Equivalent: Jurnal Ilmiah Sosial Teknik
Volume 8, Issue 4, 1885-1895
e_ISSN: 2775-0833
https://jurnalequivalent.id/index.php/jequi/index
DOI: doi.org/10.59261/jequi.v8i4.437

Comparative Analysis of Univariate and Multivariate Time Series Forecasting Models for Hard Token Inventory Planning in Banking

Risa Aprilia*
Universitas Bina Nusantara, Indonesia

Tuga Mauritsius
Universitas Bina Nusantara, Indonesia

***Corresponding author:**
Risa Aprilia, Universitas Bina Nusantara, Indonesia. ✉risa.aprilia@binus.ac.id

Article Info:

***Article history*:**
Received: August 18, 2026
Revised: September 18, 2026
Accepted: October 05, 2026
Available Online: October 22, 2026

***Keywords*:**
Demand Forecasting;
Hard Token Inventory;
Banking;
Univariate Forecasting;
Multivariate Forecasting;
Temporal Fusion Transformer;

This is an open access article under the CC BY-SA license.

Copyright © 2026 by Author. Published by Politeknik Siber Cerdika International

Abstract

Background: Hard token availability is essential for secure transaction approval and service continuity in banking, yet quota-based stock planning may lead to stockouts, overstocking, and operational inefficiencies. This study compares univariate and multivariate time series forecasting models for branch-level hard token inventory planning.

Methods: This study aims to develop a forecasting model for branch-level hard Using the CRISP-DM framework, historical operational data from 197 branches were processed into a forecasting-ready dataset. ARIMA and Prophet were developed as univariate models, while Long Short-Term Memory (LSTM) and Temporal Fusion Transformer (TFT) incorporated operational, economic, and engineered temporal features. Model performance was evaluated using MAE, RMSE, MAPE, and R^2.

Results: TFT achieved the lowest MAE (5.997) and MAPE (11.56%), as well as the highest R^2 (0.7403), indicating the strongest overall performance across three of the four evaluation metrics. Prophet recorded the lowest RMSE (10.721), while ARIMA showed the weakest overall performance.

Conclusion: The results identify TFT as the preferred model for hard token demand forecasting, based on its lower absolute and percentage prediction errors and highest R^2 among the evaluated models. Although Prophet achieved a lower RMSE, TFT demonstrated stronger overall performance across the evaluation metrics. Integrating TFT-based forecasts into DigiMS can support more data-driven stock allocation, replenishment planning, and early identification of potential stock shortages and excess inventory at bank branches.

To cite this article: Aprilia, R., & Mauritsius, T. (2026). Comparative analysis of univariate and multivariate time series forecasting models for hard token inventory planning in banking. *Equivalent: Jurnal Ilmiah Sosial Teknik*, *8*(5), 1885-1895. https://doi.org/10.59261/jequi.v8i4.437

INTRODUCTION

The rapid growth of banking has significantly transformed banking services by improving transaction speed, accessibility, and operational efficiency (Bueno et al., 2024; Hakizimana et al., 2023). Along with this transformation, banking institutions are increasingly required to maintain service continuity, transaction security, and operational reliability, particularly for business and corporate customers conducting financial transactions. In digital cash management services, hard tokens remain critical operational infrastructure, directly mitigating compliance and regulatory risks under central bank mandates (Bharath et al., 2024; Kiettikunwong & Sangsarapun, 2025). These devices function as two-factor authentication mechanisms as they function as two-factor authentication devices for transaction approval and release processes.

The availability of hard tokens has become an important operational issue because it directly affects customer onboarding, transaction activation, and the continuity of banking services. An insufficient number of hard tokens may delay service activation, interrupt customer

transactions, and reduce service quality, whereas excessive inventory may increase procurement, storage, and distribution costs. Therefore, maintaining an appropriate inventory level requires balancing procurement and holding costs against business disruption penalties; inadequate inventory incurs service activation delays and reputational costs, while excess inventory increases capital expenditure and storage costs. for banking institutions. The relationship between demand forecasting and inventory planning is also important because inaccurate demand estimates can affect subsequent replenishment and inventory control decisions (Goltsos et al., 2022).

Based on historical operational data collected from 197 branch offices, this study utilizes 128,694 daily observations of hard token demand. The exploratory analysis indicates that 27,370 observations (21.3%) experienced unmet demand (Lost Demand), implying that customer demand could not always be fulfilled by the available inventory. In addition, 149 observations recorded zero remaining stock (Stock End Original = 0), indicating that stock depletion occurred during the observation period. These findings demonstrate that hard token demand varies considerably across branches and over time, making inventory planning based solely on historical averages insufficient to anticipate future demand accurately.

