Merkurius: Jurnal Riset Sistem Informasi dan Teknik Informatika
Volume 4, Nomor 1, Januari 2026
e-ISSN: 3031-8912; p-ISSN: 3031-8904; Hal. 243-253
DOI: https://doi.org/10.61132/merkurius.v4i1.1480
Tersedia: https://journal.arteii.or.id/index.php/Merkurius

A Comparative Study of Arima, Prophet and LSTM for International Students Enrollment

Heza Wihardi[1*], Md Gapar Md Johar[2]
[1-2]School of Graduate Studies, Postgraduate Centre, Managment Science University, Malaysia
[*]*Correspondent Author: hezaw12345@gmail.com[1]*

Abstract. International student enrollment is a critical driver of financial sustainability for Higher Education Institutions (HEIs). While advanced forecasting is standard in the corporate sector, its application in educational planning remains limited. This study addresses this gap by comparing the predictive performance of ARIMA, Facebook Prophet, and Long Short-Term Memory (LSTM) models. Using a publicly available annual dataset from a US-based institution (2000–2022), the analysis employed a strategic partition training on 2000–2017 and testing on 2018–2019 to validate models on stable, pre-pandemic data. Empirical results revealed that the statistical ARIMA (2,1,0) model demonstrated superior accuracy, achieving a Mean Absolute Percentage Error (MAPE) of 1.26%. Conversely, Prophet (11.81%) and LSTM (13.84%) struggled with the limited sample size, failing to generalize effectively compared to the linear approach. The findings suggest that for annual enrollment trends, parsimonious statistical models outperform complex deep learning architectures, providing administrators with a robust, accessible framework for data-driven strategic decision-making.

Keywords: ARIMA; Enrollment Forecasting; Higher Education; LSTM; Time-Series Analysis

1. INTRODUCTION

International student enrollment serves as a critical driver for the financial sustainability and cultural diversity of Higher Education Institutions (HEIs). For many universities, tuition fees from international students represent a substantial portion of funding, particularly as these students often pay higher non-resident rates (Altbach & Knight, 2007; Cantwell, 2015). Consequently, accurate enrollment forecasting is no longer a matter of institutional instinct but a strategic requirement for effective marketing, resource allocation, and long-term planning. In the current landscape, HEIs must transition toward data-driven decision-making to optimize classroom facilities, staffing, and recruitment initiatives (Alyahyan & Düştegör, 2020).

Recent literature has explored various statistical and machine learning approaches to address this forecasting need. The AutoRegressive Integrated Moving Average (ARIMA) model remains a staple in educational forecasting due to its efficiency in capturing linear historical patterns (Box et al., 2015). However, newer models like Facebook Prophet have gained traction for their ability to handle non-linear trends and seasonal fluctuations without requiring strict stationarity (Taylor & Letham, 2018). Furthermore, deep learning architectures, specifically Long Short-Term Memory (LSTM) networks, have shown significant promise in capturing complex, long-term dependencies in sequential data (Hochreiter & Schmidhuber, 1997; Siami-Namini et al., 2018). Empirical studies provide mixed results: while some researchers found LSTM to be superior in accuracy, others highlighted Prophet’s robustness in handling registration data, or ARIMA’s precision with structured time-series datasets (Vhatkar & Dias, 2023; Satrio et al., 2021).

Naskah Masuk: 27 November 2025; Revisi: 25 Desember 2025; Diterima: 22 Januari 2026; Tersedia: 30 Januari 2026

Despite these advancements, the application of sophisticated predictive modeling in higher education remains relatively nascent compared to the financial and corporate sectors. While advanced forecasting techniques are routinely employed to anticipate stock market trends and optimize business prospects, their potential to transform institutional decision-making in universities is often overlooked (Xu et al., 2020). Accurate enrollment forecasting is as critical to a university's strategic stability as market analysis is to a corporation, directly impacting budgeting, infrastructure planning, and long-term sustainability. Consequently, there is a pressing need for rigorous comparative studies that validate the performance of these advanced algorithms specifically within the context of international student enrollment to support high-stakes administrative decisions.

