

Transactions on
Computational Modeling and Intelligent Systems

Volume 4, July 2026, Article 10092
https://doi.org/10.65112/tcmis.10092

eISSN: 3108-5202
www.tcmis.org

RESEARCH ARTICLE

A machine-learning decision support system for demand-driven nurse staffing in outpatient clinics: Design and validation at an Indonesian public hospital

Muhammad Fadhillah Hadi and Asmuliardi Muluk*

Department of Industrial Engineering, Faculty of Engineering, Universitas Andalas, Padang, Indonesia.

**Corresponding Author. Email: asmuliardi@gmail.com (A. Muluk)*

Article Information

Received: 4 June 2026
Accepted: 7 July 2026
Published: 11 July 2026

AMS 2020 Classification:
62M10, 68T05, 90B90

Abstract

Public hospitals in developing countries often plan their health workforce reactively despite highly variable daily patient demand, leading to uneven workloads and longer waiting times. This study designed, verified, and retrospectively evaluated a web-based decision support system (DSS) that couples daily demand forecasting with a transparent workload-based nurse-staffing rule and automated cross-unit reallocation, using five years of outpatient records (2021–2025; 364,849 visits) from a class-B public hospital in Batam, Indonesia. ARIMA, long short-term memory (LSTM), and Prophet models were compared for the five highest-volume clinics under a single consistent protocol across three horizons—one-step-ahead, rolling-origin seven-day-ahead, and long-horizon multi-step—using MAE, RMSE, and the scale-free MASE against a seasonal-naive benchmark, with Diebold–Mariano tests for significance. ARIMA was the most reliable model: at the operational seven-day horizon it achieved a mean MASE of 0.79 and significantly outperformed the naive benchmark in four of five clinics, whereas at long horizons all models converged toward naive-level accuracy. A retrospective counterfactual simulation, in which implied nurse requirements are derived from the same workload rule used by the DSS, indicates that forecast-driven recommendations would have reduced total staffing mismatch by 55.9% under realistic seven-working-day planning (60.3% in the one-step upper bound), versus 43.2% for seasonal-naive staffing recommendations, and reveals structural misallocation across clinics, 56% of which could be covered by same-day reallocation. Sensitivity analyses over preprocessing choices, workload parameters, and weekly-lag specifications support the robustness of these findings. System computations were verified exactly, and a single-expert review by the hospital's Head of Services established face validity of the tool. The results demonstrate that an auditable forecasting-to-staffing pipeline is feasible on routine hospital data, and that evaluation protocol and forecast horizon critically shape conclusions about model superiority.

Keywords: Decision support system, demand forecasting, ARIMA, nurse staffing, healthcare operations, time-series analysis

1. Introduction

Access to quality health care depends not only on clinical capability but on how efficiently hospitals plan and allocate their human resources [1], [2]. As outpatient demand grows and fluctuates, the mismatch between available staff and actual workload directly affects waiting times, staff fatigue, and service quality [3]. Hospitals are therefore increasingly expected to support operational decisions with structured, data-driven methods rather than experience-based intuition [4].

The case institution, a class-B public hospital in Batam, Indonesia (RSUD Embung Fatimah), is the city's only municipally owned hospital and a referral centre for the surrounding region [5]. Its outpatient volume is large and uneven: daily visits vary sharply within the week, while clinic staffing is determined in a largely static, reactive

manner. Documented service complaints and long waiting times indicate an imbalance between demand and service capacity. Such variability makes fixed daily staffing inefficient: under-staffing on peak days raises waiting times and workload, while over-staffing on quiet days wastes scarce personnel.

Decision support systems (DSS) provide an established means of structuring semi-structured operational decisions by combining data, models, and interactive interfaces [6–8]. When coupled with machine learning, a DSS can learn historical demand patterns and translate them into proactive resource recommendations [9]. Patient-arrival forecasting has been studied extensively in emergency and outpatient settings [10], [11]. However, two gaps remain. First, much of this work stops at the forecasting model and is not embedded in a deployable decision tool that converts predictions into concrete staffing actions, including reallocation across units. Second, comparative forecasting studies frequently report accuracy under evaluation protocols that do not match how the model would be used in deployment, which can overstate real-world performance.

This study addresses both gaps by designing, implementing, and validating an end-to-end web DSS for nurse staffing, and by re-examining model performance under a consistent, deployment-aligned, statistically tested evaluation. The contributions are: (i) an integrated forecasting-to-staffing pipeline with transparent, auditable logic and automated cross-unit reallocation, evaluated on five years of real hospital data; (ii) a fair, horizon-aware comparison of ARIMA, LSTM, and Prophet using scale-free metrics, a seasonal-naive benchmark, and Diebold–Mariano significance tests; and (iii) a retrospective counterfactual simulation that quantifies, under deployment-aligned planning horizons and against both fixed and seasonal-naive staffing baselines, the operational value of forecast-driven staffing. Three research questions guide the work: RQ1—what are the temporal patterns of outpatient demand across the busiest clinics? RQ2—under a consistent protocol, which model best forecasts daily visits, and at which horizons? RQ3—how can forecasts be integrated into a verifiable web DSS, and what operational value does it offer?

2. Literature review

2.1. Demand forecasting in health care

Time-series forecasting is widely used to anticipate hospital demand. Classical statistical models such as ARIMA remain strong baselines for outpatient and primary-care visit forecasting [12], [13], while deep-learning approaches such as LSTM have been applied to surgical caseloads and related volumes [14]. Recent work emphasises feature-aware and probabilistic forecasting of emergency arrivals to support capacity planning [10], [11]. Comparative reviews report that no single model family dominates across datasets, with classical models remaining competitive in a substantial share of applications, particularly on univariate series with limited data [15]. Cross-domain comparisons of ARIMA, LSTM, and Prophet similarly reach context-dependent conclusions [16–18].

2.2. Forecast evaluation and staffing methods

Prophet is an additive model designed for business time series with strong seasonality and holiday effects [19]; unlike ARIMA and LSTM it is not autoregressive, a property that materially affects how it should be evaluated. LSTM networks capture nonlinear temporal dependencies through gated recurrent units [20–22]. Sound evaluation requires metrics appropriate to the data: percentage errors such as MAPE are unstable when actual values are small, motivating scale-free alternatives such as MASE [23], [24], and formal comparison of predictive accuracy is enabled by the Diebold–Mariano test [25]. On the operations side, the Workload Indicators of Staffing Need (WISN) method links staffing requirements to measured workload and is used in Indonesian health facilities [26], [27]. Reviews of health information systems note that decision support remains under-integrated into routine workforce planning [28], [29]. The present study connects these strands by embedding a rigorously evaluated forecaster into a WISN-inspired, auditable DSS and quantifying its operational value.

