Funnel-Based Predictive Modeling for Forecasting Student Admissions in Higher Education

Obaja Marum Lumbanraja*[1]
[1,2]Institution/affiliation; Jln Kol. Masturi no.288, Kabupaten Bandung Barat, Jawa Barat, Universitas Advent Indonesia
[3]Sistem Informasi, Fakultas Teknologi Informasi, Universitas Advent Indonesia
e-mail: ***[1]obaja.lumbanraja@unai.edu**

Abstract

Forecasting student admissions remains a challenge due to fluctuating online engagement and complex administrative processes. Existing predictive models rarely integrate website behavioral data with institutional admission funnels, resulting in lower accuracy. This study bridges that gap by combining web analytics from Google Analytics 4 (GA4) with administrative enrollment funnel data from the admission of new students (Penerimaan Mahasiswa Baru/PMB) system to develop a unified predictive framework. The approach strengthens forecasting by aligning digital behavior with verified enrollment milestones. A quantitative explanatory design was employed, applying Pearson correlation to identify linear relationships and Seasonal ARIMA (SARIMA) to model cyclical admission trends. The dataset includes GA4 metrics sessions, engagement rate, bounce rate, and events per session and PMB funnel stages from account creation to confirmed enrollment. Results reveal strong correlations ($r > 0.9$, $p < 0.001$) between digital engagement and mid-funnel conversions, while SARIMA achieved its highest accuracy for early-stage predictions (MAPE ≈ 19%). Forecasts for final outcomes were less accurate, reflecting administrative variability. These findings confirm that web engagement metrics are reliable leading indicators of student interest and mid-stage commitment. This research establishes a reproducible pipeline unifying web analytics (GA4) with institutional funnel data (PMB), providing empirical evidence that digital engagement is a reliable leading indicator of early and mid-stage commitment, thereby forming a novel and adaptable foundation for data-driven enrollment planning.

Keywords— Predictive Modeling, Funnel Analytics, Student Admissions, SARIMA, Higher Education

1. INTRODUCTION

Forecasting student admissions is a persistent challenge for higher education institutions because applicant interest fluctuates with academic calendars and administrative timelines, yet decisions about staffing, budgeting, and marketing must be made ahead of time [2]. Although time-series methods have shown promise for modeling admission seasonality and improving short-term planning in university contexts, many institutions still underutilize behavioral data already generated by their own systems [3]. In practice, routinely collected indicators such as visit counts, traffic sources, and the timing/frequency of visits can be linked to concrete stages of the admission funnel (account creation, form completion, personal-data submission, and final registration marked by student ID number issuance) to yield actionable signals of intent [4]. However, few studies have operationalized this linkage between web activity and administrative conversion events over time in a unified forecasting workflow [1].

Concurrently, the admissions literature has broadened to include predictive and algorithmic approaches ranging from time-series forecasting of enrollment to machine-learning models that support admission decisions while calling for transparent, fair use of these tools in practice [8]. Related work in data-driven education demonstrates that forecasting enrollment can support planning and resource allocation when cyclical peaks recur, reinforcing the operational value of short-horizon predictions [9]. Yet across these lines of inquiry, integrating website engagement signals with institutional funnel records remains comparatively uncommon, despite evidence that such integration can enhance early and mid-funnel prediction where behavioral signals are strongest [5].

This study addresses that gap with a funnel-based predictive modeling framework that maps website engagement to admission of new students (Penerimaan Mahasiswa Baru/PMB) funnel stages like Accounts Created (Jumlah Akun), Applications Confirmed (Jumlah Formulir), and Confirmed Enrollment (Jumlah Data Diri) and models their joint temporal behavior. By aligning digital behavior (e.g., sessions, engaged sessions, engagement rate, events per session, bounce rate) with verified administrative milestones in a single time-indexed dataset, we enable inference and short-term forecasting tailored to the admission cycle [6]. By aligning digital behavior (e.g., sessions, engaged sessions, engagement rate, events per session, bounce rate) with verified administrative milestones in a single time-indexed dataset, we enable inference and short-term forecasting tailored to the admission cycle. To capture seasonality, trend, and autocorrelation in monthly data [7], the author employs Seasonal ARIMA (SARIMA). This methodology is specifically chosen over generic regression or complex machine learning models because SARIMA explicitly models trend, autocorrelation, and seasonality jointly within a single probabilistic framework. Furthermore, SARIMA is widely recommended and established for analyzing cyclical administrative processes such as university admissions, generating statistically sound, interpretable, and operationally useful short-horizon forecasts necessary for planning and resource allocation. The contribution is twofold: (i) a reproducible pipeline unifying web analytics and administrative funnel records for admission forecasting, and (ii) empirical evidence that this integration is especially informative for early and mid-funnel stages, where behavioral signals are predictive of progression [10].