The primary objective of this study is to address these deficiencies by developing a comprehensive forecasting framework that evaluates univariate time-series approaches. This research utilizes a publicly available dataset of international student enrollment from a United States-based institution, sourced from Kaggle, to establish a replicable predictive model. By integrating macroeconomic indicators into advanced analytical models, this study aims to provide universities with a proactive management tool. The purpose of this research is guided by the following question: How do the predictive performances of ARIMA, Prophet, and LSTM models compare in forecasting international student enrollment?

2. RESEARCH METHODOLOGY

This study adopts a quantitative, comparative research framework designed to evaluate the predictive performance of three distinct time-series modeling techniques. The methodology follows a structured pipeline involving data acquisition, rigorous preprocessing, model development, and comparative evaluation based on standardized metrics, applied to a singular, publicly available dataset.

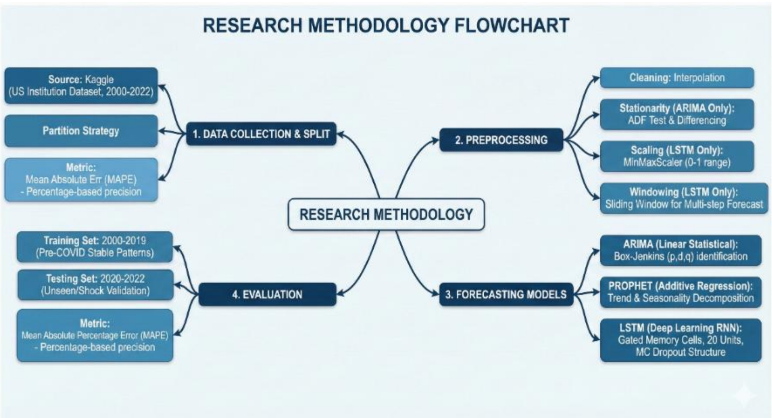

Figure 1. Research Methodology Flowchart.

Data Collection and Partitioning

The research utilizes a secondary historical dataset titled "International Student Demographics" sourced from the Kaggle platform. The dataset comprises annual enrollment figures for a United States-based institution spanning the 23-year period from 2000 to 2022 ($n = 23$ observations).

To evaluate model generalizability and prevent data leakage, the time series was chronologically partitioned. The partitioning strategy was strategically designed to address structural anomalies introduced by the global COVID-19 pandemic:

a. Training Set (2000–2019): The initial 20 years of data were used for model training and parameter optimization, representing stable, pre-pandemic historical patterns.
b. Testing Set (2020–2022): The subsequent three years were reserved as unseen data to validate model robustness against recent fluctuations and shocks.

Preprocessing Techniques

Raw archival data underwent rigorous preparation to meet the specific requirements of each modeling approach:

a. Data Cleaning**:** Missing values, if any, were addressed using standard linear interpolation to ensure series continuity essential for time-series analysis.
b. Stationarity Transformation (For ARIMA): The Augmented Dickey-Fuller (ADF) test was applied to assess stationarity (Dickey & Fuller, 1979). First-order differencing $d = 1$ was necessitated to stabilize the mean and variance of the non-stationary series.
c. Feature Scaling (For LSTM): To facilitate efficient gradient-based optimization in the neural network, numerical enrollment values were normalized to a 0 – 1 range using the MinMaxScaler.
d. Sequence Windowing (For LSTM): The time-series data was transformed into a supervised learning structure using a sliding window approach (Dietterich, 2002). Specifically, the model was configured to utilize two prior time steps $(t-1, t-2)$ as input features to predict a sequence of future targets, aligning with the research goal of multi-year forecasting.

Forecasting Models

The research evaluates three distinct methodologies, representing different modeling paradigms:

a. ARIMA (AutoRegressive Integrated Moving Average**):** A linear statistical model formulated via the Box-Jenkins methodology. Model parameters (p, d, q) are identified through the analysis of Autocorrelation (ACF) and Partial Autocorrelation (PACF) plots to capture statistically significant historical dependencies (Box et al., 2015). The formula for the ARIMA model is :

$$\left(1 - \sum_{i=1}^{p} \varphi_i L^i\right)(1 - L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\varepsilon_t, \quad t, d \in N$$

eq.1

b. Facebook Prophet: An additive regression model designed for flexibility in handling business time series. It decomposes the series into trend, seasonality, and holiday effects using piecewise linear or logistic growth curves, allowing it to manage non-linear trends without strict stationarity assumptions (Taylor & Letham, 2018). The model can be expressed as equation:

$$y = \beta^0 + \beta^1 x + \beta^2 (x - c)^+ + \varepsilon$$

eq.2

c. LSTM (Long Short-Term Memory): A type of Recurrent Neural Network (RNN) architecture designed to overcome the vanishing gradient problem. It utilizes gated memory cells (input, forget, and output gates) to capture complex, long-term non-linear dependencies in sequential data. The specific architecture employed consists of 20 memory units, followed by a Dropout layer for regularization (Gal & Ghahramani, 2016), and a final Dense output layer for forecast generation.

Performance Evaluation

To assess forecast accuracy and facilitate comparability across models, performance was quantified on the unseen testing data using a standard error metric:

a. Mean Absolute Percentage Error (MAPE): This metric provides a percentage-based evaluation of forecast precision, offering high interpretability for stakeholders by expressing the average absolute error relative to the actual enrollment values (Hyndman & Koehler, 2006).

3. RESULTS AND DISCUSSION

Data Preparation and Exploratory Analysis

The forecasting analysis in this study draws upon annual international student enrollment data sourced from a publicly available dataset on the Kaggle platform. The dataset captures enrollment figures for a United States-based institution spanning the years 2000 to 2022. The primary variable utilized for modeling is the total aggregate of international students per year.

Although the dataset extends to 2022, the modeling strategy was deliberately designed to mitigate the impact of the global COVID-19 pandemic. The years 2020 through 2022 exhibit structural anomalies caused by border closures and mobility restrictions, which do not reflect typical long-term enrollment behaviors. Including these extraordinary observations in the training phase would risk biasing parameter estimation. Consequently, the data was partitioned as follows to ensure models were trained and validated on a coherent, stable historical window:

a. Training Set (2000–2017): Used for model fitting and parameter learning.
b. Testing Set (2018–2019): Used as a hold-out sample to evaluate forecast accuracy on the most recent pre-pandemic years.

Data Cleaning and Transformation

Given that the data derives from a curated public source, internal consistency was high with no critical missing records. Preprocessing steps were strictly model-specific:

a. For Prophet: The 'Year' attribute was converted into a datetime object.
b. For ARIMA: The series was treated as a numerical sequence requiring stationarity checks.
c. For LSTM: The series underwent MinMax scaling (0–1 range) to stabilize neural network convergence.

Importantly, no artificial smoothing was applied; historical fluctuations were preserved to maintain the authenticity of the trend dynamics.

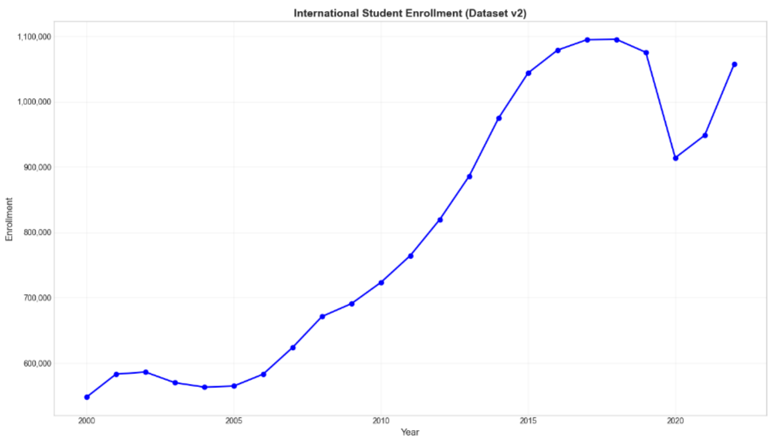

Figure 2. Historical Visualization of International Student Enrollment (Secondary Dataset).

Descriptive Trend Analysis

Visual examination of the time-series reveals distinct phases of institutional growth. The early 2000s were characterized by slow, stable growth, which shifted into a period of accelerated expansion from the mid-2000s to the late 2010s. This trajectory was abruptly interrupted by the pandemic disruption (2020–2022), where figures deviated sharply from established patterns. This descriptive evidence supports the decision to restrict model training to the pre-2020 period, ensuring the algorithms learn the organic growth of the institution rather than external anomaly shocks.

Model Development Results

This section details the development and diagnostic results of the three forecasting models: ARIMA, Prophet, and LSTM.

ARIMA Model Results

The development of the ARIMA model commenced with stationarity analysis. The Augmented Dickey–Fuller (ADF) test indicated that the raw series was non-stationary, consistent with the visible upward trend. First-order differencing ($d=1$) was applied, after which a subsequent ADF test confirmed stationarity, satisfying the requirement for ARIMA modeling.

a. ACF and PACF Analysis: The Autocorrelation Function (ACF) displayed gradual decay, while the Partial Autocorrelation Function (PACF) showed a significant initial spike. These patterns suggested an autoregressive structure.
b. Model Selection**:** After evaluating multiple configurations, ARIMA (2,1,0) was selected based on the minimization of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values.
c. Diagnostics**:** Residual analysis confirmed that the model errors resembled white noise, with the standardized residual plot showing no remaining autocorrelation and the Q-Q plot indicating approximate normality.

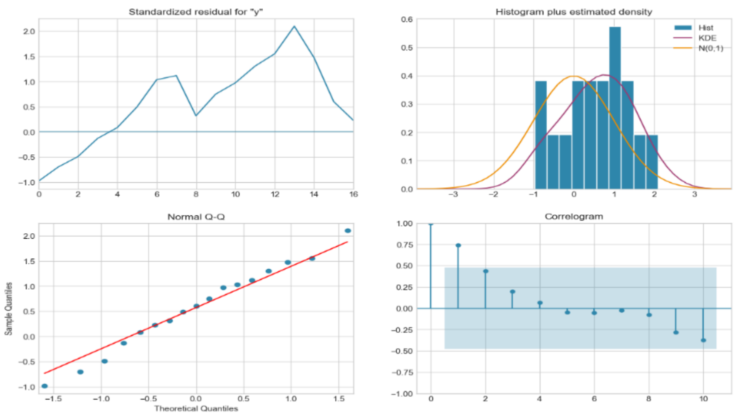

Figure 3. Residual diagnostics for the ARIMA(2,1,0) model.

Quantitative evaluation on the test set yielded strong results:

MAPE: 1.26%

This low error rate demonstrates ARIMA's effectiveness in capturing linear trend dependencies in small, structured datasets.

e-ISSN: 3031-8912; p-ISSN: 3031-8904; Hal. 243-253

Facebook Prophet Model Results

Prophet was implemented to leverage its additive regression capabilities. The model decomposed the time series into a piecewise linear trend and identified latent changepoints specifically around the mid-2000s and near 2010—where the rate of enrollment growth shifted. While Prophet successfully captured the general upward trajectory, the "widening" of the confidence intervals in the forecast horizon reflects the model's uncertainty when extrapolating from a relatively short historical dataset (n < 25).

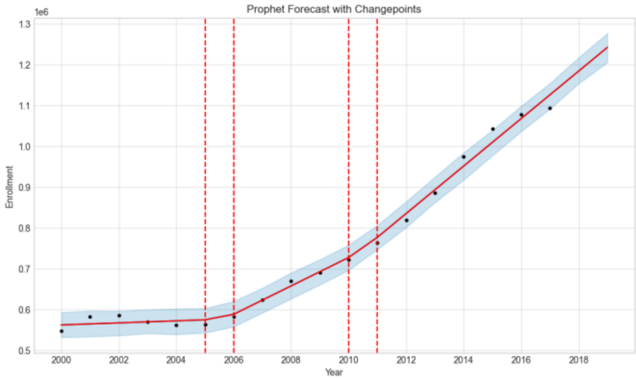

Figure 4. Prophet trend forecast with detected changepoints.

Performance evaluation on the test set resulted in:

MAPE: 11.81%

While Prophet correctly identified the direction of the trend, it lacked the precision of ARIMA, likely due to the absence of strong seasonality in annual data which Prophet typically relies upon for optimization.