3. Methodology

3.1. Study site and data

Routine outpatient registration records of RSUD Embung Fatimah covering January 2021–December 2025 were used, comprising 364,849 visits. The unit of analysis is the daily visit count per clinic. The five highest-volume clinics were analysed: Internal Medicine, Medical Rehabilitation, Cardiology, Pulmonology, and General/MCU (Table 1). Only the visit date and clinic fields were extracted from the registration system and aggregated to anonymous daily counts per clinic before analysis; no patient-identifying information was accessed, and no personal data left the hospital environment. The study received ethical approval from the Health Research Ethics Committee of RSUD Embung Fatimah, Batam (approval no. 0498/500.10.30.8/II/2026, dated 18 February 2026), and institutional permission for research and data collection was granted by the Director of RSUD Embung Fatimah (letter no. 4710/800.2.1.2/XI/2025, dated 28 November 2025), following a formal request from Universitas Andalas (letter no. B/1191/UN16.9.5.3/PK.01.06/2025).

Table 1. Outpatient dataset summary (2021–2025).

Clinic	Records	Modelled daily obs.	Existing nurses
Internal Medicine	64,004	1,202	3
Medical Rehabilitation	37,235	1,191	6
Cardiology	34,934	1,158	2
Pulmonology	25,587	1,167	2
General / MCU	24,128	778	3

3.2. Pre-processing and series characterisation

Records were aggregated to daily counts per clinic. Saturdays, Sundays, and Indonesian national holidays were excluded because the clinics operate on weekdays, and low-activity days (count ≤ 5) were removed, consistent with the original study design. The scale of these exclusions and their consequences are quantified in Table 8: weekend and holiday activity is negligible in four of the five clinics, and the low-activity rule removes only 0.5–1.7% of working days in four clinics, but 33.6% in General/MCU, whose walk-in medical check-up demand is intermittent; a robustness check retaining low-activity days is reported alongside. Each series was split 80/20 chronologically into training and test segments. Before modelling, each series was characterised using augmented Dickey–Fuller (ADF) and KPSS stationarity tests and STL decomposition [30], from which trend and seasonal strengths were computed (Table 2). Three of five series are non-stationary in level, supporting first differencing. The 2021 pandemic period constitutes a regime shift: mean daily visits in 2021 were 21–87% below the 2022–2025 average, so part of the apparent long-term growth reflects post-COVID recovery rather than organic demand growth.

Table 2. Series characteristics per clinic (training period).

Clinic	Mean/day	CV	ADF p	KPSS p	Trend str.	Seasonal str.
Internal Medicine	53.2	0.39	0.271	0.01	0.58	0.29
Medical Rehabilitation	31.2	0.33	0.262	0.01	0.79	0.38
Cardiology	30.1	0.50	0.479	0.01	0.55	0.18
Pulmonology	21.9	0.38	0.011	0.01	0.30	0.26
General / MCU	28.9	1.45	0.000	0.10	0.59	0.00

3.3. Forecasting models and model selection

Three models were implemented. ARIMA used order (7,1,1), reflecting a weekly autoregressive structure with first differencing. To verify this specification, a per-clinic order search over $p \in \{0,1,2,3,5,7\}$, $q \in \{0,1,2\}$ with d set by the ADF test was conducted using the Akaike information criterion, and residuals of the selected models were tested with the Ljung–Box statistic (Table 3). The search selected (7,1,1) for Internal Medicine, and the near-equivalent (7,1,2) for Medical Rehabilitation and Cardiology ($\Delta AIC \leq 3.5$ relative to (7,1,1)), all with white-noise residuals. For the two series classified as level-stationary by the ADF test, Pulmonology and General/MCU, the best specification within $d = 0$ was (5,0,2); because AIC values are not comparable across different orders of differencing, the competing specifications were additionally compared out of sample under the rolling seven-day protocol (Table 9), where the largest MASE difference in any clinic is 0.04. Under (7,1,1), residual autocorrelation remained for Pulmonology (Ljung–Box $p < 0.05$), which is acknowledged as a limitation. The seven-lag autoregressive window was retained—although the working-day series has a five-day week—because it nests the five-day weekly lag on the consecutive working-day index and matches the deployed system's specification; the seven-day LSTM input window follows the same convention, and Table 9 confirms that five-lag and seasonal ($m = 5$) alternatives do not change the conclusions. LSTM used a 7-day input window, one 50-unit LSTM layer (ReLU) and a dense output, trained for 20 epochs (Adam, MSE) on min–max-scaled inputs with the scaler fitted on training data only. Prophet used additive weekly and yearly seasonality with Indonesian holidays [19].

Table 3. ARIMA order selection and residual diagnostics.

Clinic	Selected order (AIC)	AIC	(7,1,1) AIC	Ljung–Box p	White residuals
Internal Medicine	(7,1,1)	7931.9	7931.9	0.535	yes
Medical Rehabilitation	(7,1,2)	6207.9	6211.4	0.936	yes
Cardiology	(7,1,2)	7068.3	7069.2	0.463	yes
Pulmonology	(5,0,2)	6497.4	6477.9	0.000	no
General / MCU	(5,0,2)	5682.5	5710.7	0.065	yes (marginal)

3.4. Evaluation protocol

All models were evaluated under one identical protocol, reported for three horizons that bracket realistic deployment. (a) One-step-ahead: each model predicts day t given all actual data up to t−1, reflecting a DSS re-run daily as new counts arrive. (b) Rolling-origin seven-day-ahead: from 17 forecast origins spaced 14 days apart across the test segment, each model forecasts the next 7 working days using parameters estimated on the training segment and data observed up to the origin; this reflects weekly staff planning [23]. Throughout, "seven-day" denotes seven consecutive working days (≈1.4 calendar weeks) on the working-day index. (c) Long-horizon multi-step: a single fit on the training segment forecasting the entire test segment. This design corrects a protocol inconsistency in which autoregressive models (ARIMA, LSTM) would receive one-step information while the non-autoregressive Prophet would not. Accuracy was measured by MAE, Eq. (1); RMSE, Eq. (2); MAPE, Eq. (3); and the scale-free MASE, Eq. (4), computed against a seasonal-naive benchmark with period m = 5 working days. MASE below 1 indicates accuracy better than seasonal-naive. Differences from ARIMA were tested with the Diebold–Mariano statistic with the Harvey small-sample correction [25].

