

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Penelitian ini menggunakan data kasus *stunting* di Kota Tangerang, khususnya di 13 kecamatan yaitu Karawaci, Pinang, Ciledug, Periuk, Cipondoh, Neglasari, Batu Ceper, Benda, Larangan, Cibodas, Karang Tengah, Tangerang, dan Jatiuwung. Data diperoleh dari Dinas Komunikasi dan Informatika Kota Tangerang dan mencakup pengukuran kasus *stunting* anak dalam rentang waktu 7 Januari 2024 sampai 3 Agustus 2025. Data yang diberikan mencakup berbagai macam variabel, seperti tanggal pengukuran, ID pengukuran, NIK anak, nama anak, jenis kelamin, tanggal lahir, berat badan, tinggi badan, kelurahan, kecamatan, kabupaten/kota, serta keterangan status *stunting*. Namun, dalam penelitian ini hanya beberapa variabel yang digunakan, yaitu tanggal pengukuran, kecamatan, dan keterangan *stunting* (*stunting* atau tidak). Data harian tersebut kemudian diagregasi menjadi data mingguan agar lebih mudah dianalisis dan lebih mudah untuk melihat tren kasus *stunting* per minggu di tiap kecamatan.

Pemilihan objek penelitian ini berdasarkan pada fakta bahwa hingga saat ini belum tersedia sistem prediksi tren *stunting* yang bisa membantu pemerintah Kota Tangerang dan pihak terkait dalam mengambil keputusan secara lebih cepat dan tepat. Selain itu, penelitian terdahulu umumnya menggunakan pendekatan *multivariate* yang melibatkan banyak faktor penyebab, sedangkan penelitian ini fokus pada pendekatan *univariate*, yaitu hanya menggunakan data historis kasus *stunting* itu sendiri untuk prediksi. Pendekatan ini juga sejalan dengan program nasional yang baru diluncurkan pada tahun 2025 [51], di mana penanganan *stunting* saat ini masih bersifat reaktif, yaitu dilakukan setelah kasus *stunting* terjadi. Oleh karena itu, diperlukan alat yang bisa membantu dalam memprediksi yang bersifat proaktif agar intervensi dapat dilakukan lebih awal untuk mencegah peningkatan kasus *stunting*.

3.2 Metode Penelitian

Penelitian ini merupakan penelitian kuantitatif, karena data yang dianalisis berbentuk angka (numerik) [52]. Penelitian ini menggunakan enam variasi model yaitu ARIMA Manual, Auto ARIMA, LSTM Manual, LSTM Optuna, GRU Manual, dan GRU Optuna untuk pemodelan data *stunting* mingguan yang telah melalui proses data *smoothing*. Pendekatan ini diterapkan untuk memprediksi tren *stunting* dengan tujuan mendapatkan model prediksi yang paling akurat dan andal.

