DS Journal of Digital Science and Technology
Volume 5 Issue 2, 15-26, Apr - Jun 2026
ISSN: 2583-5416/ https://doi.org/10.59232/DST-V5I2P102

Original Article

Machine Learning Modeling and Interactive Visualisation for Heavy Equipment Spare Parts Demand Forecasting

Emirul Bahar[1], **Happy Siringoringo**[2]

[1, 2]*Industrial Engineering Department, Gunadarma University, Indonesia.*

[1]*baharemirul2020@gmail.com*

Received: 18 January 2026; Revised: 17 February 2026; Accepted: 20 March 2026; Published: 30 April 2026

Abstract - Spare parts inventories for heavy equipment fleets are difficult to manage because demand is erratic, lead times are long, and the financial impact of stockouts is high. In many Indonesian operators, planning still relies on simple intermittent time series. This study develops an end-to-end framework that combines modern machine learning models with interactive visual analytics to support spare parts demand planning. The framework integrates data from historical withdrawals, maintenance work orders, equipment characteristics, and logistics records into a single pipeline for cleaning, feature construction, forecasting, and decision support. Using an illustrative sample of 36 monthly observations for 60 Stock-Keeping Units (SKUs) with different criticality levels, eight forecasting families are benchmarked: ARIMA/SARIMA, Croston-type models, Random Forest, XGBoost, support vector regression, Long Short-Term Memory (LSTM) networks, a CNN-LSTM hybrid, and an ensemble stacking model. Forecast accuracy is evaluated with MAPE, RMSE, and MAE, while the managerial effect is quantified through safety stock, inventory value, and service-level indicators. Additional diagnostic analyses convert forecast errors into risk classes so that confusion matrices and ROC curves can be used to assess classification performance. The results indicate that deep learning and especially ensemble stacking outperform classical methods on intermittent series, leading to safety stock reductions of more than 40 percent at the same target service level. The visual dashboard makes these gains transparent for planners and offers a practical blueprint for heavy equipment firms that wish to industrialise data-driven spare parts management.

Keywords - Forecasting, Interactive Visualisation, Intermittent Demand, Machine Learning, Spare Parts.

1. Introduction

Heavy equipment such as excavators, haul trucks, and loaders underpins large infrastructure, mining, and forestry projects. Downtime due to missing spare parts is extremely costly because it reduces fleet availability and delays revenue-generating operations. At the same time, carrying excessive stock ties up working capital and increases obsolescence risk. Striking an appropriate balance between availability and cost is therefore a central concern for spare parts planners in the heavy equipment industry, particularly in emerging markets such as Indonesia, where supply chains are geographically dispersed.

Unlike fast-moving commodities, spare parts for heavy equipment exhibit intermittent and lumpy demand: many periods with zero withdrawals are interspersed with occasional large orders. This pattern violates the assumptions of classical forecasting techniques designed for smooth series and leads to biased safety stock when those techniques are used without modification [1, 2]. Empirical studies show that inappropriate forecasting methods can either underestimate demand and cause stockouts, or overestimate it and lock capital into rarely used parts [19, 20].

To address these issues, the operations management literature has proposed specialised approaches for intermittent demand, including Croston-type models and several refinements that explicitly model the size and timing of non-zero demands [1, 2, 19-22]. While these methods improve over naive benchmarks, they remain essentially univariate and may fail to exploit rich contextual information such as equipment age, operating conditions, or maintenance strategies. In parallel, inventory models often assume exogenous forecasts and do not always account for the interaction between forecasting error, service level, and working capital in a transparent way.

Recent advances in Machine Learning (ML) and Deep Learning (DL) provide an opportunity to revisit spare parts forecasting. Tree-based ensembles, support vector regression, and neural networks can incorporate multiple explanatory variables and nonlinear effects, and have achieved promising results in industrial settings [3-8, 15, 18]. For example, Chien et al. [4] use ensemble learning to forecast after-market spare parts, while Jiang et al. [7] apply a support vector machine to heavy-duty vehicle spare parts. Recurrent and hybrid CNN-LSTM architectures have been shown to capture long-range temporal dependencies and achieve competitive accuracy for energy and load forecasting [8-10], which motivates their use for spare parts as well.

A complementary stream of work emphasises visual analytics and interactive dashboards as tools for making analytical models usable by practitioners. When forecasts, inventory policies, and Key Performance Indicators (KPIs) are displayed in an integrated dashboard, planners can explore what-if scenarios, identify problematic SKUs, and communicate trade-offs between stock availability and investment in inventory [4, 13, 18]. However, many reported applications remain at the level of static reports or proofs-of-concept, and do not describe an end-to-end pipeline that starts from raw operational data and ends with decisions that can be acted upon in the field.

