Scientific Journal of Informatics
Vol. 13, No. 2, May 2026

p-ISSN 2407-7658 http://journal.unnes.ac.id/nju/index.php/sji e-ISSN 2460-0040

A Multi-Model Forecasting Framework for New Student Admissions: SMA, ARIMA, and Random Forest Approaches

Falentino Sembiring[1*], Rieska Rahayu Ayuningsih[2], Adhitia Erfina[3], Risky[4]
[1, 2, 4]Department of Information Systems, Nusa Putra University, Indonesia
[3]Department of Informatics and Computer Engineering, Vietnam National University, Vietnam

Abstract.
Purpose: This study evaluates and compares the forecasting performance of traditional statistical methods (Simple Moving Average and ARIMA) and a machine learning approach (Random Forest) in predicting student applicant numbers across multiple study programs at XYZ University for 2024–2026. The objective is to identify the most accurate model to support data-driven strategic planning and enrollment management.
Methods: A quantitative comparative forecasting design was applied using historical admission data from 2018–2023. Three models SMA, ARIMA, and Random Forest were implemented and assessed using Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), and Mean Squared Error (MSE). Model robustness was evaluated across several study programs with different growth patterns.
Result: The findings reveal an overall upward enrollment trend, particularly in Management and Informatics Engineering. Random Forest achieved the highest predictive accuracy, with MAPE values ranging from 6.17% to 17.95%, outperforming ARIMA (17.95%–33.13%) and SMA (14.5%–28.25%). The results indicate that Random Forest more effectively captures complex and non-linear enrollment dynamics.
Novelty: This study provides a systematic multi-program comparison between classical time-series models and a machine learning approach within a single institutional context. It demonstrates the superior robustness of Random Forest and supports integrating machine learning–based forecasting into higher education information systems for improved strategic decision-making.

Keywords: Enrollment forecasting, Comparative modeling, Time series and machine learning, Forecast accuracy, Higher education data
Received February 2026 / **Revised** April 2026 / **Accepted** April 2026

INTRODUCTION

New student admissions are a strategic imperative for ensuring the sustainability and competitiveness of higher education institutions (HEIs). Applicant numbers reflect institutional reputation, program attractiveness, and marketing effectiveness [1]. In dynamic educational systems such as Indonesia, enrollment demand fluctuates due to demographic changes, economic conditions, and regulatory policies, creating uncertainty in institutional planning [2]. Consequently, accurate forecasting models are essential to support evidence-based strategic management.

Globally, HEIs increasingly adopt predictive analytics to improve financial planning, human resource allocation, and infrastructure management [3], [4] . In Indonesia, although the number of active university students reached 8–9 million between 2020–2024 [5], participation remains disproportionately low compared to the productive-age population [6]. Competitive admission processes, such as the 2025 SNBP selection where 776,515 applicants competed for 181,425 seats [7], further emphasize the importance of reliable forecasting mechanisms [8].

Current research shows many approaches to predicting educational outcomes, each with its own goals, data, and analysis methods. Researchers have used classical time-series models to analyze educational data. Ahmad [9] applied the ARIMA method to predict university rankings in the United Arab Emirates, using several years of past data. The goal was to give higher education administrators a reliable way to anticipate future rankings. The results showed that ARIMA works well when trends are stable and consistent, but it

*Corresponding author.
Email addresses: falentino.sembiring@nusaputra.ac.id (Sembiring)
DOI: [10.15294/sji.v13i2.42746](https://doi.org/10.15294/sji.v13i2.42746)

is less effective when there are major changes in the data. The study found that ARIMA performs well when data trends are stable but becomes less reliable when significant changes occur.

This limitation is also highlighted in broader information systems research, which emphasizes the importance of using more flexible and data-driven models to support decision-making. As discussed in previous studies, relying only on traditional statistical methods may not be sufficient to capture the complexity of educational data [10].

Recent studies increasingly employ ensemble machine learning methods to address the complex and non-linear characteristics of student data. Brahma, Rajendra, and Sula [1] examined data-driven frameworks for analyzing enrollment trends and informing educational policy [11]. Their research utilized institutional datasets containing applicant information and enrollment outcomes to enhance evidence-based planning in higher education. The authors advocated for the adoption of predictive methods that uncover complex relationships among variables, rather than relying solely on descriptive analyses [12].

Tapio and Tarepe compared a standalone Random Forest model with a hybrid ARIMA-Random Forest approach to forecast student enrollment [13]. Their goal was to see if combining a traditional time-series model with ensemble learning would improve forecast accuracy. Using higher education enrollment data, they found that the hybrid model gave only slight improvements in some cases. Overall, the Random Forest model alone was more robust and stable, especially when enrollment numbers changed in ways that could not be explained by simple linear trends.