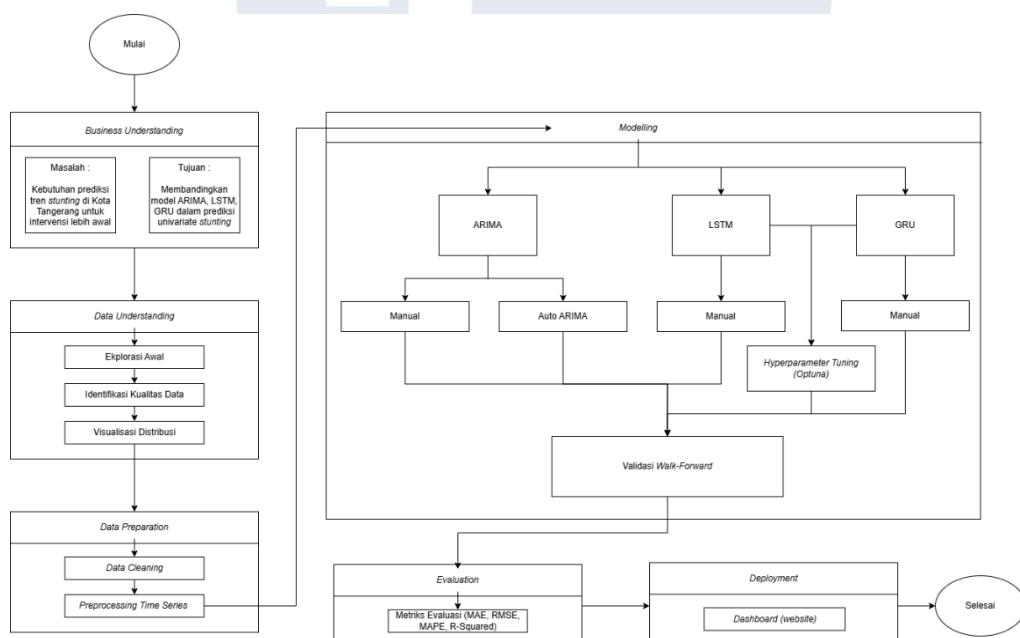
3.2.1 Alur Penelitian

Alur penelitian merupakan langkah-langkah yang digunakan untuk menjalankan proses penelitian secara terstruktur dan terarah. Dalam penelitian ini, alur tersebut melibatkan penggunaan metode *data mining* untuk membantu dalam menganalisis data *stunting*. Metode *data mining* sendiri adalah cara atau langkah-langkah yang digunakan untuk menganalisis data dalam jumlah besar untuk menemukan pola, tren, atau informasi yang berguna [53]. Beberapa pendekatan *data mining* yang umum digunakan antara lain CRISP-DM, *Knowledge Discovery In Databases* (KDD), dan *Sample, Explore, Modify, Model, and Assess* (SEMMA). Pendekatan-pendekatan ini membantu dalam memastikan proses analisis data dilakukan secara sistematis dan terstruktur, sehingga memudahkan dalam pengambilan keputusan yang tepat. Adapun perbandingan metode *data mining* dapat dilihat pada Tabel 3.1.

Tabel 3.1 Perbandingan Metode *Data Mining* [54]

Metode	CRISP-DM	KDD	SEMMA
Tahapan	1. <i>Business Understanding</i> 2. <i>Data Understanding</i> 3. <i>Data Preparation</i> 4. <i>Modelling</i> 5. <i>Evaluation</i> 6. <i>Deployment</i>	1. <i>Preprocessing</i> 2. <i>Preprocessing</i> 3. <i>Transformation</i> 4. <i>Data Mining</i> 5. <i>Evaluation</i>	1. <i>Sample</i> 2. <i>Explore</i> 3. <i>Modify</i> 4. <i>Model</i> 5. <i>Assess</i>
Kelebihan	Fleksibel, dapat diterapkan di berbagai industri, dan berfokus pada pemahaman bisnis di awal [55].	Berfokus pada menemukan pola serta informasi penting dari data besar [56].	Berfokus pada transformasi dan persiapan data yang baik [57].
Kekurangan	Prosesnya memakan waktu serta sumber daya yang cukup besar [55].	Prosesnya cukup kompleks, kurang menekankan pada tujuan bisnis [56].	Kurang menekankan pada aspek bisnis, sulit diterapkan untuk proyek yang sangat besar [57].

Dari ketiga metodologi, yang paling tepat untuk penelitian ini adalah CRISP-DM, karena memiliki beberapa keunggulan yang sesuai dengan kebutuhan. CRISP-DM sangat fleksibel dan bisa diterapkan di berbagai jenis industri dan proyek, sehingga prosesnya mudah disesuaikan dengan karakteristik data dan tujuan penelitian. Salah satu keunggulan CRISP-DM adalah penekanannya pada tahap *business understanding* sejak awal, sehingga membantu dalam memastikan bahwa penelitian selalu fokus pada kebutuhan bisnis dan sehingga bisa membantu dalam pengambilan keputusan yang tepat berdasarkan hasil analisis. Selain itu, CRISP-DM memiliki tahapan *deployment*, sehingga model yang dihasilkan tidak hanya diuji, tetapi juga bisa diterapkan dalam konteks nyata. Hal ini sangat penting agar hasil penelitian memberikan nilai tambah dan manfaat bagi pihak terkait.



