

Developing a Forecasting Framework for Spare Parts for Maintenance Service Providers

Emad W. Jamal and Abdul Zubar Hameed

Department of Industrial Engineering, Faculty of Engineering,
King Abdulaziz University, P.O. Box 80204, Jeddah 21589,
Saudi Arabia.

Abstract— The increasing complexity of operations in the maintenance service industry necessitates sophisticated approaches to forecasting spare parts demand. This study introduces a robust forecasting framework using advanced machine learning techniques to address the inefficiencies and inaccuracies associated with traditional forecasting methods. Focusing on ATM supply chain operations, we implement and evaluate three predictive models: Linear Regression, Long Short-Term Memory (LSTM) networks, and an Enhanced Time-Series model with Bidirectional LSTM and Dropout Regularization. Data preprocessing was rigorously performed on a comprehensive dataset, including information from maintenance activities, part specifications, and supply chain parameters. The development of the Spare Part Operational Demand Index (SPODI) was pivotal, integrating multiple predictive factors into a single, actionable metric to refine the predictive maintenance strategy. This study is one of the first to empirically examine the efficacy of LSTM networks in the context of spare parts demand forecasting, revealing their superior ability to capture complex, long-term patterns. The Linear Regression model served as a baseline, demonstrating modest predictive power primarily for lower to medium usage levels. In contrast, the LSTM network showed significant enhancements in accuracy, particularly for higher demand values, through its ability to learn from temporal sequences. The Enhanced Time-Series model, incorporating Bidirectional LSTM and Dropout Regularization, surpassed other models, achieving the highest accuracy, as demonstrated by its close adherence to the perfect prediction line and the sharply peaked residuals distribution. Our findings highlight the importance of the 'SPODI Score' within the LSTM-based models, validating its role as a key indicator of demand. The progression from Linear Regression to Enhanced LSTM models underlines a transformative shift towards data-driven forecasting, with machine learning algorithms offering a potent tool for predictive accuracy and efficient inventory management. This study contributes substantial insights into the field of predictive maintenance, emphasizing the need for dynamic, AI-driven approaches to spare parts demand forecasting, ultimately enhancing service quality and cost efficiency in the maintenance service sector.

Keywords—component; formatting; style; styling; insert (key words)

I. INTRODUCTION

Effective company performance hinges on the efficient execution of all internal and external activities, alignment with strategic guidelines, and continuous resource optimization. Central to these activities is the management of purchase orders, which formalize the acquisition of goods necessary for business operations. Accurate and detailed purchase orders ensure error-free transactions and support sector-specific demands within the company, underpinning competitiveness and cost optimization (Boylan & Syntetos, 2010; Sheikh et al., 2000; Rego & Mesquita, 2011). The shift towards data-driven forecasting, especially machine learning algorithms, is revolutionizing resource optimization and predictive modeling in purchase orders. Unlike traditional methods, these algorithms adjust predictions based on new data, embodying a dynamic approach to forecasting that is essential in the rapidly evolving business environment (Dombi et al., 2018; Kalchschmidt et al., 2003; Zhang et al., 2021).

In the realm of spare parts management, the challenge of accurately predicting and analyzing demand is a critical concern that has garnered substantial attention within the academic and industrial sectors. The literature reveals a significant gap in the effective prediction mechanisms for spare parts demand, which is inherently tied to a multitude of factors, resulting in a high degree of randomness. The conventional approach adopted by many enterprises leans heavily on the experiential acumen of production operators to devise demand plans, a practice that often lacks a robust forecasting framework, leading to inefficiencies such as the "bullwhip effect" where the amplification of demand uncertainty results in excessive inventory accumulation and sluggish turnover rates (Käki, 2007; Hemeimat et al., 2016). This underscores the imperative for establishing a more scientific and comprehensive spare parts inventory management system, one that is grounded in predictive accuracy to ensure optimal supply levels without incurring unnecessary capital and maintenance costs (Auweraer & Boute, 2019).

The discourse extends into the strategic outsourcing of maintenance and its integration with spare parts management, highlighting the evolution of maintenance from a peripheral to a core activity within various sectors, notably in avionics. Outsourcing, driven by motivations such as cost reduction, skill enhancement, and service level optimization, introduces a nuanced framework for managing spare parts, encompassing operative maintenance, maintenance management, contractual management, and spare parts management. This multi-faceted

II. CASE STUDY

approach not only addresses the logistical aspects but also emphasizes the decision-making frameworks that govern the stocking, inventory levels, and component classification of spare parts, crucial for mitigating risks associated with lumpy demand and ensuring efficient service provision (Macchi et al., 2016; Tarakci, 2007; Harland, 2003).