$$\mathrm{MAE} = (1/n)\ \Sigma\ |y_t - \hat{y}_t| \tag{1}$$

$$\mathrm{RMSE} = [\ (1/n)\ \Sigma\ (y_t - \hat{y}_t)^2\ ]^{1/2} \tag{2}$$

$$\mathrm{MAPE} = (100\%/n)\ \Sigma\ |(y_t - \hat{y}_t)/y_t| \tag{3}$$

$$\mathrm{MASE} = \mathrm{MAE} / [\ (1/(N-m))\ \Sigma\ |y_i - y_{i-m}|\ ] \tag{4}$$

3.5. Workload-based staffing rule

Forecasts are converted into nurse requirements using a transparent, WISN-inspired rule. For predicted daily volume A, total standard activity time per patient T_s (minutes) and effective working time per shift T_e (minutes), the required number of nurses is given by Eq. (5). This is a simplified workload-to-capacity ratio derived from WISN principles rather than a full WISN computation. Of the WISN components, the rule retains the activity (service) standard—the measured time per patient for the direct nursing tasks of an outpatient encounter—and the effective working time per shift; it omits WISN's annual available-working-time calculation (leave, absence, training), the category and individual allowance standards for non-patient activities, and the annual standard-workload aggregation, because the DSS makes a daily, per-clinic decision for which these annualised components act as approximately constant scaling factors. The simplification is stated explicitly as a limitation, and a sensitivity analysis over T_s and T_e is reported with the results. Default parameters elicited from the hospital in the underlying operational study are T_s = 35 minutes (initial assessment 10, nursing documentation 15, physician support 10) and T_e = 420 minutes of effective time per shift; both are user-adjustable in the DSS.

$$N = \lceil\ (A \times T_s) / T_e\ \rceil \tag{5}$$

3.6. DSS architecture

The DSS was implemented as an interactive web application (Streamlit). Per-clinic models are trained offline and serialised; at run time the application generates the daily forecast, applies Eq. (5) with user-adjustable activity times and shift length, and displays recommended versus existing staffing with a gap indicator. When a clinic shows a deficit, the system scans the remaining clinics for a same-day surplus and proposes a cross-unit reallocation. Reallocation proposals are advisory: they indicate where a same-day surplus exists but do not yet encode nurse skill compatibility, per-clinic minimum-staffing floors, or transfer feasibility, which remain

managerial checks before any transfer is executed. Figure 1 summarises the research methodology and Figure 2 the system architecture.

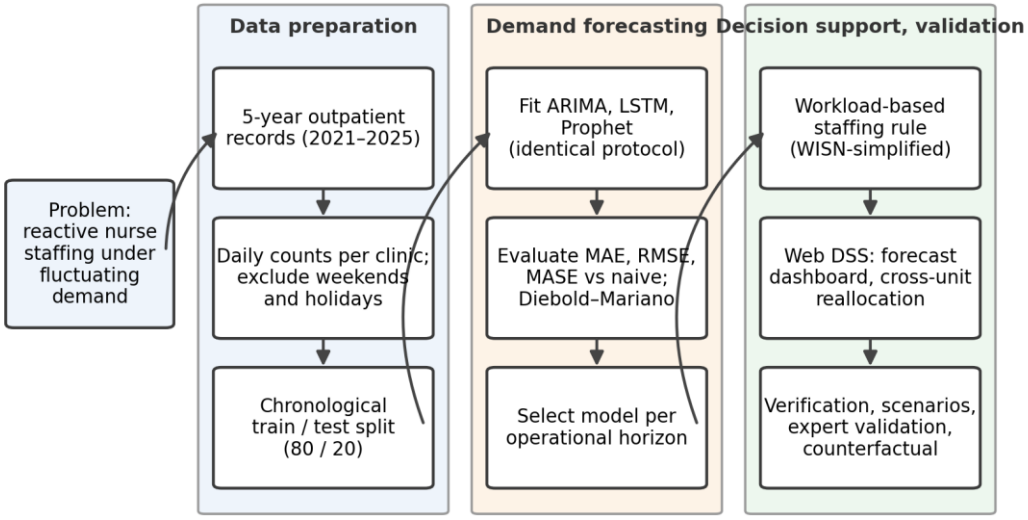


Figure 1. Research methodology flowchart.

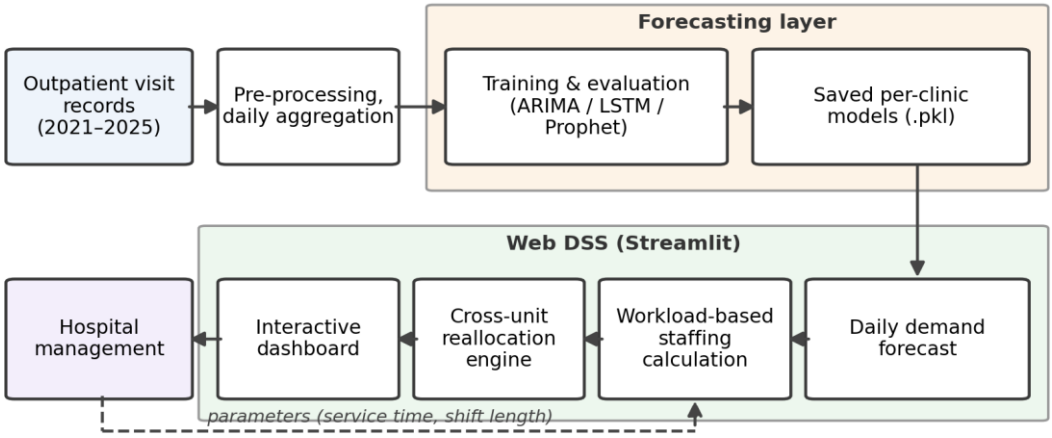


Figure 2. DSS system architecture.

3.7. Verification, validation, and counterfactual evaluation

Four complementary checks were designed: (i) computational verification, comparing system output against manual calculations of Eq. (5); (ii) functional scenario testing covering prediction, holiday handling, gap analysis, and reallocation; (iii) single-expert face validation with the hospital's Head of Services; and (iv) a retrospective counterfactual simulation of operational value. In the simulation, the daily nurse requirement implied by actual demand, $N^a = \lceil(\text{actual} \times T_s)/T_e\rceil$, is compared over the test period against alternative staffing policies, with mismatch measured in nurse-days as $\Sigma|N(\text{policy}) - N^a|$. Four policies are evaluated: the hospital's current fixed allocation; DSS recommendations in one-step mode, an information-rich upper bound corresponding to a system re-run each day for the next day; DSS recommendations under rolling seven-working-day planning, in which forecasts for the coming week are generated at weekly origins from data observed up to each origin with parameters estimated on the training segment—the horizon at which rosters are realistically fixed; and staffing recommendations derived from seasonal-naive forecasts under the same weekly protocol, as the appropriate naive benchmark. The implied requirement is a workload-rule proxy for need rather than an observed outcome such as waiting time or overtime; the analysis is therefore a simulation of operational value, not operational validation, and is labelled accordingly throughout.