Currently, hard token replenishment planning is primarily based on historical usage and predefined stock allocation policies. Although this approach provides a practical baseline, it may not adequately capture dynamic demand patterns influenced by branch operational activities, customer growth, transaction intensity, and regional economic conditions. Forecasting-based inventory planning provides a more proactive approach by estimating future demand before replenishment decisions are made, thereby strengthening the connection between demand forecasting and inventory control (Goltsos et al., 2022). Previous demand forecasting studies have also demonstrated the application of statistical and machine learning approaches to support inventory and operational planning (Kalaya et al., 2023; Nguyen et al., 2024).

Several forecasting approaches have been proposed in previous studies. Statistical models such as Autoregressive Integrated Moving Average (ARIMA) have been widely applied to time series forecasting because of their ability to model temporal dependencies using historical observations. Noor et al. (2013), including BINUS researcher Astari Retnowardhani, applied ARIMA for crime forecasting and demonstrated its use as a forecasting method to support decision-making. More recent studies have also examined ARIMA for material and demand forecasting, including forecasting teaching material requirements using historical data (Wicaksono & Mauritsius, 2025). In addition, Prophet has been applied to demand forecasting and inventory-related problems, particularly when trend and seasonal patterns need to be modeled (Doğan, 2025).

Although statistical models can provide effective forecasting based on historical patterns, demand may also be affected by multiple operational and external variables. Multivariate forecasting approaches can incorporate these additional predictors to capture relationships that cannot be represented using historical demand alone. Deep learning models such as Long Short-Term Memory (LSTM) have been widely investigated for time series forecasting and inventory management because of their ability to learn nonlinear temporal dependencies (Kim et al., 2022). Mauritsius (2025) further demonstrated the application of LSTM and other deep learning architectures for forecasting using multiple predictor variables, highlighting the potential of deep learning approaches for capturing complex relationships in time-dependent data.

Another advanced deep learning approach is the Temporal Fusion Transformer (TFT), which was specifically designed for multi-horizon time series forecasting involving multiple types of input variables. TFT combines recurrent processing, self-attention mechanisms, feature selection, and gating mechanisms to capture temporal relationships at different scales while providing greater interpretability of the forecasting process (Lim et al., 2021). Its ability to incorporate multiple covariates makes TFT particularly relevant to forecasting problems in which demand is influenced by operational and external factors.

However, most existing forecasting studies focus on demand forecasting in retail, manufacturing, energy, and supply chain environments, while the application of forecasting models to operational inventory management in banking, particularly for authentication devices such as hard tokens, remains relatively limited. Existing studies also tend to evaluate statistical

and deep learning approaches separately or within specific application domains. Consequently, there is still a need to comparatively evaluate univariate statistical forecasting and multivariate deep learning approaches within the same banking operational context.

Furthermore, this study integrates operational and regional economic variables into branch-level hard token demand forecasting. The multivariate framework incorporates Customer Information File (CIF) Registration, transaction frequency, transaction volume, and Gross Regional Domestic Product (GRDP/PDRB), together with engineered temporal features. This approach is intended to provide a more comprehensive representation of the operational and temporal factors associated with hard token demand.

Therefore, this study aims to develop and compare four forecasting models consisting of Autoregressive Integrated Moving Average (ARIMA), Facebook Prophet, Long Short-Term Memory (LSTM), and Temporal Fusion Transformer (TFT) to predict branch-level hard token demand. The study follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, which provides a structured process consisting of business understanding, data understanding, data preparation, modeling, evaluation, and deployment (Schröer et al., 2021).

The contributions of this study are threefold. First, it provides a comparative evaluation of statistical and deep learning forecasting models within the context of banking inventory management. Second, it proposes a multivariate forecasting framework that integrates historical demand, operational indicators, regional economic variables, and engineered temporal features to improve forecasting performance. Third, the selected forecasting approach is proposed as a decision-support mechanism for proactive hard token inventory planning, enabling banking operations to reduce stockout and overstock risks while improving operational efficiency and service continuity.