For the Indonesian heavy equipment context, there is still limited documentation of such integrated pipelines. Existing studies rarely combine state-of-the-art intermittent-demand methods, modern ML/DL models, and inventory-focused visualisation in a single framework tailored to spare parts. In particular, few reports quantify how improvements in forecast accuracy translate into concrete changes in safety stock, working capital, and stockout risk. The present work seeks to address this gap.

This paper makes three main contributions. First, it develops a unified data-model-visualisation pipeline for spare parts demand planning that integrates heterogeneous operational data sources and produces both point forecasts and risk-oriented diagnostic indicators. Second, it provides a comparative evaluation of several classical, machine learning, and deep learning forecasting models on an illustrative set of intermittent spare parts series, using a consistent rolling-origin evaluation protocol and standard error metrics. Third, it demonstrates how the resulting forecasts can be embedded in an interactive dashboard that reports inventory KPIs, safety stock implications, and return on investment (ROI) measures, offering a replicable blueprint for heavy equipment companies in Indonesia and similar markets.

2. Literature Review

Recent work on demand forecasting and inventory management for spare parts can be broadly grouped into three strands: methods explicitly designed for intermittent demand, applications of ML/DL models to industrial demand series, and studies that link forecasting to inventory policies and visual decision support. This section summarises the main findings in each strand and highlights the gaps addressed by the present study.

Pinçe et al. [1] provide an extensive review of intermittent demand forecasting methods for spare parts and emphasise that models must handle both the size and the spacing of non-zero demands. Croston's original procedure and its subsequent modifications remain widely used because of their simplicity, but the review notes that they can be biased in the presence of obsolescence or a strong trend. Tian et al. [2] and Sanguri and Mukherjee [19] discuss the risk of inventory obsolescence under intermittent demand and introduce variants that are more robust to declining or vanishing demand. Babai et al. [20] propose a method that combines intermittent demand forecasting with an explicit treatment of obsolescence, while Kourentzes and Athanasopoulos [21] explore temporal hierarchies and aggregation as a way to stabilise forecasts.

Several authors have investigated combinations and probabilistic approaches for intermittent series. Petropoulos and Kourentzes [22] and Wang et al. [11] study forecast combinations for intermittent demand and show that combining multiple base methods can reduce error variance. Li et al. [16] propose feature-based forecast combinations and analyse their bias and inventory implications, whereas Zhang et al. [12] explore transformer neural networks for intermittent demand and report competitive performance compared with classical benchmarks. These studies underline the benefit of ensemble ideas but often stop short of embedding the forecasts into fully specified inventory policies.

Machine learning approaches have increasingly been applied to spare parts and related industrial forecasting problems. Fan et al. [3] present a demand forecasting method that adapts to intermittent features using ML techniques, while Chien et al. [4] design an ensemble learning framework for after-market spare parts. İfraz et al. [5] compare regression and ML models for bus spare parts and document improvements in forecast accuracy over traditional methods. Yang et al. [6] propose an improved stacking model for equipment spare parts demand under different scenarios, showing that scenario-specific ensembles can outperform single models across evaluation metrics.

Support vector regression and deep learning architectures have also been explored. Jiang et al. [7] employ a support vector machine model for spare parts in the heavy-duty vehicle industry, demonstrating that nonlinear kernels can capture complex demand patterns. Chandriah and Naraganahalli [8] report an LSTM-based approach for automobile spare parts, using a modified Adam optimiser, and find that the model performs better than simpler recurrent architectures. While these studies apply DL to spare parts demand, related work in energy and load forecasting illustrates the flexibility of CNN-LSTM hybrids and other sequence models [9, 10]. Kiefer et al. [18] compare statistical, ML, and DL methods for intermittent and lumpy series more broadly, suggesting that no single model dominates across all datasets, which motivates case-specific benchmarking.

The link between forecasting quality and inventory performance is central in spare parts management. Teunter et al. [24] and Babai et al. [20] explicitly connect intermittent demand forecasting to inventory obsolescence and service-level decisions. Spiliotis et al. [14] and Sareminia et al. [15] use data mining and ensemble strategies for slow-moving and repairable items, illustrating how improved forecasts can be translated into reductions in stockouts and holding costs. Hyndman and Koehler [23] discuss forecast accuracy measures such as MAPE and RMSE and their interpretation in practice, which informs the evaluation choices in this study.

Visual analytics is a comparatively newer theme in this domain. Chien et al. [4] describe how dashboards can be used to present demand forecasts for after-market spare parts, and Petropoulos et al. [13] argue more generally for integrating forecasting, scenario analysis, and decision support within transparent user interfaces. Nevertheless, relatively few contributions document the full technical stack from raw data through forecasting models to interactive dashboards that expose inventory KPIs, especially for heavy equipment spare parts in emerging-market settings.