Alhamrani Universal (AU) exemplifies the need for improved forecasting in its ATM supply chain operations. Despite its substantial market share, AU faces challenges in maintaining accurate inventory levels for ATM spare parts, leading to increased costs and compromised service delivery. This study focuses on analyzing and enhancing AU's forecasting models to bridge the gap between theoretical and practical outcomes, thereby reducing inventory costs and improving service quality. The escalating complexity of operations necessitates the adoption of sophisticated data-driven forecasting methods. Technological advances bring about intricate challenges involving multiple factors such as customer demand, resource availability, and equipment performance, which require a profound understanding for effective management (Salmeron, 2019). Traditional forecasting methods fall short in accuracy, leading to inefficiencies and elevated costs. In contrast, machine learning algorithms enhance forecasting accuracy by incorporating real-time changes across numerous variables, enabling more precise resource allocation and bolstering operational performance (Miller S., 2021; Morariu, 2020). Furthermore, data-driven techniques facilitate proactive decision-making by predicting future demands and identifying potential risks in advance, thus improving the decision-making process and overall service quality (Data-Driven Decision Making: Management with Analytics, n.d.). The integration of machine learning into maintenance service operations promises to elevate accuracy and drive long-term business success.

Advancements in AI and machine learning, particularly Long Short-Term Memory (LSTM) networks (Figure 1), are highlighted for their potential to improve maintenance service management significantly. The study will evaluate the efficacy of LSTM networks in capturing complex, long-term patterns in spare part demand forecasting, offering a potential path to improved decision-making and long-term business success (Salmeron, 2019; Miller S., 2021; Morariu, 2020).

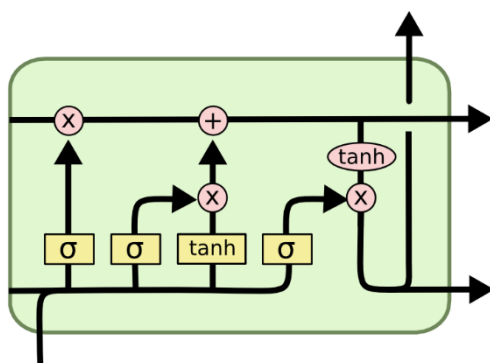


Figure 1 LSTM architecture model illustration.

A. Data Set Description

The challenge presented by the dataset on ATM machine parts is critical to the operational efficiency of banking services. At its core, the problem consists of forecasting the monthly demand for each individual part required for ATM maintenance, a task pivotal for ensuring that ATMs remain functional with minimal downtime. This predictive task is not merely about maintaining an optimal inventory level but also about preventing potential disruptions in ATM services that can affect customer satisfaction and trust. The complexity of this problem is highlighted by the variability in part failure rates, which can be influenced by factors such as usage frequency, environmental conditions, and the inherent durability of the parts. Addressing this challenge necessitates a comprehensive understanding of both the operational dynamics of ATMs and the logistical considerations of supply chain management. By accurately predicting part demand, banks can streamline their maintenance operations, reduce costs associated with emergency procurement, and ultimately enhance the reliability of their ATM networks.

Regarding the data parameters provided, the dataset encompasses a broad spectrum of information crucial for analyzing and predicting ATM part requirements. The 'SC' parameter stands for Supply Chain, indicating the logistics and processes involved in procuring, storing, and distributing the parts. Each parameter—ranging from 'Completion Date', which marks the finalization of a maintenance activity, to 'SLA Response', denoting the service level agreement compliance time—plays a significant role in understanding the lifecycle and demand of ATM parts. 'Part No' and 'Part Description' offer specifics about the components, while 'Qty' (Quantity) reflects the consumption rate, crucial for demand forecasting. 'TID Description' provides insights into the ATM models or locations, potentially influencing part failure rates. 'Employee ID' and 'Employee Name' link maintenance activities to personnel, offering a layer of detail that could be instrumental in identifying training needs or performance-related trends. Together, these parameters form a comprehensive dataset that, when effectively analyzed, can significantly optimize inventory management and maintenance scheduling in the ATM service ecosystem.

B. Data Pre-processing

The code outlines a series of data preprocessing steps essential for preparing the dataset for analysis. Initially, data is collected and then loaded into a Pandas DataFrame from a specified file path, setting the stage for subsequent manipulations. To maintain the dataset's integrity, any rows with missing values are removed, ensuring that the analysis remains accurate and reliable. Following this, the index is reset to ensure continuity, which is vital for data integrity and future access. The text data within the 'Part No' and 'Part Description' columns undergo standardization by trimming leading and trailing spaces and converting all text to uppercase, thus ensuring consistency and preventing duplicates due to case sensitivity or space discrepancies.

Further preprocessing includes converting 'Reported Date' and 'Completed Date' into a datetime format, which is crucial for conducting time-based analyses. The duration between these

dates is calculated to gauge the time taken for part replacement or maintenance, subsequently formatted into 'hh:mm' for uniformity and straightforward interpretation. This duration is then converted into decimal hours, enabling quantitative analysis. To ensure a focused analysis on typical cases, the dataset is filtered to remove the top 10% of outliers in quantities and durations based on their respective quantiles. Data types are adjusted for 'Employee ID', 'Qty', and 'SLA Response' columns to align with their content and to facilitate proper analysis. The 'Activity Item' column is renamed to 'Machine' for enhanced clarity and representation. Lastly, the year and month are extracted from the 'Reported Date' into separate columns, allowing for a nuanced examination of trends and patterns over time, enhancing the depth and quality of the subsequent analyses. Each step is meticulously designed to refine the dataset, making it suitable for developing predictive models. This preprocessing not only enhances the dataset's quality but also tailors it for specific analytical purposes, such as forecasting part demand.