4. Experimental tests and results

4.1. Demand patterns and series characteristics (RQ1)

Aggregate visits show a fluctuating, gradually rising trajectory over 2021–2025 (Figure 3), part of which reflects post-pandemic recovery (Section 3.2). At the operational scale a consistent weekly seasonality dominates: demand peaks on Monday and declines toward Friday in every clinic (Figure 4); for Internal Medicine, mean daily visits fall from 63.9 (Monday) to 40.3 (Friday). STL decomposition of the Internal Medicine series (Figure 5) confirms a marked trend component and a stable weekly seasonal component. General/MCU is an outlier: with a coefficient of variation of 1.45 and seasonal strength of 0.00 (Table 2), it is intrinsically hard to forecast, which anticipates the model results below.

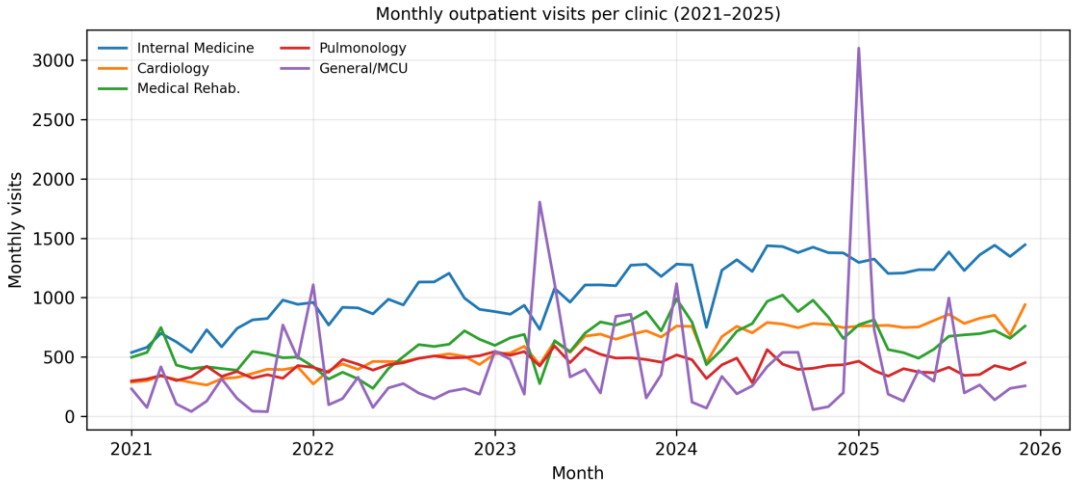


Figure 3. Monthly outpatient visits per clinic, 2021–2025.

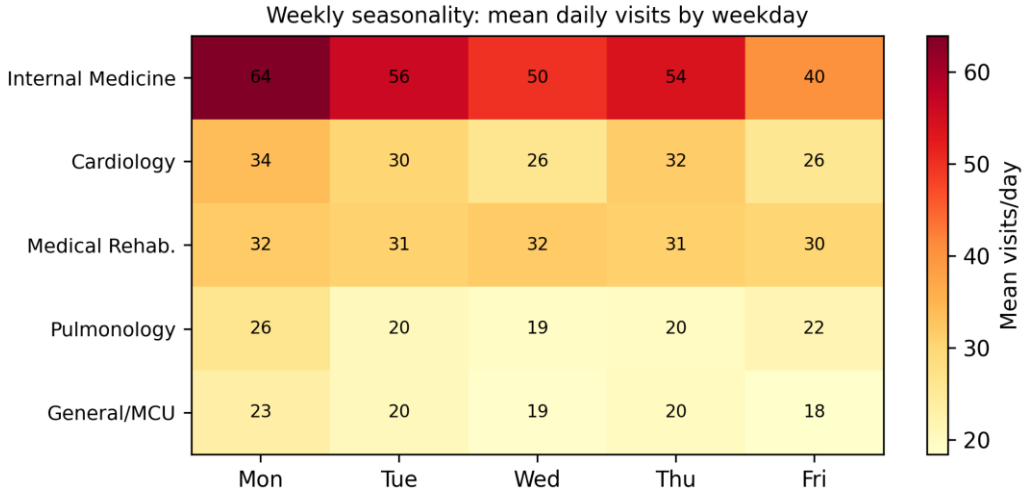


Figure 4. Weekly seasonality: mean daily visits by weekday and clinic.

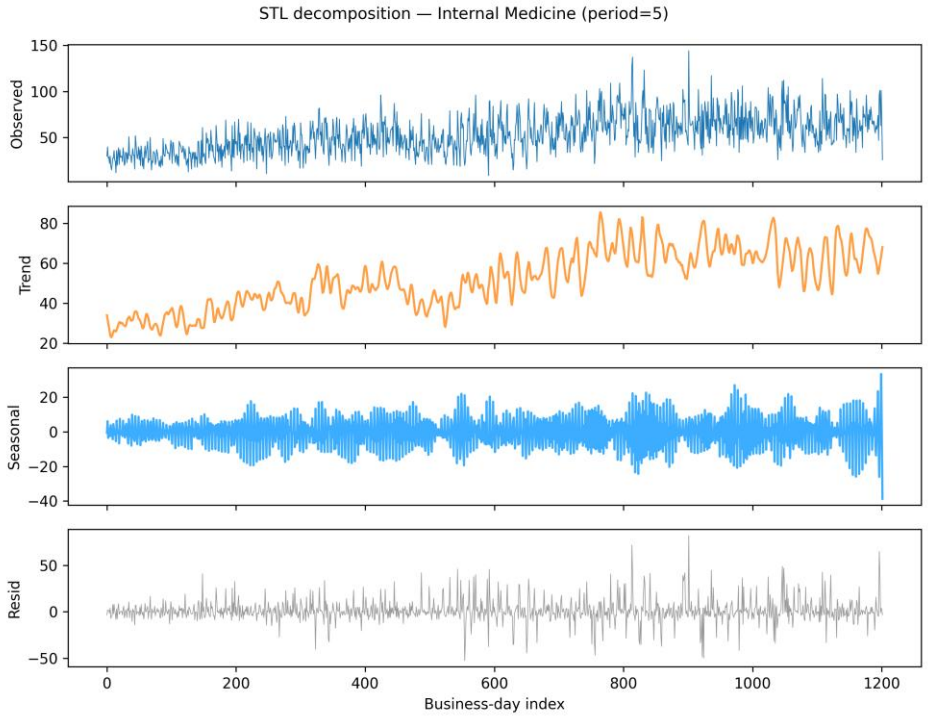


Figure 5. STL decomposition of the Internal Medicine series (period = 5 working days).

