

Machine Learning Applications in Forecasting Loan Disbursement Patterns

Paula Joy L. Dela Cruz

College of Computer Studies, Quezon City University, San Bartolome, Novaliches, Quezon City,
Philippines

DOI: <https://doi.org/10.51583/IJLTEMAS.2026.150900052>

Received: 20 September 2026; Accepted: 25 September 2026; Published: 08 October 2026

ABSTRACT

Forecasting loan disbursement is essential for financial institutions to manage liquidity effectively and mitigate risks. Traditional forecasting methods often struggle with the complex and seasonal dynamics present in financial data. This study compares several forecasting models including ARIMA, Linear Regression, Long Short-Term Memory (LSTM) networks, and Prophet using seven years of loan disbursement data. Results indicate that LSTM networks produce the most accurate forecasting. The findings show the potential of machine learning to improve planning and decision-making for lenders, while maintaining strict considerations for data privacy and fairness.

Keywords: Forecasting, Loan Disbursement, Machine Learning, Financial Services, Time Series

INTRODUCTION

Accurate forecasting of loan disbursement is a critical function for financial institutions, enabling them to manage liquidity, optimize resource allocation, and mitigate credit risk. Given the complex nature of financial time series, which often exhibit seasonality, volatility, and nonlinear behavior, traditional forecasting techniques frequently fall short in capturing these dynamics with sufficient accuracy (Schaffer, Dobbins, & Pearson, 2021; Khan & Gupta, 2020).

Conventional models like the Autoregressive Integrated Moving Average (ARIMA) have been staples in time series analysis for decades, primarily due to their interpretability and established statistical foundations (Khan & Gupta, 2020). However, ARIMA and related linear models typically assume stationarity and linearity, which limits their effectiveness in environments with complex dependencies and nonlinear patterns. As a result, their predictive capabilities may degrade for datasets exhibiting strong nonlinearities, sudden changes, or multivariate influences.

Recent advances in machine learning and deep learning offer promising alternatives capable of addressing many of these limitations. Models such as Long Short-Term Memory (LSTM) networks, a class of recurrent neural networks, can model long-range dependencies and nonlinear relationships in sequential data, making them well-suited for financial forecasting tasks (Auffarth, 2021; Prajapati et al., 2021). The growing body of research supports LSTM's superior performance across various domains involving complex time series, including epidemiology, energy, and finance, where traditional models struggle to capture underlying signals (Dama & Sinoquet, 2021; Wahedi et al., 2022).

Furthermore, hybrid approaches that combine classical statistical methods with machine learning are gaining traction, offering a balance between interpretability and accuracy (Wahedi et al., 2022). These methods enable the modeling of both linear and nonlinear components of a time series, leveraging the strengths of both approaches.

This study seeks to explore and compare the forecasting accuracy of models such as ARIMA, Linear Regression, Prophet, and LSTM networks in predicting loan disbursement patterns. By utilizing seven years of

detailed loan data, the research aims to identify the model best suited to handle the complexities inherent in loan disbursement forecasting. Ethical considerations, including data privacy and fairness, are also prioritized, ensuring the responsible application of these forecasting techniques in financial decision-making.

METHODOLOGY

The study used monthly loan data spanning from 2018 to 2024 obtained from a financial institution. This dataset encompassed total monthly disbursement amounts, loan counts, borrower demographic profiles, loan categories, and relevant macroeconomic indicators. Prior to model training, the raw dataset was cleaned to eliminate missing entries, correct formatting inconsistencies, and normalize feature scales across variables. Because initial statistical testing confirmed the time series was non-stationary exhibiting time-varying mean and variance a logarithmic transformation followed by first-order differencing was applied to stabilize the variance and eliminate underlying trends. Stabilizing the series proved particularly critical for the neural network pipeline, as it significantly improved gradient stability and accelerated model convergence during training.

To evaluate predictive performance across different algorithmic paradigms, the study implemented four distinct forecasting models:

- **ARIMA (AutoRegressive Integrated Moving Average):** A classical statistical model that combines autoregressive terms, differencing steps, and moving average components. It predicts future values by taking a linear combination of the target variable's past values (lags) and past prediction errors, making it effective for linear trends and well-defined, stationary behavior.
- **Linear Regression:** A foundational statistical baseline that estimates a linear relationship between input features (such as time steps, lag variables, and economic indicators) and the target loan disbursement amount. It calculates a constant set of weights via ordinary least squares to quantify direct, straightforward correlations over time.
- **Prophet:** An additive time-series forecasting model developed by Meta designed to decompose time series into trend, seasonality, and holiday components. It handles non-linear trends using piecewise linear or logistic growth curves and fits seasonal effects through Fourier series, allowing it to adapt to structural breaks in loan demand and integrate external covariates.
- **Long Short-Term Memory (LSTM):** A specialized architecture of Recurrent Neural Networks (RNNs) built specifically to process sequential data while overcoming the vanishing gradient problem. LSTMs utilize memory cells controlled by three gating mechanisms an input gate, a forget gate, and an output gate which allow the network to selectively retain or discard information over long time sequences. This mechanism enables the model to identify complex, non-linear dependencies and long-term economic patterns within financial time series.

To evaluate model performance without data leakage, the study employed a time-series cross-validation scheme using an expanding window approach. Predictive accuracy was assessed across four standard metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2). Throughout the analytical process, strict data ethics and governance standards were maintained. All personal records were fully anonymized prior to analysis in compliance with data privacy regulations, and model inputs were strictly restricted to objective financial indicators such as reported income, verified credit risk scores, and historical repayment performance ensuring predictions remained fair and free from proxy-driven demographic bias.

RESULTS AND DISCUSSION

The predictive performance of four forecasting models, namely Linear Regression, ARIMA, Prophet, and Long Short-Term Memory (LSTM), was assessed using the loan disbursement dataset. The results are presented in Table 1.

Table 1: Predictive Performance of Time Series Forecasting Models

Model	RMSE	MAE	MAPE	R ²
Linear Regression	14.32	11.20	8.45%	0.68
ARIMA	11.85	9.15	6.80%	0.77
Prophet	9.94	7.82	5.62%	0.83
LSTM	6.12	4.68	3.21%	0.91

As shown in Table 1, the LSTM model produced the lowest values for RMSE, MAE, and MAPE, while obtaining the highest R² value of 0.91. These results indicate that the model generated more accurate forecasts and was able to explain a larger proportion of the variation in loan disbursement data compared with the other forecasting approaches.

Among the benchmark models, Prophet achieved the second-highest performance with an R² value of 0.83, followed by ARIMA with 0.77. Linear Regression recorded the lowest R² value of 0.68 and the highest error values across all evaluation metrics. This suggests that the traditional statistical models were less effective in representing the underlying patterns present in the dataset.

The findings may be attributed to the ability of LSTM networks to learn sequential relationships from historical observations. Since loan disbursement data often exhibit changing trends and temporal dependencies, the LSTM model appears to be better suited for capturing these characteristics than the other models evaluated in this study.

Overall, the results demonstrate that the LSTM model provided the most accurate forecasts for the loan disbursement dataset. The consistent improvement observed across all evaluation metrics supports its suitability for time-series forecasting in financial applications.

In order to place these findings in a broader context, related studies on time-series forecasting were reviewed. Table 2 summarizes the methodologies, datasets, and key findings reported in previous research and provides a basis for comparing the results of the present study with those reported in the literature.

Thematic Matrix Table 2: Overview of Models and Findings in Time Series Forecasting Studies

Study	Application Domain	Models / Techniques Used	Development Tools / Data	Key Findings / Insights
Maleki et al. (2020)	COVID-19 case forecasting (global)	TP-SMN-AR (flexible autoregressive)	COVID-19 global time series; non-Gaussian modeling framework	Accurately forecasted cases using distribution-flexible models; captured heavy tails
Schaffer et al. (2021)	Health policy intervention assessment	ARIMA	Intervention data with seasonal patterns and autocorrelation	ARIMA effectively evaluated policy impacts in seasonal time series
Qi et al. (2020)	Meteorological impact on COVID-19	Generalized Additive Model (GAM)	COVID-19 cases + weather data in 30 Chinese provinces	GAM revealed temperature/humidity effects using 14-day lags and spatial

	spread			interactions
Khan & Gupta (2020)	COVID-19 case prediction in India	NAR, ARIMA, Neural Network	COVID-19 case data (Jan–Apr 2020); model comparison	NAR outperformed ARIMA ($R^2 = 0.97$ vs 0.95); better for nonlinear trend capture
Park et al. (2023)	Greenhouse environmental condition forecasting	ANN, NARX, RNN-LSTM	Multivariate greenhouse data (temperature, CO ₂ , humidity)	RNN-LSTM most accurate across 5–30 min horizons; $R^2 > 0.96$ for temperature prediction

For instance, Maleki et al. (2020) successfully forecasted global COVID-19 cases using a flexible autoregressive model (TP-SMN-AR) designed to handle heavy-tailed distributions commonly observed in real-world data. Schaffer et al. (2021) demonstrated the effectiveness of ARIMA in evaluating health policy interventions characterized by seasonal patterns. Similarly, Khan and Gupta (2020) reported that Neural Autoregressive (NAR) models outperformed ARIMA in capturing nonlinear COVID-19 trends in India.