C. The Spare Part Operational Demand Index (SPODI)

After the data is cleansed, the code proceeds with several steps to organize and extract meaningful insights from the dataset, which can be described as follows:

The DataFrame is first sorted by 'Part No', 'Supply Chain (SC)', and 'Completed Date' to systematically arrange the records. This sorting is crucial for tracking the usage and replacement patterns of the ATM parts. A new column, 'Next Damage Date', is then calculated by shifting the 'Completed Date' of the subsequent record within each group of 'Part No' and 'SC'. This step is instrumental in determining when a part might fail next, which is a key factor in predicting future part demands. The 'Time Until Damage' is introduced to measure the interval between part replacements and is then converted into a more practical unit, specifically days, resulting in the 'Time Until Damage Days' column. To encapsulate the average time until a part is damaged again, the mean 'Time Until Damage Days' for each part and supply chain combination is calculated, giving us an average that can be used for further analysis.

In preparation for the Spare Part Operational Demand Index (SPODI), the dataset is sorted again by 'Part No', 'Year', 'Month', and 'Reported Date'. This multilevel sorting is pivotal for detailed temporal analysis. The monthly and yearly usage of each part is aggregated to reflect the frequency of part demand over different timeframes. Average duration is computed to reflect the typical time parts spend in service before being replaced, and employee involvement is quantified to indicate which parts are most frequently handled by maintenance staff. The code also accounts for the influence of the supply chain, if applicable, by determining the size of the grouped dataset for each 'Part No' and 'SC'. This reflects the impact of supply Center dynamics on part replacement frequency. All these metrics are then merged into a single DataFrame to form the basis of the SPODI score, which is a composite index designed to indicate the demand for each part. Each component of the SPODI score—average duration, employee involvement, supply chain influence, and average time until damage—is weighted according to its perceived importance in determining part demand. The Spare Part Operational Demand Index (SPODI) is an innovative composite metric crafted to accurately forecast the

demand for ATM spare parts. The motivation behind developing the SPODI index is to refine the predictive maintenance strategy by integrating several predictive factors into a single, actionable figure. This index assists in optimizing stock levels in warehouses, ensuring that the supply chain operates smoothly, and that ATM downtime is minimized by having the necessary spare parts available when needed. By leveraging historical data and operational metrics, the SPODI index anticipates the frequency and urgency of part replacements, thus facilitating informed decision-making for maintenance scheduling and inventory management. To determine the relative importance of each parameter within the SPODI index, a correlation matrix is utilized. This matrix is a statistical tool that helps to identify the strength and direction of the linear relationships between the variables. By plotting a heatmap of these correlations, we gain visual insights into how each factor is associated with others, particularly their influence on spare part demand.

From this correlation matrix, the weighting factors for the SPODI index calculation can be derived. For example, if the 'Average Duration' has a high positive correlation with the actual part usage, it might be assigned a higher weight in the SPODI score formula. Conversely, parameters with lower or negative correlations might receive smaller weights. The final SPODI score for each part can be calculated using a formula like the following:

$$SPODI \text{ Score} = \omega_1 \times \text{Avg.duration} + \omega_2 \times \text{Employee Involvement} + \omega_3 \times \text{SC Influence} + \omega_4 \times \text{Avg Time Until Damage Eq. 1}$$

where ω_1 , ω_2 , ω_3 , and ω_4 are the weights derived from the correlation coefficients.

When analyzing the performed correlation matrix heatmap Figure 2, it is vital to interpret the values correctly to understand their implications for the SPODI score. Each cell in the heatmap represents the correlation coefficient between two variables, which can range from -1 to 1. A coefficient of 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no correlation.

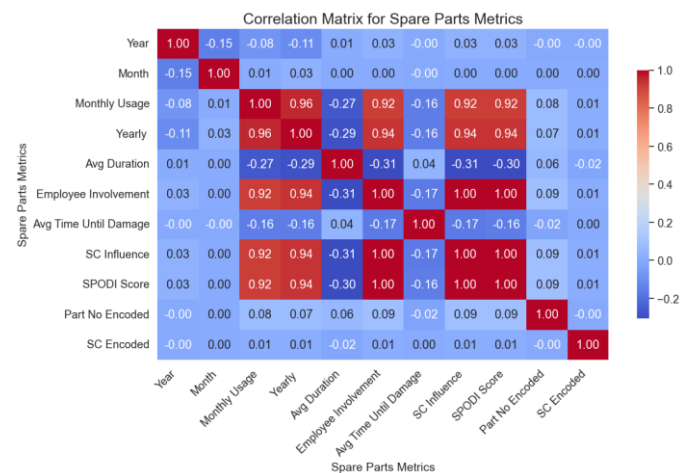


Figure 2 Correlation matrix of the spare parts metrics

Using the correlation coefficients provided from the correlation matrix heatmap, we note the following relationships:

Average Duration: A coefficient of -0.27 suggests a weak inverse relationship with monthly usage, implying that as the average duration increases, the monthly usage of the spare part slightly decreases.

Average Time Until Damage: Similarly, a coefficient of -0.16 indicates a weak inverse relationship with monthly usage.

