



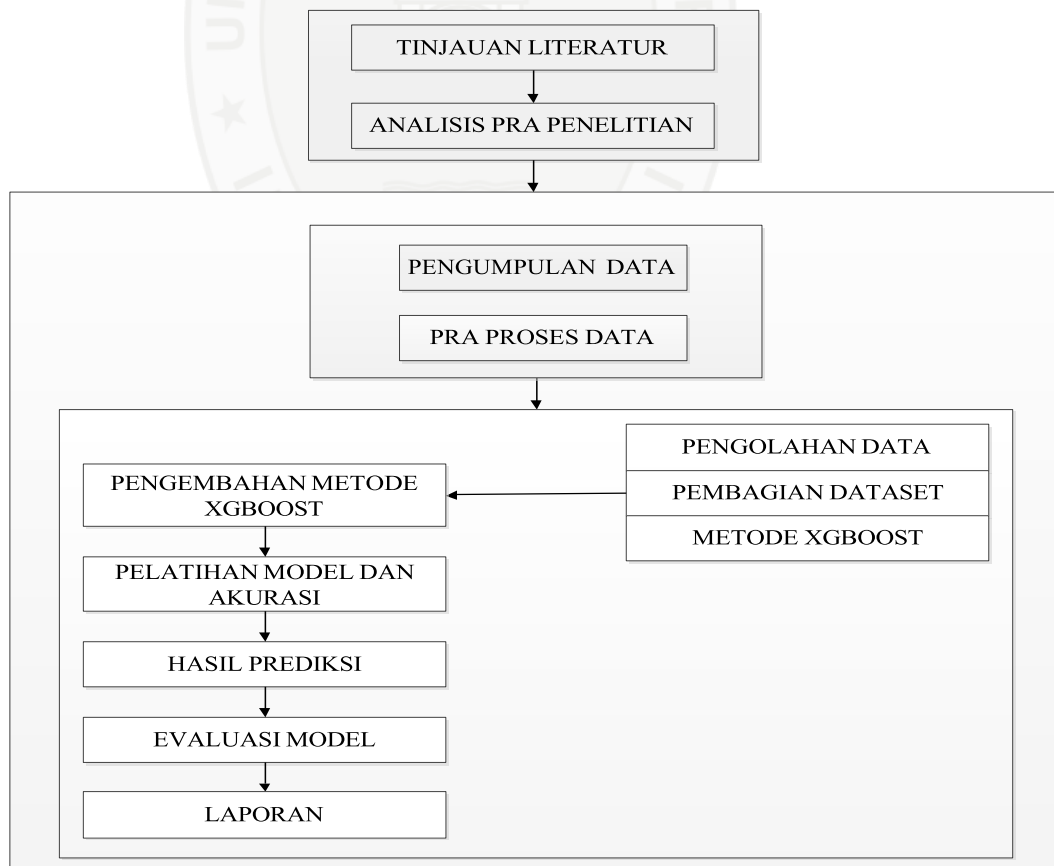
BAB III

METODE PENELITIAN

Pada penelitian ini, objek yang dikembangkan adalah data harga sawit di Kabupaten Indragiri Hilir, Riau dari tahun 2021 hingga 2025. Pengolahan data yang dilakukan menggunakan KDD (*Knowledge Discovery in Database*). Gambar 3.1 merupakan tahapan penelitian.

4.1 Kerangka Penelitian

Dalam penelitian ini, penulis menerapkan sejumlah tahapan untuk mencapai tujuan yang diinginkan. Tahapan-tahapan tersebut diperlihatkan pada Gambar 3.1:



Gambar 3.1 Tahapan Penelitian

Gambar 3.1 Tahapan Penelitian, yang menyajikan gambaran terstruktur mengenai langkah-langkah utama yang diambil penulis dalam proses penelitian ini. Gambar tersebut membantu memperjelas alur dan strategi yang digunakan dalam pencapaian tujuan penelitian.

1 Studi Literatur

Studi literatur dilakukan untuk mendapatkan dasar teori yang mendukung penelitian ini dalam kerangka teori prediksi time-series, XGBoost tidak hanya mengandalkan nilai harga sebelumnya, tetapi juga dapat mengintegrasikan fitur visualisasi waktu seperti minggu ke, bulan, tren naik-turun, serta perbedaan nilai antara periode tertentu. Fitur-fitur ini memberikan informasi tambahan yang membantu model menangkap pola musiman dan pergerakan jangka pendek maupun panjang. Model juga dapat diperluas dengan memasukkan variabel eksternal seperti harga minyak mentah atau harga ekspor CPO apabila data tersedia.

