.

Table 1. Demand Pattern Composition at Plant SDU

Pattern	ADI	CV2	Share (%)	Characteristic
Intermittent	> 1.37	<= 0.49	70%	Slow-moving, stable quantity when demand occurs
Lumpy	> 1.37	> 0.49	20%	Slow-moving, high quantity variation when demand occurs
Erratic	<= 1.37	> 0.49	2%	Fast-moving, high quantity variation
Smooth	<= 1.37	<= 0.49	8%	Fast-moving, stable quantity

Model Evaluation: Smooth Demand Pattern

Smooth items make up 8% of SKUs and represent the one category where MA12 is reasonably well-suited. These are fast-moving items with stable demand quantities, and the 12-month average does a reasonable job of tracking them. The Ensemble posts the lowest RMSE at 24.605, a 6.7% drop from MA12's 26.376, and trims SPEC by 1.5%. Croston and TSB both rank slightly ahead of MA12 on the combined four-metric average, but the gap is 0.25 rank points. Given that size of difference, replacing MA12 for smooth items would add operational complexity for negligible gain. Tree-based machine learning methods failed badly here, with RMSE running 156% to 161%

above the MA12 baseline. Smooth demand simply does not produce the high-variance, dense-observation conditions that tree-based models need in order to find meaningful splits.

Model Evaluation: Lumpy Demand Pattern

Lumpy items, 20% of SKUs, expose the most serious failure mode of MA12. Consider an item that records zero demand in ten of the past twelve months and then 150 units across the remaining two. MA12 returns a forecast of 25 units every month going forward, a figure that is almost never right. In practice the item either sees no demand (stock accumulates unused) or a large order arrives that the low forecast cannot cover. The Ensemble resolves this by separating the question of whether demand will occur from the question of how much it will be, through the SBA component, and by learning sequential timing patterns across thousands of SKU histories, through the LSTM. The result is an RMSE of 8.286 against MA12's 14.758 and a SPEC of 3.534 against MA12's 5.544, translating to a 43.9% RMSE reduction and a 36.3% SPEC reduction. These are the largest improvements observed across any pattern in the study. N-Beats contributes by isolating periodic components in the demand signal that would otherwise be buried in noise. Machine learning models underperformed on lumpy items despite having access to the full 231-feature set, because the sparse non-zero observations give tree-based methods too little to learn from.

Model Evaluation: Intermittent Demand Pattern

Intermittent items are the commercially dominant category at 70% of SKUs. The Ensemble achieves an average rank of 1.25 across the four metrics, the strongest result of any model across any pattern in this evaluation. RMSE falls from MA12's 10.487 to 7.269, a 30.7% cut, and SPEC drops from 4.342 to 3.707, a 14.6% reduction. The most telling figure may be directional accuracy: the Ensemble is right about the direction of demand 56.5% of the time, compared to 31.2% for MA12. The mechanism behind this is the LSTM component picking up the sequential structure of zero and non-zero periods across the 55-month training window. An intermittent item that demands five units roughly every third month will receive a MA12 forecast of 1.67 units per month regardless of where in its cycle it currently sits. The LSTM, trained across thousands of similar series, learns to condition the forecast on how many consecutive zero months have just passed, which is information that MA12 simply does not use.

Model Evaluation: Erratic Demand Pattern

Erratic items (2% of SKUs) are the hardest to forecast because demand arrives frequently but in quantities that jump around unpredictably. The Ensemble achieves the best RMSE at 23.206, an 18.5% gain over MA12's 28.480. Its SPEC, however, runs 5.9% above the MA12 baseline. The reason is that SPEC penalises stockout errors three times more heavily than overstock errors, and the Ensemble produces accurate central forecasts rather than systematically over-shooting. Simpler methods like SES and ADIDA tend to over-forecast slightly, which under this asymmetric weighting structure actually reduces their SPEC score even though their point accuracy is lower. For practitioners whose primary concern is stockout avoidance on erratic items, SES or ADIDA is the better input. Where minimising total forecast error matters more, the Ensemble remains the stronger choice.

Summary of Selected Methods

Table 2 brings together the recommended method for each demand pattern alongside the key error metric improvements over MA12.