Employee Involvement and SC Influence: Both have a strong positive correlation of 0.92 with monthly usage, suggesting that as employee involvement or SC influence increases, the usage of the part also increases significantly.

To calculate the weighting factors for the SPODI index, we will first normalize these coefficients by converting them to their absolute values and summing them up, which will form the denominator of our weight calculations. Each weight is then the absolute value of the individual correlation coefficient divided by this sum. Utilizing the correlation matrix values, the weighting factors for each parameter were calculated using the normalization equation. The resulting weights, which reflect the relative importance of each parameter in the Spare Part Operational Demand Index (SPODI), are as follows: Average Duration received a weight of approximately 0.119, indicating a lesser impact on the SPODI score due to its negative correlation. Average Time Until Damage, also negatively correlated, was assigned a weight of about 0.070. Employee Involvement and Supply Center (SC) Influence, both with strong positive correlations, emerged as the most impactful factors with equal weights of approximately 0.405 each. These weights serve as multipliers in the SPODI score formula (Eq. 1), quantitatively expressing the influence of each factor in predicting the monthly demand for spare parts.

D. Applying Machine Learning Algorithms

1. Linear Regression Model

To fit the data and evaluate the predictive model, the study employs a systematic approach using the Python library scikit-learn, which is renowned for its comprehensive machine learning capabilities. The process begins with the selection of variables for the model, where 'Part No Encoded', 'Year', 'Month', and 'SPODI Score' are chosen as independent variables (X) due to their potential impact on predicting 'Monthly Usage', the dependent variable (y). This choice is predicated on the hypothesis that these factors collectively influence the monthly demand for spare parts.

The dataset is then partitioned into training and testing sets using the 'train_test_split' function, with an allocation of 80% for training and 20% for testing, ensuring a representative sample for model training while reserving a subset for unbiased evaluation. The 'random_state' parameter is set to ensure reproducibility of results.

A Linear Regression model is applied and fitted with the training data, assuming a linear relationship between the selected features and the monthly usage of spare parts. Following the model training, predictions are made on the test set.

The model's performance is assessed through two primary metrics: the Mean Squared Error (MSE) and the R-squared (R^2) value. The MSE provides a measure of the average squared difference between the observed actual outcomes and the

predictions by the model, offering insight into the magnitude of the error. It can be calculated through the following equation:

$$MSE = \frac{1}{M} \sum_{m=1}^M (Y_m - \hat{Y}_m)^2 \quad \text{Eq. 2}$$

where \hat{Y}_m represents the forecasted value of CCI for a specific month during the test period; Y_m denotes the actual CCI value for that month within the test dataset; and M signifies the total count of observations within the test dataset. A lower MSE value for a predictive model indicates its high accuracy in forecasting. The R^2 statistic, on the other hand, offers a measure of the proportion of variance in the dependent variable that is predictable from the independent variables, serving as an indicator of the model's explanatory power.

2. Long Short-Term Memory (LSTM) networks

In advancing the model fitting process, a sophisticated approach involving Long Short-Term Memory (LSTM) networks (Graves, Long Short-Term Memory, 2012), part of the TensorFlow Keras framework, is employed. This method is particularly suited for predicting time-series data due to LSTM's ability to remember long-term dependencies. Initially, both the features and target variables are normalized using the MinMaxScaler to a range between 0 and 1, ensuring that the LSTM model receives data within a scale it can efficiently process.

The LSTM network requires the input data to be in a specific format of [samples, time steps, features]. For this implementation, each data point is treated as a sequence with one time step and multiple features. Following the scaling, the dataset is divided into training and test sets, maintaining an 80-20 split, with a random state set for reproducibility of results.

The Sequential model architecture is defined with an LSTM layer, containing 50 units to capture the temporal dependencies within the data, followed by a Dense layer for output prediction. The model is compiled with the Adam optimizer and mean squared error loss function, aiming to minimize the difference between predicted and actual values.

To mitigate the risk of overfitting, an Early Stopping callback is utilized. This halts the training process when the validation loss ceases to decrease, ensuring the model retains generalizability. The model undergoes training over 100 epochs with a batch size of 32, with performance validated against a portion of the training set. Post-training, predictions are made on the test set and rescaled to their original range for accurate evaluation. The model's effectiveness is quantified through the MSE and R^2 metrics, providing insight into its predictive accuracy and the variance explained by the model, respectively. This advanced fitting process exemplifies the application of deep learning techniques in time-series forecasting, showcasing their potential for high accuracy in predictive tasks.

E. Enhanced Time-Series Forecasting with Bidirectional LSTM and Dropout Regularization

To further enhance the predictive model, an augmented LSTM architecture (Graves, Fernández, & Schmidhub, 2005) is implemented, incorporating Bidirectional LSTM and Dropout layers, leveraging the TensorFlow Keras library. This advanced configuration is designed to capture both forward and backward dependencies in the data, offering a more nuanced understanding of temporal patterns.

The preprocessing phase involves scaling both the features (X) and the target variable (y) using MinMaxScaler to normalize the data within the range [0, 1]. This normalization facilitates the model's learning process by providing data within a consistent scale. The scaled features are reshaped to match the LSTM's expected input format of [samples, time steps, features], treating each data point as a sequence of one time step with multiple features.