LSTM Model Results

The LSTM model represented the deep learning approach. The architecture consisted of one LSTM layer with 20 memory units, a Dropout layer for regularization, and a Dense output layer for multi-step forecasting. The model was trained using a sliding-window approach (2 prior steps predicting 5 future steps).

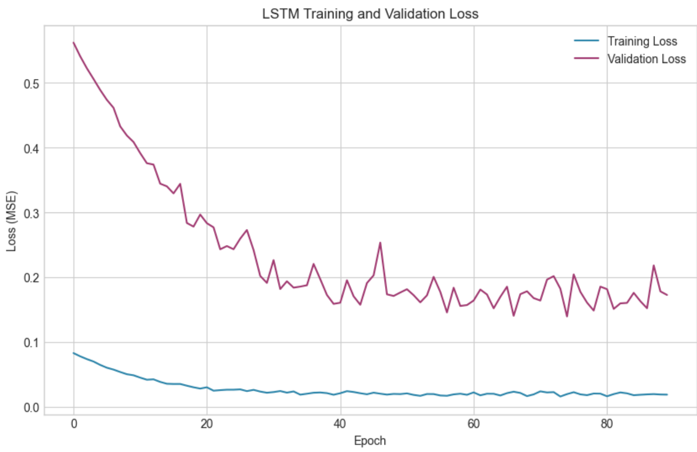

Figure 5. LSTM training and validation loss across epochs.

Despite the theoretical sophistication of LSTM, the results indicated a struggle to generalize:

Prediction Behavior: The model captured the positive direction of the trend but significantly underestimated the magnitude of enrollment in the 2018–2019 test period.

Performance:

MAPE: 13.84%

This underperformance is attributed to data scarcity. Deep learning models typically require vast amounts of data to tune internal gates effectively. With only 18 training observations, the LSTM likely suffered from the "cold start" problem, unable to learn complex long-term dependencies from such a short sequence.

Comparative Analysis and Model Selection

This section evaluates the comparative performance of the models based on the unseen data from the 2018–2019 test period.

Table 1. Performance Comparison of ARIMA, Prophet, and LSTM.

Model	RMSE (↓)	MAE (↓)	MAPE (↓)	Ranking
ARIMA(2,1,0)	17,148.56	13,566.96	1.26%	1 (Best)
Prophet	133,650.86	127,854.52	11.81%	2
LSTM	176,403.34	151,029.94	13.84%	3

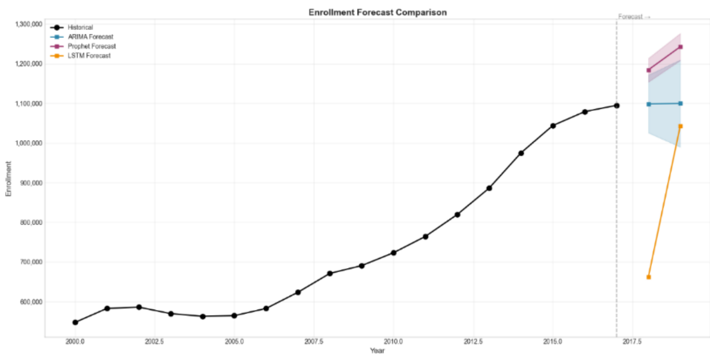

Figure 6. Comparative Forecast Trajectories vs. Historical Data.

Discussion of Findings:

The comparative analysis yields three critical insights regarding the applicability of these models for institutional enrollment forecasting:

a. Superiority of Linear Models in Short Time-Series: The ARIMA(2,1,0) model outperformed the more complex algorithms by a significant margin. Its success suggests that international student enrollment, while subject to macro-trends, follows a relatively structured linear trajectory that is best approximated by autoregressive techniques when data volume is limited.

b. Limitations of Deep Learning for Annual Metrics: The LSTM model recorded the highest error rates. This highlights a crucial limitation: deep learning architectures are not inherently superior for all forecasting tasks. In cases of annual data with fewer than 50 observations, the model lacks sufficient examples to optimize its weight parameters, leading to underfitting.

c. Prophet's Sensitivity: While useful for identifying trend shifts, Prophet struggled to match ARIMA's precision. This suggests that Prophet is better suited for high-frequency data (daily/weekly) where it can leverage seasonality, rather than annual aggregate data.