Gambar 3.1 Alur Penelitian CRISP - DM

Gambar 3.1 merupakan alur penelitian menggunakan *framework* CRISP-DM. *Framework* ini merupakan salah satu metodologi yang umum digunakan dalam proyek *data mining* karena memberikan alur kerja yang sistematis dan terstruktur. Dengan mengikuti tahapan-tahapan dalam CRISP-DM, penelitian ini dapat dilakukan secara lebih terarah. Adapun tahapan-tahapannya sebagai berikut:

1) *Business Understanding*

Tahapan yang pertama adalah *business understanding*. Tujuan dari tahapan ini adalah untuk mengidentifikasi masalah dan tujuan yang ada. Dalam penelitian ini, masalah yang dihadapi adalah belum adanya *dashboard* prediksi yang dapat membantu pemerintah Kota Tangerang dalam melihat tren *stunting* di setiap kecamatan. Data yang tersedia hanya berupa catatan historis (07 Januari 2024 - 03 Agustus 2025) tanpa adanya proyeksi ke depan, sehingga sulit digunakan untuk perencanaan intervensi dini. Tujuan dari penelitian ini adalah mengembangkan dan membandingkan beberapa model prediksi *univariate time series* untuk memperkirakan tren *stunting* mingguan pada setiap kecamatan di Kota Tangerang. Enam variasi model yang dibandingkan adalah ARIMA Manual, Auto ARIMA, LSTM Manual, LSTM dengan optimasi Optuna, GRU Manual, dan GRU dengan optimasi Optuna. Hasil model kemudian ditampilkan dalam bentuk *dashboard* agar mudah dipahami dan dapat digunakan sebagai dasar pengambilan keputusan.

2) *Data Understanding*

Tahapan selanjutnya adalah *data understanding*. Tahapan ini bertujuan untuk menganalisis dan memahami dataset yang digunakan. Data yang digunakan diperoleh dari Dinas Komunikasi dan Informatika Kota Tangerang dengan periode waktu 07 Januari 2024 hingga 03 Agustus 2025 yang berisi kasus *stunting* di Kota Tangerang dengan jumlah *rows* 109,208 dan 12 kolom. Langkah awal yang dilakukan dengan eksplorasi data untuk melihat jumlah baris, kolom, serta informasi *dataset*. Selanjutnya dilakukan identifikasi kualitas data seperti melakukan pengecekan data duplikat, *missing value*, dan kecamatan di luar wilayah Kota Tangerang. Tujuan dilakukan hal ini adalah agar data yang digunakan benar-benar relevan. Selain itu, dilakukan visualisasi distribusi data dan statistik deskriptif untuk memahami pola dan tren awal kasus *stunting*. Hasil dari tahap ini menjadi dasar dalam proses pembersihan dan persiapan data pada tahap berikutnya.

3) *Data Preparation*

Tahapan berikutnya adalah *data preparation* yang bertujuan untuk menyiapkan data sebelum dilakukan pemodelan. Langkah pertama yang dilakukan adalah melakukan data *cleansing*, yaitu menghapus data duplikat, menghilangkan nilai kosong, dan memfilter hanya wilayah yang termasuk ke dalam 13 kecamatan di Kota Tangerang. Selain itu, data juga difokuskan pada kategori anak dengan status *stunting* agar analisis lebih spesifik. Setelah data dibersihkan, dilakukan *preprocessing time series* untuk mengubah data menjadi format yang sesuai dengan analisis deret waktu. Proses ini meliputi *resampling* data ke dalam bentuk mingguan. Setelah melakukan analisis awal, ditemukan bahwa data yang digunakan sangat bersifat acak, sehingga penelitian ini melakukan penerapan data *smoothing* untuk mengurangi fluktuasi acak pada data. Setelah data *smooth* didapat, dilakukan normalisasi menggunakan *MinMaxScaler* pada model LSTM dan GRU agar proses pelatihan model menjadi lebih stabil. Sementara untuk model ARIMA, dilakukan *log transform* untuk menstabilkan variasi dan mengurangi perubahan ekstrem pada data. Hasil dari tahap ini menghasilkan data mingguan yang bersih dan halus yang kemudian digunakan dalam proses pemodelan pada tahap berikutnya.