The model is structured as a Sequential model, beginning with a Bidirectional LSTM layer with 50 units, enabling the model to learn from sequences in both forward and backward directions. This layer is followed by a Dropout layer with a rate of 0.2 to prevent overfitting by randomly omitting a fraction of the neurons during training. Subsequently, a second LSTM layer with 100 units captures complex dependencies without returning sequences, leading to a final Dropout layer and a Dense output layer for prediction. The model is compiled with the Adam optimizer (P. Kingma & Ba, 2014) and mean squared error loss function, targeting minimization of the difference between predicted and actual values. Training is guided by an Early Stopping callback set to monitor validation loss, with a patience of 20 epochs, to halt training when improvements cease, thereby ensuring the model does not overfit.

After training for up to 200 epochs with a batch size of 64, the model's predictions are inversely transformed to their original scale for accurate evaluation. The model's performance is also assessed using the MSE and R² metrics, measuring the average squared difference between the predicted and actual values, and the proportion of variance explained by the model, respectively.

III. RESULTS

The study embarked on evaluating three distinct modeling approaches to predict monthly usage of ATM machine parts, revealing varied performance levels as shown in Figure 1. Initially, a Linear Regression model was developed, yielding MSE of 193.755 and an R-squared of 0.85. This baseline model, despite its straightforward application, suggested room for improved accuracy.

Transitioning to LSTM networks, it showed a significant enhancement in forecasting accuracy, with a reduction in MSE to 98.07 and an increase in R-squared to 0.92 after 100 training epochs. This improvement highlighted the LSTM's capability to effectively learn from temporal data sequences, offering a more accurate prediction model.

The final model, Enhanced Time-Series Forecasting with Bidirectional LSTM and Dropout Regularization achieved the best performance in our study. After 200 epochs of training, it recorded an MSE of 71.24 and an R-squared of 0.94, demonstrating the highest forecasting accuracy among the evaluated models. The incorporation of bidirectional LSTM layers allowed for a comprehensive analysis of temporal dependencies, while dropout regularization helped prevent overfitting.

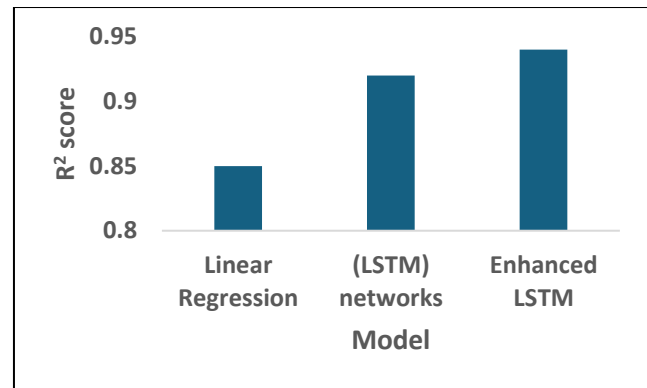


Figure 1 Comparison of the prediction capability of each model

A. Linear Regression Model Analysis

The linear regression model analysis began with the examination of the relationship between the actual and predicted monthly usage of ATM spare parts in Figure 2(a). The diagonal dashed line represents the line of perfect prediction. Observations clustered around this line indicate accurate predictions. The spread of data points along the diagonal shows that the model is generally effective in predicting lower and medium values of monthly usage. However, as the actual monthly usage increases, the predictions become more dispersed, suggesting that the model's accuracy decreases for higher values of spare part usage. This dispersion points to potential limitations of the Linear Regression model in capturing the complexity of higher usage patterns, potentially due to non-linear relationships not accounted for in the model.

The histogram in Figure 2(b) showcases the distribution of residuals — the differences between the predicted and actual values. The distribution appears to be roughly normal and centered around zero, which suggests that the model does not have a systematic bias; it is just as likely to overestimate as to underestimate the monthly usage. The presence of longer tails indicates some predictions with large errors, which could be due to outliers in the data or instances where the model's assumptions do not hold. This can be a concern for the maintenance service provider, as significant underestimation or overestimation of spare part demand could lead to either stockouts or excess inventory.

The feature importance graph in Figure 2(c) reveals the weight of each feature in the Linear Regression model. 'Part No Encoded' has the most significant negative influence on the model's predictions, suggesting that certain parts are likely associated with decreased usage. This could imply that the model recognizes specific parts that have a lower turnover rate, which could be indicative of their longer life span or lower failure rate. 'Year' also has a negative coefficient, indicating a trend of decreasing usage over time, possibly due to improvements in part durability or ATM technology. Surprisingly, 'SPODI Score' has a relatively small positive coefficient, which contrasts with expectations given its intended role as a comprehensive indicator of demand.