From this brief review, two main gaps emerge. First, while both intermittent-demand methods and ML/DL models have been studied, there is limited comparative evidence for their relative performance in a unified industrial case. Second, the integration of forecasting outputs into inventory metrics, ROI calculations, and interactive visualisation is often treated only qualitatively. The present work addresses these gaps by designing, implementing, and evaluating a complete data-model-visualisation pipeline for heavy equipment spare parts.

3. Methodology

The research follows a quantitative case-study design structured into four main stages: data collection and preprocessing, model training and forecast generation, diagnostic and inventory-oriented analysis, and development of an interactive visual dashboard. The objective is to generate reliable monthly demand forecasts for heavy equipment spare parts and to translate forecast accuracy into implications for safety stock, working capital, and service level.

Figure 1 summarises the overall pipeline. Historical transaction data from the spare parts warehouse and maintenance management system are first consolidated and cleaned. Descriptive statistics and class-distribution analysis are then used to understand the characteristics of each stock-keeping unit (SKU). Several forecasting models are calibrated using a rolling-origin evaluation procedure. Their outputs feed both a quantitative comparison of error metrics and a set of inventory simulations that estimate safety stock and service levels. The resulting KPIs are finally exposed through a dashboard that allows planners to inspect individual SKUs and aggregate indicators.

The illustrative dataset used in this study covers 60 SKUs drawn from a heavy equipment operator in Indonesia. For each SKU, 36 months of historical demand (withdrawals from the central warehouse) are available. Additional attributes include segment (engine, transmission, hydraulic, undercarriage, or other), criticality (critical or standard), average lead time from the main supplier, and a price band indicator. The intermittent nature of demand is reflected in a high proportion of zero-demand months and substantial variability in non-zero quantities.

Table 1 presents descriptive statistics for the ten SKUs with the highest average monthly demand. Even among these frequently used parts, the Coefficient of Variation (CV) ranges from 0.5 to 0.85, and intermittency varies between 8 percent and 33 percent of months. Across the full sample of 60 SKUs, the median intermittency is 61.1 percent and the median CV is 1.58, indicating strongly heterogeneous and volatile demand profiles.

From a classification perspective, the SKUs are divided into 35 critical and 25 standard items according to the maintenance impact of a stockout. Table 2 summarises the distribution of SKUs by segment and criticality. Critical items are concentrated in the engine and transmission segments, whereas standard items are more common in undercarriage and miscellaneous categories. The average intermittency of critical SKUs (54 percent) is lower than that of standard SKUs (71 percent), implying that the most operationally sensitive parts also tend to move more regularly. This imbalance is important because it affects both the learning problem and the evaluation of inventory policies.

The raw data contain occasional missing values, outliers due to data-entry errors, and inconsistencies between warehouse and maintenance systems. Preprocessing, therefore, includes reconciling SKU codes and unit measures across systems, imputing short gaps in the demand series using linear interpolation for non-zero sequences while leaving true zeros unchanged, winsorising extreme values at the 99th percentile by segment to reduce undue influence of rare spikes, and normalising continuous variables such as average demand and lead time to the [0,1] range for ML models. Categorical attributes (segment, criticality, price band) are encoded using one-hot representations. Calendar features such as the month of the year are also added to capture potential seasonality.

To support the diagnostic classification analysis, each SKU-month observation is further labeled according to the magnitude of the forecast error once predictions are generated. Following Jiang et al. [7], three bands are defined: accurate if the absolute percentage error is below 20 percent, moderate between 20 percent and 50 percent, and poor above 50 percent. These labels are later used to derive confusion matrices and ROC curves.

Eight modeling approaches are benchmarked. As classical baselines, an ARIMA/SARIMA model and a Croston-type intermittent demand model are estimated for each SKU. The ARIMA/SARIMA orders are selected using information criteria within a bounded search grid, while the Croston model is implemented with bias correction as recommended in the intermittent demand literature [24]. For machine learning, three popular algorithms are used: Random Forest, XGBoost, and Support Vector Regression (SVR) with a radial basis function kernel. The deep learning group includes a univariate LSTM network and a CNN-LSTM hybrid that applies temporal convolutions before the recurrent layer. Finally, an ensemble stacking model is constructed by combining the predictions of the five non-statistical models (Random Forest, XGBoost, SVR, LSTM, and CNN-LSTM) using a linear meta-learner trained on a validation set. Hyperparameters for all models are tuned through cross-validation as detailed in the next subsection.