2 Analisis Kebutuhan Data

Data yang digunakan pada penelitian ini merupakan data mingguan harga TBS dari tahun 2021 hingga 2025 di Kabupaten Indragiri Hilir, Riau. Data ini digunakan untuk menganalisis tren harga dan membangun model prediksi berdasarkan berbagai faktor yang mempengaruhi harga.

Tabel 3.1 Dataset Harga Sawit

Lokasi 1		Lokasi 2		Lokasi 3		Lokasi 4		Lokasi 5	
Tanggal	Harga	Tanggal	Harga	Tanggal	Harga	Tanggal	Harga	Tanggal	Harga
02/12/2025	2500	11/08/2025	2000	12/04/2025	2500	23/02/2025	2100	27/02/2025	3055
03/12/2025	1200	01/09/2025	1900	14/04/2025	2450	24/02/2025	2150	04/03/2025	3070
04/12/2025	3000	03/09/2025	2000	20/04/2025	2450	14/03/2025	2100	10/11/2025	2850
05/12/2025	2400	04/09/2025	2000	22/04/2025	2500	16/03/2025	2150	11/11/2025	2850
06/12/2025	3030	05/09/2025	2100	26/04/2025	2450	18/03/2025	2100	12/11/2025	2850
07/12/2025	2850	07/09/2025	2000	28/04/2025	2450	21/03/2025	2150	07/12/2025	2875
08/12/2025	3030	11/09/2025	1900	30/04/2025	2450	05/04/2025	2000	09/12/2025	2650
09/12/2025	3030	12/09/2025	2000	15/05/2025	3250	06/04/2025	2000	16/05/2026	2650





Tabel 3.1 Lanjutan

10/12/2025	2850	19/09/2025	2000	07/06/2025	3945	11/05/2025	2650	17/05/2026	2650
11/12/2025	2850	05/10/2025	1900	20/06/2025	2500	20/06/2025	2500	18/05/2026	2650
12/12/2025	3000	06/10/2025	2000	26/06/2025	2500	21/07/2025	2545	17/06/2026	2700
13/12/2025	2400	08/10/2025	1900	29/06/2025	3000	23/07/2025	2870	12/07/2026	2700
14/12/2025	2400	10/10/2025	1900	13/07/2025	2750	01/08/2025	2000	15/07/2026	2550
15/12/2025	3000	11/10/2025	1900	18/07/2025	2500	03/09/2025	2000	17/07/2026	2700
16/12/2025	2400	13/10/2025	2450	08/11/2025	2400	06/10/2025	2450	18/07/2026	2400
17/12/2025	2500	14/10/2025	1900	11/11/2025	2350	27/10/2025	2400	19/07/2026	2550
18/12/2025	2500	19/10/2025	2100	13/11/2025	2350	19/10/2025	2700	22/06/2026	2723
19/12/2025	1200	18/10/2025	2000	16/11/2025	2350	02/11/2025	2600	23/07/2026	2870
20/12/2025	3000	23/10/2025	2100	21/11/2025	2350	16/11/2025	2575		
21/12/2025	2500	25/10/2025	1900	24/11/2025	2350	26/11/2025	2000		
22/12/2025	3185	31/10/2025	2000	25/11/2025	2350	30/11/2025	2625		
23/12/2025	3185	06/11/2025	2100	26/11/2025	2350	05/06/2026	2575		
24/12/2025	2500	17/11/2025	2100	29/11/2025	2350	18/06/2026	2700		
25/12/2025	1200	20/11/2025	2000	01/12/2025	2350	15/07/2026	2300		
26/12/2025	3000	23/11/2025	2000	04/12/2025	2350	19/07/2026	2550		
27/12/2025	2500	26/11/2025	2000	08/12/2025	2350	22/06/2026	2723		
28/12/2025	3000	01/12/2025	2100	10/12/2025	2350	23/07/2026	2870		

Data yang digunakan dalam penelitian ini merupakan data primer yang diperoleh secara langsung dari petani kelapa sawit yang berada di beberapa lokasi perkebunan di Kabupaten Indragiri Hilir, Provinsi Riau. Pengambilan data dilakukan melalui proses wawancara dan pencatatan harga jual Tandan Buah Segar (TBS) yang diterima oleh petani pada saat melakukan penjualan hasil panen kepada pengepul maupun pabrik kelapa sawit.