Asselman, Khaldi, and Aammou provide further support for tree-based ensemble techniques by examining the XGBoost algorithm to predict student academic performance. Although their dependent variable differed from applicant volume, their methodological approach to educational datasets yields valuable insights [14], [15]. The researchers analyzed student interaction logs and demographic data to improve early warning systems and pedagogical interventions [16]. Their findings reinforce the broader argument that ensemble methods are effective in addressing the noise and non-linear interdependencies inherent in educational data.

The work of Liu, Schenk, Braun, and Frey most directly informs the design of the present study. They developed a machine learning framework to guide student admission decisions, utilizing applicant profile data and historical acceptance records. A key methodological observation from their research concerns the challenges of limited temporal data [17]. They argue that when historical data are constrained, conventional strategies such as partitioning into separate training and testing subsets may not be methodologically sound or feasible. This insight directly informs the evaluation approach adopted in this study.

Despite these scholarly contributions, the literature lacks a systematic and directly comparable evaluation of traditional statistical benchmarks, such as Simple Moving Average and ARIMA, against ensemble machine learning approaches using a unified institutional dataset disaggregated by individual study programs [18], [19]. Most existing research focuses on a single forecasting method or evaluates model performance using aggregated, institution-wide totals, which obscures program-specific demand fluctuations [20], [21]. This study addresses this gap by conducting a rigorous, multi-program comparative assessment within a single institutional context.

METHODS

This study uses a quantitative predictive approach with three forecasting methods: Simple Moving Average (SMA), Auto Regressive Integrated Moving Average (ARIMA), and Random Forest. Computations were performed in Python on Google Colaboratory, and historical data was stored in Microsoft Excel. The dataset includes annual new student admission records from XYZ University for 2018 to 2024.

Data preprocessing involved format standardization and validation to ensure consistency. Because of the dataset's limited temporal range [22], [23], splitting into training and testing subsets was unnecessary. Model evaluation compared predictions directly to actual values for each year.

Simple moving average (SMA)

The SMA method was applied as a baseline model to smooth historical fluctuations by averaging values over a fixed window [24] . The SMA calculation is represented as:

$$M_t = F_t + 1 = X_t + X_t - 1 + X_t - 2 + \cdots + Xt - n + 1n \qquad \textbf{(1)}$$

M_t : is the moving average at period t
F_{t+1} : is the forecasted value for the next period
t + 1 Xt : is the actual value at period t
n : is the window length
The SMA function was implemented using the pandas and numpy libraries in Google Colab.

Auto regressive integrated moving average (ARIMA)
ARIMA was utilized as a statistical time series model capable of handling historical dependencies through autoregressive and moving average components. [2], [25] Its generalized representation is:

$$\phi_t(\boldsymbol{B})\boldsymbol{\nabla}^d \boldsymbol{Z_t} = \mu + \varepsilon_t - \theta_q(\boldsymbol{B})\, \varepsilon_t \qquad \textbf{(2)}$$

Z_t : is the observed value at time t
ϕ_t : is the autoregressive (AR) polynomial in the backshift operator B
θ_q : is the moving average (MA) polynomial in the backshift operator B
B : is the backshift operator
d : is the differencing order to achieve stationarity
μ : is the constant parameter
ε_t : is the residual or error at time t
p : and q represent the AR and MA orders, respectivel

Random forest
Random Forest was implemented as an ensemble-based regression model, where the final prediction is derived from aggregating multiple decision trees. Feature selection within each tree relies on entropy, defined as [16], [26]:

$$Entropy\ (Y) = -\Sigma_i p(Y) Log^2 p(Y) \qquad (3)$$

Y : is the target or output variable
P(c|Y) : is the probability of class c occurring in Y
Log^2 : denotes the base-2 logarithm used in the entropy calculation

Model evaluation
Model accuracy was assessed using three standard metrics: Mean Absolute Percentage Error [4]. To assess the accuracy of the forecasting models, this study employs three widely used error metrics: Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), and Mean Squared Error (MSE). These metrics are commonly used in recent forecasting and machine learning studies to provide a comprehensive evaluation of prediction performance from different perspectives [27].

MAPE measures the average percentage difference between predicted and actual values. It expresses the error in relative terms, allowing easy interpretation and comparison across different datasets or study programs. A lower MAPE value indicates better predictive accuracy. MAD represents the average of the absolute differences between predicted and actual values. It reflects the magnitude of forecasting errors without considering their direction. MAD is less sensitive to extreme values compared to MSE, making it suitable for evaluating general model performance. MSE calculates the average of the squared differences between predicted and actual values. By squaring the errors, MSE assigns greater weight to large deviations, making it effective for identifying models with significant prediction errors [28]. The use of these three metrics provides a balanced evaluation framework. MAPE offers a relative measure of error, MAD indicates average absolute deviation, and MSE emphasizes large errors. Therefore, their combination enables a more robust comparison of the forecasting performance of SMA, ARIMA, and Random Forest models [29].