Table 2. Summary of Selected Forecasting Method by Demand Pattern

Pattern	Current	Proposed	RMSE Diff	SPEC Diff	Decision
Smooth (8%)	MA12	MA12 retained	-6.7%	-1.5%	No change; gain too marginal
Lumpy (20%)	MA12	Ensemble	-43.9%	-36.3%	Replace with Ensemble
Intermittent (70%)	MA12	Ensemble	-30.7%	-14.6%	Replace with Ensemble
Erratic (2%)	MA12	Ensemble + higher SS	-18.5%	+5.9%	Ensemble with elevated Z-score (90% SL)

CONCLUSION

Plant SDU's simultaneous problems of excess inventory and inadequate service level both trace back to using MA12 across a portfolio that is 90% intermittent or lumpy by pattern. A single averaging method, applied uniformly, cannot correct for the structural mismatch between how it weights history and how these demand types actually behave. The empirical evaluation across 23 model variants and four demand patterns produced clear results. MA12 is adequate for the 8% of SKUs that are smooth and should be kept there. For the remaining 92%, an Ensemble of SES, SBA, LSTM, and N-Beats with pattern-specific weights outperforms all alternatives on the combined four-metric ranking. For lumpy items the RMSE gain is 43.9% and the SPEC gain is 36.3%. For intermittent items the RMSE gain is 30.7%, SPEC falls 14.6%, and directional accuracy jumps from 31.2% to 56.5%. For erratic items the Ensemble improves RMSE by 18.5% but raises SPEC slightly, pointing to the need for supplementary safety stock calibrated to a 90% service level target for this small category.

Feature importance analysis identified recent lag demand, rolling demand statistics, zero-count frequency, active equipment population, coal price, and lead time variability as the key drivers of accuracy. Each of these features has a direct operational interpretation and can be incorporated into an automated monthly replenishment pipeline without manual intervention. The practical takeaway is that matching the forecasting method to the demand pattern, rather than applying one method to all items, is achievable with existing data and standard modelling tools, and the improvement in forecast calibration is large enough to affect both the service level and the inventory position at the same time. Future work should extend this pattern-specific framework to other plants in the HCS network, test customer-level demand signals from maintenance schedules, and explore dynamic re-classification systems that update each SKU's demand category as its consumption profile evolves.

REFERENCES

Box, G. E. P., & Jenkins, G. M. (1970). Time series analysis: Forecasting and control. Holden-Day.

Chatfield, C. (1995). The analysis of time series: An introduction (5th ed.). Chapman & Hall/CRC.

Christopher, M. (2016). Logistics and supply chain management (5th ed.). FT Publishing International.

Corsten, D., & Gruen, T. (2004). Stock-outs cause walkouts. Harvard Business Review, 82(5), 26-28.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). Forecasting with exponential smoothing: The state space approach. Springer.

Katyare, P., Joshi, S. S., & Kulkarni, M. (2024). Machine learning based material demand prediction of construction equipment for maintenance. International Journal of Industrial Engineering Research and Development, 15(1), 45-56.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3149-3157.

Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. Applied Soft Computing, 11(2), 2664-2675.

Kiefer, D., Grimm, F., Bauer, M., & van Dinther, C. (2021). Demand forecasting intermittent and lumpy time series. International Journal of Production Economics, 235, 108076. https://doi.org/10.1016/j.ijpe.2021.108076

Kourentzes, N. (2013). Intermittent demand forecasts with neural networks. International Journal of Production Economics, 143(1), 198-206. https://doi.org/10.1016/j.ijpe.2012.12.004

Nielsen, M. A. (2015). Neural networks and deep learning. Determination Press.

Nikolopoulos, K., Syntetos, A. A., Boylan, J. E., Petropoulos, F., & Assimakopoulos, V. (2011). An aggregate-disaggregate intermittent demand approach (ADIDA) to forecasting. Journal of the Operational Research Society, 62(3), 544-554. https://doi.org/10.1057/jors.2010.32

Prestwich, S. D., Rossi, R., Tarim, S. A., & Hnich, B. (2014). Mean-based error measures for intermittent demand forecasting. International Journal of Production Research, 52(22), 6782-6791. https://doi.org/10.1080/00207543.2014.917771

Syntetos, A. A., Boylan, J. E., & Croston, J. D. (2005). On the categorization of demand patterns. Journal of the Operational Research Society, 56(5), 495-503. https://doi.org/10.1057/palgrave.jors.2601841

Trading Economics. (2025). Coal prices. Retrieved August 9, 2025, from https://tradingeconomics.com/commodity/coal

Waters, C. D. J. (2003). Inventory control and management (2nd ed.). Wiley.