IMPLEMENTATION OF DATA MINING METHODS FOR PREDICTING STUNTING IN KEDIRI CITY BASED ON GEOSPATIAL VISUALIZATION

Nanda Alviodita[1)], I Kadek Dwi Nuryana[2)]

[1)]Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya, Surabaya, Indonesia
Email: nandaalviodita.22006@mhs.unesa.ac.id
[2)]Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya, Surabaya, Indonesia
Email: dwinuryana@unesa.ac.id

Abstract

Stunting remains a significant public health issue in Indonesia, including Kediri City, where 17.6% of children were affected in 2024. The limited availability of predictive stunting data is one of the main obstacles in implementing more targeted interventions. This study applies the CRISP-DM framework using Holt's Double Exponential Smoothing (Holt DES) and ARIMA methods to analyze historical stunting trends and forecasts case numbers for January 2025. Data were obtained from four community health centers (Puskesmas) in Kediri City (Sukorame, Pesantren I, Ngletih, and Mrican), comprising 4,281 stunting measurement records from 18 sub-districts (Kelurahan), classified by the height-for-age (TB/U) indicator. The dataset was divided into 10 months of training data (January-October 2024) and 2 months of testing data (November-December 2024). Evaluation results indicate that ARIMA achieved the lowest MAPE of 2.98% outperforming Holt DES at 3.83%, both classified as excellent. ARIMA forecasts 337 stunting cases for January 2025, which were distributed across 18 sub-districts based on historical proportions. The results are presented in an interactive Tableau dashboard with geospatial visualization, enabling decision-makers to identify sub-district with the highest predicted stunting burden in Kediri City.

Keywords: stunting, forecasting, ARIMA, Holt's Double Exponential Smoothing, geospatial visualization, Tableau, data mining

Abstrak

Stunting masih menjadi salah satu permasalahan kesehatan masyarakat yang signifikan di Indonesia, termasuk di Kota Kediri, dengan prevalensi sebesar 17,6% pada tahun 2024. Keterbatasan ketersediaan data prediksi stunting menjadi salah satu kendala utama dalam pelaksanaan intervensi yang lebih tepat sasaran. Penelitian ini menerapkan kerangka kerja Cross-Industry Standard Process for Data Mining (CRISP-DM) dengan menggunakan metode Holt's Double Exponential Smoothing (Holt DES) dan Autoregressive Integrated Moving Average (ARIMA) untuk menganalisis tren historis stunting serta memprediksi jumlah kasus pada Januari 2025. Data diperoleh dari empat Pusat Kesehatan Masyarakat (Puskesmas) di Kota Kediri, yaitu Sukorame, Pesantren I, Ngletih, dan Mrican, yang mencakup 4.281 catatan pengukuran stunting dari 18 kelurahan berdasarkan indikator tinggi badan menurut umur (TB/U). Dataset dibagi menjadi data pelatihan selama 10 bulan (Januari–Oktober 2024) dan data pengujian selama 2 bulan (November–Desember 2024). Hasil evaluasi menunjukkan bahwa metode ARIMA menghasilkan nilai Mean Absolute Percentage Error (MAPE) terendah sebesar 2,98%, lebih baik dibandingkan metode Holt DES yang menghasilkan nilai MAPE sebesar 3,83%. Kedua metode tersebut termasuk dalam kategori akurasi prediksi sangat baik. Metode ARIMA memprediksi sebanyak 337 kasus stunting pada Januari 2025, yang kemudian didistribusikan ke 18 kelurahan berdasarkan proporsi historis masing-masing wilayah. Hasil prediksi disajikan melalui dashboard Tableau interaktif yang dilengkapi dengan visualisasi geospasial, sehingga memungkinkan para pengambil kebijakan untuk mengidentifikasi kelurahan dengan prediksi jumlah kasus stunting tertinggi di Kota Kediri.

Kata Kunci: stunting, peramalan, ARIMA, Holt's Double Exponential Smoothing, visualisasi geospasial, Tableau, penambangan data.

PENDAHULUAN

Stunting is characterized by a child's height falling below the age-specific standard due to prolonged chronic malnutrition. The World Health Organization (WHO) defines stunting as a length-for-age or height-for-age (TB/U) measurement below -2 standard deviations (SD) from the median of the WHO child growth standards, indicating short stature relative to age [1]. Stunting impedes not only physical growth but also cognitive development and learning ability [4].