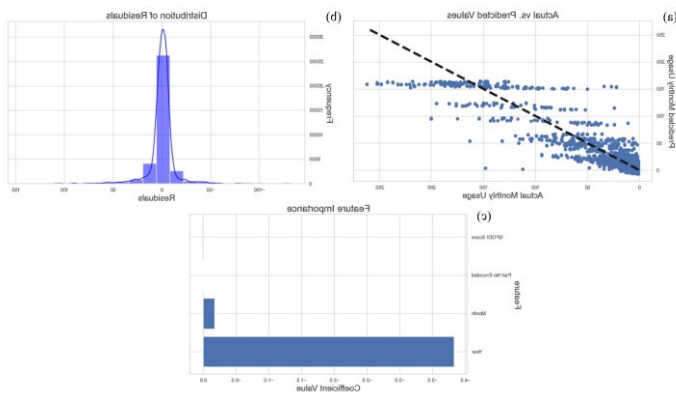


Figure 2 Summary of linear regression analysis

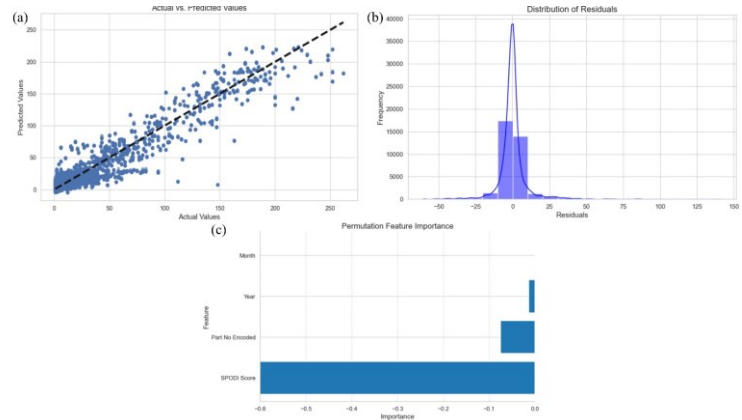


Figure 3 Summary of LSTM results

B. Long Short-Term Memory (LSTM) network

The implementation of the LSTM network for forecasting the monthly usage of ATM machine parts marks a significant advancement over the linear regression model. The LSTM's architecture is inherently suited for the time-series nature of the dataset, given its proficiency in capturing long-term temporal dependencies, a critical feature for accurately predicting spare parts demand.

Upon examination of the LSTM model results, the actual vs. predicted values scatter plot presented in Figure 3(a) indicates a strong alignment along the line of perfect prediction. This alignment suggests that the LSTM model can closely approximate the actual demand, particularly notable for higher values where linear regression had shortcomings. The model's capacity to handle nonlinear relationships and temporal sequences evidently contributes to its enhanced accuracy.

The residuals distribution chart shown in Figure 3(b) exhibits a narrow, concentrated peak, which implies that the LSTM predictions are typically close to the actual values with fewer and less significant outliers than those observed in the linear regression model. This distribution indicates a better fit to the data and suggests a higher consistency in the model's predictive capability.

Feature importance derived from the LSTM model via permutation (Figure 3(c)) highlights the dominant role of the 'SPODI Score' in predicting demand. This finding corroborates the premise that integrating multiple operational metrics into a composite score effectively captures the complexities of spare parts usage. The dominance of the 'SPODI Score' validates the effort invested in creating this index as a cornerstone for the forecasting model.

In conclusion, the LSTM network provides a robust framework for forecasting spare parts demand. Its strengths in dealing with time-series data, accounting for non-linear patterns, and leveraging the constructed 'SPODI Score' demonstrate a significant leap forward in predictive accuracy. This positions LSTM networks as a powerful tool for maintenance service providers, enabling more efficient inventory management and reducing the risk of service disruptions due to spare parts shortages.

C. Enhanced Time-Series Forecasting with Bidirectional LSTM and Dropout Regularization

In the pursuit of a sophisticated forecasting framework for ATM spare parts, the Enhanced Time-Series Forecasting model using Bidirectional LSTM and Dropout Regularization has demonstrated commendable accuracy and robustness, as evidenced by Figure 4. The scatter plot (Figure 4(a)) presents a dense concentration of points closely hugging the dashed diagonal line, which signifies the line of perfect prediction. This illustrates that the Enhanced LSTM model has a high degree of precision in forecasting the monthly demand for spare parts. The model excels at predicting both lower and higher ranges of actual monthly usage, with a particularly impressive capture of trends in the upper usage spectrum, where linear models might falter. The model's ability to grapple with complex, non-linear patterns in the time series data is a testament to the efficacy of the Bidirectional LSTM architecture.

The residuals histogram (Figure 4(b)) indicates a normal distribution with a sharp peak and minimal skew, centered around zero. The symmetry of this distribution around the origin suggests that the model predictions are unbiased and errors are evenly distributed on either side of the perfect prediction. The presence of a very narrow peak signifies that the majority of predictions are extremely close to the actual values, with few instances of large prediction errors.

In the feature importance bar graph (Figure 4(c)), the 'SPODI Score' emerges with a dominant negative weight, which may initially appear counterintuitive. However, considering the nature of permutation importance, this indicates that the 'SPODI Score' has a substantial impact on the predictive power of the model; when its values are permuted, the accuracy of the model's predictions changes significantly. This highlights the pivotal role of the 'SPODI Score' in driving the forecast accuracy for spare part demand and agrees with the findings of applying LSTM networks. Conversely, the 'Part No Encoded' feature exhibits the least importance, indicating that the model does not rely heavily on this feature for its predictions.

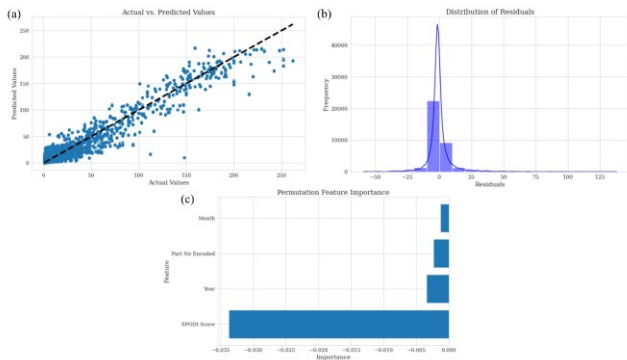


Figure 4 Enhanced Time-Series Forecasting with Bidirectional LSTM analysis summary

D. Enhancing the predictive ability of the models

The Linear Regression model laid the groundwork, offering a straightforward approach that proved to be generally effective for lower to medium ranges of monthly usage. However, the model's predictive ability waned as usage increased, a limitation highlighted by the dispersion of data points at higher usage levels. This deviation signaled the model's struggle with complex patterns that were not linearly correlated, which are essential to grasp for maintaining service levels in the face of variable demand.

The transition to the LSTM network marked a significant step forward. Unlike its predecessor, the LSTM network could account for long-term temporal dependencies, a feature quintessential to the time-series nature of the demand data. The actual vs. predicted values from the LSTM model exhibited a tighter clustering around the line of perfect prediction across the full range of usage, including the higher values where the Linear Regression model faltered. The residuals distribution revealed a marked improvement, showcasing a narrowed concentration indicative of a better fit and higher consistency.

The pinnacle of the forecasting models evaluated was the Enhanced Time-Series model with Bidirectional LSTM and Dropout Regularization. This model not only maintained the LSTM's proficiency in understanding time-related dependencies but also introduced bidirectional data processing to capture patterns across both past and future sequences. This comprehensive temporal understanding, coupled with dropout to avoid overfitting, resulted in the most accurate model. It demonstrated an impressive precision in forecasting, robust even at the extremes of the usage spectrum, and outperformed the other models in terms of residuals distribution and adherence to the perfect prediction line.

Across the models, the importance of features varied. Notably, the 'SPODI Score' emerged with significant influence in the advanced models, validating its design as a comprehensive indicator of demand. It was in the LSTM-based models that the 'SPODI Score' realized its potential, indicating that the sophistication of these models could unlock the full predictive power of the composite score.

IV. CONCLUSION

The comparison of the traditional Linear Regression approach with the advanced LSTM and Enhanced Time-Series Forecasting models underscores a transformative shift towards more dynamic, data-driven forecasting methods. The LSTM network, with its ability to navigate time-series data and unearth non-linear patterns, provides a more refined and accurate prediction model. This is further improved upon by the Enhanced LSTM framework, which employs Bidirectional LSTM and Dropout Regularization, delivering a high-precision tool capable of adeptly managing the complexity inherent in spare part demand forecasting.

The culmination of these advancements positions the LSTM networks, particularly the Enhanced Time-Series Forecasting model, as a significant tool for maintenance service providers. By achieving higher forecasting accuracy and more efficient inventory management, these providers are better equipped to avoid service interruptions and the associated costs. This research contributes noticeable insights to the field of predictive maintenance, facilitating a move away from reliance on historical averages and expert judgment towards embracing the sophistication of machine learning algorithms.

Overall, the study underscores the importance of selecting an appropriate forecasting model tailored to the specific complexity of the demand patterns for spare parts. As the field progresses, the continued refinement of predictive frameworks such as the SPODI index and the exploration of advanced neural network architectures are anticipated to further enhance the robustness and reliability of demand forecasting in the maintenance service sector.

V. REFERENCES

- [1] GUTSTADT, L. (1965, January). A decision aid for spare parts provisioning. In Support for Manned Flight Conference (p. 269).
- [2] Cohen, M. A., Kleindorfer, P. R., & Lee, H. L. (1989). Near-optimal service constrained stocking policies for spare parts. *Operations Research*, 37(1), 104-117.
- [3] Cohen, M. A., Kleindorfer, P. R., Lee, H. L., & Pyke, D. F. (1992). Multi-item service constrained (s, S) policies for spare parts logistics systems. *Naval Research Logistics (NRL)*, 39(4), 561-577.
- [4] Moore Jr, J. R. (1971). Forecasting and scheduling for past-model replacement parts. *Management Science*, 18(4-part-1), B-200.
- [5] Dekker, R., Kleijn, M. J., & De Rooij, P. J. (1998). A spare parts stocking policy based on equipment criticality. *International Journal of production economics*, 56, 69-77.
- [6] Braglia, M., Grassi, A., & Montanari, R. (2004). Multi-attribute classification method for spare parts inventory management. *Journal of quality in maintenance engineering*.
- [7] Kamath, K. R., & Pakkala, T. P. M. (2002). A Bayesian approach to a dynamic inventory model under an unknown demand distribution. *Computers & Operations Research*, 29(4), 403-422.
- [8] Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Journal of the Operational Research Society*, 23(3), 289-303.
- [9] Eaves, A. H., & Kingsman, B. G. (2004). Forecasting for the ordering and stock-holding of spare parts. *Journal of the Operational Research Society*, 55(4), 431-437.
- [10] Muckstadt, J. A., & Thomas, L. J. (1980). Are multi-echelon inventory methods worth implementing in systems with low-demand-rate items?. *Management Science*, 26(5), 483-494.
- [11] Hausman, W. H., & Erkip, N. K. (1994). Multi-echelon vs. single-echelon inventory control policies for low-demand items. *Management science*, 40(5), 597-602.
- [12] Snerbrooke, C. C. (1968). Metric: A Multi-Echelon Technique for Recoverable Item Control. *Operations Research*, 16, 122-141.

- [13] Muckstadt, J. A. (1973). A model for a multi-item, multi-echelon, multi-indenture inventory system. *Management science*, 20(4-part-i), 472-481.
- [14] Snerbrooke, C. C. (1986). VARI-METRIC: Improved approximations for multi-indenture, multi-echelon availability models. *Operations Research*, 34(2), 311-319.
- [15] Sleptchenko, A., van der Heijden, M. C., & van Harten, A. (2002). Effects of finite repair capacity in multi-echelon, multi-indenture service part supply systems. *International Journal of Production Economics*, 79(3), 209-230.
- [16] Pillai, R. S., Rathi, A. K., & Cohen, S. L. (1998). A restricted branch-and-bound approach for generating maximum bandwidth signal timing plans for traffic networks. *Transportation Research Part B: Methodological*, 32(8), 517-529.
- [17] Hopp, W. J., Zhang, R. Q., & Spearman, M. L. (1999). An easily implementable hierarchical heuristic for a two-echelon spare parts distribution system. *IIE transactions*, 31(10), 977-988.
- [18] Caglar, D., Li, C. L., & Simchi-Levi, D. (2004). Two-echelon spare parts inventory system subject to a service constraint. *IIE transactions*, 36(7), 655-666.
- [19] Zhang, S., Huang, K., & Yuan, Y. (2021). Spare parts inventory management: A literature review. *Sustainability*, 13(5), 2460.
- [20] Bridgman, M. S., & Mount-Campbell, C. A. (1993). Determining the number of spares in an inventory/repair system which supports equipment with scheduled usage. *International journal of production economics*, 30, 501-518.
- [21] Walker, J. (1996). A graphical aid for the initial purchase of 'insurance type' spares. *Journal of the Operational Research society*, 47(10), 1296-1300.
- [22] Cobbaert, K., & Van Oudheusden, D. (1996). Inventory models for fast moving spare parts subject to "sudden death" obsolescence. *International Journal of Production Economics*, 44(3), 239-248.
- [23] Kennedy, W. J., Patterson, J. W., & Fredendall, L. D. (2002). An overview of recent literature on spare parts inventories. *International Journal of production economics*, 76(2), 201-215.
- [24] Käki, A. (2007). Forecasting in End-Of-Life Spare Parts Procurement. Master's Thesis—Helsinki University of Technology System Analysis Laboratory.
- [25] Zhu, S. (2021). Spare Parts Demand Forecasting and Inventory Management: Contributions to Intermittent Demand Forecasting, Installed Base Information and Shutdown Maintenance (No. 538).
- [26] Hemeimat, R., Al-Qatawneh, L., Arafeh, M., & Masoud, S. (2016). Forecasting spare parts demand using statistical analysis. *American Journal of Operations Research*, 6(02), 113.
- [27] Van der Auweraer, S., & Boute, R. (2019). Forecasting spare part demand using service maintenance information. *International journal of production economics*, 213, 138-149.
- [28] Vargas, C. G., & Cortés, M. E. (2017). Automobile spare-parts forecasting: A comparative study of time series methods. *International Journal of Automotive and Mechanical Engineering*, 14, 3898-3912.
- [29] Lelo, N. A., Heyns, P. S., & Wannenburg, J. (2019). Forecasting spare parts demand using condition monitoring information. *Journal of Quality in Maintenance Engineering*, 26(1), 53-68.
- [30] Chen, T. Y., Lin, W. T., & Sheu, C. (2018). A dynamic failure rate forecasting model for service parts inventory. *Sustainability*, 10(7), 2408.
- [31] Choi, B., & Suh, J. H. (2020). Forecasting spare parts demand of military aircraft: Comparisons of data mining techniques and managerial features from the case of south korea. *Sustainability*, 12(15), 6045.
- [32] Arts, J. J. (2013). Spare parts planning and control for maintenance operations. Doctor, Eindhoven University of Technology, Eindhoven.(D175).
- [33] DEĞİRMENCİOĞLU, M. (2012). Applications of spare parts management in an international tv manufacturing company (Doctoral dissertation, DEÜ Fen Bilimleri Enstitüsü).
- [34] Cui, F. (2011). ARIMA models for bank failures: Prediction and comparison.
- [35] Nair, U. (2017). Reliability assessment of large wind farms using ARIMA modeling and availability enhancement by reliability allocation (Doctoral dissertation, Cochin University of Science and Technology).