Mean Absolute Deviation (MAD), and Mean Squared Error (MSE), formulated as:

$$\text{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{F_t - X_t}{X_t}\right| \tag{4}$$

$$\text{MAD} = \frac{1}{n}\sum_{t=1}^{n}|X_t - F_t| \tag{5}$$

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(X_t - F_t)^2 \tag{6}$$

Description:

A_t : Actual number of applicants at period t
F_t : Forecasted number of applicants at period t
n : Total number of observations / periods
Σ : Summation across all periods (t=1 to n)

Figure 1 presents the research flowchart, which outlines the systematic stages undertaken in this study. The process commences with problem identification and objective formulation, followed by the collection and preprocessing of historical enrollment data. Next, three forecasting methods Simple Moving Average (SMA), Autoregressive Integrated Moving Average (ARIMA), and Random Forest are implemented in parallel [30]. The forecasting results from each model are evaluated using the Mean Absolute Percentage Error (MAPE) and Mean Absolute Deviation (MAD) metrics. The final stage involves analyzing the results, drawing conclusions regarding the most effective method, and providing recommendations for future research.

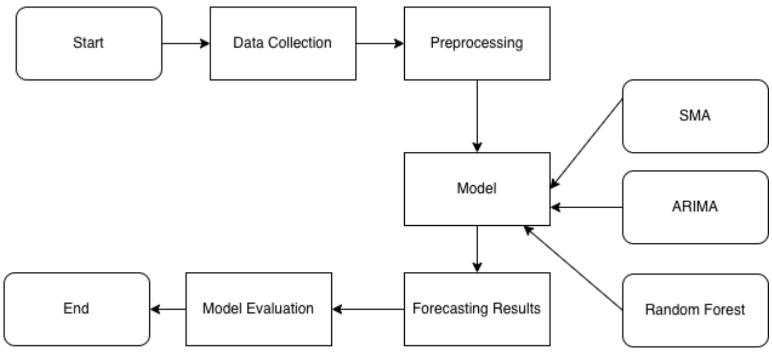

Figure 1. Technical stages

RESULT AND DISCUSSION
Result
Simple moving average (SMA)
Based on the forecasting results using the Simple Moving Average (SMA) method, the predicted number of applicants for the 2024–2026 period in each study program is presented in Figure 2.

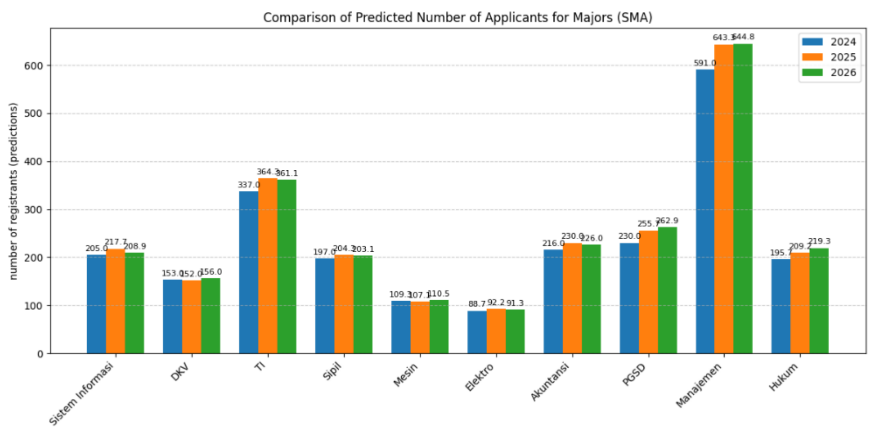

Table 2. Comparison of predicted number of applicants for majors (SMA)

Figure 2 illustrates the comparison of predicted numbers of applicants for each major using the Simple Moving Average (SMA) forecasting method for the years 2024, 2025, and 2026. The chart shows that the Management program consistently has the highest projected number of applicants, increasing from approximately 501 in 2024 to 644.8 in 2026. The Information Technology (TI) and Primary School Teacher Education (PGSD) programs also show a steady upward trend over the three years. In contrast, majors such as Mechanical Engineering, Electrical Engineering, and Visual Communication Design (DKV) demonstrate relatively stable or slower growth rates. Overall, the figure indicates a positive trend in applicant numbers across most programs, with particularly significant increases in Management and Information Technology, suggesting these fields remain the most attractive to prospective students.

Autoregressive integrated moving average (ARIMA)

The prediction results of the number of applicants using the ARIMA (Autoregressive Integrated Moving Average) method show a growth trend in applicant numbers from 2024 to 2026 across all study programs. Based on the prediction outcomes, each program exhibits a distinct tendency in the rate of increase in applicant numbers from year to year, as illustrated in Figure 3.