4) *Modelling*

Tahapan selanjutnya adalah *modelling* yang menjadi proses utama dalam penelitian ini. Terdapat enam variasi model *univariate time series* yang digunakan, yaitu ARIMA Manual, Auto ARIMA, LSTM Manual, LSTM dengan optimasi Optuna, GRU Manual, dan GRU dengan optimasi Optuna. Pemilihan model-model tersebut didasarkan pada masih terbatasnya penelitian *stunting* yang membandingkan model statistik klasik dan *deep learning*. Selain itu, berdasarkan penelusuran penelitian terdahulu di bidang *time series*, ARIMA, LSTM, dan GRU merupakan algoritma yang paling banyak digunakan. Dalam *time series*, data terbaru bergantung pada data sebelumnya, sehingga pembagian acak seperti 80% data latih dan 20% data

uji dapat menyebabkan kebocoran informasi dari masa depan ke masa lalu. Penelitian ini menggunakan validasi *walk-forward* dengan cara model dilatih menggunakan data di masa lalu, kemudian diuji pada data di masa depan secara bertahap dan berulang. Cara ini lebih menggambarkan kondisi nyata di mana model memprediksi kejadian yang akan datang berdasarkan data historis yang telah terjadi. Pada ARIMA Manual, penentuan parameter p , d , q dilakukan dengan melihat dari grafik ACF dan PACF. Sedangkan untuk Auto ARIMA menggunakan *library* pmdarima pada Python yang memilih parameter terbaik berdasarkan nilai AIC terendah. Pada model LSTM dan GRU Manual, digunakan arsitektur yang sederhana dengan *look-back* sepanjang 4 minggu. Setiap model memiliki satu layer tersembunyi (*hidden layer*) berisi 50 *neuron*, kemudian dilanjutkan dengan satu *neuron* output untuk menghasilkan satu prediksi. Model dikompilasi menggunakan *optimizer* Adam dan fungsi *loss mean squared error* (MSE), kemudian dilatih selama 50 *epoch* dengan ukuran batch 1. Sedangkan untuk model LSTM Optuna dan GRU Optuna, dilakukan optimasi *hyperparameter* menggunakan Optuna. Optuna mencari kombinasi terbaik dari beberapa komponen penting, seperti jumlah neuron (20–100), panjang *look-back* (3–8 minggu), dan jumlah *epoch* (30–100). Hasil dari keenam variasi model kemudian dibandingkan untuk mengetahui model mana yang memberikan performa terbaik dalam melakukan prediksi tren mingguan kasus *stunting*.

5) *Evaluation*

Tahapan selanjutnya adalah evaluasi. Tahap ini dilakukan untuk menilai seberapa baik performa model dalam memprediksi tren mingguan kasus *stunting* di Kota Tangerang. Evaluasi dilakukan menggunakan empat metrik yaitu MAE, RMSE, MAPE, dan *R-Squared*. Keempat metrik ini dipilih karena umum digunakan dalam penelitian prediksi deret waktu dan dapat memberikan gambaran yang lebih lengkap mengenai kemampuan model dalam memprediksi tren *stunting*.

6) *Deployment*

Tahapan yang terakhir adalah *deployment*, yaitu penerapan model terbaik agar dapat digunakan secara nyata. Setelah keenam variasi model selesai dilatih dan dievaluasi, diperoleh model dengan performa terbaik berdasarkan hasil metrik evaluasi. Model terbaik inilah yang diimplementasikan ke dalam *dashboard* berbasis *website* agar hasil prediksi dapat dipahami secara visual. *Dashboard* akan menampilkan tren historis dan prediksi mingguan *stunting* di 13 kecamatan yang ada di Kota Tangerang. *Dashboard* dikembangkan menggunakan *framework* *CodeIgniter3* karena *framework* ini juga digunakan dalam lingkungan instansi tempat penelitian dilakukan.