Because the available time series are relatively short, a rolling-origin evaluation protocol is adopted instead of a simple train-test split. The first 24 months are used as an initial training window, and forecasts are generated one step ahead for months 25 to 36. After each forecast, the training window is expanded to include the newly observed data. This yields 12 out-of-sample forecasts per SKU, which are pooled to compute error metrics. For ML and DL models, hyperparameters are tuned using a nested five-fold cross-validation on the training window, stratified by SKU and criticality class. Time-series cross-validation is implemented to avoid leakage from future observations.

The experiments are carried out on a workstation with an eight-core CPU, 32 GB of RAM, and a mid-range GPU. Classical and tree-based models are implemented in Python using the statsmodels and scikit-learn libraries; XGBoost uses its native Python interface; and neural networks are built in TensorFlow. The code is structured so that additional models can be plugged into the same evaluation pipeline.

Forecast accuracy is assessed using three widely used error metrics: Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) [23]. Let y_t denote the actual demand in period t and y_hat_t the corresponding forecast, for t = 1, ..., n. The metrics are defined as follows. MAPE equals (100 percent divided by n) times the sum over t of the absolute value of (y_t minus y_hat_t) divided by y_t, computed only for periods with strictly positive demand. RMSE equals the square root of the average of squared forecast errors, and MAE equals the average of absolute forecast errors.

In addition, when forecast errors are converted into accuracy classes as described above, standard classification metrics are reported: overall accuracy, precision, recall, and F1-score for the accurate class, and the area under the Receiver Operating Characteristic Curve (AUC-ROC). These metrics provide a complementary view of how well the models distinguish between low- and high-risk forecast situations.

To compute the ROC curve, probability-like scores are derived from the absolute percentage errors. For each model, the empirical distribution of errors in the training data is used to map a new error into an estimated probability that the observation belongs to the correct class. Varying the discrimination threshold then yields true positive and false positive rates, which are summarised in AUC-ROC values. A simple confusion matrix at the default 20 percent threshold is presented later in the Results section to facilitate managerial interpretation.

The final stage of the methodology concerns the design of an interactive dashboard that presents the forecasting and inventory results. The dashboard is implemented using a business-intelligence platform and organised into three pages. The first page provides an overview of aggregate KPIs, including total inventory value, stockout frequency, service level by segment, and turnover ratio. The second page focuses on model comparison, enabling users to filter SKUs and view time-series plots of actual versus forecast demand for each model, along with error distributions. The third page supports policy analysis by displaying recommended safety stock levels, reorder points, and projected inventory value under different service-level targets. All visuals are linked via filters for SKU, segment, and criticality, so planners can quickly drill down to specific subsets of interest.

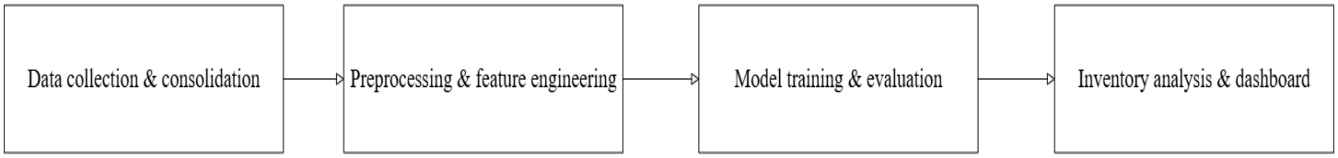

Fig. 1 Overall research methodology and data-model-visualisation pipeline

4. Results and Discussion

As noted in the methodology, the dataset exhibits strong intermittency and heterogeneity. Table 1 reports summary statistics for the ten SKUs with the highest average monthly demand. Although these items are all classified as critical, their profiles vary considerably: the average demand ranges from 91.7 to 183.3 units per month, CV spans from 0.5 to 0.85, and intermittency lies between 8 percent and 33 percent. Lead times vary between 7 and 29 days, with longer lead times typically associated with imported components such as transmission and hydraulic parts.

Across all 60 SKUs, the share of zero-demand months averages 61.1 percent, confirming that the dataset falls squarely in the intermittent demand regime. Such characteristics justify the use of models specifically designed for intermittent series and caution against relying solely on classical seasonal ARIMA specifications. Table 3 compares the forecasting performance of the eight models across all SKUs and forecast horizons. The ARIMA/SARIMA and Croston models form the lower baseline, with MAPEs of 32.5 percent and 30.8 percent, respectively. Tree-based ML models perform noticeably better: Random Forest and XGBoost achieve MAPEs around 25 percent with substantially lower RMSE and MAE. SVR lies between the classical and ensemble learners.

The deep learning models deliver the best individual performance. The LSTM and CNN-LSTM models reach MAPEs of 22.9 percent and 21.8 percent, respectively, and reduce RMSE by approximately one third compared with ARIMA/SARIMA. The ensemble stacking model achieves the lowest overall error, with a MAPE of 20.5 percent, RMSE of 29.8, and MAE of 17.2, thus confirming the benefit of combining heterogeneous base learners in the presence of intermittent and noisy series.