2. RESEARCH METHODS

This study followed the pipeline shown in Figure 1: Data Collection, Data Preprocessing, Exploratory Data Analysis (EDA), and Modeling & Forecasting. The goal was to align behavioral web analytics with verified PMB funnel events in a single monthly time series and then build seasonal forecasts tailored to UNAI's admission cycles. For forecasting, we used Seasonal ARIMA (SARIMA) because it explicitly models trend, autocorrelation, and seasonality in monthly data and is widely recommended for cyclical administrative processes such as university admissions [11]. Model adequacy and accuracy were assessed with standard time-series diagnostics and error measures (e.g., MAE, RMSE, MAPE) following best-practice guidance for forecast evaluation [12].

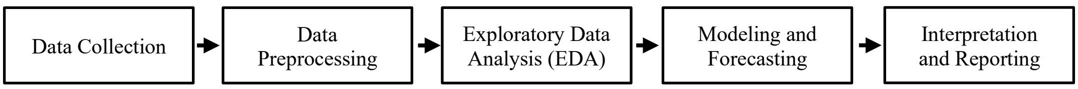

Figure 1. Block Diagram of Research Flow

2.1. Operational Definitions

To ensure consistency and conceptual precision, each variable was defined according to its source system and retained at monthly granularity to match the study's forecasting horizon (see Table 1 and Table 2).

Website engagement metrics were extracted from GA4 for the admissions portal pmb.unai.edu. In GA4, Sessions denote the number of visit instances within a month; Engaged Sessions are those lasting longer than ten seconds, including at least one conversion event, or containing two or more page views; Engagement Rate is the proportion of engaged sessions relative to total sessions; Bounce Rate captures the share of sessions that ended after a single interaction; Events per Session and Views per Session summarize interaction depth and page-consumption intensity. All GA4 definitions follow the official product documentation [11]. For data validation and visual inspection, the GA4 exports were cross-checked using Looker Studio dashboards before integration [12].

Administrative funnel variables were sourced from the PMB system. Accounts Created represents the number of prospective students who successfully created and activated an account. Applications Confirmed indicates applicants who completed and submitted the admission form and were verified by PMB administrators. Confirmed Enrollment refers to applicants who passed screening, completed required personal data, and were issued a student ID number (Nomor Induk Mahasiswa/NIM), thereby marking official enrollment.

Both datasets span January 2022 to September 2025 and were aggregated monthly to capture the institution's cyclical recruitment windows odd semester (approximately February–July) and even semester (September–January). This temporal structure is later modeled with seasonal components to reflect predictable peaks in admissions activity [13]. Data were stored in Microsoft Excel and prepared for analysis in Python (for merging, cleaning, and statistical analysis) and Orange Data Mining (for complementary EDA workflows), preserving reproducibility from raw export to model-ready table.

Table 1. Dataset

Month	Accounts Created	Applications Confirmed	Confirmed Enrollment	Sessions	Engaged sessions	Engagement rate	Bounce rate	Events per session	Views per session
2022-02	1	0	0	173	129	0,702	0,298	5,412	1,763
2022-03	64	43	46	1693	1218	0,626	0,374	10,535	4,053
2022-04	41	17	20	1254	860	0,465	0,535	4,939	1,675
2022-05	58	21	28	1388	1027	0,497	0,503	6,900	2,517
2022-06	149	95	89	2774	2038	0,658	0,342	8,472	3,126
…	…	…	…	…	…	…	…	…	…
2025-06	133	93	73	4093	2948	0,640	0,360	10,007	3,811
2025-07	195	176	167	5937	4432	3,904	2,096	55,148	19,861
2025-08	128	120	171	6243	4615	3,916	2,084	52,654	19,214
2025-09	11	11	15	1505	979	2,682	2,318	31,380	10,755

Table 2. Summary of Data Sources and Variables

Source	Data Type	Variables	Period Covered	Unit of Observation	Description
Google Analytics 4 (GA4)	Web Analytics	Sessions, Engaged Sessions, Engagement Rate, Bounce Rate, Events per Session, Views per Session	Jan 2022 - Sep 2025	Monthly	Behavioral metrics from PMB admission portal (pmb.unai.edu)
PMB Administrative System	Admission Funnel	Accounts Created Applications Confirmed Confirmed Enrollment	Jan 2022 - Sep 2025	Monthly	Stages of prospective student conversion in the admission process

2.2. Data Preprocessing

Before analysis, a structured preprocessing procedure was implemented to ensure consistency across sources and suitability for quantitative modeling (Figure 2). The process began with data cleaning, where duplicate entries were removed, missing values were reviewed and addressed, and any temporal inconsistencies were corrected. Each record was then reformatted to a standardized YYYY-MM key so that the time dimension aligned across both systems an essential step for monthly time-series analysis and later seasonal modeling [13].

Next, harmonization and normalization were applied to align units and scales. GA4 indicators that are defined as rates (e.g., Engagement Rate, Bounce Rate) were retained as such, while volume measures (e.g., Sessions, Engaged Sessions, Events, Views) were aggregated to monthly totals consistent with GA4 semantics [11]. PMB administrative counts (Accounts Created, Applications Confirmed, Confirmed Enrollment) were likewise aggregated by month to create comparable observations.

The two sources were then integrated using the shared Month-Year key to produce a single analytical table containing matched web-engagement and funnel variables for each period. Finally, a validation step was performed by cross-checking random samples against the original GA4 export and the PMB administrative views, and by visually verifying monthly totals via Looker Studio dashboards to confirm end-to-end consistency [12]. The resulting dataset serves as the foundation for the subsequent exploratory analysis and SARIMA-based forecasting.

Figure 2. Data Preprocessing Workflow

2.3. Exploratory Data Analysis (EDA)

After preprocessing, Exploratory Data Analysis (EDA) was conducted to examine the statistical structure, temporal behavior, and inter-variable relationships in the combined GA4–PMB dataset. The objective was to surface patterns that inform model choice and seasonal parameterization. All procedures were executed in Python (Jupyter Notebook) and Orange Data Mining for complementary visual analytics.

First, Descriptive and Distributional Analysis. Descriptive statistics were computed for each variable to summarize its central tendency and dispersion. Metrics such as mean (μ), median, standard deviation (σ), minimum, and maximum values were calculated using the following expressions:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i-1}^{n}(x_i - \mu)^2} \tag{2}$$

where x_i represents each observation and n is the total number of months in the dataset [15]. This step was designed to detect irregularities, missing observations, and outliers before statistical modeling. Distribution plots and boxplots were generated to ensure each variable exhibited coherent temporal and statistical behavior.

Second, Temporal and Seasonal Pattern Analysis. To investigate the presence of seasonality and recurring temporal patterns, time-series visualizations were produced for all key metrics from January 2022 to September 2025. Monthly aggregated line charts were used to inspect trends in website traffic (e.g., Sessions, Engagement Rate) and corresponding admission outcomes (Accounts Created, Applications Confirmed, Confirmed Enrollment). Seasonal cycles were analyzed using trend decomposition, where each series y_t was expressed as [11], [13]:

$$y_t = T_t + S_t + e_t \tag{3}$$

with T_t representing the long-term trend, S_t the seasonal component (12-month periodicity), and e_t the residual or irregular error term. This analysis confirmed whether the data were suitable for seasonal time-series modeling such as SARIMA, by evaluating the stability and periodic recurrence of peaks in the academic admission cycle.

Third, Correlation Analysis Framework. To quantify the linear relationships between GA4 and PMB variables, a Pearson product-moment correlation framework was applied. The coefficient rr for each variable pair was computed as:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{4}$$

Where x_i and y_i represent paired observations of two variables (e.g., Sessions and Applications Confirmed), and $\bar{x}$, $\bar{y}$ denote their respective means. Values of r range from -1 to +1, where positive coefficients indicate a direct relationship and negative coefficients indicate an inverse relationship. Statistical significance was assessed at the 0.05 level ($p < 0.05$), and results were tabulated and visualized through a correlation heatmap generated in Python using the Seaborn library. The correlation framework and inferential checks follow standard statistical guidance for time-series exploratory work [15].

These analyses established that the data exhibit stable, recurring seasonal patterns and strong early-/mid-funnel associations with web activity, justifying the use of seasonal time-series models in the next phase.

2.4. Modeling and Forecasting

The forecasting component of this study employs a Seasonal AutoRegressive Integrated Moving Average model, denoted

$$SARIMA(p, d, q)(P, D, Q)_s \quad (5)$$

Where p is the order of AR (Auto Regressive), d is the order of differencing, q is the order of MA (Moving Average), P,D,Q are seasonal parameters, and s is the seasonal period of 12 months. SARIMA is appropriate for admission data because it jointly handles trend, seasonality, and autocorrelation in a single probabilistic framework [11], [13]. One univariate model is estimated per target series Sessions, Accounts Created, Applications Confirmed, Confirmed Enrollment to capture each series' own temporal dynamics and to generate out-of-sample forecasts.

Temporal split. To emulate real decision settings and enable an honest accuracy assessment, the time series are partitioned into a training window (January 2022-April 2025) and a hold-out test window (May-September 2025). The training window is used for model identification and parameter estimation; the test window is reserved strictly for forecast evaluation.