Pada Tabel 3.1 ditampilkan data harga TBS yang berasal dari lima lokasi pengamatan. Setiap lokasi memiliki data berupa tanggal transaksi dan harga jual TBS yang diperoleh petani. Pengambilan data dari beberapa lokasi bertujuan untuk memperoleh variasi harga yang mencerminkan kondisi nyata di lapangan, mengingat harga TBS dapat berbeda pada setiap wilayah akibat perbedaan kualitas buah, umur tanaman, biaya transportasi,

jarak ke pabrik, serta kebijakan pembelian masing-masing pabrik atau pengepul.

Penggunaan data yang berasal dari beberapa lokasi petani diharapkan mampu menghasilkan model prediksi yang lebih representatif terhadap kondisi fluktuasi harga TBS di Kabupaten Indragiri Hilir, sehingga hasil prediksi yang diperoleh dapat memberikan informasi yang lebih akurat dan bermanfaat bagi petani maupun pihak yang berkepentingan dalam pengambilan keputusan terkait penjualan hasil panen.

Sebelum digunakan, data diperiksa kualitasnya melalui pengecekan urutan waktu, konsistensi tipe format angka, serta penanganan nilai kosong. Nilai atau baris yang tidak lengkap dikeluarkan terutama setelah proses penambahan fitur lag dan moving average untuk mencegah kesalahan data dan kebocoran informasi.

Fitur kalender, seperti minggu ke dan bulan, digunakan untuk menangkap ritme musiman atau pola tertentu berdasarkan waktu. Fitur lag hingga empat periode dimasukkan untuk memberikan informasi memori jangka pendek harga TBS. Fitur perubahan nilai atau delta disertakan untuk menangkap pergerakan dan momentum kenaikan atau penurunan harga dalam periode tertentu. Data tersebut dikelompokkan dan di analisis berdasarkan perminggu harga, sehingga terdapat 260 data masukan. Tabel di bawah menunjukkan data harga kelapa sawit, di Kabupaten Indragiri Hilir, Riau. Data tersebut telah diubah ke dalam bentuk time series.





1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Plagiarisme juga dilarang dan dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Tabel 3.2 Data Time Series Harga Kelapa Sawit

tanggal	harga_cpo	harga_kernel	harga_tbs	minggu_ke	bulan
2021-01-10 00.00.00	9521,12	6665,26	2174,31	1	1
2021-01-17 00.00.00	10009,09	7031,63	2297,95	2	1
2021-01-24 00.00.00	9159,37	7098,57	2117,6	3	1
2021-01-31 00.00.00	9159,37	7098,57	2117,6	4	1
2021-02-14 00.00.00	9368,98	6901,79	2151,25	6	2
2021-02-21 00.00.00	9411,14	7031,63	2172,72	7	2
2021-02-28 00.00.00	9740,11	7022,21	2241,15	8	2
2021-03-07 00.00.00	9976,89	7143,36	2297,14	9	3
2021-03-14 00.00.00	10012,44	7241,11	2310,76	10	3
2021-03-21 00.00.00	10337,55	7349,06	2384,26	11	3
2021-03-28 00.00.00	10335,21	7332,98	2388,03	12	3
2021-04-04 00.00.00	10242,19	7299,78	2361,15	13	4
2021-04-11 00.00.00	9766,74	7051,91	2240,92	14	4
2021-04-18 00.00.00	10263,09	7164,26	2358	15	4
2021-04-25 00.00.00	10334,96	7146,09	2370,57	16	4
2021-05-02 00.00.00	10921,63	7276,4	2506,52	17	5
2021-05-09 00.00.00	10961,18	7331,39	2517,13	18	5
2021-05-16 00.00.00	11140,51	7428,96	2563,75	19	5

3 Analisa Model XGBoost

a. Data Processing

Tahap pertama dalam proses pelatihan model XGBoost adalah **Data**

Processing. Tahap ini bertujuan untuk mempersiapkan dataset agar

memiliki kualitas yang baik sebelum digunakan pada proses pelatihan model. Dataset yang digunakan merupakan data historis harga Tandan Buah Segar (TBS) kelapa sawit Kabupaten Indragiri Hilir periode 2021–2025. Selanjutnya dilakukan proses *preprocessing* yang meliputi pemeriksaan data yang hilang (*missing value*), penghapusan data duplikat, normalisasi format tanggal, konversi tipe data ke dalam format numerik, serta pengurutan data berdasarkan urutan waktu (*chronological order*). Selain itu, dilakukan rekayasa fitur (*feature engineering*) dengan membentuk variabel seperti lag harga (lag-1, lag-2, dan lag-3), moving average, perubahan harga (delta), minggu ke-, bulan, dan tahun. Penambahan fitur tersebut bertujuan agar model mampu mengenali pola historis perubahan harga sawit secara lebih baik sehingga dapat meningkatkan akurasi prediksi.

b. Data Split into Training, Validation & Testing

Membuat prediksi dengan model XGBoost adalah langkah selanjutnya. Data dibagi menjadi set pelatihan dan pengujian sebelum proses prediksi dimulai. Data latih didistribusikan sebagai berikut: 70% dan 80%, sedangkan data uji didistribusikan 30% dan 20%. Tabel 3.2 menunjukkan pemisahan antara data yang digunakan untuk pelatihan dan pengujian.

Tabel 3.3 Pembagian Data

Pembagian Data	Training	Testing
70% : 30%	182	78
80% : 20%	208	52



Pelatihan model prediksi harga sawit menggunakan algoritma Extreme Gradient Boosting (XGBoost). Pada tahap ini, data yang telah melalui proses preprocessing dan pembagian data digunakan untuk melatih model agar mampu mempelajari pola hubungan antara variabel input dan target harga sawit[20].

c. XGBoost Model Training, Validation & Testing

Tahap selanjutnya adalah proses pelatihan model XGBoost. Pada tahap ini, data training dimasukkan ke dalam algoritma XGBoost untuk membangun model prediksi harga sawit. Proses pelatihan diawali dengan pembentukan sebuah pohon keputusan (*decision tree*) sebagai model awal. Hasil prediksi yang dihasilkan oleh pohon pertama kemudian dibandingkan dengan nilai aktual sehingga diperoleh nilai kesalahan (*error*). Selanjutnya algoritma XGBoost membangun pohon keputusan baru untuk memperbaiki kesalahan yang dihasilkan oleh pohon sebelumnya melalui mekanisme *gradient boosting*. Proses tersebut dilakukan secara berulang hingga jumlah iterasi atau jumlah pohon keputusan yang telah ditentukan pada parameter model tercapai. Selama proses pelatihan berlangsung, data validation digunakan untuk mengevaluasi performa model dan melakukan penyetelan parameter (*hyperparameter tuning*) agar diperoleh model dengan performa terbaik. Selain itu, diterapkan teknik *early stopping* untuk menghentikan proses pelatihan apabila performa model tidak lagi mengalami peningkatan sehingga dapat mengurangi risiko *overfitting*. Setelah proses pelatihan selesai, model kemudian diuji menggunakan data testing untuk mengetahui



kemampuan model dalam memprediksi harga TBS kelapa sawit pada data yang belum pernah dipelajari sebelumnya.

d. Compile XGBoost Model

Setelah proses pelatihan selesai dilakukan, tahap berikutnya adalah Compile XGBoost Model. Pada tahap ini dipilih model terbaik berdasarkan hasil evaluasi selama proses validasi. Model yang memiliki performa terbaik kemudian disimpan sebagai model akhir (*trained model*) yang siap digunakan pada proses prediksi. Model tersebut mencakup struktur pohon keputusan, bobot (*weight*) setiap pohon hasil proses *boosting*, serta parameter terbaik yang diperoleh selama proses pelatihan. Dengan menyimpan model yang telah dilatih, sistem tidak perlu melakukan proses pelatihan ulang setiap kali akan melakukan prediksi sehingga proses prediksi dapat dilakukan dengan lebih cepat dan efisien.

e. Prediction

Tahap terakhir adalah Prediction, yaitu proses memanfaatkan model XGBoost yang telah selesai dilatih untuk memprediksi harga Tandan Buah Segar (TBS) kelapa sawit pada periode berikutnya. Pada tahap ini, data baru atau data testing dimasukkan ke dalam model sebagai masukan. Selanjutnya model akan memproses setiap variabel input melalui seluruh pohon keputusan yang telah terbentuk selama proses pelatihan. Hasil prediksi dari setiap pohon kemudian digabungkan sehingga menghasilkan satu nilai prediksi akhir harga sawit. Nilai prediksi tersebut dibandingkan dengan



harga aktual untuk mengetahui tingkat ketepatan model menggunakan metrik evaluasi seperti *Mean Absolute Error (MAE)*, *Root Mean Square Error (RMSE)*, dan *Coefficient of Determination (R²)*. Selain itu, hasil prediksi juga dapat divisualisasikan dalam bentuk grafik perbandingan antara harga aktual dan harga prediksi sehingga memudahkan analisis terhadap performa model.

4 Evaluasi Model

Kinerja model dievaluasi menggunakan beberapa metrik. Skor R² (R-squared) dihitung untuk menilai seberapa baik model menangkap variansi dalam data. Skor R² dihitung untuk data pelatihan dan pengujian, dan rata-ratanya juga dilaporkan. Selain itu, validasi silang dilakukan menggunakan *cross_val_score* untuk mendapatkan penilaian yang lebih akurat dan andal tentang kinerja model. Evaluasi lebih lanjut mencakup analisis residu untuk memeriksa distribusi kesalahan prediksi dan mengidentifikasi potensi bias model.

Nilai RMSE yang semakin kecil menunjukkan bahwa hasil prediksi model semakin mendekati nilai sebenarnya. Perumusan RMSE dapat dituliskan sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Sementara itu, koefisien determinasi (R²) digunakan untuk mengukur seberapa besar proporsi variasi pada variabel target yang dapat dijelaskan oleh model. Nilai R² dihitung berdasarkan perbandingan antara





jumlah kuadrat selisih antara nilai aktual dan nilai prediksi dengan jumlah kuadrat selisih antara nilai aktual dan rata-rata nilai aktual. Nilai R^2 berkisar antara 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan bahwa model mampu menjelaskan sebagian besar variasi data[17].

Rumus R^2 dinyatakan sebagai berikut :

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

Dengan demikian, nilai RMSE yang rendah dan nilai R^2 yang tinggi menunjukkan bahwa model prediksi memiliki kinerja yang lebih baik dalam menjelaskan variasi hubungan antara variabel input dan output secara akurat.

5 Uji Coba

Uji coba model XGBoost dalam penelitian ini dilakukan untuk mengetahui kemampuan model dalam memprediksi harga Tandan Buah Segar (TBS) kelapa sawit berdasarkan data historis yang telah dipelajari pada tahap pelatihan. Proses pengujian diawali dengan memuat model XGBoost yang telah selesai dilatih, kemudian data *testing* yang belum pernah digunakan selama proses pelatihan dimasukkan ke dalam sistem sebagai data uji. Model selanjutnya memproses setiap data masukan dengan memanfaatkan seluruh pohon keputusan (*decision tree*) yang telah terbentuk selama proses *gradient boosting* untuk menghasilkan nilai prediksi harga sawit. Hasil prediksi kemudian dibandingkan dengan data harga aktual guna mengetahui tingkat ketepatan model. Untuk mengukur kinerja model digunakan metrik evaluasi berupa Mean Absolute Error

(MAE), Root Mean Square Error (RMSE), dan Coefficient of Determination (R^2). Nilai MAE dan RMSE digunakan untuk mengetahui besar kesalahan prediksi, sedangkan nilai R^2 digunakan untuk mengukur seberapa baik model mampu menjelaskan hubungan antara data historis dengan hasil prediksi yang dihasilkan. Selain itu, hasil prediksi juga divisualisasikan dalam bentuk grafik perbandingan antara harga aktual dan harga prediksi sehingga memudahkan analisis terhadap performa model. Apabila nilai MAE dan RMSE yang diperoleh semakin kecil serta nilai R^2 semakin mendekati 1, maka model XGBoost dapat dikatakan memiliki tingkat akurasi yang baik dalam memprediksi harga TBS kelapa sawit. Oleh karena itu, hasil uji coba ini menjadi dasar untuk menilai bahwa algoritma XGBoost mampu mempelajari pola fluktuasi harga sawit dari data historis dan menghasilkan prediksi yang dapat digunakan sebagai pendukung pengambilan keputusan bagi petani, koperasi, maupun perusahaan perkebunan dalam memperkirakan harga sawit pada periode berikutnya.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.