4.2. Forecasting accuracy across horizons (RQ2)

Table 4 reports accuracy averaged over the five clinics for the three horizons, and Figure 6 visualises the one-step versus long-horizon contrast. Three findings stand out. First, under one-step forecasting ARIMA is the most accurate model (MAE 9.8, MASE 0.75). Second, at the operational seven-day rolling horizon ARIMA remains the most reliable model (mean MASE 0.79): it attains MASE below 1 in every clinic and outperforms the seasonal-naive benchmark significantly in four of five clinics (Diebold–Mariano $p < 0.01$; Table 5, Figure 7), while never being significantly worse than any competitor. LSTM is competitive (mean MASE 0.85) and marginally—but not significantly—better than ARIMA in the two most irregular series. Prophet is unreliable (MASE 2.17 in Medical Rehabilitation and 1.27 in General/MCU); because it is not autoregressive its accuracy is identical across horizons, which also explains why protocols granting autoregressive models one-step information would disadvantage it. Third, at the long multi-step horizon all models converge (MAE ≈ 15) and perform only marginally better than—and for General/MCU not significantly different from—seasonal-naive (MASE 1.1–1.2). MAPE proved unstable (inflated by low-count days, exceeding 80% on average at the long horizon) and is therefore not used as a primary criterion [24].

Table 4. Forecasting accuracy by model and horizon (mean over five clinics; MAE / MASE).

Horizon	ARIMA	LSTM	Prophet	Seasonal-naive
One-step (daily update)	9.8 / 0.75	10.3 / 0.79	15.3 / 1.23	13.8 / 1.04
Rolling origin, h = 7	10.6 / 0.79	11.4 / 0.85	15.4 / 1.21	13.3 / 1.03
Long-horizon multi-step	15.2 / 1.13	15.0 / 1.16	15.3 / 1.23	16.9 / 1.37

Table 5. Rolling-origin seven-day-ahead MASE per clinic and Diebold–Mariano p-values (reference: ARIMA).

Clinic	ARIMA	LSTM	Prophet	Naive	p vs LSTM	p vs Prophet	p vs Naive
Internal Medicine	0.86	1.02	0.95	1.13	0.041	0.477	0.005
Medical Rehabilitation	0.71	0.76	2.17	0.91	0.431	0.000	0.003
Cardiology	0.95	1.11	0.96	1.30	0.041	0.635	0.000
Pulmonology	0.72	0.70	0.72	1.02	0.399	0.878	0.001
General / MCU	0.72	0.68	1.27	0.78	0.280	0.001	0.609

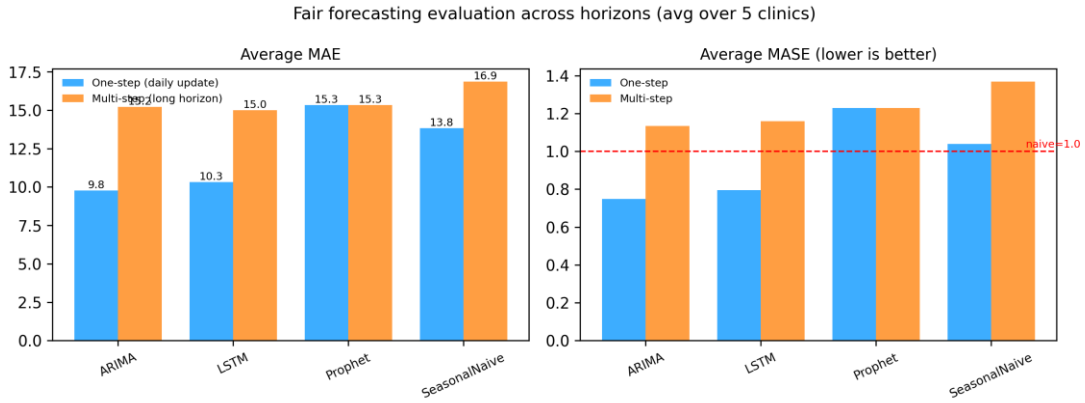


Figure 6. Fair evaluation across horizons: average MAE and MASE per model (naive reference at MASE = 1).

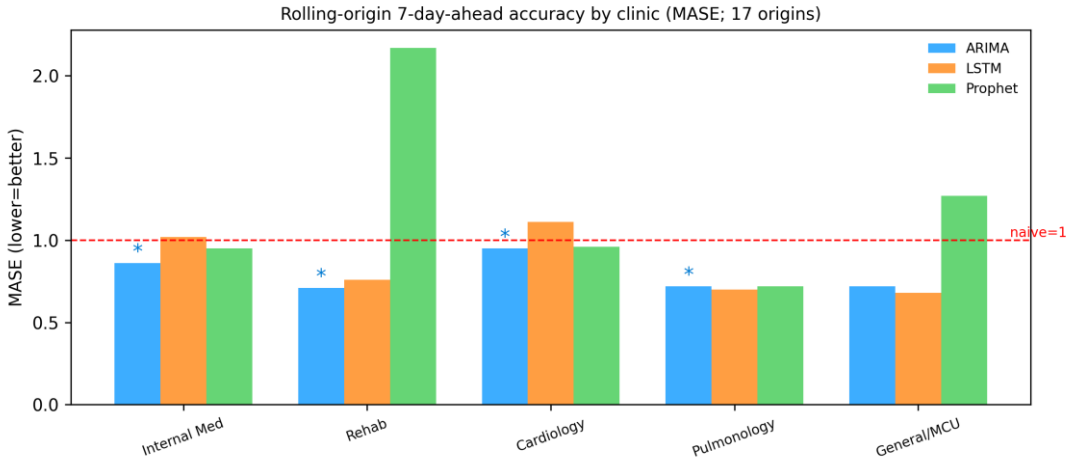


Figure 7. Rolling-origin seven-day-ahead MASE by clinic; * marks clinics where ARIMA is significantly better than seasonal-naive (DM $p < 0.05$).

4.3. Operational value: counterfactual staffing evaluation (RQ3)

Table 6 and Figure 8 compare staffing mismatch against implied need over the test period. All figures are outputs of the retrospective counterfactual simulation defined in the methodology, not observed operational outcomes. Under the current fixed allocation the five clinics accumulate 2,132 nurse-days of mismatch. Forecast-driven DSS recommendations reduce this to 941 nurse-days (a 55.9% reduction) under realistic rolling seven-working-day planning, and to 846 nurse-days (60.3%) in the one-step upper bound; staffing driven by seasonal-naive forecasts under the same weekly protocol attains 1,211 nurse-days (43.2%). At the horizon at which rosters are actually planned, the forecast-driven policy therefore retains a 12.7-percentage-point advantage over the naive staffing benchmark. The simulation also reveals structural misallocation that no daily rule can fix alone: Medical Rehabilitation (6 nurses) is over-staffed on 239 of 239 test days, while Internal Medicine (3 nurses) is under-staffed

on 232 of 241 days. This is precisely the situation the DSS's reallocation engine targets: summing across clinics, 56% of all deficit nurse-days could be covered by a same-day surplus elsewhere. The evaluation is honest about heterogeneity: in Pulmonology, where fixed staffing already matches need closely, forecast noise makes the DSS marginally worse (+5.6% mismatch at the weekly horizon), and in the intermittent General/MCU series the weekly-horizon gain shrinks to 2.3% (26.9% one-step); forecast-driven staffing adds most value where demand is volatile relative to current allocation, and least where demand is intermittent.

Table 6. Retrospective counterfactual staffing mismatch over the test period (nurse-days), by staffing policy; percentage change is relative to fixed staffing.

Clinic	Fixed staffing	DSS one-step	DSS 7-day	Seasonal-naive 7-day	Change (DSS 7-day)
Internal Medicine (3)	704 (under-staffed 232/241 days)	278	288	388	−59%
Medical Rehabilitation (6)	651 (over-staffed 239/239 days)	66	73	93	−89%
Cardiology (2)	424	215	228	304	−46%
Pulmonology (2)	89	94	94	144	+6%
General / MCU (3)	264	193	258	282	−2%
Total	2,132	846	941	1,211	−55.9%

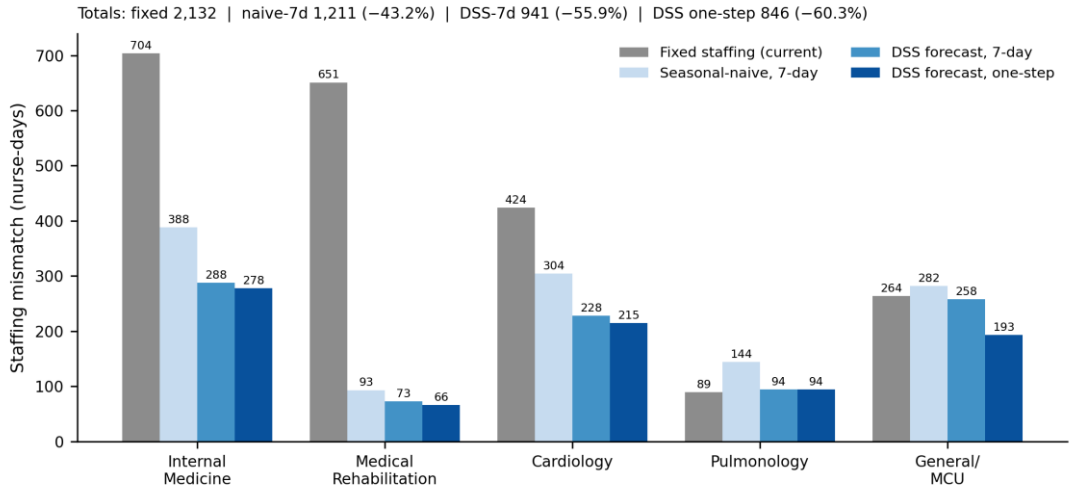


Figure 8. Retrospective counterfactual simulation: staffing mismatch versus implied need under fixed staffing, forecast-driven DSS (one-step and seven-day modes), and seasonal-naive staffing.

4.4. Sensitivity and robustness analyses

Three analyses probe the robustness of the reported results. First, preprocessing: Table 8 quantifies the excluded observations per clinic. Weekend and holiday rows are rare because the clinics operate on weekdays; the exception is General/MCU, whose occasional Saturday check-up sessions account for 419 visits on 8 weekend days over five years. The low-activity rule (count ≤ 5) removes only 0.5–1.7% of working days in four clinics but 33.6% in General/MCU. Re-running the rolling seven-day evaluation with low-activity days retained leaves ARIMA more accurate than the concurrent seasonal-naive benchmark in all five clinics, and accuracy is materially affected only in General/MCU (ARIMA MASE 0.72 → 1.33; naive 0.78 → 1.75), whose series becomes strongly intermittent; in Medical Rehabilitation the degradation is moderate (0.71 → 0.86) and elsewhere it is negligible. The exclusion therefore biases neither the model ranking nor the conclusions for the four regular clinics, but forecasts—and staffing values derived from them—for General/MCU should be read with caution.

Second, workload parameters: varying T_s over 25–45 minutes and T_e over 390–450 minutes changes the seven-day DSS mismatch reduction from 52.1% to 61.3% (55.9% at the default (35, 420)). The qualitative conclusion—

forecast-driven staffing removes roughly half or more of the mismatch—is insensitive to the WISN simplification within this range, although the absolute nurse-day figures scale with the assumed activity time.

Third, the weekly-lag specification: Table 9 compares ARIMA(7,1,1) with five-lag and seasonal five-day alternatives under the rolling seven-day protocol. The largest MASE difference in any clinic is 0.04, confirming that the seven- versus five-lag choice is immaterial out of sample and that the conclusions do not depend on this specification.

Table 8. Preprocessing sensitivity: working days per clinic, observations removed by the low-activity rule (count ≤ 5), and rolling seven-day accuracy with the rule applied (filtered) versus retained (all); MASE is computed within each preprocessing variant.

Clinic	Working days	Removed by count ≤ 5 (%)	ARIMA MASE (filtered → all)	Naive MASE (filtered → all)
Internal Medicine	1,209	7 (0.6%)	0.86 → 0.84	1.13 → 1.11
Medical Rehabilitation	1,197	6 (0.5%)	0.71 → 0.86	0.91 → 1.17
Cardiology	1,178	20 (1.7%)	0.95 → 0.96	1.30 → 1.31
Pulmonology	1,187	20 (1.7%)	0.72 → 0.75	1.02 → 0.92
General / MCU	1,172	394 (33.6%)	0.72 → 1.33	0.78 → 1.75

Table 9. Weekly-lag specification sensitivity: rolling-origin seven-day MASE by clinic.

Clinic	ARIMA (7,1,1)	ARIMA (5,1,1)	ARIMA (5,1,2)	SARIMA $(1,1,1)(1,0,1)_5$
Internal Medicine	0.86	0.90	0.89	0.91
Medical Rehabilitation	0.71	0.71	0.71	0.70
Cardiology	0.95	0.95	0.95	0.98
Pulmonology	0.72	0.70	0.70	0.69
General / MCU	0.72	0.73	0.74	0.73

4.5. System verification and expert face validation

Computational verification reproduced manual results exactly: for Cardiology with a predicted volume of 44 patients, Eq. (5) yields $\lceil(44 \times 35)/420\rceil = 4$ nurses, identical to the system output. Functional testing confirmed correct behaviour in all five scenarios (Table 7). In expert validation, the hospital's Head of Services judged the dashboard outputs clear, the staffing logic transparent, and the recommendations relevant to daily operations. Because this involved a single domain expert, it establishes face validity rather than broad user acceptance; a multi-user usability study is identified as future work.

Table 7. Functional scenario testing.

Function	**Scenario**	**Expected result**	**Status**
Patient prediction	Select working day	Forecast displayed	Pass
Holiday detection	Select weekend/holiday	"Unit closed", zero patients	Pass
Requirement calculation	Enter service times	Nurses computed via Eq. (5)	Pass
Gap analysis	Need exceeds existing	Deficit indicator raised	Pass
Reallocation	Surplus available elsewhere	Transfer suggestion shown	Pass

4.6. Discussion

The results carry a clear methodological message: conclusions about model superiority depend on the evaluation protocol and the forecast horizon. A single headline number (e.g., MAE ≈ 9.8 for ARIMA) without a stated horizon overstates what a planner experiences when forecasting a week or more ahead. For this application the operationally relevant result is the rolling seven-day evaluation, where ARIMA is accurate, statistically better than naive in four of five clinics, computationally light, and interpretable—making it a defensible default for a daily-updated DSS. The convergence of all models toward naive-level accuracy at long horizons, and the failure of all models to beat naive significantly in the structurally irregular General/MCU clinic, are honest boundaries of applicability that managers should know. The counterfactual simulation reframes the contribution: beyond forecast accuracy, the DSS's simulated value lies in correcting a structural misallocation (a chronically over-staffed unit coexisting with a chronically under-staffed one) and in enabling same-day reallocation that could absorb 56% of deficits. Importantly, most of the simulated benefit survives the move from one-step to realistic weekly planning (60.3% to 55.9%), and the forecast-driven policy continues to dominate the seasonal-naive staffing benchmark (43.2%) at the planning horizon. Limitations include: a single hospital and five clinics; exogenous drivers (outbreaks, policy and referral changes) not modelled; a simplified WISN ratio as the need proxy, so mismatch is measured against a modelled rather than observed requirement; clinic-invariant service times (T_s), although workload intensity plausibly differs across outpatient specialties—the T_s/T_e sensitivity analysis only partially mitigates this; exclusion of low-activity days, which is consequential for the intermittent General/MCU series; reallocation logic that does not yet encode nurse skill compatibility, minimum-staffing floors, or transfer feasibility; lightly tuned LSTM; rolling-origin evaluation with fixed parameters; single-expert face validation rather than a multi-user usability study; and a counterfactual that simulates rather than observes operational outcomes such as waiting times, overtime, or service quality.

5. Conclusions and future research

This study designed, verified, and retrospectively evaluated a web-based DSS that converts five years of routine outpatient records into demand-driven nurse-staffing recommendations for an Indonesian public hospital. Outpatient demand exhibits strong weekly seasonality with a Monday peak, a post-pandemic recovery trend, and one structurally irregular clinic (RQ1). Under a consistent, multi-horizon, significance-tested evaluation, ARIMA is the most reliable forecaster at operational horizons (rolling seven-day MASE 0.79; significantly better than seasonal-naive in four of five clinics), while all models converge to naive-level accuracy at long horizons (RQ2). Integrated with a transparent workload rule and reallocation logic, the forecast-driven DSS would—in a retrospective counterfactual simulation—reduce staffing mismatch by 55.9% relative to the current fixed allocation under realistic weekly planning (60.3% in the one-step upper bound, versus 43.2% for seasonal-naive staffing), and exposes structural misallocation, with 56% of deficits coverable by same-day reallocation; its computations were verified exactly and a single-expert review established face validity (RQ3). Future research will incorporate exogenous regressors and full WISN parameters, probabilistic forecasts with service-level staffing, constraint-aware reallocation (nurse skill compatibility, minimum-staffing floors, transfer feasibility), automatic per-series model selection and tuning, integration with the hospital information system, multi-user usability evaluation, prospective evaluation against observed waiting-time and workload outcomes, and shift-level scheduling optimisation.

Acknowledgments

None.

Funding

This research received no external funding.

Conflict of interest

There is no conflict of interest to disclose.

Author contributions

M. F. Hadi: Conceptualization, data curation, software, investigation, formal analysis, writing - original draft. **A. Muluk:** Conceptualization, methodology, supervision, validation, writing - review and editing.

Declaration of using AI tools

The authors used an AI assistant to support English drafting and editing and to reproduce data figures from the authors' own data; all analyses, content, and conclusions were verified by and are the responsibility of the authors. No generative-AI-created imagery was used; data charts were generated from source data and diagrams were author-designed.

Ethics approval and institutional permission

The study was approved by the Health Research Ethics Committee of RSUD Embung Fatimah, Kota Batam, Indonesia (approval no. 0498/500.10.30.8/II/2026, dated 18 February 2026). Institutional permission for research and data collection was granted by the Director of RSUD Embung Fatimah (letter no. 4710/800.2.1.2/XI/2025, dated 28 November 2025), following the formal request of the Department of Industrial Engineering, Universitas Andalas (letter no. B/1191/UN16.9.5.3/PK.01.06/2025). The study used routinely collected registration records; only visit dates and clinic identifiers were processed, records were aggregated to anonymous daily counts per clinic, and no patient-identifying information was accessed or stored.

Data and code availability

The patient-level registration data cannot be shared publicly. The anonymised aggregate daily visit counts per clinic (date, clinic, count) that support all analyses, together with the full analysis code (data preparation, model estimation, rolling-origin evaluation, counterfactual simulation, and sensitivity analyses; Python 3 with statsmodels, Prophet, and TensorFlow, fixed random seed), are available from the corresponding author upon reasonable request.

References

[1] W. D. Savira and W. Subadi, "Kualitas pelayanan kesehatan dilihat dari aspek tangible dalam rangka meningkatkan pelayanan kesehatan pada Puskesmas Hikun," *Jurnal Administrasi Publik dan Administrasi Bisnis*, vol. 6, no. 2, pp. 749–762, 2023.

[2] Badan Pusat Statistik, Profil Statistik Kesehatan 2025. Jakarta, Indonesia: Badan Pusat Statistik, 2025.

[3] R. Amran, A. Apriyani, and N. P. Dewi, "Peran penting kelengkapan rekam medik di rumah sakit," *Baiturrahmah Medical Journal*, vol. 1, no. 1, pp. 69–76, 2021.

[4] A. Komashie, S. Hinrichs-Krapels, and P. J. Clarkson, "Systems approaches to healthcare systems design and care delivery: an overview of the literature," *Proceedings of the Design Society*, vol. 1, pp. 2941–2950, 2021, doi: 10.1017/pds.2021.555.

[5] Peraturan Wali Kota Batam No. 140 Tahun 2024 tentang Rumah Sakit Umum Daerah Embung Fatimah, 2024.

[6] A. Sebaa, A. Nouicer, A. Tari, R. Tarik, and O. Abdellah, "Decision support system for health care resources allocation," *Electronic Physician*, vol. 9, no. 6, pp. 4661–4668, 2017, doi: 10.19082/4661.

[7] D. J. Power, *Decision Support Systems: Concepts and Resources for Managers*. Westport, CT, USA: Quorum Books, 2002.

[8] E. Turban, J. E. Aronson, and T. P. Liang, *Decision Support Systems and Intelligent Systems*, 7th ed. New Delhi, India: Prentice-Hall, 2007.

[9] A. Aditya et al., *Machine Learning*. Padang, Indonesia: PT Global Eksekutif Teknologi, 2023.

[10] B. M. Porto and F. S. Fogliatto, "Enhanced forecasting of emergency department patient arrivals using feature engineering approach and machine learning," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, Art. no. 377, 2024, doi: 10.1186/s12911-024-02788-6.

[11] B. Rostami-Tabar, J. Browell, and I. Svetunkov, "Probabilistic forecasting of hourly emergency department arrivals," *Health Systems*, vol. 13, no. 2, pp. 133–149, 2024, doi: 10.1080/20476965.2023.2200526.

[12] I. Basri K., "Peramalan jumlah kunjungan pasien rawat jalan menggunakan metode ARIMA, SES dan Holt-Winters di Balai Kesehatan Masyarakat XYZ," *Jurnal Tata Kelola dan Kerangka Kerja Teknologi Informasi*, vol. 7, no. 1, pp. 29–39, 2021, doi: 10.34010/jtk3ti.v7i1.5548.

[13] N. Manabung, C. Mongi, and Y. Langi, "Peramalan banyaknya pasien yang berobat di Puskesmas Bengkol dengan menggunakan metode ARIMA," *Jurnal LPPM Bidang Sains dan Teknologi*, vol. 9, no. 1, pp. 11–19, 2024.

[14] H. Bui, S. D. Ekşioğlu, A. A. Villafranca, J. A. Sanford, and K. W. Sexton, "A long short-term memory model for forecasting the surgical case volumes at a hospital," *IISE Transactions on Healthcare Systems Engineering*, vol. 13, no. 3, pp. 226–236, 2023.

[15] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, "A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks," *Future Internet*, vol. 15, no. 8, Art. no. 255, 2023, doi: 10.3390/fi15080255.

[16] I. G. T. Suryawan, I. K. N. Putra, P. M. Meliana, and I. G. I. Sudipa, "Performance comparison of ARIMA, LSTM, and Prophet methods in sales forecasting," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 4, pp. 2410–2421, 2024, doi: 10.33395/sinkron.v8i4.14057.

[17] S. Albahli, "LSTM vs. Prophet: achieving superior accuracy in dynamic electricity demand forecasting," *Energies*, vol. 18, no. 2, Art. no. 278, 2025, doi: 10.3390/en18020278.

[18] R. Anwar and L. Rassiyanti, "Analisis komparasi model peramalan Prophet dan ARIMA dalam memprediksi harga saham penutupan PT ANTM," *Lattice Journal: Journal of Mathematics Education and Applied*, vol. 5, no. 1, pp. 57–74, 2025.

[19] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018, doi: 10.1080/00031305.2017.1380080.

[20] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, 2018, doi: 10.1016/j.ejor.2017.11.054.
[21] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[22] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*. Berlin, Germany: Springer, 2012.
[23] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed. Melbourne, Australia: OTexts, 2018.
[24] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *International Journal of Forecasting*, vol. 32, no. 3, pp. 669–679, 2016, doi: 10.1016/j.ijforecast.2015.12.003.
[25] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 253–263, 1995, doi: 10.1080/07350015.1995.10524599.
[26] A. D. Wahuningtyas, R. B. M. T. Pangestu, R. P. Hariono, F. Falah, O. M. Prawiryanto, and D. W. Sutha, "Analisis kebutuhan tenaga kerja berdasarkan beban kerja dengan metode Workload Indicator of Staffing Need (WISN) unit rekam medis di Puskesmas 'X'," *MEJORA: Medical Journal Awatara*, vol. 2, no. 1, pp. 25–31, 2024, doi: 10.61434/mejora.v2i1.142.
[27] J. Natassa, N. Maimun, H. M. Ulfa, S. Wardani, and R. Isnani, "Analisis kebutuhan tenaga rekam medis menggunakan metode WISN," *Jurnal Kesehatan Komunitas*, vol. 10, no. 3, pp. 546–554, 2025, doi: 10.25311/keskom.Vol10.Iss3.1734.
[28] A. Epizitone, S. P. Moyane, and I. E. Agbehadji, "A systematic literature review of health information systems for healthcare," *Healthcare*, vol. 11, no. 7, Art. no. 959, 2023, doi: 10.3390/healthcare11070959.
[29] K. F. Nisa, T. Dompak, and K. T. Lodan, "Perbandingan sistem pelayanan kesehatan di Indonesia dan di Singapura," *Prosiding Seminar Nasional Ilmu Sosial dan Teknologi (SNISTEK)*, vol. 6, pp. 72–77, 2024.
[30] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "STL: a seasonal-trend decomposition procedure based on loess," *Journal of Official Statistics*, vol. 6, no. 1, pp. 3–73, 1990.



All open access articles published in Transactions on Computational Modeling and Intelligent Systems (http://tcmis.org) are distributed under the terms of the CC BY-NC 4.0 license (Creative Commons Attribution Non-Commercial 4.0 International Public License as currently displayed at http://creativecommons.org/licenses/by-nc/4.0/legalcode) which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original work is properly cited.