Step 1. Stationarity assessment and transformations.
We first inspect seasonality and trend through time-plots and decomposition (EDA), then test (non)stationarity using ADF and/or KPSS on raw and differenced series. When variance visibly grows with level, consider a variance-stabilizing transform (e.g., log or Box Cox), applied consistently within a series. Nonseasonal differencing d and seasonal differencing D are set to the minimum orders that yield stationarity while avoiding over differencing (checked by auto correlation decay and unit-root tests) [2], [11].
Seasonality and trend are inspected via time plots and decomposition (Section 2.3). (Non)stationarity is checked on raw and differenced series using standard unit-root diagnostics as described in SARIMA practice [11]. If variance increases with level, a variance-stabilizing transform (e.g., log or Box–Cox) is considered and applied consistently within a series. Nonseasonal differencing d and seasonal differencing D are set to the minimum orders that achieve stationarity while avoiding over differencing (verified by autocorrelation decay and unit-root checks) [11].

Step 2. Order identification.
Conditional on (d, D), search a parsimonious grid for $(p,q,P,Q) \in \{0,1,2\}$with seasonal period $s=12$. Candidate specifications are ranked by AICc, favoring simpler structures when criteria are comparable. Preliminary ACF/PACF patterns and residual behavior guide exclusions (e.g., dropping redundant MA terms when ACF tails decay rapidly) [11], [13].

Step 3. Estimation.
Candidate models are estimated by maximum likelihood using stats models' SARIMAX routine with standard regularity constraints (enforce stationarity/enforce invertibility). Convergence diagnostics and parameter significance are examined; in case of instability or near non-invertibility, refit with reduced orders or re-evaluate differencing choices.

Step 4. Diagnostics.
Adequacy is confirmed via residual diagnostics: (i) whiteness checks using the Ljung Box portmanteau statistic on multiple lags; (ii) residual ACF/PACF inspection to ensure no remaining seasonal or short-lag autocorrelation; and (iii) distributional checks (mean ≈ 0, constant variance). Models failing diagnostics are revised (e.g., add a seasonal MA or lower a nonseasonal AR) and re-evaluated [11].

Step 5. Forecast generation and intervals.
For each accepted model, produce multi-step dynamic forecasts for the five-month test horizon (May-September 2025). Alongside point forecasts, prediction intervals report 80% and 95%, which quantify forecast uncertainty and are essential for operational planning (e.g., scheduling outreach or staffing).

Step 6. Accuracy evaluation.
Forecasts are compared with actual observations using three complementary metrics: MAE (average absolute error), RMSE (root-mean-square error, emphasizing larger misses), and MAPE (percentage error facilitating cross-series comparison) [9]. Formally, for horizon n with actuals y_t and forecasts $\hat{y}_t$:

$$\text{MAE} = \frac{1}{\text{n}}\text{n} \sum |y_t - \hat{y}_t| \tag{6}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum(y_t - \hat{y}_t)^2} \tag{7}$$

$$MAPE = 100.\frac{1}{n}\sum\frac{y_t - \hat{y}_t}{y_t} \tag{8}$$

Lower values indicate superior accuracy; MAPE is interpreted with caution when actual values are near zero [12].

Step 7. Robustness and season-aware interpretation.
Because UNAI admissions close in mid-August, forecasts produced for September are interpreted as early interest for the next cycle rather than as continuing the same cohort. To assess stability, the researcher optionally conducts rolling-origin validation (moving the cutoff backward by one month and repeating the evaluation), and the researcher verify that outliers or promotional shocks do not unduly drive parameter choices. Where late-funnel prediction is weak (e.g., Confirmed Enrollment), the researcher documents the limitation and note that exogenous factors (deadlines, campaign flags) would be handled in future SARIMAX/ML extensions rather than in univariate SARIMA [11]. [13].

Implementation. All modeling is executed in Python (Jupyter Notebook) using stats models for SARIMA estimation and pandas/matplotlib for data handling and visualization. Reproducibility is ensured by versioned notebooks, fixed random seeds (where applicable), and an analysis log that records selected orders, diagnostics, and evaluation results.

In sum, this procedure follows best practices for seasonal time-series forecasting: achieve stationarity with minimal differencing, select parsimonious orders via information criteria, validate with residual diagnostics, and evaluate on an untouched hold-out window yielding forecasts that are statistically sound and operationally useful for enrollment planning [11], [12], [13].

3. RESULT AND DISCUSSION

This section presents the empirical results of the funnel-based predictive modeling and interprets the findings in the context of student admission forecasting at Universitas Advent Indonesia (UNAI). Analyses include correlation evaluation between web engagement indicators and admission funnel variables, followed by seasonal forecasting using the SARIMA model. The goal is to demonstrate how digital behavioral patterns correspond with institutional enrollment outcomes and to assess the accuracy of time-series forecasting for each funnel stage.

3.1. Correlation Analysis

The first phase examined linear associations between GA4 metrics and PMB funnel variables using Pearson's product–moment correlation, a standard approach for quantifying strength and direction of paired monthly relationships in exploratory time-series work [16]. The resulting correlation matrix and triangular heatmap (Figure 3) show a consistent pattern of strong positive associations between website activity and admission performance, supporting the use of behavioral indicators as leading signals for funnel movement [17].