Data from the 2024 Indonesian Nutritional Status Survey (SSGI) indicate that the national prevalence of stunting in Indonesia was 19.8%, while Kediri City reported a rate of 17.6% in the same year [2]. Despite this decline from the previous year, ongoing intervention remains necessary to align with the second Sustainable Development Goal (SDG), which aims to end hunger and enhance nutritional quality [3].

A significant challenge in addressing stunting is the limited availability of predictive data. Current data are typically confined to annual statistical reports, restricting their utility for trend analysis and forecasting [10]. Furthermore, substantial variation in stunting prevalence across sub-districts underscores the need for location-specific information to facilitate more targeted interventions [8].

To address these gaps, this study employs data mining techniques, specifically Holt's Double Exponential Smoothing and ARIMA, to analyze historical stunting trends and forecast case numbers for January 2025[6], [7]. The resulting forecasts are disaggregated at the sub-district level and visualized using an interactive Tableau dashboard with geospatial features [9], [15], providing a more comprehensive, location-specific understanding of stunting in Kediri City.

METHODS

This research uses data mining in accordance with the CRISP-DM (*Cross Industry Standard Process for Data* Mining) framework, which comprises six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment [13]. The CRISP-DM framework was selected due to its structured, sequential approach to data mining projects, which ensures alignment between each phase and the research objectives [13]. The overall research process is illustrated in Figure 1.

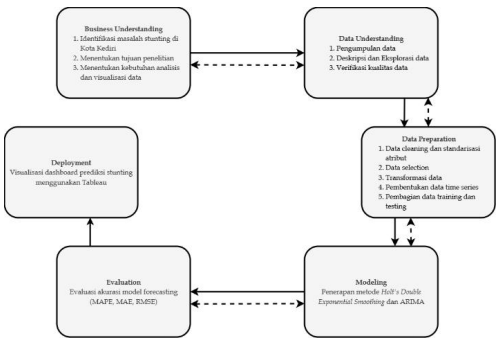


Figure 1. CRISP-DM Framework

Data were obtained from four community health centers (Puskesmas) in Kediri City: Sukorame, Pesantren I, Ngletih, and Mrican, comprising 4,281 records of child nutrition measurements collected between January and December 2024. During the Data Understanding phase, the dataset was examined for structural integrity, missing values, and anomalies in height-for-age (TB/U) measurements. In the Data Preparation phase, data cleaning involved standardizing sub-district names and removing incomplete records. Relevant variables (NIK, TB/U, measurement date, and sub-district) were selected, aggregated by month using the n_distinct() function in R, and structured as a monthly time series [7]. The Modeling phase applied two forecasting methods: Holt's Double Exponential Smoothing and ARIMA, both trained on data from January to October 2024 [6], [7]. ARIMA model parameters were determined automatically using the auto.arima() function, guided by the Akaike Information Criterion (AIC) [12]. During the Evaluation phase, model performance was assessed on November–December 2024 data using RMSE, MAE, and MAPE, with MAPE serving as the primary criterion for model selection [7]. In the Deployment phase, the selected model forecasted stunting cases for January 2025, distributing results across 18 sub-districts according to historical proportions, and visualized outcomes in an interactive Tableau dashboard with geospatial mapping [9], [15], [16].

RESULTS AND DISCUSSION

This section presents and discusses the results of data analysis at each stage of the research, from stunting trend analysis and model development to forecasting and geospatial visualization. The discussion is conducted systematically to show how stunting patterns in Kediri City were identified and how the forecasting results were translated into actionable, location-based information.

4.1. Data Understanding

The data understanding phase was conducted to examine the structure, quality, and characteristics of the dataset prior to analysis. The dataset comprises 4,281 stunting measurement records from 18 sub-district across four community health centers (Puskesmas) in Kediri City. Only children classified as Short (Pendek) or Very Short (Sangat Pendek) based on the TB/U indicator were included, resulting in 3,334 Short cases and 947 Very Short cases. The attributes selected for this study are presented in Table 1.

Table 1. Dataset Attribute

Attribute	Data Type	Description
NIK	*Character*	Unique child identifier number
Desa/kel	*Character*	Sub-district name of the child's residence
TB/U	*String*	Heigh-for-age nutritional status (Short/Very Short)
Tanggal	*POSIx*	Date of

Pengukuran	measurement

Based on Table 1, four attributes were selected as directly relevant to the forecasting process: NIK (to count unique stunting cases per period), Desa/kel (sub-district for spatial distribution), TB/U (to identify stunting category), and Tanggal Pengukuran (to construct the monthly time series). Kelurahan Pesantren had the highest number of stunting records (383 cases), while Kelurahan Dermo had the lowest (97 cases), indicating significant variation in stunting burden across sub-districts.

Descriptive statistical analysis of monthly stunting case counts is presented in Table 2.

Table 2. Descriptive Statistics

Min	Max	Mean	Median	Std Dev
318	406	356.75	356	25.47

Based on Table 2, the monthly stunting case counts ranged from a minimum of 318 cases (November 2024) to a maximum of 406 cases (April 2024), with a mean of 356.75 and a standard deviation of 25.47. The relatively modest standard deviation indicates moderate variation across monthly periods, confirming the suitability of time series methods for this dataset.

4.2. Data Preparation

The data preparation phase involved four main step: data cleaning, data selection, data transformation, and time series construction. In the data cleaning step, sub-district name formats were standardized (e.g., variations such as "PESANTREN", "Kel. Pesantren", and "pesantren" were unified), whitespace was removed from NIK values, and records with missing sub-district information were excluded. In the data selection step, four attributes were selected from the original 15 available in the dataset: NIK (to count unique stunting cases per period), Desa/kel (to distribute forecasting results at the sub-district level), TB/U (to identify stunting category), Tanggal Pengukuran (to construct the monthly time series). In the data transformation step, monthly periods were formed using the floor_date() function, data were grouped by month using group_by(bulan), and the number of unique stunting cases per month was calculated using the n_distinct(NIK) function in R [7]. The resulting aggregated data contained two variables: bulan (month) as the time marker and jumlah (case count) as the primary forecasting variable. Finally, in the time series construction step, the aggregated monthly data were converted into a time series object using the ts() function with frequency= 12. The dataset was then divided into training data (January-October 2024, 10 periods) and testing data (November-December 2024, 2 periods) [7].

4.3. Stunting Trend Analysis

A trend analysis was conducted to identify patterns in stunting cases in Kediri City in 2024, prior to the forecasting process. Monthly aggregated data were converted into a time series object using the ts() function with a frequency of 12, reflecting the 12 monthly observations for the year [7]. Table 3 presents the monthly distribution of stunting cases.

Table 3. Monthly Stunting Cases

Month	Number of Cases
January	378
February	378
March	383
April	406
May	359
June	356
July	335
August	356
September	339
October	336
November	318
December	337

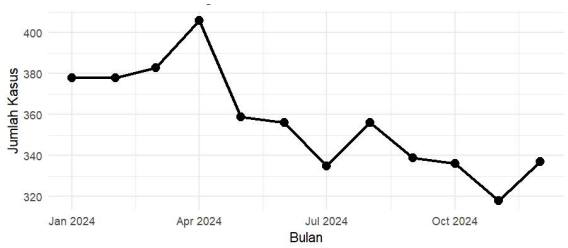


Figure 2. Monthly Stunting Trend, 2024

Based on Table 3 and Figure 2, the number of stunting cases throughout 2024 fluctuated across observation periods. The highest number of cases occurred in April 2024 with 406 cases, while the lowest occurred in November 2024 with 318 cases. The average number of stunting cases throughout 2024 was 356.75 cases per month. The pattern indicates that stunting cases were not constant but experienced increases and decreases over time.

The presence of such variation across periods confirms that the data exhibit a trend pattern, making it suitable for forecasting to estimate stunting cases in subsequent periods. According to time series forecasting theory, trend identification is an important preliminary step before the forecasting process is carried out [7]. Therefore, this study applied Holt's Double Exponential Smoothing and ARIMA to model the historical stunting pattern throughout 2024 [6], [7].

4.4. Forecasting Results on Test Data

Data from January to October 2024 served as training data to construct the forecasting model, whereas data from November and December 2024 were used as test data to evaluate the model's predictive capability [7]. The actual values and predictions produced by each method are shown in Table 4.

Table 4. Test Period

Month	Actual	Holt DES	ARIMA
November 2024	318	329	336
December 2024	337	323	336

Based on Table 4, Holt's Double Exponential Smoothing produced forecasts of 329 cases for November 2024 and 323 cases for December 2024 [6]. In comparison, ARIMA predicted 336 cases for both November and December 2024 [12]. Compared with the actual data, Holt's DES showed differences of 11 cases in November and 14 cases in December. In contrast, ARIMA showed a deviation of 18 cases in November but only 1 case in December.

4.5. Model Evaluation

The models were evaluated using the RMSE, MAE, and MAPE metrics to quantify prediction error for each method [7]. Evaluation values were calculated by comparing actual data with predicted results on the testing dataset. The MAPE interpretation criteria used in this study are shown in Table 5, while the evaluation results are presented in Table 6.

Table 5. MAPE Interpretation Criteria

MAPE Value	Criteria
<10%	Excellent
10%-20%	Good
21%-50%	Adequate
>50%	Poor

Table 6. Forecasting Evaluation Results

Method	RMSE	MAE	MAPE(%)
Holt DES	12.71	12.59	3.83
ARIMA	12.75	9.50	2.98

Based on Table 6, Holt's DES yielded an RMSE of 12.71, an MAE of 12.59, and a MAPE of 3.83%. In comparison, ARIMA produced an RMSE of 12.75, an MAE of 9.50, and a MAPE of 2.98%. The lower MAE and MAPE values demonstrate that ARIMA produced fewer prediction errors than Holt DES. Although Holt DES achieved a marginally lower RMSE, model selection prioritized MAPE as this metric express prediction error as a percentage and is more interpretable [7].

Referring to Table 5, both methods achieved MAPE values below 10%, placing them in the Excellent category. Thus, both methods demonstrated excellent predictive performance. However, ARIMA achieved greater accuracy with the lowest MAPE of 2.98%, and was consequently selected as the optimal model for generating the January 2025 forecast.

4.6. Forecasting for January 2025

Based on the model evaluation results, ARIMA was selected as the best model, yielding the lowest MAPE of 2.98%. The ARIMA model was subsequently used to forecast the number of stunting cases in January 2025 using the entire 2024 historical dataset. The forecasting results are shown in Table 7.

Table 7. Forecasting Results for January 2025

Methods	Forecast (January 2025)
Holt DES	318
ARIMA	337

The forecasting results in Table 7 indicate that ARIMA predicated 337 stunting cases for January 2025, while Holt DES predicted 318 cases. Since ARIMA was selected as the best model, 337 cases was adopted as final forecast. This value was derived from the historical pattern of stunting cases throughout 2024 as modelled using ARIMA [12]. The ARIMA forecast of 337 closely matches the December 2024 actual count (337 cases), suggesting stability in the underlying trend. This prediction provides an initial indication of stunting conditions for the subsequent period, supporting the planning of more targeted stunting intervention programs in Kediri City.

4.7. Sub-district-Level Distribution of Predictions

After obtaining a total forecast of 337 cases for January 2025, the predicted results were distributed to the sub-district level based on historical proportions of unique stunting cases (counted by NIK) recorded in each sub-district during 2024 [6]. Additionally, the predicted measurement date for each sub-district was estimated using the mode of historical measurement dates. The complete distribution results are presented in Table 8.

Table 8. Historical Proportions and Predicted Stunting Distribution by Sub-district, January 2025

Sub-District	Historical Cases (2024)	Proportion(%)	Predicted Cases (Jan 2025)	Predicted Date
Pesantren	83	9.53	32	08 Jan 2025
Bananan	66	7.58	26	05 Jan 2025
Bujel	58	6.66	22	16 Jan 2025
Betet	57	6.54	22	11 Jan 2025
Pojok	55	6.31	21	16 Jan 2025
Blabak	53	6.08	21	12 Jan 2025
Ngampel	53	6.08	21	15 Jan 2025
Mrican	51	5.86	20	04 Jan 2025
Bandar Lor	50	5.74	19	16 Jan 2025
Bangsal	47	5.40	18	09 Jan 2025
Gayam	47	5.40	18	04 Jan 2025
Mojoroto	46	5.28	18	16 Jan 2025
Bawang	43	4.94	17	13 Jan 2025
Sukora	43	4.94	17	16 Jan

 ISSN 3110-8245 (E)

me				2025
Tempurejo	45	5.17	17	13 Jan 2025
Ketami	28	3.21	11	05 Jan 2025
Ngletih	25	2.87	10	03 Jan 2025
Dermo	21	2.41	8	05 Jan 2025

Based on Table 8, Pesantren sub-district had the highest historical number of unique stunting cases (83 cases, 9.53%), resulting in the highest predicted case count of 32 cases for January 2025. In contrast, Dermo sub-district had the lowest historical count (21 cases, 2.41%), corresponding to a prediction of only 8 cases. Differences in predicted case numbers across sub-district are attributable to differences in historical proportions sub-district with higher historical counts produce proportionally larger predictions.

This finding indicates that historical patterns continue to influence the estimation of stunting cases in subsequent periods, and that sub-district-level distribution provides location specific information useful for targeted intervention planning [6], [8], [17].

4.8. Geospatial Dashboard Visualization

The predicted stunting cases for January 2025 were subsequently presented in an interactive dashboard using Tableau Public, integrated with GeoJSON administrative boundary data for Kediri City sub-district [9], [15]. The dashboard is accessible at: https://public.tableau.com/app/profile/nanda.alviodita4134/viz/DashboardPrediksiStunting2025/Dashboard1?publish=yes. The dashboard visualization is shown in Figure 3.

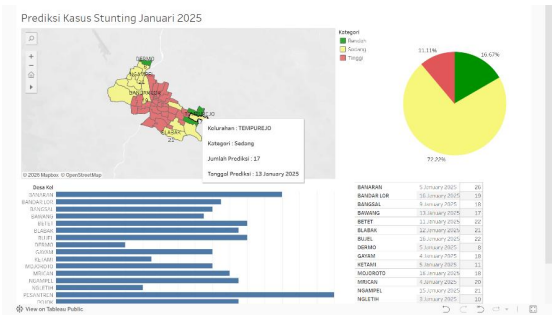


Figure 3. Dashboard Visualization

The dashboard was designed to facilitate users understanding of the prediction results through visual, informative, and interactive presentation [16]. The dashboard consists of four main components: (1) a choropleth map displaying each sub-district's predicted stunting severity category, differentiated by color (green= low, yellow= medium, red= high) [8]; (2) a pie chart showing the proportion of sub-districts within each prediction category; (3) a bar chart comparing predicted case counts across all 18 sub-districts; and (4) a detail table listing sub-district names, predicted measurement dates, and predicted case counts [15]. The dashboard is also equipped with a tooltip feature that displays detailed information sub-district name, prediction category, predicted case count, and predicted measurement date when a sub-district is selected on the map [16]. An interactive cross-filtering feature updates all dashboard components simultaneously when a sub-district is selected.

Based on the dashboard visualization, most sub-districts fall into the medium category, while the high (Pesantren, Banaran, Bujel) and low (Dermo,Ngletih, Ketami) categories account for smaller proportions. Sub-districts in the high category warrant priority attention in stunting intervention planning. The integration of interactive dashboard with geospatial visualization demonstrates that location-based visual presentation can effectively support health monitoring and evidence-based decision-making in stunting management [9], [11], [14].

CONCLUSION

This study successfully applied the CRISP-DM framework using Holt's Double Exponential Smoothing (Holt DES) and ARIMA to analyze stunting trends and forecast case numbers in Kediri City. Evaluation results indicate that ARIMA achieved the lowest MAPE of 2.98%, outperforming Holt DES at 3.83%, both classified as Excellent. The ARIMA model predicted 337 stunting cases for January 2025, which were distributed across 18 sub-districts based on historical proportions, with Pesantren sub-district receiving the highest allocation of 32 cases. The forecasting results were presented through an interactive Tableau dashboard with geospatial visualization, enabling decision-makers to identify sub-districts with the highest predicted stunting burden and supporting more targeted intervention planning. Future research should consider extending the historical data period, comparing ARIMA with advanced methods such as SARIMA or machine learning algorithms, and incorporating causal predictors to improve forecast accuracy.

REFERENCES

A. Marsya, D. Nasution, and R. F. Wijaya, "Pemanfaatan Business Intelligence untuk Visualisasi Data dan Pemetaan Kasus Keluarga Berisiko Stunting Menggunakan Tableau," Journal of Science and Social Research, vol. 8, no. 2, pp. 2058-2064, 2025, doi: 10.54314/jssr.v8i2.3160.

F. A. Mufarroha, D. R. Anamisa, A. Jauhari, and T. Septiyanto, "Utilizing single exponential smoothing for early detection and forecasting of stunting cases in Madura," AIP Conference Proceedings, p. 020013, 2025, doi: 10.1063/5.0241211.

H. Hizriansyah, "Perancangan Model Dashboard untuk Pelaporan dan Visualisasi Data Kesehatan Sebagai Sistem Monitoring di Dinas Kesehatan Gunungkidul," Journal of Information Systems for Public Health, vol. 8, no. 1, pp. 1-9, 2023, doi: 10.22146/jisph.72268.

L. Andriani, S. Nisa, and Y. Aprihatin, "Pengaruh Negatif Stunting terhadap Perkembangan Kognitif dan

Motorik pada Anak Balita," As-Shiha: Journal of Medical Research, vol. 4, no. 1, 2023, doi: 10.69922/asshiha.v4i1.47.

N. A. I. Lestari, R. Kusumawati, and F. Fatimah, "Analisis Sebaran Spasial Penyakit Stunting pada Kelurahan Kedung Halang," BINA: Jurnal Pembangunan Daerah, vol. 2, no. 1, pp. 52-58, 2023, doi: 10.62389/bina.v2i1.52.

N. Aminudin, S. Andriyanto, S. Muharni, and D. Feriyanto, "Visualisasi Data Interaktif untuk Analisis Tren Stunting Pendek dan Sangat Pendek pada Balita di Kabupaten Pringsewu," Metta: Jurnal Ilmu Multidisiplin, vol. 5, pp. 45-51, 2025, doi: 10.37329/metta.v5i2.4046.

R. J. Hyndman and G. Athanasopoulos, "9.5 Non-seasonal ARIMA models," in Forecasting: Principles and Practice, 3rd ed., OTexts, 2021. [Online]. Available: https://otexts.com/fpp3/non-seasonal-arima.html

R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 3rd ed., OTexts, 2021. [Online]. Available: https://otexts.com/fpp3/

R. Khoiriyah and S. R. Nadhiroh, "Tren Prevalensi Stunting Kota Pasuruan Jawa Timur Tahun 2019-2023," Jurnal Promotif Preventif, vol. 7, no. 2, pp. 342-353, 2024, doi: 10.47650/jpp.v7i2.1256.

R. T. Gessese et al., "Forecasting under-five stunting in Ethiopia using classical and machine learning time series models," PLOS ONE, vol. 21, no. 4, p. e0345000, 2026, doi: 10.1371/journal.pone.0345000.

R. Wirth and J. Hipp, "CRISP-DM: Towards a Standard Process Model for Data Mining," in Proc. 4th Int. Conf. Practical Application of Knowledge Discovery and Data Mining, 2000, pp. 29-39.

S. Few, Information Dashboard Design: Displaying Data for At-a-glance Monitoring, Analytics Press, 2013.

S. T. Atmoko, D. Winarti, and E. Yandani, "Pengembangan Sistem Informasi Geografis untuk Visualisasi Spasial dan Pemantauan Interaktif Kasus Stunting di Kabupaten Dharmasraya Berbasis Website," JEKIN - Jurnal Teknik Informatika, vol. 5, no. 2, pp. 960-970, 2025, doi: 10.58794/jekin.v5i2.1606.

Survei Status Gizi Indonesia (SSGI), "Laporan Nasional SSGI 2024," Kementerian Kesehatan Republik Indonesia, Jakarta, 2024.

T. Hidayat, M. Jajuli, and Susilawati, "Clustering Daerah Rawan Stunting di Jawa Barat menggunakan Algoritma K-Means," INFOTECH: Jurnal Informatika dan Teknologi, vol. 4, no. 2, pp. 137-146, 2023, doi: 10.37373/infotech.v4i2.642.

World Health Organization and United Nations Children's Fund, "Global Nutrition Targets 2030: Stunting Brief," World Health Organization, 2025.

Y. Yuwanti, F. M. Mulyaningrum, and M. M. Susanti, "Faktor-Faktor yang Mempengaruhi Stunting pada Balita di Kabupaten Grobogan," Jurnal Keperawatan dan Kesehatan Masyarakat Cendekia Utama, vol. 10, no. 1, pp. 74-84, 2021, doi: 10.31596/jcu.v10i1.704.