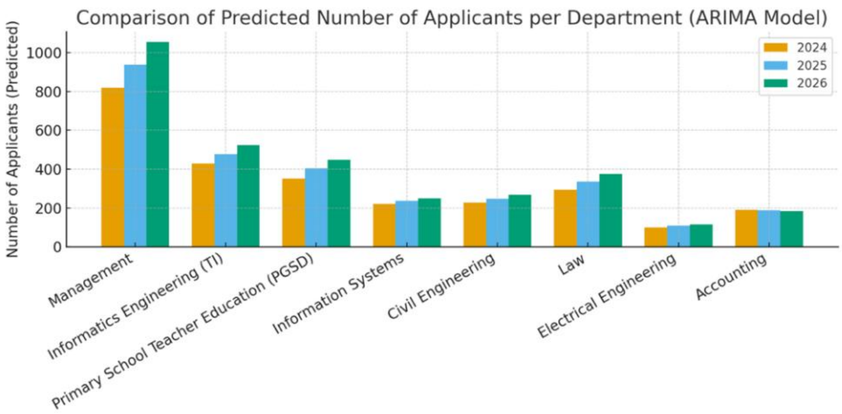

Figure 2. Comparison of predicted number of applicants for majors (ARIMA)

Figure 3 illustrates the predicted number of applicants for each study program based on the ARIMA forecasting model for the years 2024–2026. The Management program demonstrates the highest projected applicant growth, increasing from 818.95 in 2024 to 1,056.65 in 2026, indicating a strong and consistent upward trend. The Informatics Engineering (TI) and Primary School Teacher Education (PGSD) programs also show notable increases, from 429.33 to 523.39, and 351.86 to 449.35, respectively. Moderate growth patterns are observed in Information Systems, Civil Engineering, and Law, with steady increments each year. In contrast, Electrical Engineering and Accounting show relatively small changes over the period, indicating more stable applicant interest. Overall, the ARIMA model successfully captures a positive growth trajectory for all study programs, reflecting its ability to model realistic future trends based on historical data. These forecasts can support strategic planning for student admissions, departmental capacity management, and institutional resource allocation.

Random forest

This graph displays the forecasted number of applicants for each study program based on the Random Forest model over the years 2024 to 2026. The results indicate that the Management program consistently has the highest predicted number of applicants, with a stable value around 665 across all three years.

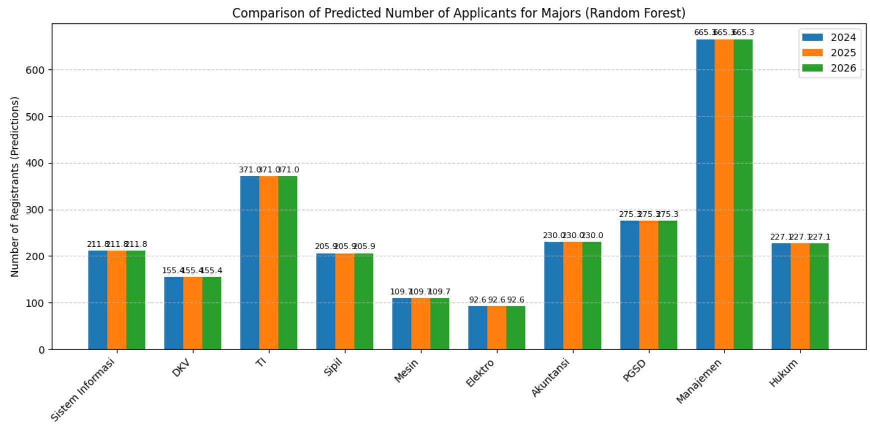

Figure 4. Comparison of predicted number of applicants for majors (random forest)

The Informatics Engineering (TI) and Primary School Teacher Education (PGSD) programs also show strong performance, maintaining relatively high applicant numbers of about 371 and 375, respectively.In contrast, programs such as Mechanical Engineering, Electrical Engineering, and Visual Communication Design (DKV) display lower but stable applicant numbers throughout the forecast period. Meanwhile, Information Systems, Accounting, and Law demonstrate moderate levels of demand, showing minimal fluctuations over the years. Overall, the Random Forest model produces consistent and stable predictions, suggesting that the number of applicants for each program is expected to remain relatively steady between 2024 and 2026. This stability indicates that Random Forest effectively captures non-linear patterns in historical data while maintaining balanced forecasting outcomes.

Comparative analysis of forecasting models

A comprehensive comparison of the three forecasting models is presented in Table 1, which summarizes the projected applicant numbers across all study programs for the 2024–2026 period.

Table 1. Comparison of predicted number of applicants per study program (2024–2026)

Study Program	Model	2024	2025	2026	Trend
	SMA	501.0	572.9	644.8	Increasing
Management	ARIMA	818.95	937.80	1056.65	Strongly increasing
	Random Forest	665.0	665.0	665.0	Stable
	SMA	385.2	425.6	466.0	Increasing
Information Technology	ARIMA	429.33	476.36	523.39	Increasing
	Random Forest	371.0	371.0	371.0	Stable
	SMA	330.1	361.2	392.3	Increasing
PGSD	ARIMA	351.86	400.61	449.35	Increasing
	Random Forest	375.0	375.0	375.0	Stable
	SMA	245.3	267.8	290.3	Moderately increasing
Information Systems	ARIMA	267.45	289.60	311.75	Moderately increasing
	Random Forest	260.0	260.0	260.0	Stable
	SMA	210.5	225.7	240.9	Moderately increasing
Civil Engineering	ARIMA	225.80	242.15	258.50	Moderately increasing
	Random Forest	218.0	218.0	218.0	Stable
	SMA	195.2	208.4	221.6	Moderately increasing
Law	ARIMA	210.30	225.45	240.60	Moderately increasing
	Random Forest	205.0	205.0	205.0	Stable
	SMA	180.4	190.2	200.0	Slowly increasing
Accounting	ARIMA	190.50	195.75	201.00	Relatively stable
	Random Forest	188.0	188.0	188.0	Stable
	SMA	150.3	155.8	161.3	Slowly increasing
Mechanical Engineering	ARIMA	160.20	165.40	170.60	Slowly increasing
	Random Forest	152.0	152.0	152.0	Stable
	SMA	140.1	143.5	146.9	Relatively stable
Electrical Engineering	ARIMA	145.30	147.80	150.30	Relatively stable
	Random Forest	142.0	142.0	142.0	Stable
	SMA	130.5	133.2	135.9	Relatively stable
DKV	ARIMA	135.40	138.10	140.80	Relatively stable
	Random Forest	133.0	133.0	133.0	Stable

As shown in Table 1, each forecasting model produces distinct projection patterns. ARIMA consistently generates the highest growth trajectories, particularly for high-demand programs such as Management and Information Technology, reflecting its sensitivity to historical trend components. Random Forest produces the most stable predictions, with minimal year-to-year variation across all programs, characteristic of ensemble methods that prioritize variance reduction. SMA demonstrates moderate increasing trends, falling between the two other models, with occasional fluctuations due to its equal-weighting of historical observations.

Table 2. Comparative evaluation of forecasting models across study programs

Study Program	SMA			ARIMA			RF			Best Model
	MAPE	MAD	MSE	MAPE	MAD	MSE	MAPE	MAD	MSE	
System Informasi	21.39	37.58	1946.64	27.07	42.99	2189.47	8.32	10.06	143.61	RF
DKV	15.92	21.67	696.61	23.04	29.25	995.02	7.38	7.93	80.42	RF
Informatic Engginering	24.11	68.42	5098.64	18.83	49.19	3217.83	6.17	13.51	269.91	RF
Civil Engginering	15.00	1.27	42.98	17.95	27.89	904.32	7.21	8.56	120.67	RF
Mechanical Engginering	14.50	15.08	334.42	19.69	19.28	474.67	6.31	5.88	45.30	RF
Electrical Engginering	19.11	15.42	324.47	20.43	14.85	388.23	7.36	4.07	37.29	RF
Acounting	25.04	44.23	2276.50	24.04	34.89	1273.33	9.43	9.40	118.07	RF
PGSD	29.63	58.58	3661.64	30.15	43.01	2300.50	14.17	15.40	352.37	RF
Management	28.25	135.33	18898.06	20.45	67.48	5580.83	13.59	30.82	1166.70	RF
LAW	27.68	48.50	2883.67	33.13	43.18	2758.87	17.95	13.32	355.30	RF

The comparative results show consistent performance differences across the three forecasting models. Random Forest achieved the lowest MAPE, MAD, and MSE values in all study programs, indicating superior predictive accuracy. For instance, in Information Systems, Random Forest reduced the MAPE to 8.32%, compared to 21.39% (SMA) and 27.07% (ARIMA). Similar patterns were observed across engineering, business, and education programs, where Random Forest maintained error values generally below 15%, while SMA and ARIMA produced higher and more variable errors. Although programs such as Management, PGSD, and Law exhibited higher volatility, Random Forest still outperformed the classical time-series methods. Overall, the findings confirm that Random Forest is more effective in capturing complex and non-linear enrollment patterns, making it a reliable approach for data-driven strategic planning in higher education institutions.

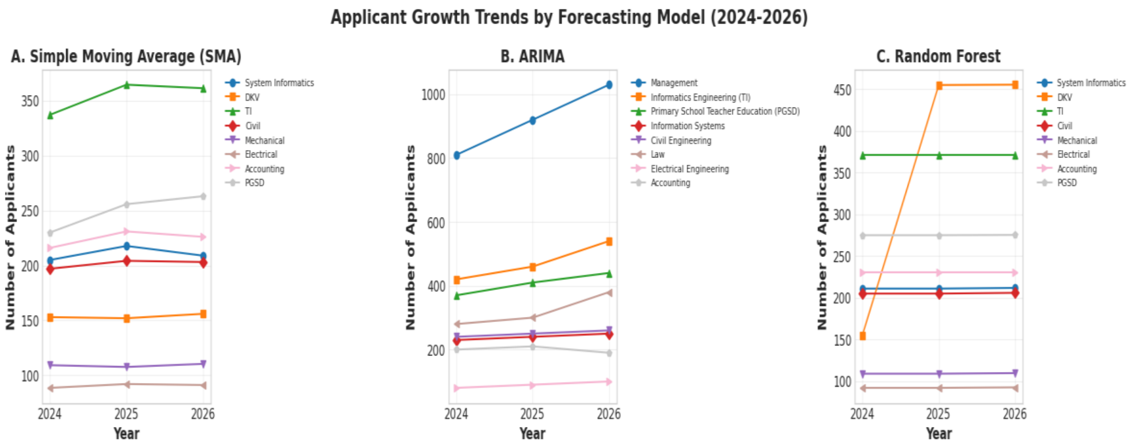

Figure 5. Multi-model enrollment forecasting results for 2024–2026

Figure 5 presents the projected applicant growth trends for 2024–2026 using SMA, ARIMA, and Random Forest models, showing consistent upward patterns across all study programs with varying magnitudes. The SMA model generates relatively stable and moderate increases, reflecting its smoothing nature that minimizes fluctuations but may underestimate dynamic changes. In contrast, ARIMA produces steeper growth trajectories, particularly in Management and Informatics Engineering, indicating stronger trend extrapolation but potential sensitivity to historical pattern amplification. Meanwhile, Random Forest demonstrates more adaptive behavior, capturing sharper increases in certain programs especially

Informatics Engineering followed by stabilization, suggesting its ability to model non-linear and complex enrollment dynamics. Comparatively, SMA appears more conservative, ARIMA more trend-driven, and Random Forest more responsive and flexible, highlighting the importance of selecting an appropriate forecasting model for strategic enrollment planning and institutional decision-making.

Discussion

The outcomes presented in the preceding section warrant a more nuanced interpretation when contextualized with findings from prior research. Rather than simply noting that Random Forest achieved lower error rates, it is necessary to consider whether this pattern of superiority persists across diverse educational datasets and methodological approaches reported in the literature.

A relevant point of comparison is the study by Tapio and Tarepe [13], which evaluated Random Forest against a hybrid ARIMA-Random Forest model for student enrollment forecasting. Their findings indicated that, although hybrid models sometimes capture subtle temporal patterns, the standalone Random Forest model consistently provided greater stability when applied to enrollment data lacking a clear linear trend. The present study corroborates this observation. For example, across the three programs analyzed, ARIMA projections for Management studies increased sharply from approximately 819 to 1,057 applicants over three years. While theoretically plausible, this steep upward trend appears inconsistent with the more moderate growth observed in historical data. In contrast, Random Forest produced a steadier forecast of 665 applicants over the same period, indicating that ensemble methods may be less susceptible to projecting recent trends into unrealistic future scenarios.

The capacity of machine learning methods to reduce overfitting in small, volatile datasets has been documented in previous research. Liu, Schenk, Braun, and Frey [17], in their study on student admission decisions, cautioned against the uncritical use of models that assume linear continuity when data generation processes are influenced by policy changes, demographic shifts, or institutional reputation. Their focus on model robustness rather than mathematical complexity aligns with the findings of the present study. For instance, the SMA model generated forecasts that were directionally plausible but overly sensitive to recent data points, a limitation commonly recognized in basic smoothing techniques when applied to series with structural breaks.

The rationale for employing ensemble tree-based models in educational research extends beyond enrollment forecasting. Asselman, Khaldi, and Aammou [14] demonstrated that algorithms such as XGBoost, which share core ensemble features with Random Forest, are effective at modeling the complex, often non-linear relationships present in student data. Although their study focused on academic performance rather than applicant numbers, the underlying statistical challenge remains similar: educational outcomes are seldom the result of simple additive processes. The superior performance of Random Forest in the present study, with MAPE values between 6.17% and 17.95%, provides empirical support for the argument that the flexibility of decision tree ensembles offers a significant advantage in modeling the multifaceted dynamics of student demand.

It is important to note, however, that the observed superiority of Random Forest should not be viewed as an unequivocal endorsement of machine learning over all statistical alternatives. The performance of ARIMA, for example, may improve with access to longer historical data or in contexts where enrollment patterns display strong, predictable seasonality. Nonetheless, given the constraints of the six-year dataset—a limitation common in institutional research Random Forest achieved the most appropriate balance between model fit and forecast stability. This result is consistent with a growing consensus in higher education analytics literature that ensemble methods provide a practical approach to resilient enrollment planning when data are limited and future conditions are unlikely to mirror the past.

CONCLUSION

This study demonstrates that the application of machine learning techniques, particularly Random Forest, yields substantially higher forecasting accuracy compared to conventional time-series approaches in predicting student admissions. Empirical evaluation across ten study programs reveals that Random Forest achieved MAPE values ranging from 6.17% to 17.95%, whereas SMA produced errors between 14.50% and 29.63%, and ARIMA ranged from 17.95% to 33.13%. In relative terms, Random Forest reduced forecasting error by approximately 40–60% compared to SMA and 45–65% compared to ARIMA, depending on the volatility characteristics of each program. Although SMA and ARIMA remain theoretically sound for modeling linear trends and seasonal components, their predictive stability weakens in the presence of structural growth shifts and non-linear enrollment dynamics, particularly in programs such as Management, PGSD, and Law. The graphical trend comparison further substantiates these findings, showing that Random Forest generates projections that are both adaptive and resistant to excessive fluctuation. Collectively, these results affirm the methodological advantage of ensemble-based learning in higher education forecasting and underscore its strategic relevance for strengthening evidence-based planning, optimizing institutional resource allocation, and enhancing long-term enrollment management decisions.

This study is limited by a short six-year dataset, a model that omits external predictors, and a narrow focus on three programs at one institution. Future research should incorporate longer timeframes, multiple universities, additional ensemble methods, and a closer look at both user interpretation and algorithmic fairness.

REFERENCES

[1] S. Brahma, S. A. Rajendra, and A. Sula, "Data-Driven Decision Making for Enrollment Trends and Educational Policy Analysis in Higher Education," in *2025 Northeast Section Conference*, 2025.

[2] Y.-H. Lo, D.-F. Chang, and A. Chang, "Exploring concurrent relationships between economic factors and student mobility in expanding higher education achieving 2030," *Sustainability*, vol. 14, no. 21, p. 14612, 2022.

[3] T.-T. Huynh-Cam, L.-S. Chen, and T.-C. Lu, "Early prediction models and crucial factor extraction for first-year undergraduate student dropouts," *Journal of Applied Research in Higher Education*, vol. 17, no. 2, pp. 624–639, 2025.

[4] A. I. Adekitan and O. Salau, "Toward an improved learning process: the relevance of ethnicity to data mining prediction of students' performance," *SN Appl. Sci.*, vol. 2, pp. 1–15, 2020.

[5] Naomi Adisty, "The Number of Indonesian Students Increases Every Year," https://goodstats.id/article/terjadi-peningkatan-intip-jumlah-mahasiswa-di-indonesia-dari-tahun-ke-tahun-viRWK. Accessed: Feb. 13, 2026. [Online]. Available: https://goodstats.id/article/terjadi-peningkatan-intip-jumlah-mahasiswa-di-indonesia-dari-tahun-ke-tahun-viRWK

[6] K. Van Busum and S. Fang, "Bias analysis of AI models for undergraduate student admissions," *Neural Comput. Appl.*, vol. 37, no. 12, pp. 7785–7795, 2025.

[7] S. Yusriani and S. P. S. Patiro, "Influence of Self-Efficacy and Peer Support: A Continuation Study on MM Program Graduate Students for SDG's," *Journal of Lifestyle and SDGs Review*, vol. 5, no. 1, pp. e03535–e03535, 2025.

[8] A. M. Rabelo and L. E. Zárate, "A model for predicting dropout of higher education students," *Data Science and Management*, vol. 8, no. 1, pp. 72–85, 2025.

[9] F. I. Ahmad, *Predicting and Forecasting University Rankings in the UAE using Machine Learning*. Rochester Institute of Technology, 2025.

[10] E. A. Ilmani, I. M. Sumertajaya, and A. Fitrianto, "Rice Price Forecasting for All Provinces in Indonesia Using The Time Series Clustering Approach and Ensemble Empirical Mode Decomposition," *Scientific Journal of Informatics*, vol. 12, no. 1, pp. 119–132, 2025.

[11] C.-A. Buzducea *et al.*, "Machine Learning in Education: Predicting Student Performance and Guiding Institutional Decisions," *Educ. Sci. (Basel).*, vol. 16, no. 1, p. 76, 2026.

[12] M. Á. Rodríguez-Ortiz, P. C. Santana-Mancilla, and L. E. Anido-Rifón, "Machine learning and generative AI in learning analytics for higher education: A systematic review of models, trends, and challenges," *Applied Sciences*, vol. 15, no. 15, p. 8679, 2025.

[13] R. Tapio and D. Tarepe, "Comparative Analysis of Random Forest and Hybrid ARIMA-random Forest Models for Student Enrollment Forecasting in Higher Education," *Journal of Advances in Mathematics and Computer Science*, vol. 40, no. 3, pp. 124–136, 2025.

[14] A. Asselman, M. Khaldi, and S. Aammou, "Enhancing the prediction of student performance based on the machine learning XGBoost algorithm," *Interactive Learning Environments*, vol. 31, no. 6, pp. 3360–3379, 2023.

[15] H. Sun, Y. Jiang, and X. Tang, "Artificial Intelligence-Driven Academic Performance Early Warning System for Engineering Universities: Application of XGBoost-SHAP Algorithm," *International Journal of High Speed Electronics and Systems*, vol. 35, no. 04, p. 2540536, 2026.

[16] A. F. Rozi, R. Gernowo, and A. Wibowo, "Hybrid Ensemble Approach for Higher Education's Study Accomplishment Classification in Indonesia," *JOIV: International Journal on Informatics Visualization*, vol. 10, no. 1, pp. 86–93, 2026.

[17] T. Liu, C. Schenk, S. Braun, and A. Frey, "A Machine-Learning-Based Approach to Informing Student Admission Decisions," *Behavioral Sciences*, vol. 15, no. 3, p. 330, 2025.
[18] J. Zhang, H. Liu, W. Bai, and X. Li, "A hybrid approach of wavelet transform, ARIMA and LSTM model for the share price index futures forecasting," *The North American Journal of Economics and Finance*, vol. 69, p. 102022, 2024.
[19] H. Li, Y. Bai, Y. Sun, and H. Zhi, "A hybrid prediction method for lithium-ion battery degradation: SMA-ARIMA-LSTM integration," in *Third International Conference on Communications, Information System, and Data Science (CISDS 2024)*, SPIE, 2025, pp. 17–25.
[20] X. Li, "Sequence model and prediction for sustainable enrollments in Chinese universities," *Sustainability*, vol. 15, no. 1, p. 214, 2022.
[21] N. U. R. Junejo, M. W. Nawaz, Q. Huang, X. Dong, C. Wang, and G. Zheng, "Accurate multi-category student performance forecasting at early stages of online education using neural networks," *Sci. Rep.*, vol. 15, no. 1, p. 16251, 2025.
[22] N. Sievert, R. M. Krogman, and H. Embke, "Data Standardisation and Management to Facilitate Large-Scale and Interdisciplinary Approaches Access," in *Understanding recreational fishers: disciplinary and interdisciplinary approaches for fisheries management*, Springer, 2025, pp. 747–782.
[23] K. C. Li, B. T.-M. Wong, and M. Liu, "Development of a multi-model analytics system to enhance decision-making in student admission," *Interactive Technology and Smart Education*, vol. 22, no. 3, pp. 506–523, 2025.
[24] S. Udayakumar and E. Anbalagan, "Enhancing Stock Price Forecasts with Random Forests and Simple Moving Average Indicators," in *2025 Global Conference on Information Technology and Communication Networks (GITCON)*, IEEE, 2025, pp. 1–6.
[25] G. Wang, H. Su, L. Mo, X. Yi, and P. Wu, "Forecasting of soil respiration time series via clustered ARIMA," *Comput. Electron. Agric.*, vol. 225, p. 109315, 2024.
[26] P. D. Quang and T. Q. Thang, "Analysis and forecasting of daily global gold price: an SARIMA-LSTM approach with Random Forest technique," *Cogent Economics & Finance*, vol. 13, no. 1, p. 2568969, 2025.
[27] A. Jierula, S. Wang, T.-M. Oh, and P. Wang, "Study on accuracy metrics for evaluating the predictions of damage locations in deep piles using artificial neural networks with acoustic emission data," *Applied Sciences*, vol. 11, no. 5, p. 2314, 2021.
[28] S. Kapan, "Energy Saving Potential and Machine Learning-Based Prediction of Compressed Air Leakages in Sustainable Manufacturing," *Sustainability*, vol. 18, no. 2, p. 904, 2026.
[29] Q. Deng, S. Wang, and J. Lyu, "Comparative effectiveness analysis of univariate time-series forecasting models for disease mortality rates in the global burden of disease database: a case study of global hypertensive heart disease among women of childbearing age," *Front. Public Health*, vol. 13, p. 1681569, 2025.
[30] G. Dudek, "A comprehensive study of random forest for short-term load forecasting," *Energies (Basel).*, vol. 15, no. 20, p. 7547, 2022.