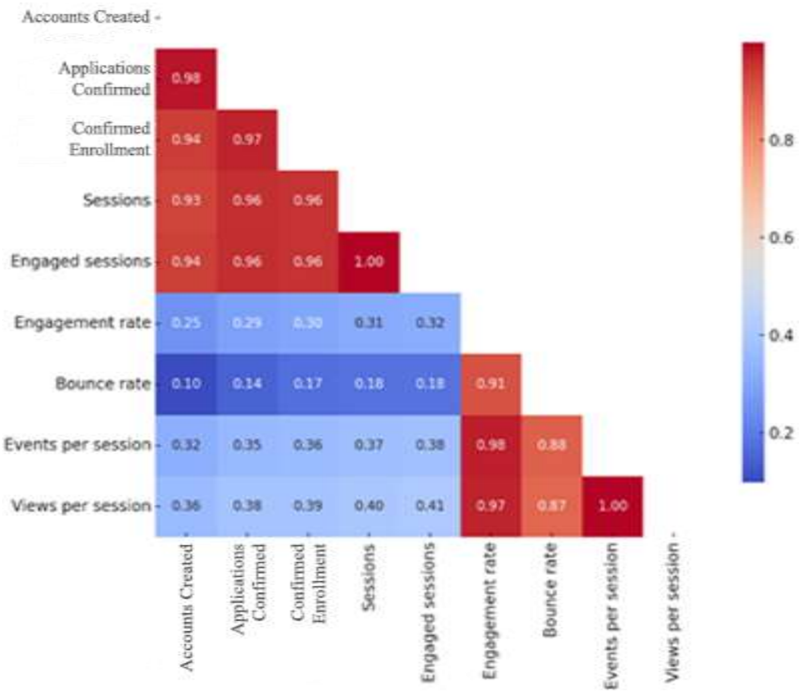

Figure 3. Triangular Correlation Heatmap

As summarized in Table 3, Sessions, Engaged Sessions, and Events per Session display very strong positive correlations with mid-funnel outcomes particularly Accounts Created and Applications Confirmed ($r \approx 0.94$–0.96; $p < 0.001$) indicating that higher visit volume and deeper on-site interaction are closely aligned with progression through the application process [18]. The result is consistent with prior evidence that repeated interaction and event-rich sessions are predictive of conversion in education-oriented digital funnels [19].

The Engagement Rate also shows strong positive relationships ($r \approx 0.74$–0.85) with both form submission and data-completion metrics, suggesting that longer, more meaningful sessions contribute to higher conversion likelihood [20]. Conversely, Bounce Rate exhibits a strong negative association ($r \approx -0.68$ to -0.75) with the same outcomes, reflecting the intuitive link between single-touch exits and reduced completion probability [21].

Table 3. Pearson Correlation Summary

GA4 Metric	PMB Funnel Variable	Pearson r	p-value	Interpretation
Sessions	Applications Confirmed	0.956	<0.001	Very Strong Positive
Engaged Sessions	Confirmed Enrollment	0.960	<0.001	Very Strong Positive
Events per Sessions	Applications Confirmed	0.942	<0.001	Very Strong Positive
Engagement Rate	Applications Confirmed	0.843	<0.001	Strong Positive
Bounce Rate	Applications Confirmed	-0.745	<0.001	Strong Negative

Taken together, these results indicate that traffic intensity and interaction quality jointly mirror the advancement of prospective students across funnel stages [22]. The magnitude and direction of the observed coefficients provide empirical justification for integrating GA4 behavioral signals with administrative PMB records in a unified predictive pipeline [23]. This alignment motivates the next step seasonal time-series forecasting where behavioral metrics are used to project short-horizon admission outcomes under the academic calendar's cyclicality [24], with robustness considerations informed by sliding-window evaluations and related operational studies [25].

3.2. SARIMA Forecasting Performance

Following the correlation analysis, a SARIMA model was estimated for each target series (Sessions, Accounts Created, Applications Confirmed, Confirmed Enrollment) to produce short-term forecasts for May–September 2025. The specification captured UNAI's cyclical admission calendar with recurring peaks in February–July and September–January, consistent with seasonal modeling guidance for monthly administrative data [24].

Model evaluation. Table 4 reports MAE, RMSE, and MAPE computed on the hold-out window. For Sessions, Accounts Created, and Applications Confirmed, MAPE < 25% indicates high short-horizon accuracy and supports the feasibility of using historical patterns to anticipate demand within the admission cycle [26]. In contrast, Confirmed Enrollment shows MAPE > 50%, which is expected because final registration is strongly influenced by nonseasonal administrative drivers (e.g., document verification and manual confirmations) that are not fully described by univariate seasonal dynamics [22].

Table 4. SARIMA Forecast Accuracy (May-Sep 2025)

Series	MAE	RMSE	MAPE (%)	Predictive Strength Category	Interpretation
Sessions	≈ 270	≈ 335	19.2	High (Excellent Fit)	High accuracy; captures web-traffic seasonality effectively
Accounts Created	≈ 215	≈ 298	21.0	High (Good Fit)	Reliable early-funnel prediction
Applications Confirmed	≈ 190	≈ 275	18.5	High (Strong Fit)	Strong mid-funnel predictive power
Confirmed Enrollment	≈ 120	≈ 190	51.3	Low (Weak Fit)	Less accurate due to administrative delay and nonseasonal shocks

While the RMSE values appear large in absolute terms, their relative scale to average monthly levels is < 20% for Sessions, Accounts Created, and Applications Confirmed evidence of practically useful accuracy for planning purposes [26]. The higher RMSE observed for Confirmed Enrollment reflects late-funnel variability that is driven by timing shocks rather than stable seasonality; such effects are better accommodated by exogenous regressors (e.g., deadline or campaign flags) or rolling re-estimation, which we note as directions for future SARIMAX/robustness work [25].

Overall, the results confirm that seasonal time-series structure is strong for early- and mid-funnel indicators, whereas the late-funnel outcome requires additional operational signals beyond SARIMA's autoregressive seasonal form [19].

3.3. Comparative Trend Visualization

The overlay of actual vs. forecasted values with 80%/95% intervals (Figure 4) shows that SARIMA reproduces the dominant seasonal peaks in the admission calendar especially the April–July 2025 build-up prior to the mid-August closure consistent with seasonal monthly modeling of administrative processes [24].

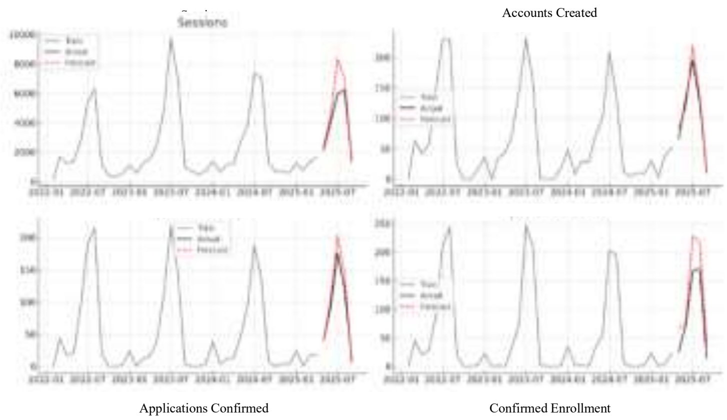

Figure 4. SARIMA Forecast vs Actual Values (May-Sep 2025) for Key Funnel Stages

For Sessions, Accounts Created, and Applications Confirmed, the forecast tracks the observed series closely during the rise-to-peak phase and the taper that follows, indicating low bias around the main seasonal cycle.

Values after mid-August reflect the early-interest phase for the next academic cycle (September 2025 onward), so the visualization is interpreted as the start of a new funnel rather than a continuation of the same cohort [22]. As expected, prediction intervals are tightest near the forecast origin and widen gradually toward September, reflecting increasing uncertainty with horizon length an inherent property of recursive multi-step time-series prediction and a practical consideration for planning lead times [26].

For Confirmed Enrollment, the line plot exhibits larger deviations and wider bands, reinforcing that late-funnel outcomes are influenced by nonseasonal administrative shocks (e.g., verification timing), which visual diagnostics make immediately apparent and which motivate exogenous-variable extensions (Section 4) [25].

3.4. Discussion and Implications

The integration of behavioral web analytics with administrative enrollment data provides a clearer view of how prospective students progress across funnel stages, validating a single pipeline for analysis and forecasting [23]. The correlation and forecasting results together indicate that sessions, engaged sessions, and events per session operate as leading indicators of subsequent admission activity, with periods of heightened interaction preceding increases in account creation and confirmed forms [17].

From a theoretical standpoint, the findings support the idea that digital behavioral signals can approximate intent in education contexts, extending funnel-based predictive modeling commonly used in marketing into higher-education forecasting [19]. The observed alignment between temporal fluctuations in web metrics and admission cycles further substantiates the use of seasonal time-series models for academic recruitment processes [24].

Operationally, the SARIMA forecasts replicated UNAI's cyclical recruitment pattern (peaks in February–July and September–January), enabling the admissions office to anticipate waves of demand and time campaigns, social media outreach, and staffing to match expected surges [24]. In practice, short-horizon forecasts and their intervals can function as an early-warning signal prompting targeted communication or remarketing when digital interest begins to taper [26].