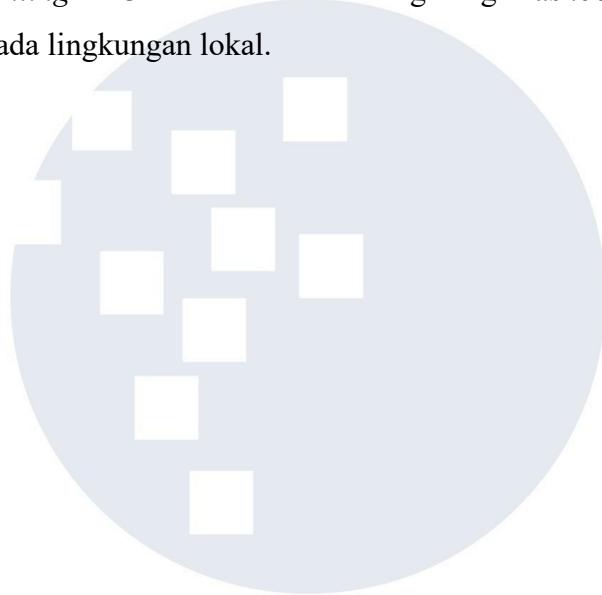
3.3 Teknik Pengumpulan Data

Penelitian ini menggunakan data sekunder yang diberikan langsung dari Dinas Komunikasi dan Informatika Kota Tangerang. Data tersebut tidak diambil sendiri, melainkan diberikan secara resmi oleh pihak Dinas Komunikasi dan Informatika Kota Tangerang. Data ini berisi informasi tentang kasus *stunting* yang ada di Kota Tangerang pada periode 07 Januari 2024 hingga 03 Agustus 2025. Dengan adanya data ini, proses analisis dapat dilakukan secara lebih akurat karena sumbernya jelas dan terpercaya. Pemberian data ini juga menjadi bentuk dukungan dari pemerintah daerah terhadap kegiatan penelitian akademik, khususnya dalam pengembangan solusi berbasis data untuk membantu dalam memantau dan mencegah *stunting* di Kota Tangerang.

3.4 Teknik Analisis Data

Teknik analisis data pada penelitian ini menggunakan *framework* CRISP-DM sebagai kerangka kerja utama. Penelitian ini membangun dan membandingkan enam variasi model yaitu ARIMA Manual, Auto ARIMA, LSTM Manual, LSTM dengan optimasi Optuna, GRU Manual, dan GRU dengan optimasi Optuna. Proses pelatihan dan validasi model dilakukan menggunakan bahasa pemrograman *Python* pada *Jupyter Notebook*. Setelah seluruh model selesai dilatih, dilakukan proses evaluasi performa untuk menilai seberapa baik masing-masing model dalam memprediksi tren mingguan kasus *stunting* di Kota Tangerang. Evaluasi performa model menggunakan empat metrik yaitu MAE, RMSE, MAPE, dan *R-Squared*.

Pemilihan keempat metrik ini bertujuan agar evaluasi dilakukan secara menyeluruh. Dalam konteks prediksi kasus *stunting* yang memiliki pola data acak dan fluktuatif, keempat metrik ini memberikan gambaran yang lebih komprehensif mengenai keandalan model. Model dengan performa terbaik dipilih sebagai model terbaik. Model terbaik ini kemudian diimplementasikan ke *dashboard* berbasis *website* menggunakan *framework CodeIgniter3* yang menampilkan hasil prediksi tren mingguan kasus *stunting* di 13 kecamatan Kota Tangerang. *Dashboard* yang dibuat masih dijalankan pada lingkungan lokal.



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA