Bigint Computing Journal
Volume 4, Issue 2, 2026, pp. 68-73
DOI: https://doi.org/10.55537/bigint.v4i2.1852
ISSN: 3032-5374
https://journal.aira.or.id/bigint/index

Streamlit-Based Durian Yield Forecasting Using an ARIMA Model

Nurul Ifkah Lolona Silalahi*, Triase

Department of Information Systems, Faculty of Science and Technology, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

ABSTRACT
Durian production in Dairi Regency, North Sumatra, fluctuates substantially across quarters, creating uncertainty for harvest planning and distribution. This study develops a lightweight decision-support system that integrates an AutoRegressive Integrated Moving Average (ARIMA) forecasting pipeline with a Streamlit web application. The source dataset contains 41,060 agricultural harvest records from 2020–2024, aggregated into 20 quarterly regional observations. The raw series was non-stationary according to the Augmented Dickey–Fuller test (ADF = −0.449, p = 0.901), while first-order differencing produced a stationary series (ADF = −4.120, p = 0.0009). Automated model search selected ARIMA (4,0,1), with AIC = 401.649 and BIC = 407.624. A chronological 80/20 holdout evaluation on the four quarters of 2024 produced an RMSE of 1.16 tons, MAE of 1.15 tons, and MAPE of 4.33%, recalculated from the reported quarter-level forecasts. The Streamlit implementation integrates data management, stationarity diagnosis, automated parameter selection, and forecast visualization. The results indicate that an interpretable ARIMA baseline can provide useful short-horizon regional forecasts when historical data are limited, although the short series requires cautious generalization and further validation.
KEYWORDS: ARIMA, Agricultural Forecasting, Durian Production, Streamlit, Time Series

*Corresponding author: Nurul Ifkah Lolona Silalahi, nurulsilalahi637@gmail.com

Article History: Received: 2026-07-31; Accepted: 2026-08-15

1. INTRODUCTION

Crop-yield forecasting supports production planning, supply-chain coordination, and food-system decision making, particularly when production varies across seasons [1]. Reviews of agricultural prediction research show that data-driven forecasting is increasingly used as a decision-support function, but performance depends strongly on data volume, temporal resolution, and validation design [1]. Operational forecasting systems also benefit from workflows that connect predictive models to accessible interfaces rather than leaving statistical outputs as isolated analytical scripts [2], [3].

For relatively short univariate series, classical statistical models remain valuable because they are interpretable, computationally efficient, and can be estimated without the large training sets typically required by modern machine-learning or deep-learning architectures [4], [5]. ARIMA is a standard framework for modeling autocorrelation and stochastic dynamics and can effectively model trends and fluctuations in univariate time series without requiring additional independent variables [6]. Automated search procedures can further select parsimonious model orders using information criteria [7], [8]. This makes ARIMA a reasonable baseline for the Dairi dataset, which contains only 20 quarterly observations after aggregation [6], [7].

The empirical context of this study is durian production in Dairi Regency. The agricultural office records harvest information at a granular farmer level, but the operational data are not directly structured as a forecasting series. Converting these records into a chronological quarterly series and integrating the model into a web application can provide a practical bridge between statistical forecasting and routine agricultural management. More complex methods such as recurrent neural networks can be powerful in agriculture, but their advantages are usually more defensible when substantially larger datasets or richer explanatory variables are available [4], [5]. Recent crop-yield studies further demonstrate that prediction performance can improve when climatic and biophysical information is incorporated through machine-learning or hybrid modeling approaches [9]–[11].

This study addresses that gap by developing a Streamlit-based forecasting system for quarterly durian production. The contribution comprises three integrated elements. First, 41,060 harvest records are transformed into a regional quarterly time series. Second, an automated ARIMA workflow integrates stationarity diagnostics, model selection, and chronological holdout evaluation. Third, the analytical workflow is deployed through an interactive web interface for non-specialist users. The study also explicitly reports the limitations created by the short time series and by differences between the ADF diagnostic and the model order selected by the automated search.

2. METHODOLOGY

2.1. Data Collection and Quarterly Aggregation

The dataset was obtained from the Dairi Regency Agricultural Office and covers the 2020–2024 period. It contains 41,060 harvest records with fields describing farmer identity, sub-district, cultivated land area, mature trees, extension officer, harvest quarter, harvest year, and recorded yield. For forecasting, individual records were aggregated by quarter to form one regional production value for each period. The resulting sequence contains 20 contiguous observations from 2020-Q1 to 2024-Q4.

$$Y_t = \Sigma_i \, y_{i,t} \tag{1}$$

where Y_t is total regional durian production in quarter t and $y_{i,t}$ is the recorded production of observation i in the same quarter. Aggregation creates a regular univariate time series but also reduces the effective sample size, which is a central limitation of the analysis.

2.2. Stationarity Diagnosis and ARIMA Model Selection

Stationarity was assessed with the Augmented Dickey–Fuller (ADF) unit-root test [12]. A p-value above 0.05 was interpreted as insufficient evidence to reject a unit root, while first-order differencing was examined as a diagnostic transformation [12]. The ARIMA(p,d,q) model represents autoregressive order p, differencing order d, and moving-average order q, following the Box–Jenkins formulation [6]. Automated model identification followed a stepwise search over low-order p and q values using the principle of information-criterion minimization [7], [8].

$$\varphi(B)(1 - B)^d Y_t = c + \theta(B)\varepsilon_t \tag{2}$$

Because the series contains only five years of quarterly observations, the search was configured as non-seasonal to reduce parameter proliferation. The ADF result was used as a diagnostic rather than imposed as a hard constraint on d; consequently, the automated search could select a model whose differencing order differed from the stand-alone ADF indication. This distinction is reported transparently because information criteria and unit-root tests answer different model-selection questions [8], [13].

2.3. Forecast Evaluation

Evaluation used a chronological 80/20 split: the first 16 quarters (2020-Q1 to 2023-Q4) formed the training set, and the four quarters of 2024 formed the test set. Time-ordered holdout evaluation was preferred to random resampling because observations are serially dependent. Out-of-sample testing and time-series cross-validation principles emphasize preserving temporal order when estimating generalization performance [14], [15].

$$RMSE = \sqrt{[(1/n)\, \Sigma_t (Y_t - \hat{Y}_t)^2]} \tag{3}$$

$$MAE = (1/n)\, \Sigma_t \, |Y_t - \hat{Y}_t| \tag{4}$$

$$MAPE = (100/n)\, \Sigma_t \, |(Y_t - \hat{Y}_t) / Y_t| \tag{5}$$

RMSE, MAE, and MAPE were calculated from the quarter-level holdout forecasts, where lower values generally indicate better forecasting accuracy [16]. Forecast-accuracy measures should nevertheless be interpreted in relation to data scale and evaluation design rather than used as stand-alone evidence of universal model superiority [13], [16].

2.4. Streamlit Implementation

The statistical workflow was integrated into a Streamlit application that supports harvest-data inspection, preprocessing, stationarity testing, automated model selection, and forecasting visualization. Streamlit was selected because it allows Python-based analytical pipelines to be exposed through an interactive web interface with minimal deployment overhead [17]. The interface was localized for the intended Indonesian users while the analytical outputs retain standard statistical terminology.

3. RESULTS AND DISCUSSION

3.1. Quarterly Production Pattern and Stationarity

The aggregated series shows a recurring within-year pattern across the 20 quarters. Production is generally lower in Quarter II and substantially higher in Quarter IV. The minimum observed value is 18.90 tons in 2021-Q2, while the maximum is 38.50 tons in 2024-Q4. Figure 1 presents the complete quarterly trajectory.

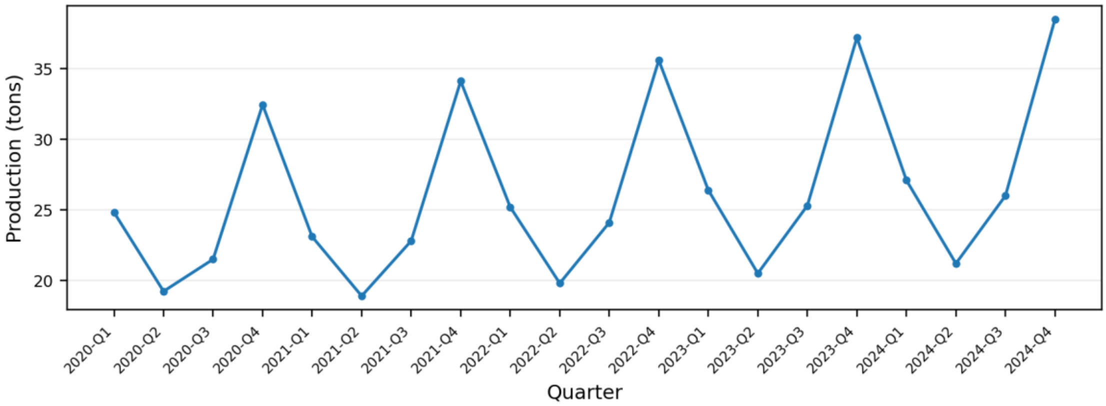

Figure 1. Historical Quarterly Durian Production in Dairi Regency (2020–2024)

The raw series produced an ADF statistic of −0.449 with p = 0.901, indicating that the unit-root null hypothesis could not be rejected. After first-order differencing, the ADF statistic decreased to −4.120 with p = 0.0009, providing evidence of stationarity. Table 1 summarizes the diagnostic results.

Table 1. Augmented Dickey–Fuller Stationarity Test Results

Series	ADF statistic	p-value	5% critical value	Decision
Raw series (d=0)	−0.449	0.901	−3.00	Non-stationary
First difference (d=1)	−4.120	0.0009	−3.00	Stationary

3.2. Model Selection and Holdout Forecast Accuracy

The automated search selected ARIMA(4,0,1) as the preferred specification, with AIC = 401.649 and BIC = 407.624 as reported by the implemented model-selection module. This selected order differs from the first-difference indication of the ADF test; therefore, the result is treated as an automated information-criterion choice rather than as proof that the raw process is stationary.

Table 2. Out-of-Sample Forecasts for the 2024 Testing Period

Testing period	Actual yield (tons)	Forecast (tons)	Absolute error (tons)	Percentage error
2024-Q1	27.10	26.05	1.05	3.87%
2024-Q2	21.20	19.85	1.35	6.37%
2024-Q3	26.00	24.90	1.10	4.23%
2024-Q4	38.50	37.40	1.10	2.86%

Across the four holdout quarters, forecasts track the observed direction and remain within 1.35 tons of the actual value. Recalculation from Table 2 yields RMSE = 1.16 tons, MAE = 1.15 tons, and MAPE = 4.33%. Figure 2 shows that the largest proportional error occurs in 2024-Q2, while the forecast remains close to the observed peak in 2024-Q4.

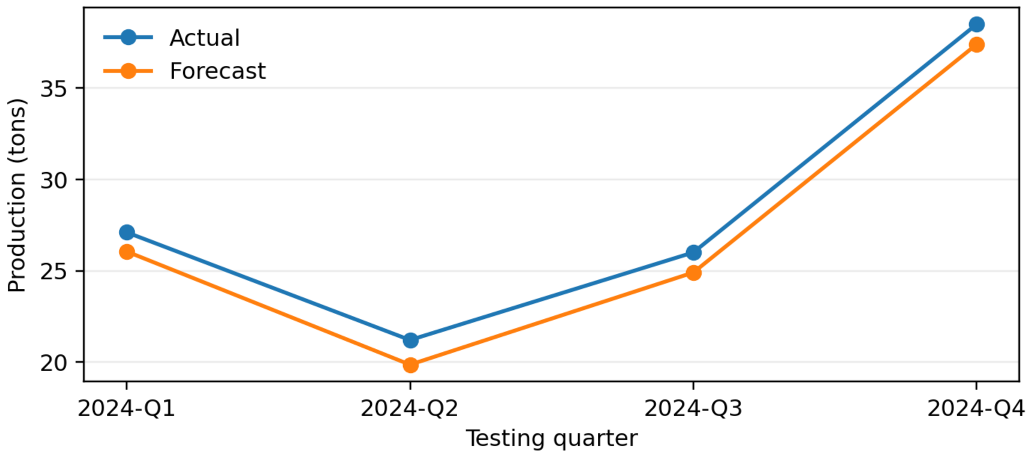

Figure 2. Actual and ARIMA Forecast Production During the 2024 Holdout Period

3.3. UML System Design

The system architectural design is structured using Unified Modeling Language (UML) diagrams, specifically Use Case and Activity diagrams, to define functional interactions and procedural workflows.

Use Case Diagram

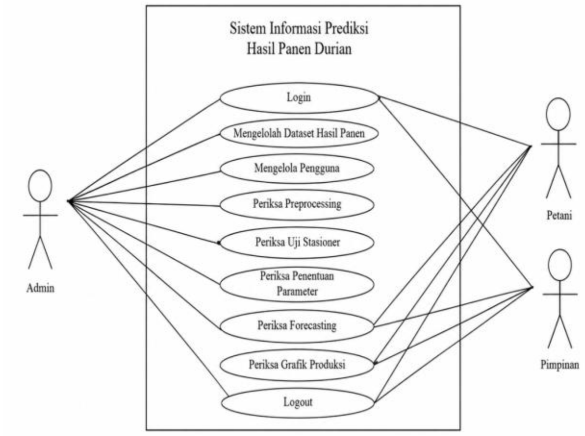

Figure 3. UML Use Case Diagram of Proposed Streamlit Forecasting System

A Use Case Diagram describes system functionality and the interactions between the system and its actors. As shown in Figure 3, the system involves three primary actors: Admin, Farmer (Petani), and Head/Management (Pimpinan). The Admin is responsible for managing harvest-yield datasets and users, as well as performing preprocessing, stationarity testing, ARIMA parameter determination, forecasting, and production-graph monitoring. The Farmer and Head/Management actors primarily interact with the system to access forecasting results and production information. These use cases define the main functional scope of the proposed system and indicate the responsibilities assigned to each actor.

Activity Diagram

An Activity Diagram illustrates a system workflow by showing the sequence of activities, decision points, and possible outcomes throughout a business or system process.

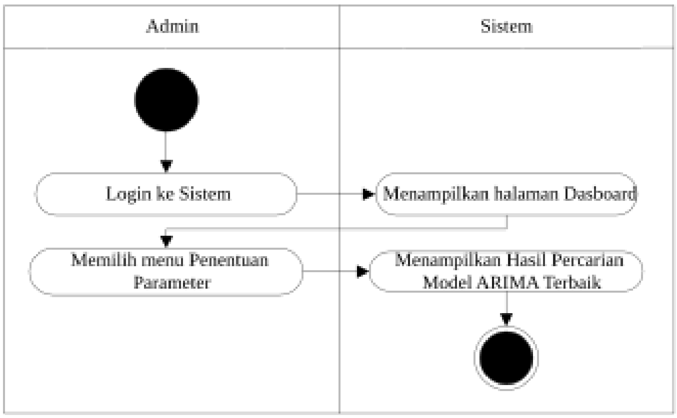

Figure 4. UML Activity Diagram of the Automated ARIMA Parameter Determination Workflow

Figure 4 illustrates the workflow for determining the optimal ARIMA model parameters. The process begins when the Admin logs into the system and accesses the Parameter Determination menu. The system then performs an automated search by evaluating candidate ARIMA models based on their Akaike Information Criterion (AIC) values. After the evaluation process is completed, the system displays the ARIMA model with the selected parameter configuration and the corresponding results to the Admin.

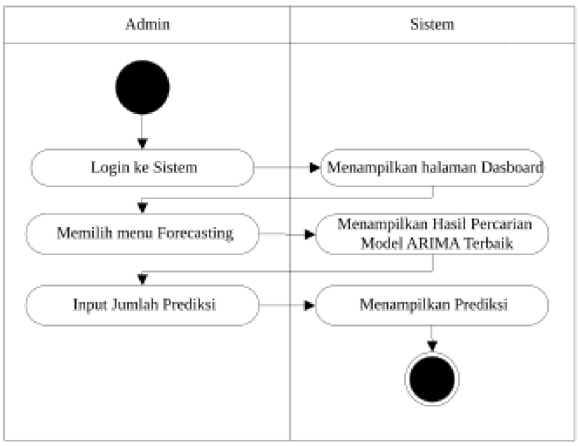

Figure 5. UML Activity Diagram: Multi-Quarter Harvest Yield Forecasting Workflow

Figure 5 illustrates the workflow for generating harvest-yield forecasts using the selected ARIMA model. The process begins with Admin authentication, followed by navigation to the Forecasting menu. The Admin then specifies the desired forecast horizon by entering the number of future quarters to be predicted. The system processes the input using the selected ARIMA model and subsequently displays the forecasted harvest yield. The resulting predictions can then be presented through the system interface to support interpretation of the projected production.

3.4. Streamlit-Based Forecasting Interface

The forecasting pipeline was implemented as an interactive Streamlit application. The forecasting screen presents the selected ARIMA specification, model information criteria, and a visual comparison between historical production and projected values. The interface allows agricultural administrators to execute the analytical workflow without directly operating statistical code. Figure 6 shows the deployed forecasting view.

Figure 6. Streamlit Forecasting Interface for Durian Production

3.5. Comparative Discussion and Methodological Limitations

The results support the use of a classical statistical baseline in a small-data agricultural setting. Large reviews of crop-yield prediction emphasize that more complex machine-learning models can be effective when richer climatic, soil, and remote-sensing inputs are available [1], [2]. However, studies focused on small-data yield forecasting also show that model complexity must be justified by the available sample and validation evidence [3]. In the present case, only 20 quarterly observations are available after aggregation, so a transparent ARIMA baseline is easier to estimate and interpret than a high-capacity neural model [6], [7].

Three limitations are particularly important. First, the four-quarter holdout is too small to establish stable generalization, and rolling-origin evaluation should be adopted when additional years become available [14], [15]. Second, the raw series fails the ADF test, whereas the automated AIC search selects d=0. This mismatch should be investigated with residual diagnostics, alternative differencing constraints, and direct comparisons with ARIMA models that enforce d=1 [7], [8], [12]. Third, the model is univariate and does not incorporate rainfall, temperature, orchard age, pest pressure, or management practices. Agricultural forecasting research consistently indicates that such explanatory variables can materially improve yield prediction when reliable data are available [1], [2], [4].

Accordingly, the reported MAPE should be interpreted as preliminary evidence on one historical split rather than as a definitive accuracy claim. Future work should compare ARIMA with seasonal and exogenous-variable models, use longer time series, evaluate multiple forecast origins, and quantify prediction intervals. Forecasting research broadly recommends that model choice be evaluated through reproducible out-of-sample performance rather than in-sample fit alone [13], [16].

4. CONCLUSION

This study developed a Streamlit-based ARIMA decision-support workflow for quarterly durian production forecasting in Dairi Regency. A total of 41,060 harvest records were aggregated into 20 quarterly observations covering 2020–2024. The raw time series was non-stationary under the ADF test, while first differencing produced a stationary diagnostic result. Automated model selection identified ARIMA(4,0,1) with AIC 401.649 and BIC 407.624. Based on the four-quarter 2024 holdout, the reported forecasts correspond to an RMSE of 1.16 tons, MAE of 1.15 tons, and MAPE of 4.33%.

The main practical contribution is the integration of the forecasting workflow into an accessible web interface that combines data inspection, stationarity diagnosis, model selection, and forecast visualization. Nevertheless, the short series, single holdout period, and inconsistency between the ADF indication and the automated d=0 selection limit the strength of the statistical conclusion. Longer historical data, rolling-origin validation, residual diagnostics, seasonal models, and climatic exogenous variables are recommended before the system is used for higher-stakes production planning.

ACKNOWLEDGMENT

The authors thank Universitas Islam Negeri Sumatera Utara and the Dairi Regency Agricultural Office for supporting the study and providing access to the agricultural harvest records used in the analysis.

DECLARATIONS

Artificial Intelligence Use

Generative AI was used only to assist with language editing during manuscript revision. The authors reviewed the final text and remain responsible for the article content, analysis, and conclusions.

Author Contributions

Nurul Ifkah Lolona Silalahi: Conceptualization, methodology, software development, data curation, formal analysis, visualization, and writing—original draft. Triase: Supervision, methodology review, validation, and writing—review and editing. All authors reviewed and approved the final manuscript.

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

Ethics Approval / Ethical Clearance

Not applicable. This study used secondary administrative agricultural records and did not involve experimental research with human participants or animals.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request, subject to the data-access restrictions of the Dairi Regency Agricultural Office and applicable privacy considerations.

REFERENCES

[1] T. van Klompenburg, A. Kassahun, and C. Catal, “Crop yield prediction using machine learning: A systematic literature review,” Computers and Electronics in Agriculture, vol. 177, Art. no. 105709, 2020, doi: 10.1016/j.compag.2020.105709.

[2] D. Paudel, H. Boogaard, A. de Wit, S. Janssen, S. Osinga, C. Pylianidis, and I. N. Athanasiadis, “Machine learning for large-scale crop yield forecasting,” Agricultural Systems, vol. 187, Art. no. 103016, 2021, doi: 10.1016/j.agsy.2020.103016.

[3] M. Meroni, F. Waldner, L. Seguini, H. Kerdiles, and F. Rembold, “Yield forecasting with machine learning and small data: What gains for grains?,” Agricultural and Forest Meteorology, vols. 308–309, Art. no. 108555, 2021, doi: 10.1016/j.agrformet.2021.108555.

[4] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, “Machine learning in agriculture: A review,” Sensors, vol. 18, no. 8, Art. no. 2674, 2018, doi: 10.3390/s18082674.

[5] A. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” Computers and Electronics in Agriculture, vol. 147, pp. 70–90, 2018, doi: 10.1016/j.compag.2018.02.016.

[6] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control, 5th ed. Hoboken, NJ, USA: Wiley, 2015.

[7] R. J. Hyndman and Y. Khandakar, “Automatic time series forecasting: The forecast package for R,” Journal of Statistical Software, vol. 27, no. 3, pp. 1–22, 2008, doi: 10.18637/jss.v027.i03.

[8] H. Akaike, “A new look at the statistical model identification,” IEEE Transactions on Automatic Control, vol. 19, no. 6, pp. 716–723, 1974, doi: 10.1109/TAC.1974.1100705.

[9] J. H. Jeong, J. P. Resop, N. D. Mueller, D. H. Fleisher, K. Yun, E. E. Butler, D. J. Timlin, K.-M. Shim, J. S. Gerber, V. R. Reddy, and S.-H. Kim, “Random forests for global and regional crop yield predictions,” PLOS ONE, vol. 11, no. 6, Art. no. e0156571, 2016, doi: 10.1371/journal.pone.0156571.

[10] A. Crane-Droesch, “Machine learning methods for crop yield prediction and climate change impact assessment in agriculture,” Environmental Research Letters, vol. 13, no. 11, Art. no. 114003, 2018, doi: 10.1088/1748-9326/aae159.

[11] M. Shahhosseini, G. Hu, I. Huber, and S. V. Archontoulis, “Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt,” Scientific Reports, vol. 11, Art. no. 1606, 2021, doi: 10.1038/s41598-020-80820-1.

[12] D. A. Dickey and W. A. Fuller, “Distribution of the estimators for autoregressive time series with a unit root,” Journal of the American Statistical Association, vol. 74, no. 366a, pp. 427–431, 1979, doi: 10.1080/01621459.1979.10482531.

[13] F. Petropoulos et al., “Forecasting: Theory and practice,” International Journal of Forecasting, vol. 38, no. 3, pp. 705–871, 2022, doi: 10.1016/j.ijforecast.2021.11.001.

[14] C. Bergmeir, R. J. Hyndman, and B. Koo, “A note on the validity of cross-validation for evaluating autoregressive time series prediction,” Computational Statistics & Data Analysis, vol. 120, pp. 70–83, 2018, doi: 10.1016/j.csda.2017.11.003.

[15] L. J. Tashman, “Out-of-sample tests of forecasting accuracy: An analysis and review,” International Journal of Forecasting, vol. 16, no. 4, pp. 437–450, 2000, doi: 10.1016/S0169-2070(00)00065-0.

[16] 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.

[17] Streamlit, “Streamlit documentation.” Accessed: Aug. 7, 2026. [Online]. Available: https://docs.streamlit.io/