Methodologically, SARIMA proved effective for early- and mid-funnel targets (Sessions, Accounts Created, Applications Confirmed), where MAPE values were below 25%, reflecting stable seasonal structure [24].

Despite the strong results for early and mid-funnel prediction, this study is subject to several limitations that guide future research.

First, the dependency on a single institution's dataset inherently limits the generalizability of the quantitative findings. While the unified GA4–PMB pipeline itself is replicable, the precise parameters and the timing of cyclical peaks (February–July and September–January) are specific to UNAI's unique admission calendar and administrative definitions. Institutions with vastly different application structures or seasonal cycles may require alternative model parameterizations or modeling approaches.

Second, the study highlights a critical limitation in the modeling approach: the weak predictive accuracy observed for the late-funnel outcome, Confirmed Enrollment (MAPE 51.3%). This outcome reflects its high sensitivity to nonseasonal administrative factors (such as document verification timing and manual confirmations) and timing shocks that cannot be fully captured by the univariate SARIMA model. This emphasizes the necessity for future extensions, specifically transitioning to a SARIMAX model [25] or hybrid machine-learning approaches, to incorporate exogenous predictors like deadlines, scholarship announcements, or campaign flags to enhance precision at the final conversion stage.

Finally, while a fixed hold-out window was used for accuracy evaluation, future studies are encouraged to employ rolling-origin validation (moving the cutoff backward by one month and repeating the evaluation). This robustness check is necessary to fully assess the model's stability and reliability across shifting seasonal patterns and varying forecast horizons over a longer data collection period.

In summary, a GA4–PMB integration within a SARIMA-based framework offers a scalable decision-support approach for continuous monitoring and short-term projection of admission trends turning recruitment management into a more predictive and proactive practice grounded in quantitative evidence [23].

4. CONCLUSION

This study demonstrates the practical and methodological potential of linking web analytics and institutional data for evidence-based admission forecasting in higher education. By integrating behavioral data from GA4 with administrative funnel records from the PMB system, the research establishes a replicable framework for analyzing and predicting student enrollment dynamics using time-series methods.

The use of SARIMA modeling proved that seasonal and temporal patterns in digital engagement can be systematically quantified to support short-term planning and decision-making. Beyond forecasting, the integration process itself highlights the feasibility of building institutional data ecosystems that connect marketing behavior with operational outcomes.

Rather than providing a static prediction, the model functions as a strategic monitoring tool one that can evolve as universities incorporate new data sources such as campaign schedules, scholarship cycles, or policy changes. Future extensions using SARIMAX or hybrid machine-learning approaches are recommended to capture these exogenous influences and enhance predictive accuracy.

In essence, this research moves the discussion of student admissions from retrospective reporting toward predictive, data-driven management, offering a foundation for universities seeking to align digital engagement analytics with strategic enrollment objectives.

5. ACKNOWLEDGMENTS

The author expresses sincere gratitude to Universitas Advent Indonesia (UNAI) for its institutional support and commitment to advancing data-driven research and innovation in higher education. Deep appreciation is also extended to Maria P. Viledy and Gailine O. Lumbanraja for their constant encouragement, patience, and love throughout the completion of this study. Their unwavering support made this work possible.

REFERENCES

[1] Vaishnavi Punde, Shekhar Pawar, “Admission Prediction Using Time Series Analysis,” International Journal of Innovative Science and Research Technology, 2024. doi: 10.38124/ijisrt/ijisrt24apr2377.

[2] Yu (April) Chen, Ran Li, Linda Serra Hagedorn, “Undergraduate International Student Enrollment Forecasting Model: An Application of Time Series Analysis,” Journal of International Students, 2019. doi: 10.32674/JIS.V9I1.266.

[3] Na’im Akbar, “Modelling and forecasting the number of students enrolled in the College of Administrative Sciences at Kuwait University using Multiplicative SARIMA Model,” Al-Mağallah al-ʿilmiyyaẗ li Qitāʿ Kulliyyāẗ al-tiğāraẗ bi Ğāmiʿaẗ al-Azhar, 2023. doi: 10.21608/jsfc.2023.367828.

[4] Fachri Ayudi Fitrony, Laksmita Dewi Supraba, Tessa Rantung, I Made Artha Agastya, Kusrini Kusrini, “Analysis to Predict the Number of New Students At UNU Pasuruan using Arima Method,” Jurnal Sistem Informasi dan Komputer, 2025. doi: 10.32736/sisfokom.v14i1.2251.

[5] Olayinka Abiola-Adams, Bisayo Oluwatosin Otokiti, Florence Ifeanyichukwu Olinmah, Dennis Edache Abutu, Isaac Okoli, Cyril Imohiosen, “Building Performance Forecasting Models for University Enrollment Using Historical and Transfer Data Analytics,” Journal of Frontiers in Multidisciplinary Research, 2021. doi: 10.54660/.jfmr.2021.2.1.162-168.

[6] Francisco de Borja Valdés Bertrand, “A machine learning modeling prediction of enrollment among admitted college applicants at University of Santo Tomas,” Nucleation and Atmospheric Aerosols, 2022. doi: 10.1063/5.0100174.

[7] Jamal Hussain, David Rosangliana, Vanlalruata, “Student Gross Enrolment Ratio Forecasting: A Comparative Study Using Statistical Method and Machine Learning,” International Journal of Information and Education Technology, 2023. doi: 10.18178/ijiet.2023.13.3.1824.

[8] Treena Basu, Ron Buckmire, Osei K. Tweneboah, “An Application of Machine Learning to College Admissions: The Summer Melt Problem,” Journal of Machine Learning for Modeling and Computing, 2022. doi: 10.1615/jmachlearnmodelcomput.2022046289.

[9] Julius Cesar Mamaril, Melvin A. Ballera, “Multiple educational data mining approaches to discover patterns in university admissions for program prediction,” International Journal of Informatics and Communication Technology, 2022. doi: 10.11591/ijict.v11i1.pp45-56.

[10] Jung-Pin Wu, Ming-Shr Lin, Chi-Lun Tsai, “A predictive model that aligns admission offers with student enrollment probability,” Education Sciences, 2023. doi: 10.3390/educsci13050440.

[11] Eric Huguenin, “Google Analytics 4 Properties: A Complete Guide,” MeasureSchool, accessed Oct. 2025. [Online]. Available: https://measureschool.com/google-analytics-4-properties/

[12] Google, “Visualize Google Analytics data in Looker Studio,” Looker Studio Help, accessed Oct. 2025. [Online]. Available: https://support.google.com/looker-studio

[13] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 3rd ed. Melbourne, Australia: OTexts, 2021. [Online]. Available: https://otexts.com/fpp3/seasonal-arima.html

[14] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," International Journal of Forecasting, vol. 22, no. 4, pp. 679–688, 2006. doi: 10.1016/j.ijforecast.2006.03.001.

[15] D. C. Montgomery and G. C. Runger, Applied Statistics and Probability for Engineers, 7th ed. Hoboken, NJ, USA: Wiley, 2018.

[16] H. A. Mengash, "Using data mining techniques to predict student performance to support decision making in university admission systems," IEEE Access, 2020. doi: 10.1109/ACCESS.2020.2981905.

[17] K. C. Li, B. Wong, and H. T. Chan, "Predictive analytics for university student admission: a literature review," in Lecture Notes in Computer Science, 2023. doi: 10.1007/978-3-031-35731-2_22.

[18] S. Shilbayeh and A. A. Abonamah, "Predicting student enrollments and attrition patterns in higher educational institutions using machine learning," The International Arab Journal of Information Technology, 2021. doi: 10.34028/18/4/8.

[19] K. Li, B. T. Wong, and M. Liu, "Development of a multi-model analytics system to enhance decision-making in student admission," Interactive Technology and Smart Education, 2025. doi: 10.1108/itse-12-2024-0328.

[20] V. Punde and S. Pawar, "Admission Prediction Using Time Series Analysis," International Journal of Innovative Science and Research Technology, 2024. doi: 10.38124/ijisrt/ijisrt24apr2377.

[21] F. Mohamed and B. Bukhatwa, "Predicting Student Enrollment at the Education College of Benghazi via Mathematical Modeling Approaches," Journal of Qadisiyah Computer Science and Mathematics, 2024. doi: 10.29304/jqcsm.2024.16.41803.

[22] Y. Zhao and A. Otteson, "A Practice in Enrollment Prediction with Markov Chain Models," 2024. doi: 10.48550/arXiv.2405.14007.

[23] G. A. N. Pongdatu and Y. H. Putra, "Seasonal Time Series Forecasting Using SARIMA and Holt–Winter's Exponential Smoothing," Microelectronics Systems Education, 2018. doi: 10.1088/1757-899X/407/1/012153.

[24] A. Vieira, I. Pereira de Sousa, and S. Dória-Nóbrega, "Forecasting daily admissions to an emergency department considering single and multiple seasonal patterns," Healthcare Analytics, 2023. doi: 10.1016/j.health.2023.100146.

[25] M. Bose and K. Mali, "Handling Seasonal Pattern and Prediction Using Fuzzy Time Series Model," 2020. doi: 10.1007/978-981-15-1041-0_4.

[26] A. A. Rizal, "Comparative Evaluation of Forecasting Methods for Tourist Arrival Prediction: A Sliding Window-Based Analysis," CogITo Smart Journal, vol. 11, no. 1, pp. 152–166, 2025. doi: 10.31154/cogito.v11i1.947.152-166.