

Baseline Projection of Outpatient Visits Using Simple Linear Regression: A Case Study at RSU Ananda Purwokerto

Anisa Ul Khasanah ^{1*}, Sulistiyasni ², Muh. Hery Santoso ³

¹ Sekolah Tinggi Manajemen Informatika dan Komputer Widya Utama: khasanaha57@gmail.com

² Sekolah Tinggi Manajemen Informatika dan Komputer Widya Utama: sulistiyasnipwt@swu.ac.id

³ Sekolah Tinggi Manajemen Informatika dan Komputer Widya Utama: muhhery@swu.ac.id

Abstract

Anticipating outpatient demand can support hospital planning, although forecasts based on short and fluctuating annual series require cautious interpretation. This study aimed to develop a transparent baseline projection of outpatient visits at RSU Ananda Purwokerto using simple linear regression. The analysis used eight annual aggregate observations from 2018–2025; the 2025 value represented visits from January through October, whereas the preceding observations covered complete calendar years. Calendar year was used as the predictor and outpatient visits as the outcome. Manual calculations were compared with the Linear Regression output in RapidMiner, and in-sample fit was evaluated using the Mean Absolute Percentage Error (MAPE). The fitted model produced a slope of -2,103.57 visits per year and projected 62,590 visits in 2026, declining to 54,176 in 2030. However, the slope was not statistically significant ($p = .369$), and the model explained only 13.6% of the observed variation ($R^2 = .136$). The in-sample MAPE was 15.66%, which met the study's descriptive “good” category but did not constitute independent validation of future forecast accuracy. Therefore, the projections should be interpreted as preliminary baseline scenarios rather than certain future totals. The model may support initial operational discussion, but decisions should also consider current hospital information and should be updated using complete, preferably monthly, observations.

Keywords: *Outpatient Visits, Simple Linear Regression, Hospital Demand Forecasting, MAPE, Rapidminer*

Introduction

Hospitals are healthcare institutions that must maintain service quality while using limited human, physical, and financial resources efficiently. Outpatient services are particularly important because they often represent the first point of contact between patients and a hospital, generate a substantial share of routine service activity, and influence the workload of registration units, clinics, pharmacies, laboratories, and medical record units. The quality of these services depends not only on clinical competence but also on management's ability to anticipate demand and allocate resources at the appropriate time (Setiatin & Syahidin, 2021; Wardani et al., 2024). Outpatient utilization may also differ according to access and equity conditions across service settings (Areru et al., 2021). For this reason, the number of outpatient visits is a relevant operational indicator for planning personnel, facilities, medicines, appointment schedules, and administrative support.

Variation in outpatient visits can create operational pressure when hospital capacity does not match patient demand. When visits increase unexpectedly, healthcare workers may experience heavier workloads, waiting times may lengthen, and the risk of delays or administrative errors may increase (Barsasella & Faturrohmi, 2025; Ratumurun & Manik, 2025). When visits decline, hospitals may face underused personnel, rooms, and supporting facilities, which can reduce operational efficiency. Demand forecasting therefore provides management with an evidence-based reference for balancing service capacity and expected utilization. Forecasts are not intended to eliminate uncertainty; rather, they help decision makers prepare alternative staffing, scheduling, and resource-allocation strategies before changes in demand occur (Subramanian, 2021).

RSU Ananda Purwokerto is a general hospital located in West Purwokerto and serves patients from Purwokerto and surrounding areas. Its outpatient visit records show substantial year-to-year variation. The hospital recorded 65,756 visits in 2018 and 83,319 visits in 2019, before the COVID-19 pandemic. Visits then reached 84,499 in 2020 and 89,287 in 2021, declined to 68,750 in 2022, and fell further to 45,031 in 2023. Hospital records indicate that the 2023 decline coincided with the termination of cooperation with BPJS Kesehatan. Visits subsequently increased to 69,771 in 2024, while the recorded value for 2025 reached 70,039 through October (RSU Ananda Purwokerto, 2025). These figures illustrate a pattern that is neither consistently increasing nor consistently decreasing.

The observed variation occurred across several distinct contexts: the pre-pandemic period, the COVID-19 period, the post-pandemic recovery period, and the period following a change in BPJS cooperation. These contexts are relevant when interpreting the data because healthcare utilization can change when access arrangements, service availability, population behavior, or health-financing policies change. However, the historical coincidence between a change in visits and a particular event does not by itself establish a causal effect. In the present study, year is the only explanatory variable in the regression model. Consequently, COVID-19 conditions and the change in BPJS cooperation are treated as contextual information rather than variables whose effects are statistically estimated. This distinction is necessary to prevent the forecast from being interpreted as a causal evaluation of hospital policy.

The development of hospital information systems and data analytics enables routinely collected visit records to be used for planning rather than solely for administrative reporting. Historical visit data can reveal general tendencies and provide an initial estimate of future service volumes, while modern health analytics can extend this role through data-mining and machine-learning decision support (Beam & Kohane, 2018; Fernandes Prabhu et al., 2025; Haryanto et al., 2024). Studies of patient-arrival forecasting have applied regression, statistical time-series, machine-learning, and hybrid approaches, with model choice depending on data frequency, local demand patterns, available predictors, and the planning objective (Jiang et al., 2023; Wargon et al., 2009). Jones et al. (2008), for example, compared methods for forecasting daily emergency-department volume. Luo et al. (2017) combined autoregressive integrated moving average and simple exponential smoothing models to forecast daily outpatient visits. Huang et al. (2020) developed a hybrid approach combining ARIMA and adaptive filtering, while Yu et al. (2017) used decomposition and an artificial neural-network ensemble. Large forecasting comparisons likewise show that no single method is uniformly superior across all series (Makridakis et al., 2018).

Simple linear regression is one of the most accessible methods for describing a directional relationship between time and service volume. In this approach, time is used as the independent variable and the number of visits as the dependent variable. The slope summarizes the average change in visits for each additional time period, while the fitted equation can be extended to produce baseline projections. The method is transparent, computationally simple, and easy for hospital managers to interpret. These characteristics make it useful when the objective is to construct an initial trend-based forecast and when the available dataset is limited. The model can also be reproduced manually and checked using analytical software such as RapidMiner, thereby making the calculation process easier to communicate to nontechnical decision makers. Nevertheless, reproducible software implementation should be distinguished from evidence of predictive validity, and scalable forecasting practice still requires appropriate model evaluation (Taylor & Letham, 2018).

Nevertheless, ease of application does not guarantee that a linear model adequately represents hospital demand. Outpatient visits may contain nonlinear trends, seasonal variation, temporary shocks, and structural changes. A simple regression on time does not explicitly model autocorrelation, seasonality, or the effects of external variables. Its projections are also sensitive

to unusual observations and to differences in observation length. This issue is especially relevant in the present dataset because the 2025 value represents visits only through October, whereas the preceding observations represent complete calendar years. Therefore, the resulting model should be understood as a preliminary baseline that requires careful data preparation, diagnostic assessment, and periodic updating, rather than as a definitive representation of future demand.

Forecast evaluation is equally important because a fitted trend may produce numerical projections even when the statistical evidence for that trend is weak. Measures such as the Mean Absolute Percentage Error (MAPE) describe the average relative difference between actual and fitted or forecast values, but percentage-based errors have known limitations and should be interpreted in relation to the data and decision context (Armstrong & Collopy, 1992; Hyndman & Koehler, 2006). An error measure calculated from the same observations used to estimate the model reflects in-sample fit and does not automatically establish out-of-sample forecasting performance. Time-ordered validation is preferable when sufficient observations are available because ordinary resampling may not respect temporal dependence (Bergmeir et al., 2018). Coefficient significance, goodness of fit, residual behavior, and the evaluation design should therefore be considered together rather than relying on a single performance measure (Steyerberg et al., 2010). Prediction uncertainty should also be acknowledged, particularly when projections extend several years beyond a small historical sample (Gneiting & Katzfuss, 2014).

Previous Indonesian studies provide evidence that regression and trend models can support hospital planning. Baihaqi et al. (2019) applied simple linear regression to predict hospital visits based on service type and patient age. Nugroho et al. (2024) used simple linear regression to estimate outpatient visits at Pati Islamic Hospital and emphasized its usefulness for resource planning. Sabran et al. (2024) applied linear trend analysis to outpatient visits at Haji General Hospital of East Java Province, while Shirvia et al. (2025) combined simple linear regression and RapidMiner-based processing in a study at RSU Bunda Purwokerto. These studies show that interpretable baseline models remain relevant for healthcare institutions that require straightforward forecasting tools.

For hospital management, the practical value of a forecast depends on how it is translated into decisions. A projected change in visits may inform the number and distribution of registration staff, nurses, physicians, clinic sessions, waiting-room capacity, and supporting supplies. It may also provide an early signal for reviewing access barriers, referral relationships, patient experience, or service promotion. Nevertheless, managerial decisions should not be based on a single point forecast alone. Forecast values need to be reviewed together with recent operational information and plausible uncertainty because actual utilization may depart from the historical trend. This is particularly important when the available records include shocks or policy changes. Thus, the model in this study is intended to support, rather than replace, managerial judgment and routine monitoring of outpatient services.

However, the available studies also reveal several areas requiring further attention. Results from one hospital cannot be transferred directly to another because hospitals differ in patient populations, referral patterns, service portfolios, insurance cooperation, operational policies, and dominant demand predictors (Jiang et al., 2023). In addition, previous applications have often emphasized forecast values and categorical interpretations of prediction error without sufficiently discussing incomplete observations, limited sample size, statistical significance, or the distinction between fitted accuracy and future predictive validity. International studies have frequently used higher-frequency data and more complex time-series or hybrid models (Bai et al., 2023; Huang et al., 2020; Luo et al., 2017; Wargon et al., 2009), whereas many hospitals may initially require a simpler and more transparent baseline before more advanced methods can be implemented.

The research gap addressed in this study therefore concerns both the specific operational context of RSU Ananda Purwokerto and the need for a cautious application of an interpretable

baseline model. The study analyzes the hospital's annual outpatient visit records for 2018-2025 and uses simple linear regression to generate projections for 2026-2030. The contribution is not the introduction of a new regression method, but the structured application of an established method to hospital planning, accompanied by an interpretation of the model's direction, error measure, statistical evidence, and practical limitations. The pre-pandemic, pandemic, post-pandemic, and BPJS-related contexts are used to explain why historical observations may fluctuate, not to claim causal effects that are not estimated by the model.

Accordingly, this study aims to: (1) describe the annual pattern of outpatient visits at RSU Ananda Purwokerto; (2) estimate a simple linear regression model with time as the explanatory variable; (3) generate baseline projections of outpatient visits for 2026-2030; (4) evaluate the model using the available statistical and error measures; and (5) discuss the implications and limitations of the projections for hospital resource planning. The results are expected to provide an initial quantitative reference for workforce allocation, clinic scheduling, facility preparation, and service evaluation. They are also expected to support periodic forecasting updates and future comparisons with time-series, machine-learning, or multivariable models when more detailed data become available.

Method

This study used a quantitative, retrospective observational design to develop a baseline projection of annual outpatient visits at RSU Ananda Purwokerto. The design was observational because no intervention or treatment was assigned to patients or hospital units; the analysis used visit totals that had already been recorded in the Hospital Information System. Simple linear regression was selected to quantify the average linear association between calendar year and the annual number of outpatient visits and to extrapolate that historical trend to 2026–2030. Accordingly, the model was intended as a transparent planning baseline rather than as a causal model of the effects of the COVID-19 pandemic, BPJS cooperation, or other service policies.

The study setting was RSU Ananda Purwokerto, Jalan Pemuda No. 30, West Purwokerto, Banyumas, Indonesia. The unit of analysis was the annual aggregate number of outpatient visits, not an individual patient. The observation period covered 2018–2025 and produced eight annual observations. The outcome variable was the total number of recorded outpatient visits in each observation period, while the explanatory variable was calendar year. These variables were chosen because the study sought to obtain an initial description of the long-term direction of outpatient utilization using the aggregate data available to the hospital.

The empirical data consisted of retrospective administrative records obtained from the Medical Records Unit and the Hospital Information System of RSU Ananda Purwokerto. Because the records were generated before the present analysis and contained annual aggregate counts, they were treated as secondary administrative data. Scientific articles, electronic books, and relevant previous studies were used as supporting literature for the selection of the forecasting method, interpretation of outpatient demand, and evaluation of prediction error. The hospital record was therefore the source of the numerical analysis, whereas the literature provided the conceptual and methodological basis.

All available annual outpatient visit totals for 2018–2025 were included, so the study used a census of the eight annual observations rather than drawing a statistical sample. The descriptive context was classified as pre-pandemic for 2018–2019, the COVID-19 pandemic period for 2020–2021, and the post-pandemic period from 2022 onward. The termination of BPJS Kesehatan cooperation was recorded as contextual information associated with the 2023 service environment. These labels were not entered as independent variables and were not tested as causal determinants in the regression model.

Table 1. Annual Outpatient Visit Data at RSU Ananda Purwokerto

Year	Number of Visits	Context
2018	65,756	Pre-COVID-19
2019	83,319	Pre-COVID-19
2020	84,499	COVID-19 pandemic period
2021	89,287	COVID-19 pandemic period
2022	68,750	Post-COVID-19 period
2023	45,031	Post-COVID-19 period / BPJS cooperation terminated during the year
2024	69,771	Post-COVID-19 period
2025	70,039	Post-COVID-19 period / January-October only

Data were collected through document review, observation, interviews, and a literature review. Document review was used to obtain the annual outpatient visit totals from the hospital information records. Observation provided an overview of the outpatient service and record-management process. A semistructured interview with the person responsible for the Medical Records Unit was used to clarify the origin, coverage, and interpretation of the visit records, including the status of the 2025 observation. The literature review covered patient-demand forecasting, simple linear regression, healthcare data analytics, and prediction-error assessment. No names, medical diagnoses, identification numbers, or other patient-level information were used in the analysis.

Before modeling, the annual totals were arranged chronologically and the period represented by each record was identified. Values were analyzed as visit counts, with punctuation in the source table interpreted as a thousands separator. Particular attention was given to the 2025 value of 70,039 because it covered January through October only, whereas the 2018–2024 values represented complete calendar years. The 2025 value was retained to reproduce the dataset and model used in the study, but it was explicitly treated as a partial-year observation. It was not annualized because monthly visit totals for estimating November and December were not available. Consequently, its direct comparability with complete annual observations is limited. Simple linear regression was estimated with the actual calendar year as the predictor, using the following equation:

$$Y = a + bX$$

Where:

Y = Predicted Number of Outpatient Visits

X = Calendar Year

a = Intercept

b = Regression Coefficient or Estimated Annual Slope

The intercept represents the model-based value when X equals zero and therefore has no direct operational interpretation in this specification. The slope represents the estimated average change in outpatient visits associated with a one-year increase in calendar time. The analysis proceeded by compiling the annual data, identifying observation coverage, reviewing descriptive changes, estimating the intercept and slope, producing point projections for 2026–2030, and comparing the manual coefficients with those generated by the Linear Regression operator in RapidMiner. In RapidMiner, year was assigned as the numerical predictor and outpatient visits as the numerical label. RapidMiner was used to verify computational consistency rather than as a separate forecasting method. Statistical significance of the slope was assessed at the 5% level ($\alpha = .05$). Assumptions of linearity, independence, constant residual variance, and approximate residual normality were considered when interpreting the model, but the small dataset limits robust diagnostic assessment.

Prediction error was summarized using the Mean Absolute Percentage Error (MAPE), which calculates the average absolute difference between an observed value and its fitted value relative to the observed value. MAPE was selected because its percentage scale is readily interpretable by

hospital managers. In this study, MAPE was calculated using the same 2018–2025 observations employed to estimate the regression model, according to the following formula:

$$MAPE = \frac{\sum \left| \frac{A_t - F_t}{A_t} \right|}{n} \times 100\%$$

Where:

A_t = Actual Outpatient Visits in Period T

F_t = Fitted Outpatient Visits in Period T

n = Number of Evaluated Observation Periods

For descriptive interpretation, MAPE values of less than or equal to 10% were classified as very good, values greater than 10% and up to 20% as good, values greater than 20% and up to 50% as reasonable, and values above 50% as weak. These cutoffs were retained as practical descriptive guidelines rather than universal statistical standards because the suitability of percentage errors depends on the data and decision context (Armstrong & Collopy, 1992; Hyndman & Koehler, 2006). Because the calculation used the model-development observations, it measures in-sample fit and not independent future forecasting performance. Model adequacy was therefore not determined from MAPE alone; the slope p-value, limited number of observations, incomplete 2025 coverage, historical fluctuations, and absence of external predictors were also considered.

The final outputs were the fitted regression equation, the estimated annual direction of change, point projections for 2026–2030, coefficient consistency between manual calculation and RapidMiner, and the in-sample MAPE. The workflow provided an initial assessment for operational planning but did not estimate the independent effects of COVID-19, BPJS policy, staffing, physician availability, economic conditions, or patient satisfaction. The projections were therefore treated as exploratory baseline estimates and should be updated when complete 2025 data and more frequent monthly observations become available. Because only aggregate records were analyzed, no individual patient identity or clinical information was processed or reported.

Results and Discussion

The analysis used eight annual outpatient-visit observations for 2018–2025. The values for 2018–2024 represented complete calendar years, whereas the 2025 value of 70,039 covered January through October only. Calendar year was specified as the independent variable and the number of outpatient visits as the dependent variable. The data were analyzed manually and with the Linear Regression operator in RapidMiner. Because one observation represented only ten months and the series contained substantial year-to-year fluctuations, all estimates were interpreted as exploratory results from a limited dataset.

The fitted simple linear regression produced an intercept of 4,324,426.14 and a slope of -2,103.57. The resulting equation was:

$$Y = 4,324,426.14 - 2,103.57X$$

Table 2. Summary of Simple Linear Regression Analysis Results

Analysis Component	Result
Number of observations	8 annual observations
Data period	2018–2025 (2025 through October)
Constant (a)	4,324,426.14
Regression coefficient (b)	-2,103.57
Model equation	$Y = 4,324,426.14 - 2,103.57X$
Slope p-value	.369
MAPE	15.66%
MAPE category	Good (in-sample)
R^2	.136

As summarized in Table 2, the negative slope indicates an estimated average decrease of approximately 2,104 outpatient visits for each additional calendar year. Manual calculation and RapidMiner produced equivalent coefficients, with only rounding differences: 4,324,426.14 versus 4,324,426.143 for the intercept and -2,103.57 versus -2,103.571 for the slope. Recalculation from the eight reported observations produced $R^2 = .136$ and a two-sided slope p-value of .369. Thus, calendar year explained approximately 13.6% of the observed variation, and the slope was not statistically significant at $\alpha = .05$. The 95% confidence interval for the slope, approximately -7,405 to 3,198 visits per year, included zero. These statistics do not support treating the fitted decrease as a certain underlying trend.

The fitted equation was then evaluated at calendar years 2026–2030 to obtain point projections. These projections extend the estimated straight-line pattern beyond the observed period; they do not account for changes in financing arrangements, service capacity, patient preferences, disease patterns, or other external conditions. The resulting values are presented in Table 3.

Table 3. Projected Outpatient Visits at RSU Ananda Purwokerto for 2026–2030

Year	Projected Outpatient Visits
2026	62,590
2027	60,487
2028	58,383
2029	56,280
2030	54,176

Table 3 shows point projections of 62,590 visits in 2026, 60,487 in 2027, 58,383 in 2028, 56,280 in 2029, and 54,176 in 2030. The values decrease by approximately 2,104 visits per year because they are generated directly from the fitted linear slope. These figures describe the continuation of the estimated historical line and should not be interpreted as guaranteed annual totals. In particular, the 2030 projection is consistently reported as 54,176 visits.

The in-sample MAPE was 15.66%. Under the descriptive thresholds adopted in this study, this value falls within the 10%–20% category commonly labeled good. However, the value was calculated from the same eight observations used to estimate the coefficients. It therefore describes how closely the fitted line reproduces the model-development data and does not constitute an independent assessment of forecast performance for 2026–2030. The MAPE result must be considered together with the low R^2 , nonsignificant slope, short annual series, and incomplete 2025 observation.

Discussion

The main result is that simple linear regression provides a transparent baseline equation for describing the direction of annual outpatient visits at RSU Ananda Purwokerto. Its practical strength lies in the ease with which the coefficients and point projections can be calculated, reproduced, and communicated. This is consistent with Nugroho et al. (2024), who showed that simple linear regression can support an initial assessment of outpatient demand. Nevertheless, the present findings also show why an easily implemented model should not automatically be interpreted as a strong forecasting model. The regression summarizes the available annual observations, but the statistical evidence for a stable linear decline is weak.

The slope of -2,103.57 represents the average annual change implied by the fitted line, not a fixed decrease that must occur every year. Actual visits increased from 2018 to 2021, declined substantially in 2022 and 2023, and increased again in 2024 and the first ten months of 2025. This pattern is not consistently linear. The observed changes coincided with different service contexts, including the COVID-19 period and changes in BPJS Kesehatan cooperation, but these factors were not entered as predictors. Consequently, the present regression cannot quantify their effects or establish that they caused the observed fluctuations.

The projections decrease from 62,590 visits in 2026 to 54,176 visits in 2030. If future conditions follow the fitted historical line, this scenario may help the hospital anticipate changes in staffing, clinic schedules, registration services, waiting-room capacity, and supporting facilities. The projections are nevertheless conditional baseline values. They should be reviewed against current operational information and updated as new observations become available, especially because the model does not include seasonality, structural changes, or external explanatory variables. This cautious use is consistent with the role of forecasting as decision support rather than a substitute for managerial judgment (Shirvia et al., 2025; Subramanian, 2021).

From a managerial perspective, the point projections are more appropriately used for scenario planning than as fixed targets. Hospital management can compare actual monthly or annual visits with the baseline and investigate material deviations. Visits below the baseline may prompt a review of access, referral relationships, financing cooperation, clinic availability, patient experience, and service communication. Visits above the baseline may require adjustments to personnel, appointment schedules, registration capacity, medicines, and waiting areas. This approach allows the model to function as an early reference while preserving flexibility when actual demand differs from the historical trend.

The MAPE of 15.66% provides limited but useful information about in-sample fit. It indicates that the absolute fitted error averaged 15.66% of the observed visit total across the eight periods. The largest mismatch occurred around a year with a pronounced historical fluctuation, illustrating that a single straight line cannot fully represent abrupt changes in hospital utilization. Therefore, labeling the MAPE category as good should be understood only within the selected descriptive threshold. It does not override the low explanatory power of the model or guarantee comparable errors for future observations.

The inferential results reinforce the need for caution. With $R^2 = .136$, approximately 86.4% of the variation in the reported annual visit totals remained unexplained by calendar year. The slope p-value of .369 exceeded .05, and its confidence interval crossed zero, so the available data did not provide sufficient statistical evidence of a systematic linear decrease. In addition, the analysis contained only eight observations, and the 2025 value represented January–October rather than a complete year. The projections should therefore be regarded as preliminary and should not be used alone for high-stakes staffing or investment decisions. The comparison between manual calculation and RapidMiner confirms computational consistency for the same model specification. Both approaches produced the same intercept and slope after rounding. This agreement verifies the arithmetic implementation, but it does not independently validate the predictive adequacy of the model. Software can reproduce a correctly specified calculation even when the available data are limited or the fitted relationship is weak. RapidMiner is therefore best described as a supporting analytical tool that facilitates processing and replication (Lathifah & Dana, 2024), rather than as evidence that the future projections are accurate.

Despite these limitations, forecasting outpatient demand remains relevant to hospital resource planning. Anticipated visit volumes can inform workforce deployment, clinic schedules, service rooms, administrative capacity, medicines, and supporting facilities (Huang et al., 2020; Jones et al., 2008; Luo et al., 2017; Wang & Wang, 2025). Reviews of patient-arrival models emphasize that locally relevant predictors, suitable model choice, and explicit performance evaluation are essential because demand patterns differ across institutions (Jiang et al., 2023; Wargon et al., 2009). A baseline model may be particularly useful when it is updated routinely and compared with observed demand. More reliable evaluation would require complete and preferably monthly data, a longer observation period, time-ordered out-of-sample validation, prediction intervals, and comparisons with time-series or multivariable models. External variables such as BPJS arrangements, physician availability, service access, and patient satisfaction could also be incorporated when valid measurements are available.

Overall, the analysis shows that the 2018–2025 secondary administrative records can be summarized using a simple linear regression and reproduced in RapidMiner. The model yields a declining baseline projection for 2026–2030 and an in-sample MAPE of 15.66%, but the low R^2 , nonsignificant slope, partial 2025 record, and pronounced historical fluctuations restrict the strength of the conclusion. The results may support initial operational discussion at RSU Ananda Purwokerto, provided that they are combined with current hospital information, presented with appropriate uncertainty, and updated when complete and higher-frequency data become available.