METHOD

This study adopted the Cross-Industry Standard Process for Data Mining (CRISP-DM) as the primary methodological framework to transform an operational banking problem into an analytical forecasting solution. CRISP-DM provides a structured and iterative process consisting of six major phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment (Schröer et al., 2021). The framework has been widely applied to structure data mining and analytics projects because it provides a systematic connection between business objectives and analytical modeling activities (Acuña-Cid et al., 2025).

The main objective of this research is to improve the forecasting accuracy of hard token demand at the branch level to support adaptive stock planning and distribution decisions in banking operations. To achieve this objective, the study applied a comparative forecasting approach consisting of univariate models, namely Autoregressive Integrated Moving Average (ARIMA) and Prophet, and multivariate deep learning models, namely Long Short-Term Memory (LSTM) and Temporal Fusion Transformer (TFT). Comparative evaluation between statistical and machine learning approaches is commonly used to identify the forecasting method that is most appropriate for a particular time series characteristic (Kontopoulou et al., 2023).

The univariate approach utilized historical demand data as the primary forecasting input, whereas the multivariate approach integrates additional internal operational and external economic variables to capture more complex temporal dependencies and relationships. The use of additional explanatory variables is particularly relevant for demand forecasting problems in which demand may be influenced by factors beyond its own historical values (Liu & Vakharia, 2024; Nguyen et al., 2024).

Data Transformation Workflow

The forecasting dataset was developed through a structured transformation pipeline to ensure data consistency and model compatibility.

The initial dataset consisted of branch-level hard token monitoring data containing stock, usage, and demand-related attributes. The target variable used in forecasting was Real_Demand, representing actual demand after incorporating both fulfilled demand and simulated unmet demand.

The formulation is expressed as:

RealDemand= UsageRecorded + LostDemand

The Lost_Demand variable represents hard token demand that could not be fulfilled due to inventory constraints and was obtained from manual records maintained by branch personnel. Real_Demand was calculated as the sum of Usage_Recorded and Lost_Demand, providing a more comprehensive representation of operational demand than recorded usage alone. Data consistency was assessed through calculation verification, although the accuracy of unmet demand depends on the reliability of branch-level records.

Data preprocessing included duplicate removal, categorical standardization, chronological sorting, and branch-level filtering. Central warehouse records were excluded because they do not represent operational branch demand. Temporal transformation was subsequently performed by extracting and encoding relevant time attributes to support time-series modeling. The YearMonth attribute was constructed from the year and month fields to facilitate temporal analysis and data management.

YearMonth= f (Year,Month)

Feature Engineering and Multivariate Integration

To enhance forecasting capability beyond historical demand patterns, multivariate feature engineering was conducted by integrating internal operational and external explanatory variables. Feature engineering is an important stage in time series forecasting because appropriately constructed temporal and explanatory features can provide additional information for learning demand patterns (Karmaker, 2025).

The external variable included in this study was *Produk Domestik Regional Bruto* (PDRB) as a proxy for regional economic conditions. Meanwhile, the internal operational variables included CIF_REGISTER, FREKUENSI_TRX_MTD, and VOLUME_TRX_MTD, representing customer activity and transactional intensity. The incorporation of operational and external variables is consistent with previous demand forecasting studies that use additional predictors to improve the representation of demand-driving factors (Nguyen et al., 2024).

In addition, temporal and statistical features were engineered to improve predictive representation. Lag-based features were created to capture short-term historical dependency:

$$Lag_1 = Demand_{t-1}$$

$$Lag_3 = Demand_{t-3}$$

Lag features provide historical context to forecasting models by allowing the model to incorporate previous demand observations as predictors of future demand (Karmaker, 2025).

Rolling mean transformation was used to smooth short-term fluctuations:

$$RollingMean_{3t} = \frac{Demand_{t-1} + Demand_{t-2} + Demand_{t-3}}{3}$$

Rolling statistics can provide a smoothed representation of recent demand patterns and are commonly used as engineered predictors in time series forecasting (Karmaker, 2025).

To represent cyclical seasonality, month values were encoded using sinusoidal transformation:

$$Month_{sin} = sin + \frac{2\pi \,.Month}{12}$$

$$Month_{cos} = cos + \frac{2\pi \,.Month}{12}$$

Sinusoidal temporal encoding allows cyclical characteristics such as monthly seasonality to be represented continuously rather than as independent categorical values (Bansal et al., 2025). These engineered features allowed the multivariate models to capture seasonality, economic influence, and branch-level operational dynamics more effectively

Experimental Setup and Data Splitting

The forecasting experiments used historical observations spanning January 2024 to December 2025, covering 24 months of data. For both univariate and multivariate forecasting, the observations were arranged chronologically, and data splitting preserved their temporal order to reflect the forecasting of future demand based on historical observations and to prevent information leakage from future periods into the training process.