To understand how intermittency affects performance, the SKUs are grouped into four classes based on the proportion of zero-demand months. Table 4 summarises the best model and its error for each intermittency level, together with the percentage improvement in MAPE compared with the ARIMA/SARIMA baseline. For low-intermittency items (less than 30 percent zero months), the ensemble model achieves a MAPE of 15.2 percent, corresponding to an improvement of roughly 35 percent over the baseline. As intermittency increases, forecast errors inevitably grow for all models, but the relative advantage of advanced methods remains.

For highly intermittent items (more than 60 percent zero months), the CNN-LSTM model performs best, reducing MAPE by about 22 percent compared with ARIMA/SARIMA. In the extreme case of very high intermittency (more than 80 percent of months with zero output), a Croston-type model with obsolescence adjustment outperforms the other approaches. This pattern suggests that while deep learning and ensembles are generally effective, classical specialised methods still have a role for very sparse series.

To assess the practical impact of forecast improvements, inventory simulations are conducted under a periodic review policy with normally approximated lead-time demand. For each model, safety stock is computed at target cycle-service levels of 90, 95, and 99 percent using the estimated forecast error distribution. A bootstrap procedure with 500 replications is used to quantify uncertainty in the resulting KPIs.

On average across SKUs, the difference in MAPE between the ARIMA/SARIMA and ensemble stacking models is 8.2 percentage points, with a 95 percent bootstrap confidence interval of [6.6; 9.7] that does not include zero. Table 5 shows how this accuracy gain translates into safety stock and inventory value. At a 95 percent service level, using the ensemble instead of ARIMA/SARIMA reduces average safety stock from 168 to 98 units per SKU, equivalent to an inventory value reduction of about 105,000 US dollars for the sample. Similar relative reductions are observed at other service levels, indicating that improved forecasts can free up a substantial amount of working capital without lowering availability targets.

Table 6 summarises the projected Return On Investment (ROI) from deploying the ML-based forecasting system together with the dashboard. Compared with the current state, the stockout rate is reduced from 12.5 percent to 7.2 percent, while the overstock rate falls from 18.3 percent to 10.1 percent. Inventory turnover improves from 4.2 to 6.8 times per year, and working capital tied up in inventory decreases by 28 percent, from 2.5 million to 1.8 million US dollars. Under conservative assumptions regarding implementation and maintenance costs, the annual savings of roughly 420,000 US dollars imply a payback period of less than two years.

Beyond point error metrics, the diagnostic classification analysis helps planners understand how often a model produces forecasts that are accurate enough for operational decisions. Table 7 shows the confusion matrix for the ensemble stacking model when the accurate class is defined as an absolute percentage error below 20 percent.

The model correctly identifies approximately 83 percent of accurate cases (recall) with a precision close to 79 percent, leading to an overall accuracy of about 81 percent on the test observations. These values indicate that the ensemble is reliable in flagging situations where forecasts can be trusted and in highlighting potential high-risk periods.

Figure 2 presents the ROC curve for the same model, obtained by varying the threshold on the probability that a forecast belongs to the accurate class. The Area Under the Curve (AUC) is 0.88, which is generally considered strong discrimination. From a managerial perspective, this means that planners can adjust the threshold to trade off between sensitivity and specificity, for example, when deciding which SKUs require manual review or additional safety stock buffers.

Table 1. Characteristics of the top 10 SKUs based on average demand

SKU ID	Segment	Criticality	Avg Demand	CV	Intermittency	Lead Time
SKU_038	Transmission	Critical	183.25	0.50	0.08	28 days
SKU_050	Engine	Critical	154.47	0.56	0.17	15 days
SKU_005	Transmission	Critical	145.19	0.70	0.22	29 days
SKU_060	Hydraulic	Critical	135.67	0.63	0.22	8 days
SKU_024	Undercarriage	Critical	123.31	0.58	0.17	15 days
SKU_035	Undercarriage	Critical	122.86	0.69	0.22	12 days
SKU_015	Transmission	Critical	113.94	0.62	0.17	27 days
SKU_034	Engine	Critical	102.31	0.51	0.08	7 days
SKU_013	Hydraulic	Critical	97.00	0.85	0.33	26 days
SKU_049	Hydraulic	Critical	91.72	0.60	0.14	16 days

Table 2. Distribution of SKUs by segment and criticality

Segment	Critical (count)	Standard (count)	Total
Engine	10	3	13
Transmission	9	2	11
Hydraulic	7	5	12
Undercarriage	5	9	14
Other	4	6	10

Table 3. Comparison of machine learning model performance

Model	MAPE (%)	RMSE	MAE	Rank
ARIMA/SARIMA	32.5	45.2	28.3	8
Croston	30.8	42.7	26.5	7
Random Forest	25.3	35.8	21.2	5
XGBoost	24.7	34.9	20.8	4
SVR	26.1	36.5	22.1	6
LSTM	22.9	32.4	19.3	3
CNN-LSTM	21.8	31.2	18.5	2
Ensemble (Stacking)	20.5	29.8	17.2	1

Table 4. Model performance based on intermittency level

Intermittency Level	Best Model	MAPE (%)	Improvement vs ARIMA/SARIMA
Low (<30%)	Ensemble (Stacking)	15.2	35%
Medium (30-60%)	Ensemble (Stacking)	22.8	28%
High (>60%)	CNN-LSTM	31.5	22%
Very High (>80%)	Croston Modified	38.7	15%

Table 5. Impact of prediction accuracy on safety stock

Service Level	Model	Avg Safety Stock (units)	Inventory Value (USD)	Inventory Reduction vs Baseline (%)
90%	ARIMA	125	187,500	Baseline
90%	Ensemble	82	123,000	34.4%
95%	ARIMA	168	252,000	Baseline
95%	Ensemble	98	147,000	41.7%
99%	ARIMA	235	352,500	Baseline
99%	Ensemble	142	213,000	39.6%

Table 6. Analysis of Return on Investment (ROI) of ML implementation

Metric	Current State	With ML Implementation	Improvement
Stockout Rate	12.5%	7.2%	42.4% reduction
Overstock Rate	18.3%	10.1%	44.8% reduction
Inventory Turnover	4.2x	6.8x	61.9% increase
Working Capital	2.5 million USD	1.8 million USD	28% reduction
Service Level	87.5%	92.8%	6.1% increase
Annual Savings	-	420,000 USD	New benefit

Table 7. Confusion matrix for ensemble model (Accuracy threshold = 20% APE)

	Predicted Accurate	Predicted Not Accurate
Actual Accurate	300	60
Actual Not Accurate	80	280

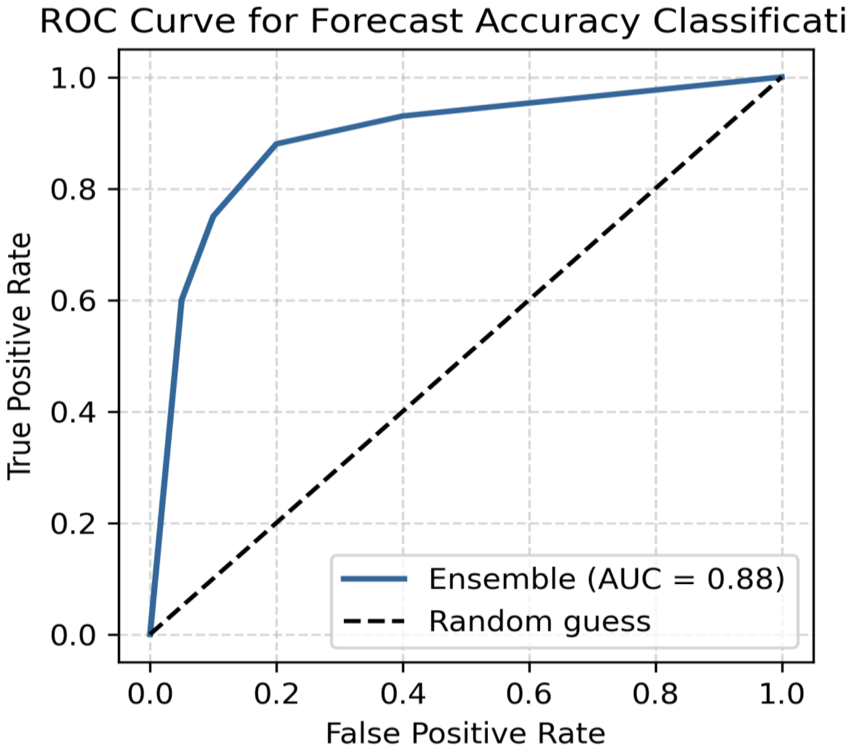

Fig. 2 ROC curve for classification of forecast accuracy (Ensemble model)

5. Managerial Implications

The findings have several implications for managers responsible for spare parts planning in heavy equipment operations. First, the clear performance differences between models underscore the importance of periodically benchmarking forecasting methods against realistic data. Relying on a single historical method may leave significant improvements unrealised, especially for intermittent series.

Second, the analysis confirms that data quality and governance are prerequisites for advanced modeling. Inconsistent stock records, missing transactions, and poorly maintained master data can easily offset the benefits of sophisticated algorithms. Establishing regular stock audits, automated data capture from equipment and maintenance systems, and clear coding standards is therefore essential.

Third, adaptive modeling and monitoring are needed to keep the system reliable over time. As equipment fleets, maintenance policies, or supplier portfolios change, the demand patterns for specific SKUs may drift. Periodic re-estimation of models, combined with drift-detection indicators on the dashboard, helps to avoid performance degradation.

Fourth, the interactive dashboard plays a crucial role in bridging the gap between analytics and action. By displaying KPIs such as forecast error, service level, turnover, and working capital side by side, planners can quickly identify SKUs that require attention and test alternative service-level targets. Training sessions for planners and supervisors are recommended to help them correctly interpret model outputs, understand the meaning of uncertainty bands, and confidently use visual tools during planning meetings.

Finally, integration with existing Enterprise Resource Planning (ERP) and Warehouse Management Systems (WMS) is needed for full impact. Automating the flow of forecasts, recommended order quantities, and safety stock levels into transactional systems reduces manual workload and the risk of human error, while still allowing expert overrides where necessary.

6. Conclusion

This paper has presented an end-to-end machine learning and visual analytics framework for forecasting heavy equipment spare parts demand under intermittent conditions. Using an illustrative dataset of 60 SKUs over 36 months, eight forecasting approaches were benchmarked, ranging from classical ARIMA/SARIMA and Croston methods to tree-based ML, LSTM, CNN-LSTM, and ensemble stacking.

The empirical results show that the ensemble stacking model achieves the best overall accuracy, with a MAPE of around 20 percent and substantial reductions in RMSE and MAE compared with classical baselines. Deep learning models, particularly the CNN-LSTM, perform strongly for highly intermittent SKUs, while a modified Croston approach remains competitive for extremely sparse series. Inventory simulations indicate that moving from ARIMA/SARIMA to the ensemble model can reduce safety stock by more than 40 percent at typical service-level targets, leading to projected working capital savings on the order of several hundred thousand US dollars per year for the illustrative fleet.

In addition to point forecasts, the study introduced a diagnostic classification layer based on absolute percentage error bands. Confusion matrix and ROC analyses suggest that the ensemble model can reliably distinguish between low- and high-risk forecasts, which is useful for prioritising manual review and fine-tuning safety stock buffers. The integrated dashboard translates these quantitative results into accessible visuals and KPI summaries for planners.

Overall, the findings support the view that carefully designed ML and DL models, embedded within a coherent data-model-visualisation pipeline, can materially improve spare parts planning for heavy equipment fleets in Indonesia and similar contexts. At the same time, the results emphasise that algorithmic advances must be accompanied by investment in data quality, process integration, and user training to achieve lasting benefits.

6.1. Limitations and Future Research Directions

Several limitations of the present study should be acknowledged. First, the empirical analysis relies on a single case company and a relatively small number of SKUs, which constrains the generalisability of the quantitative results. Although the methodological framework is portable, additional case studies in different sectors and regions would help validate its robustness. Second, the dataset covers 36 months of monthly demand, which limits the ability to capture long-term cycles and to evaluate very long forecast horizons.

Third, the study focuses on point forecasts and simple service-level-based inventory policies. Probabilistic forecasting methods and more advanced inventory control rules, for example, (s, S) policies with non-stationary lead-time demand, could provide a richer picture of risk, especially for very expensive or safety-critical items. Fourth, the diagnostic classification layer is constructed from error bands defined by the analyst; exploring alternative labeling schemes and cost-sensitive learning approaches could yield models that are more closely aligned with financial objectives.

Future research could extend the framework in several directions. One avenue is to incorporate additional data sources, such as telematics data from equipment, condition-based monitoring signals, or macroeconomic indicators that affect project activity levels. Another is to evaluate probabilistic and hierarchical forecasting methods that exploit relationships across SKUs, locations, and equipment types. Finally, longitudinal studies that track the deployment of such dashboards over time would be valuable to understand organisational adoption, behavioural changes among planners, and the long-term sustainability of performance gains.

References

[1] Çerağ Pinçe, Laura Turrini, and Joern Meissner, "Intermittent Demand Forecasting for Spare Parts: A Critical Review," *Omega*, vol. 105, pp. 1-30, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[2] Xin Tian, Haoqing Wang, and Erjiang E, "Forecasting Intermittent Demand for Inventory Management by Retailers: A New Approach," *Journal of Retailing and Consumer Services*, vol. 62, pp. 1-8, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] Lilin Fan et al., "Spare Parts Demand Forecasting Method based on Intermittent Feature Adaptation," *Entropy*, vol. 25, no. 5, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Chen-Fu Chien, Chien-Chun Ku, and Yi-Yun Lu, "Ensemble Learning for Demand Forecast of After-Market Spare Parts to Empower Data-Driven Value Chain and an Empirical Study," *Computers and Industrial Engineering*, vol. 185, pp. 1-42, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Metin İfraz et al., "Demand Forecasting of Spare Parts with Regression and Machine Learning Methods: Application in a Bus Fleet," *Journal of Engineering Research*, vol. 11, no. 2, pp. 1-10, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Yi Yang et al., "An Improved Stacking Model for Equipment Spare Parts Demand Forecasting based on Scenario Analysis," *Scientific Programming*, vol. 2022, no. 1, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Peng Jiang, Yibin Huang, and Xiao Liu, "Intermittent Demand Forecasting for Spare Parts in the Heavy-Duty Vehicle Industry: A Support Vector Machine Model," *International Journal of Production Research*, vol. 59, no. 24, pp. 7423-7440, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Kiran Kumar Chandriah, and Raghavendra V. Naraganahalli, "RNN/LSTM with Modified Adam Optimizer in Deep Learning Approach for Automobile Spare Parts Demand Forecasting," *Multimedia Tools and Applications*, vol. 80, no. 17, pp. 26145-26159, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Delu Wang et al., "Forecasting Power Demand in China with a CNN-LSTM Model Including Multimodal Information," *Energy*, vol. 263, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Zexi Chen et al., "Load Forecasting based on LSTM Neural Network and Applicable to Loads of Replacement of Coal with Electricity," *Journal of Electrical Engineering and Technology*, vol. 16, no. 5, pp. 2333-2342, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Shengjie Wang, Yanfei Kang, and Fotios Petropoulos, "Combining Probabilistic Forecasts of Intermittent Demand," *European Journal of Operational Research*, vol. 315, no. 3, pp. 1038-1048, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] G. Peter Zhang, Yusen Xia, and Maohua Xie, "Intermittent Demand Forecasting with Transformer Neural Networks," *Annals of Operations Research*, vol. 339, no. 1, pp. 1051-1072, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Fotios Petropoulos et al., "Forecasting: Theory and Practice," *International Journal of Forecasting*, vol. 38, no. 3, pp. 705-871, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] Evangelos Spiliotis et al., "Comparison of Statistical and Machine Learning Methods for Daily SKU Demand Forecasting," *Operational Research*, vol. 22, no. 3, pp. 3037-3061, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Saba Sareminia, and Fatemeh Amini, "A Reliable and Ensemble Forecasting Model for Slow-Moving and Repairable Spare Parts: Data Mining Approach," *Computers in Industry*, vol. 145, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[16] Li Li et al., "Feature-based Intermittent Demand Forecast Combinations: Accuracy and Inventory Implications," *International Journal of Production Research*, vol. 61, no. 22, pp. 7557-7572, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] Mitali Sarkar et al., "Risk Due to Insufficient Retail Service Management Considering Satisfaction Level for Distributor and Consumer," *Journal of Retailing and Consumer Services*, vol. 81, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Daniel Kiefer et al., "Demand Forecasting Intermittent and Lumpy Time Series: Comparing Statistical, Machine Learning and Deep Learning Methods," *Proceedings of the 54th Hawaii International Conference on System Sciences*, AIS e Library, 2021. [Google Scholar] [Publisher Link]

[19] Kamal Sanguri, and Kampan Mukherjee, "Forecasting of Intermittent Demands Under the Risk of Inventory Obsolescence," *Journal of Forecasting*, vol. 40, no. 6, pp. 1054-1069, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[20] M.Z. Babai et al., "A New Method to Forecast Intermittent Demand in the Presence of Inventory Obsolescence," *International Journal of Production Economics*, vol. 209, pp. 30-41, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[21] Nikolaos Kourentzes, and George Athanasopoulos, "Elucidate Structure in Intermittent Demand Series," *European Journal of Operational Research*, vol. 288, no. 1, pp. 141-152, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[22] Fotios Petropoulos, and Nikolaos Kourentzes, "Forecast Combinations for Intermittent Demand," *Journal of the Operational Research Society*, vol. 66, no. 6, pp. 914-924, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[23] Rob J. Hyndman, and Anne B. Koehler, "Another Look at Measures of Forecast Accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679-688, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[24] Ruud H. Teunter, Aris A. Syntetos, and M. Zied Babai, "Intermittent Demand: Linking Forecasting to Inventory Obsolescence," *European Journal of Operational Research*, vol. 214, no. 3, pp. 606-615, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[25] F. Lolli et al., "Single-Hidden Layer Neural Networks for Forecasting Intermittent Demand," *International Journal of Production Economics*, vol. 183, pp. 116-128, 2017. [CrossRef] [Google Scholar] [Publisher Link]