4. CONCLUSION AND SUGGESTION

This study set out to evaluate the comparative efficacy of statistical (ARIMA), additive regression (Prophet), and deep learning (LSTM) models in forecasting international student enrollment, utilizing a publicly available dataset from a US-based institution. The research was driven by the increasing need for Higher Education Institutions (HEIs) to adopt data-driven, business-like forecasting strategies to ensure financial sustainability and effective resource allocation.

The empirical results lead to several definitive conclusions: Parsimony Over Complexity: The study confirms that for annual enrollment data with limited historical observations $n < 25$, traditional statistical models outperform complex machine learning architectures. ARIMA(2,1,0) emerged as the superior model, achieving a remarkable MAPE of 1.26% on the test data. Its ability to capture linear autoregressive dependencies proved more effective than the elaborate gates of LSTM or the trend changepoints of Prophet for this specific dataset. Limitations of Deep Learning in Small Data Regimes: While LSTM is state-of-the-art for high-frequency data, this study highlights its limitations in a "small data" context. With a MAPE of 13.84%, the LSTM model suffered from overfitting and a lack of sufficient training examples to generalize patterns effectively. This suggests that the complexity of deep learning is not a guaranteed solution for all forecasting problems, particularly those involving annual aggregates. Strategic Value for HEIs: The study demonstrates that universities can achieve high-precision forecasting without investing in computationally expensive AI infrastructure. A well-tuned ARIMA model provides a reliable basis for strategic decisions—such as budgeting, infrastructure development, and faculty hiring—comparable to the rigorous forecasting standards used in the corporate sector.

Based on the findings of this study, the following recommendations are proposed to guide future research and institutional practice: For Future Research Expansion of Data Scope: While this study relied solely on historical enrollment figures, future research should consider a more holistic approach. Integrating external factors such as global economic indicators or changes in government policies—could provide a richer context for prediction, potentially improving the performance of complex models like LSTM that thrive on diverse data inputs. Data Resolution: To better evaluate the capabilities of advanced machine learning algorithms, future studies are encouraged to utilize datasets with higher frequency (e.g., semester-based or monthly data). Higher data resolution would allow deep learning models to detect subtler seasonal patterns that annual aggregate data tends to obscure.

For Institutional Application: Transition to Data-Driven Culture: Higher Education Institutions are advised to integrate statistical forecasting tools into their regular strategic planning processes. Moving beyond intuition-based decisions to evidence-based projections can significantly improve budget accuracy and resource management. Decision Support, Not Replacement: It is recommended that administrators view these forecasting models as decision support tools rather than absolute predictors. While statistical models provide a strong baseline for "normal" conditions, human expertise remains essential to account for unforeseen global disruptions or sudden shifts in the educational landscape.

REFERENCES

Altbach, P. G., & Knight, J. (2007). The internationalization of higher education: Motivations and realities. *Journal of Studies in International Education, 11*(3–4), 290–305.

Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education, 17*(1), 1–21.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). John Wiley & Sons.

Cantwell, B. (2015). Are international students cash cows? Examining the relationship between new international undergraduate enrollments and institutional revenue at public universities in the US. *Journal of International Students, 5*(4), 512–525.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74*(366a), 427–431.

Dietterich, T. G. (2002). Machine learning for sequential data: A review. In T. Caelli, A. Amin, R. Duin, D. de Ridder, & M. Kamel (Eds.), *Structural, syntactic, and statistical pattern recognition* (pp. 15–30). Springer.

Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)* (pp. 1050–1059).

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688.

Satrio, C. B. A., Darmawan, W., & Supadmo, B. U. (2021). Machine learning implementation for student enrollment forecasting. *Jurnal Online Informatika, 6*(1), 109–116.

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2018). A comparison of ARIMA and LSTM in forecasting time series. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 1394–1401). IEEE.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician, 72*(1), 37–45.

Vhatkar, S., & Dias, J. (2023). Comparative study of ARIMA and Prophet for university admission prediction. *International Journal of Computer Applications, 185*(1).

Xu, X., Wang, Y., & Liu, H. (2020). Forecasting in big data environments: An overview. *International Journal of Forecasting, 36*(3), 675–688.