A chronological 80:20 training–testing split was applied to the forecasting process, with the first 80% of observations allocated to training and the remaining 20% to testing. For ARIMA and Prophet, the split was performed separately for each branch. LSTM used the same chronological split, with an additional validation subset for early stopping and monitoring model performance. For TFT, the training period was determined using a temporal cutoff based on the time index.

All models used a daily forecasting frequency. However, their forecasting horizons differed. ARIMA and Prophet generated multi-step daily forecasts covering the respective testing periods, whereas LSTM generated one-step-ahead predictions, forecasting demand for the next day based on a historical lookback window of 30 time steps. TFT was also configured for one-step-ahead daily forecasting, using a maximum encoder length of 30 time steps and a maximum prediction length of one time step.

Data normalization was performed according to the requirements of each deep learning model. For LSTM, MinMaxScaler was applied to the input features and target variable after the training and testing datasets were separated. The scaler was fitted exclusively on the training data, and the learned parameters were subsequently applied to the testing data using the transformation function. This procedure was adopted to prevent information leakage from the test set during normalization (Bhanja & Das, 2018).

For TFT, GroupNormalizer was used to normalize the target variable according to branch identity, accounting for differences in demand scales across branches. Unlike the LSTM preprocessing pipeline, TFT normalization was implemented through the TimeSeriesDataSet configuration. The model-specific normalization procedures were implemented as part of their respective forecasting pipelines.

Forecasting Model Development

This study compares four forecasting models representing statistical and deep learning approaches: ARIMA, Prophet, LSTM, and TFT. The comparison between conventional statistical forecasting and machine learning or deep learning approaches is relevant because different models may capture different types of temporal patterns and nonlinear relationships (Kontopoulou et al., 2023). The univariate forecasting models consisted of ARIMA and Prophet.

ARIMA was selected because it is a widely established statistical approach for time series forecasting and can model temporal dependencies through autoregressive, differencing, and moving-average components (Kontopoulou et al., 2023).

$$\text{ARIMA}\ (p, d, q)$$

where p denotes autoregressive lag, d differencing order, and q moving average order. ARIMA has also been applied to operational forecasting problems, including material requirement forecasting, demonstrating its applicability to demand-related planning problems (Wicaksono & Mauritsius, 2025).

In this study, model selection resulted in ARIMA(5,1,1)(0,0,0)[0] as the best-performing ARIMA specification based on the model selection procedure.

Prophet was applied as a univariate forecasting model based on additive decomposition of trend, seasonality, holiday effects, and error components. Prophet has been applied in demand

forecasting and inventory-related applications, particularly where trend and seasonal patterns are important forecasting components (Doğan, 2025).

Prophet was applied as an additive decomposition forecasting model:

$$y(t)= g(t) + s(t) + h(t) + \epsilon t$$

where g(t) represents trend, s(t) seasonality, h(t) holiday effects, and ϵt residual error.

In this study, Prophet was trained using historical daily hard token demand to provide a univariate benchmark against the ARIMA model. Forecasts were generated over the corresponding testing horizon for each branch.

For multivariate forecasting, LSTM was used to capture nonlinear temporal dependencies across multiple variables. LSTM is a recurrent neural network designed for sequential data and has been widely applied in time series forecasting (Kim et al., 2022). It was chosen because deep learning models can better capture complex nonlinear patterns compared to traditional statistical methods Kontopoulou et al. (2023), and has been previously applied in similar forecasting problems (Mauritsius, 2025).

In this study, the LSTM model used a single layer with 64 neurons, ReLU activation, 0.2 dropout, and a dense output layer. A historical lookback window of 30 time steps was used for sequence generation, allowing the LSTM model to learn temporal dependencies from recent demand and explanatory variables. The model used operational, economic, and temporal features as inputs to generate one-step-ahead demand predictions., and hyperparameter tuning was performed. The model used operational, economic, and temporal features as inputs.

The Temporal Fusion Transformer (TFT) was used as an advanced multivariate model for interpretable multi-horizon forecasting. TFT combines recurrent networks, attention mechanisms, and variable selection to model complex temporal relationships (Lim et al., 2021). It is suitable for forecasting problems involving multiple influencing factors such as historical demand and external variables (Lim et al., 2021).

In this study, TFT was trained using the same multivariate feature set, including demand history, operational indicators, economic variables, and engineered temporal features. This allows the model to capture nonlinear interactions and long-term dependencies in branch-level demand.

Model Evaluation

Model performance was evaluated using several forecasting accuracy metrics to assess prediction reliability and compare the effectiveness of univariate and multivariate forecasting models. The evaluation focused on Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2).

MAE was used to measure the average absolute difference between predicted and actual demand values, providing a straightforward interpretation of forecasting error magnitude. RMSE was applied to emphasize larger prediction errors because this metric is more sensitive to extreme deviations, making it suitable for identifying models that struggle under high-demand fluctuations. In addition, MAPE was used to evaluate percentage-based forecasting accuracy, allowing performance comparison across observations with varying demand scales. Finally, R^2 was included to assess how well each model explains the variance of actual demand patterns.

These evaluation metrics were selected to provide a comprehensive comparison between statistical forecasting models (ARIMA and Prophet) and deep learning models (LSTM and Temporal Fusion Transformer/TFT). A model with lower MAE, RMSE, and MAPE values indicates stronger predictive accuracy, while a higher R^2 value suggests better capability in explaining demand variability. Through this evaluation approach, the study identifies the most reliable forecasting model for supporting branch-level hard token inventory planning and proactive stock management within DigiMS.

Deployment Strategy

The best-performing forecasting model was proposed for deployment within DigiMS (Digital Management System) as a forecasting-driven decision support mechanism.

The integration is intended to support branch-level demand prediction, proactive quota adjustment, early replenishment alerts, and inventory risk monitoring. Through this deployment, DigiMS is expected to evolve from a reactive operational monitoring system into a predictive inventory planning platform

RESULTS AND DISCUSSION

Results

Data Transformation and Feature Engineering Results

The initial phase of the experiment focused on transforming raw branch-level hard token demand data into a forecasting-ready dataset. The original dataset consisted of historical stock, usage, and operational records distributed across multiple branches. To improve temporal consistency, the Month variable was converted into numerical representation and combined with the Year attribute to construct YearMonth, which served as the primary time index for forecasting analysis.

Following temporal transformation, demand data were aggregated at monthly branch level to ensure consistency across forecasting models. The target variable used in this study was Real_Demand, which was derived by combining recorded usage and simulated unmet demand. This formulation provided a more realistic representation of actual branch-level token requirements.

For multivariate forecasting, feature engineering was performed to enrich predictive information beyond historical demand patterns. Several lag-based variables such as lag_1 and lag_3 were generated to capture short-term temporal dependency. Rolling statistical features such as rolling_mean_3 were introduced to smooth local fluctuations and represent short-term demand trends. In addition, cyclical seasonality was encoded using month_sin and month_cos transformations to preserve periodic temporal behavior.

To further strengthen multivariate representation, the dataset was integrated with both internal operational variables and external economic indicators. Internal variables included CIF_REGISTER, FREKUENSI_TRX_MTD, and VOLUME_TRX_MTD, while external variables included PDRB as a proxy for regional economic activity. These engineered features enabled the deep learning models to capture broader demand drivers beyond historical usage.

The final multivariate dataset provided richer explanatory structure and supported more advanced temporal learning compared to the univariate forecasting dataset.

Univariate Forecasting Performance

The first forecasting experiment was conducted using univariate time series models, namely ARIMA and Prophet, where prediction relied solely on historical demand patterns.

ARIMA showed strong forecasting consistency in modeling branch-level demand. The model was effective in capturing short-term linear temporal relationships and autoregressive demand continuity. Similarly, Prophet demonstrated stable performance through its additive decomposition approach, which successfully modeled trend and seasonality components. However, because both models relied only on historical demand data, their ability to adapt to external operational fluctuations remained limited.

The results indicate that univariate models can serve as strong baseline predictors, particularly for stable or moderately seasonal inventory demand.

Multivariate Forecasting Performance

To improve forecasting capability, multivariate deep learning models were implemented using the enriched feature set. The models included Long Short-Term Memory (LSTM) and Temporal Fusion Transformer (TFT).

The LSTM model showed relatively weak forecasting performance. Although LSTM is capable of capturing sequential dependencies, its simpler recurrent architecture may have struggled to model highly heterogeneous branch behavior, abrupt demand changes, and nonlinear relationships among operational variables.