Conclusion

This study demonstrates that simple linear regression can be used to construct a transparent baseline projection of outpatient visits at RSU Ananda Purwokerto. Using the available 2018–2025 administrative records, the model estimated an average decrease of approximately 2,104 visits per year and generated point projections of 62,590 visits in 2026 and 54,176 visits in 2030. These values describe the continuation of the fitted historical line and should not be interpreted as fixed future totals. The statistical results limit the strength of the forecast conclusion. Calendar year explained 13.6% of the observed variation ($R^2 = .136$), and the estimated slope was not statistically significant at the 5% level ($p = .369$). The MAPE of 15.66% fell within the study's descriptive good category, but it represented in-sample fit because it was calculated from the same observations used to estimate the model. In addition, the dataset contained only eight annual observations, and the 2025 record covered January–October rather than a complete calendar year. Consequently, the model is more appropriately regarded as an exploratory planning baseline.

For hospital management, the projections may provide an initial reference for reviewing workforce allocation, clinic schedules, registration capacity, medicines, waiting areas, and other outpatient facilities. The model does not establish that COVID-19 conditions or changes in BPJS Kesehatan cooperation caused the observed visit fluctuations because those factors were not included as predictors. Operational decisions should therefore combine the baseline projections with recent service information, financing arrangements, physician availability, patient access, and other relevant indicators. Future analysis should use complete and preferably monthly visit data, a longer observation period, out-of-sample evaluation, and prediction intervals. Comparisons with time-series, machine-learning, or multivariable approaches may determine whether they provide more reliable forecasts. Regardless of the method selected, the model should be updated periodically so that outpatient planning remains responsive to actual changes in hospital demand.

Acknowledgment

-

References

- Areru, H. A., Dangisso, M. H., & Lindtjørn, B. (2021). Low and unequal use of outpatient health services in public primary health care facilities in southern Ethiopia: A facility-based cross-sectional study. *BMC Health Services Research*, 21, Article 776.
<https://doi.org/10.1186/s12913-021-06846-x>
- Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8(1), 69–80.
[https://doi.org/10.1016/0169-2070\(92\)90008-W](https://doi.org/10.1016/0169-2070(92)90008-W)
- Baihaqi, W. M., Dianingrum, M., & Ramadhan, K. A. N. (2019). Regresi linier sederhana untuk memprediksi kunjungan pasien di rumah sakit berdasarkan jenis layanan dan umur pasien.

- Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, 5(2), 55–62.
<https://doi.org/10.24014/coreit.v5i2.7067>
- Bai, L., Ding, G., Gu, S., Bi, P., Su, B., Qin, D., Xu, G., & Liu, Q. (2023). Predicting monthly hospital outpatient visits based on meteorological environmental factors using the ARIMA model. *Scientific Reports*, 13, Article 2812. <https://doi.org/10.1038/s41598-023-29897-y>
- Barsasella, D., & Faturrohmi, Z. A. (2025). Review of inpatient case-mix workforce needs with health workload analysis (ABK-KES) in the case-mix unit of Permata Bunda General Hospital, Tasikmalaya. *Jurnal Persada Husada Indonesia*, 12(2), 18–28.
<https://doi.org/10.56014/jphi.v12i2.428>
- Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *JAMA*, 319(13), 1317–1318. <https://doi.org/10.1001/jama.2017.18391>
- Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120, 70–83. <https://doi.org/10.1016/j.csda.2017.11.003>
- Fernandes Prabhu, D., Gurupur, V., Stone, A., & Trader, E. (2025). Integrating artificial intelligence, electronic health records, and wearables for predictive, patient-centered decision support in healthcare. *Healthcare*, 13(21), Article 2753.
<https://doi.org/10.3390/healthcare13212753>
- Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1, 125–151. <https://doi.org/10.1146/annurev-statistics-062713-085831>
- Haryanto, H., Winarto, H., & Juliane, C. (2024). Application of data mining techniques in healthcare: Identifying inter-disease relationships through association rule mining. *Journal of World Science*, 3(5), 520–529. <https://doi.org/10.58344/jws.v3i5.597>
- Huang, Y., Xu, C., Ji, M., Xiang, W., & He, D. (2020). Medical service demand forecasting using a hybrid model based on ARIMA and self-adaptive filtering method. *BMC Medical Informatics and Decision Making*, 20, Article 237. <https://doi.org/10.1186/s12911-020-01256-1>
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.
<https://doi.org/10.1016/j.ijforecast.2006.03.001>
- Jiang, S., Liu, Q., & Ding, B. (2023). A systematic review of the modelling of patient arrivals in emergency departments. *Quantitative Imaging in Medicine and Surgery*, 13(3), 1957–1971.
<https://doi.org/10.21037/qims-22-268>
- Jones, S. S., Thomas, A., Evans, R. S., Welch, S. J., Haug, P. J., & Snow, G. L. (2008). Forecasting daily patient volumes in the emergency department. *Academic Emergency Medicine*, 15(2), 159–170. <https://doi.org/10.1111/j.1553-2712.2007.00032.x>
- Lathifah, U., & Dana, R. D. (2024). Implementasi metode linear regression untuk prediksi harga properti real estate menggunakan RapidMiner. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 1129–1137. <https://doi.org/10.36040/jati.v8i1.8919>
- Luo, L., Luo, L., Zhang, X., & He, X. (2017). Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. *BMC Health Services Research*, 17, Article 469. <https://doi.org/10.1186/s12913-017-2407-9>

- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34(4), 802–808. <https://doi.org/10.1016/j.ijforecast.2018.06.001>
- Nugroho, B. I., Fatkhurrozak, Andriani, W., & Gunawan. (2024). Prediksi jumlah kunjungan pasien rawat jalan menggunakan metode regresi linier sederhana. *PROSISKO: Jurnal Pengembangan Riset dan Observasi Sistem Komputer*, 11(2), 284–293. <https://doi.org/10.30656/prosisko.v11i2.8756>
- Ratumurun, S. J., & Manik, M. J. (2025). Analysis of factors influencing outpatient waiting time: A literature review. *Jurnal Penelitian Keperawatan Kontemporer*, 5(5), 77–90. <https://doi.org/10.59894/jpkk.v5i5.1087>
- RSU Ananda Purwokerto. (2025). *Data rekam medis kunjungan pasien rawat jalan tahun 2018–2025*. Unit Rekam Medis RSU Ananda Purwokerto.
- Sabran, S., Azizah, S. N., & Rachmawati, E. (2024). Prediksi kunjungan pasien rawat jalan dengan metode analisis trend linear di RSUD Haji Provinsi Jawa Timur tahun 2023–2025. *Jurnal Manajemen dan Administrasi Rumah Sakit Indonesia (MARSI)*, 8(1), 26–32. <https://doi.org/10.52643/marsi.v8i1.4046>
- Setiatin, S., & Syahidin, Y. (2021). Analisis lama waktu tunggu pelayanan pasien rawat jalan di bagian pendaftaran Rumah Sakit X Bandung. *Akrab Juara: Jurnal Ilmu-Ilmu Sosial*, 6(5), 148–158. <https://www.akrabjuara.com/index.php/akrabjuara/article/view/1692>
- Shirvia, B., Riyandari, L., & Safitri, R. (2025). Prediksi kunjungan pasien rawat jalan dengan interpretasi analisis SWOT menggunakan algoritma regresi linear sederhana: Studi kasus RSU Bunda Purwokerto. *Jurnal Ilmiah Komputasi*, 24(2), 243–252. <https://doi.org/10.32409/jikstik.24.2.3787>
- Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., & Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology*, 21(1), 128–138. <https://doi.org/10.1097/EDE.0b013e3181c30fb2>
- Subramanian, L. (2021). Effective demand forecasting in health supply chains: Emerging trend, enablers, and blockers. *Logistics*, 5(1), Article 12. <https://doi.org/10.3390/logistics5010012>
- Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37–45. <https://doi.org/10.1080/00031305.2017.1380080>
- Wang, M., & Wang, M. (2025). Forecasting the demand for human resources in a hospital using the grey forecasting model. *Management Science and Business Decisions*, 5(1), 15–21. <https://doi.org/10.52812/msbd.110>
- Wardani, Y., Sopia, & Zagladi, A. N. (2024). Application of the workload indicators of staffing need method in achieving hospital service quality in developing countries: Systematic literature review. *Jurnal Pendidikan Administrasi Perkantoran (JPAP)*, 12(3), 391–405. <https://doi.org/10.26740/jpap.v12n3.p391-405>
- Wargon, M., Guidet, B., Hoang, T. D., & Hejblum, G. (2009). A systematic review of models for forecasting the number of emergency department visits. *Emergency Medicine Journal*, 26(6), 395–399. <https://doi.org/10.1136/emj.2008.062380>
- Yu, L., Hang, G., Tang, L., Zhao, Y., & Lai, K. K. (2017). Forecasting patient visits to hospitals using a WD&ANN-based decomposition and ensemble model. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(12), 7615–7627. <https://doi.org/10.12973/ejmste/80308>