Likewise, Park et al. (2023) found that recurrent neural network models, particularly RNN-LSTM, achieved the highest forecasting accuracy in predicting multivariate environmental conditions. Qi et al. (2020) employed Generalized Additive Models (GAMs) to examine the effects of meteorological factors on COVID-19 transmission, highlighting the importance of incorporating relevant external variables in forecasting models.

The reviewed studies suggest that the selection of forecasting models should be based on the characteristics of the dataset and the objectives of the analysis. While ARIMA remains useful for modeling seasonal and structured time-series data, neural network-based approaches offer advantages when nonlinear relationships and long-term temporal dependencies are present. Alternative methods such as TP-SMN-AR and GAM further demonstrate the value of considering data distribution and external influencing factors in forecasting applications.

Consistent with these observations, the results of the present study indicate that the LSTM model is well-suited for forecasting loan disbursement patterns. The model achieved the lowest forecasting errors and the highest coefficient of determination among the evaluated approaches, suggesting that it captured temporal patterns within the dataset more effectively than the benchmark models.

The findings further suggest that machine learning techniques, particularly LSTM networks, have considerable potential for financial forecasting applications. Improved accuracy in forecasting loan disbursements may assist lending institutions in resource planning, monitoring lending activities, and supporting informed decision-making. Future studies may enhance model performance by incorporating additional economic indicators and real-time external variables to better reflect changing market conditions.

Overall, the results indicate that the LSTM model provides a reliable approach for forecasting loan disbursement trends. Its strong predictive performance highlights the potential of machine learning methods to complement traditional forecasting techniques in financial applications. Further improvements may be achieved through the use of richer datasets and additional explanatory variables, enabling forecasting systems that are more responsive to evolving economic conditions.

REFERENCES

1. Auffarth, Ben (2021). Machine learning for time-series with Python: forecast, predict, and detect anomalies with state-of-the-art machine learning methods. Packt Publishing

2. Bandara, K., Hyndman, R., & Bergmeir, C. (2022). MSTL: A Seasonal-Trend Decomposition Algorithm for Time Series with Multiple Seasonal Patterns. *International Journal of Operational Research*, 1(1), 1. <https://doi.org/10.1504/ijor.2022.10048281>
3. Dama, F., & Sinoquet, C. (2021). Time Series Analysis and Modeling to Forecast: a Survey. arXiv (Cornell University). <https://arxiv.org/pdf/2104.00164.pdf>
4. Khan, F. U., & Gupta, R. (2020). ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. *Journal of Safety Science and Resilience*, 1(1), 12–18. <https://doi.org/10.1016/j.jnlssr.2020.06.007>
5. Nelson, D. M., Pereira, A. C. M., & de Oliveira, R. A. (2017). Stock market's price movement prediction with LSTM neural networks. In *International Joint Conference on Neural Networks (IJCNN)* (pp. 1419–1426).
6. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PloS One*, 13(3), e0194889. <https://doi.org/10.1371/journal.pone.0194889>
7. Prajapati, S., Swaraj, A., Lalwani, R., Narwal, A., Verma, K., Singh, G., & Kumar, A. (2021). Comparison of traditional and hybrid time series models for forecasting COVID-19 cases. *Research Square (Research Square)*. <https://doi.org/10.21203/rs.3.rs-493195/v1>
8. Schaffer, A. L., Dobbins, T., & Pearson, S. (2021). Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions. *BMC Medical Research Methodology*, 21(1). <https://doi.org/10.1186/s12874-021-01235-8>
9. Wahedi, H., Wrona, K., Heltoft, M., Saleh, S., Knudsen, T. R., Bendixen, U., Nielsen, I., Saha, S., & Borup, G. S. (2022). Improving accuracy of time series forecasting by applying an ARIMA-ANN hybrid model. In *IFIP advances in information and communication technology* (pp. 3–10). https://doi.org/10.1007/978-3-031-16407-1_1
10. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)