In contrast, TFT demonstrated significantly stronger forecasting capability. Through the integration of attention mechanisms, variable selection networks, temporal encoding, and gating

structures, TFT was able to dynamically prioritize relevant features and model both short-term and long-term dependencies. This resulted in better adaptability to fluctuating branch-level hard token demand.

The findings indicate that multivariate learning becomes increasingly beneficial when demand is influenced by operational activity, economic conditions, and nonlinear temporal interaction.

Comparative Performance Evaluation

To evaluate the forecasting effectiveness, both univariate and multivariate models were compared using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2). The comparative results are presented in Table 1.

Table 1
Comparative Performance Metrics of Machine Learning Models

Model	MAE	RMSE	MAPE	R^2
ARIMA	18.84	26.82	19.41%	- 0.0102
Prophet	6.288498	10.7213	12.925%	0.173
LSTM	7.5629	12.630	16.32%	0.712
TFT	5.9972	11.9943	11.56%	0.7403

The comparative analysis indicates that forecasting performance varied across the evaluated models. TFT achieved the lowest MAE (5.997) and MAPE (11.56%), indicating the smallest average absolute and percentage prediction errors, respectively. These results suggest that TFT provided more accurate hard token demand forecasts based on these evaluation metrics. Meanwhile, Prophet achieved the lowest RMSE of 10.721, indicating lower penalized prediction errors and fewer or smaller large deviations compared with the other evaluated models.

The Temporal Fusion Transformer (TFT) also demonstrated the strongest goodness of fit, achieving the highest R^2 value of 0.7403. This indicates that TFT captured variations in hard token demand more effectively than the other evaluated models according to R^2. Although TFT did not achieve the lowest RMSE, its performance across MAE, MAPE, and R^2 highlights its potential to capture complex temporal patterns and relationships among multivariate input features.

Overall, TFT was selected as the preferred model for hard token demand forecasting because it achieved the best results in three of the four evaluation metrics, namely MAE, MAPE, and R^2. Although Prophet demonstrated better performance in terms of RMSE, TFT provided a stronger overall balance between prediction accuracy and goodness of fit. These findings support the integration of TFT-based forecasts into DigiMS to improve data-driven inventory planning, support stock allocation and replenishment decisions, and anticipate potential mismatches between hard token demand and stock availability.

Operational Implications for DigiMS

The forecasting results provide direct implications for banking inventory planning. By integrating forecasting outputs into DigiMS, hard token planning can shift from a predominantly static quota-based mechanism toward a more predictive and adaptive decision-support approach.

The comparative evaluation indicates that the forecasting models demonstrated different strengths across the evaluation metrics. TFT achieved the lowest MAE and MAPE, with values of 5.997 and 11.56%, respectively, indicating the smallest absolute and percentage prediction errors among the evaluated models. Prophet achieved the lowest RMSE of 10.721, while TFT obtained the highest R^2 value of 0.7403, indicating its stronger ability to capture variations in hard token demand. These results support the selection of TFT as the preferred forecasting model based on its overall performance across the evaluation metrics, while highlighting the importance of considering multiple criteria when evaluating forecasting accuracy.

The forecasting outputs can support early stock shortage alerts, branch-level demand prediction, quota optimization, and proactive replenishment planning. By providing forward-looking estimates of hard token demand, the forecasting system can help Banking Operations improve supply responsiveness, reduce stockout risk, minimize excessive allocation, and strengthen inventory efficiency.

Considering the multivariate nature of the proposed forecasting framework and TFT's strong overall performance, including the lowest MAE and MAPE and the highest R^2, TFT is considered the preferred candidate for integration into DigiMS as a decision-support model. Although Prophet achieved a lower RMSE, TFT demonstrated a stronger overall balance across the evaluation metrics. Its integration could support more accurate hard token demand forecasting, data-driven stock allocation, and replenishment planning to help anticipate potential stock shortages and excess inventory at bank branches.

Benchmarking Against Existing Literature

The findings are generally consistent with previous studies showing that forecasting performance depends on data characteristics and the forecasting approach employed (Kontopoulou et al., 2023). In this study, ARIMA produced the weakest overall performance, while Prophet achieved the lowest RMSE of 10.721. These results indicate differences in the models' ability to represent demand patterns in the dataset, highlighting the importance of evaluating forecasting models using multiple performance metrics.

TFT achieved the lowest MAE (5.997) and MAPE (11.56%), indicating the smallest absolute and percentage prediction errors among the evaluated models. This result highlights TFT's potential to capture temporal patterns in hard token demand while leveraging multivariate input features. The architecture of TFT, which integrates variable selection, gating mechanisms, and attention-based temporal modeling, is designed to capture complex relationships in multivariate time-series forecasting (Lim et al., 2021). These capabilities may help incorporate operational and external variables to improve demand forecasting.

In addition, TFT achieved the highest R^2 value of 0.7403, indicating the strongest goodness of fit among the evaluated models. Although Prophet achieved a lower RMSE, TFT performed better in terms of MAE, MAPE, and R^2. This demonstrates that a model's performance can vary depending on the evaluation metric and that lower errors across certain metrics do not necessarily imply superiority across all measures.

Overall, TFT was selected as the preferred model because it achieved the strongest results across three of the four evaluation metrics. These findings support the use of multivariate forecasting for banking inventory planning and suggest that TFT-based forecasts could help improve stock allocation, replenishment decisions, and the anticipation of potential mismatches between hard token demand and stock availability.

Discussion

Limitations and Threats to Validity

Several limitations should be considered when interpreting the results of this study. First, the forecasting models were developed using historical branch-level hard token demand data within a limited observation period. Although the dataset captured temporal demand patterns and operational variation, longer historical windows may improve model robustness and generalization.

Second, the multivariate forecasting framework incorporated selected operational and economic variables such as transaction activity, customer registration, lag features, rolling statistics, and regional economic indicators (PDRB). However, other potentially influential factors such as regulatory policy changes, manual operational intervention, unexpected demand spikes, or macroeconomic shocks were not explicitly modeled.

Third, the LSTM and TFT architectures were trained using practical hyperparameter configurations. While these settings produced stable experimental results, more exhaustive hyperparameter optimization may further improve predictive performance, particularly for deep learning models.

Another limitation relates to branch heterogeneity. Demand behavior across branches

may vary significantly due to customer characteristics, transaction volume, and regional business activity. Although TFT partially addressed this through multivariate learning and temporal attention, certain localized operational behaviors may still remain difficult to capture.

Finally, the deployment implications proposed in this study were conceptually designed for integration into DigiMS within the Banking operational environment. Therefore, external generalization to other banking institutions or inventory systems should be interpreted cautiously, as forecasting behavior may differ under alternative operational structures.

Despite these limitations, the study provides a practical forecasting framework and demonstrates strong potential for predictive inventory planning in banking operations.

CONCLUSION

This study developed and compared univariate and multivariate time-series forecasting approaches to support branch-level hard token inventory planning in banking using the CRISP-DM framework. The comparative evaluation showed that forecasting performance varied across the four evaluation metrics. TFT achieved the lowest MAE (5.997) and MAPE (11.56%), as well as the highest R^2 value (0.7403), indicating the strongest overall performance across three of the four metrics. Prophet achieved the lowest RMSE (10.721), demonstrating that no single model outperformed all others across every evaluation criterion. Based on its overall performance, TFT was selected as the preferred model for branch-level hard token demand forecasting. Its multivariate architecture offers potential for capturing complex temporal patterns and relationships among operational, economic, and other relevant input features.

From a practical perspective, the proposed forecasting framework provides a foundation for transforming hard token inventory management from a predominantly reactive process into a more predictive and data-driven decision-support mechanism. Integrating TFT-based forecasting outputs into DigiMS could support branch-level demand prediction, more informed stock allocation, proactive replenishment, improved supply responsiveness, and reduced risks of stockout and overstock. Nevertheless, the findings should be interpreted in light of the limited historical observation period, the selected external variables, branch-level demand heterogeneity, the hyperparameter configurations evaluated, and the conceptual stage of DigiMS deployment. Future research should incorporate longer historical datasets, additional operational and macroeconomic variables, more extensive hyperparameter optimization, and validation in a live operational environment to strengthen forecasting robustness and practical applicability.

ACKNOWLEDGEMENT

The dataset supporting the findings of this study has been deposited in Zenodo https://doi.org/10.5281/zenodo.23416363. The record is publicly accessible; however, access to the data files is restricted due to confidentiality and privacy considerations. The data presented in this study are available from the corresponding author upon reasonable request, subject to applicable confidentiality and privacy requirements.

AUTHOR CONTRIBUTION STATEMENT

Risa Aprilia: Conceptualization, data curation, methodology, formal analysis, software, visualization, investigation, and writing—original draft. Tuga Mauritsius: Conceptualization, methodology, supervision, and validation. Both authors contributed to the interpretation of the research findings and approved the final manuscript for publication.

REFERENCES

Acuña-Cid, H. A., Ahumada-Tello, E., Ovalle-Osuna, Ó. O., Evans, R., Hernández-Ríos, J. E., & Zambrano-Soto, M. A. (2025). CRISP-NET: Integration of the CRISP-DM model with network analysis. *Machine Learning and Knowledge Extraction*, *7*(3), 1–38. https://doi.org/10.3390/make7030101

Bansal, A., Balaji, K., & Lalani, Z. (2025). Temporal encoding strategies for energy time series prediction. *arXiv preprint arXiv:2503.15456*.

Bhanja, S., & Das, A. (2018). Impact of data normalization on deep neural network for time series forecasting. *arXiv preprint arXiv:1812.05519*.

Bharath, A., Paduraru, A., & Gaidosch, T. (2024). *Cyber resilience of the central bank digital currency ecosystem*. International Monetary Fund.

Bueno, L. A., et al. (2024). Impacts of digitization on operational efficiency in the banking sector: Thematic analysis and research agenda proposal. *International Journal of Information Management Data Insights*, *4*(1), 100230. https://doi.org/10.1016/j.jjimei.2024.100230

Doğan, A. (2025). Time series-based demand forecasting: A comparative analysis of Holt-Winters, SARIMA, and Prophet models on retail inventory data. *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, *24*(53), 641–668. https://doi.org/10.46928/iticusbe.1725584

Goltsos, T. E., Syntetos, A. A., Glock, C. H., & Ioannou, G. (2022). Inventory–forecasting: Mind the gap. *European Journal of Operational Research*, *299*(2), 397–419. https://doi.org/10.1016/j.ejor.2021.07.040

Hakizimana, S., Wairimu, M.-M. C., & Muathe, S. (2023). Digital banking transformation and performance-where do we stand. *International Journal of Management Research and Emerging Sciences*, *13*(1), 179–196.

Kalaya, P., Termsuksawad, P., & Wasusri, T. (2023). Forecasting and inventory planning for irregular demand patterns. *International Journal of Knowledge and Systems Science*, *14*(1).

Karmaker, S. (2025). Feature Engineering in Time Series Forecasting. *Preprint, novembre*.

Kiettikunwong, N., & Sangsarapun, W. (2025). G-Token implications and risks for the financial system under state-issued digital instruments in Thailand. *Journal of Risk and Financial Management*, *18*(10), Article 555. https://doi.org/10.3390/jrfm18100555

Kim, M., Lee, J., Lee, C., & Jeong, J. (2022). Framework of 2D KDE and LSTM-based forecasting for cost-effective inventory management in smart manufacturing. *Applied Sciences*, *12*(5), 2380. https://doi.org/10.3390/app12052380

Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023). A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. *Future Internet*, *15*(8), Article 255. https://doi.org/10.3390/fi15080255

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*(4), 1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

Liu, R., & Vakharia, V. (2024). Optimizing supply chain management through BO-CNN-LSTM for demand forecasting and inventory management. *Journal of Organizational and End User Computing*, *36*(1), 1–25. https://doi.org/10.4018/JOEUC.335591

Mauritsius, T. (2025). Extending the forecasting horizon of daily new COVID-19 cases using non-pharmaceutical measures and the effective reproduction number (Rt): A deep learning-based framework. *Decision Science Letters*, *14*(2), 283–302. https://doi.org/10.5267/j.dsl.2025.1.008

Nguyen, T. N., Hossain, M. Z., Khan, M. M., Sharif, K. S., Das, R., & Haque, M. S. (2024). Product demand forecasting for inventory management with Freight Transportation Services Index using advanced neural networks algorithm. *American Journal of Computing and Engineering*, *7*(4), 50–58. https://doi.org/10.47672/ajce.2432

Noor, N. M. M., Retnowardhani, A., Abd, M. L., & Saman, M. Y. M. (2013). Crime forecasting using ARIMA model and fuzzy alpha-cut. *Journal of Applied Sciences*, *13*(1), 167–172. https://doi.org/10.3923/jas.2013.167.172

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, *181*, 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Wicaksono, B. A., & Mauritsius, T. (2025). Forecasting teaching materials using the autoregressive integrated moving average (ARIMA) method case study: Universitas Terbuka. *International Journal of Engineering Trends and Technology*, *73*(6), 476–484. https://doi.org/10.14445/22315381/IJETT-V73I6P139