



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.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

BAB II

TINJAUAN LITERATUR

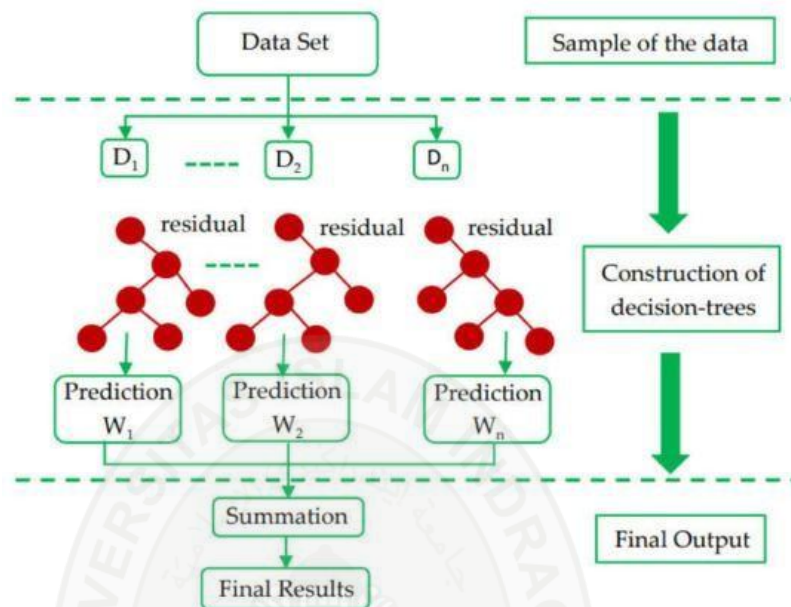
Berdasarkan kondisi tersebut, penelitian ini berfokus kepada bagaimana menerapkan sistem prediksi harga sawit menggunakan algoritma XGboost. Membangun system berbasis web/aplikasi yang dapat mengintegrasikan model XGboos untuk melakukan prediksi harga sawit secara otomatis dari data harga TBS sawit 2021–2025 yang diunggah pengguna.

Sebagai bahan acuan penulis mengambil beberapa tinjauan literatur yang berhubungan pada sistem prediksi harga sawit berbasis web/aplikasi dengan menggunakan algoritma Xgboost sebagai alat bantu yang dapat memprediksi harga jual kelapa sawit tersebut.

2.1 Sistem Prediksi

Prediksi adalah suatu proses untuk memperkirakan atau memperhitungkan kejadian yang akan datang berdasarkan data, informasi, atau pola yang ada saat ini. Prediksi dapat diterapkan dalam berbagai bidang, seperti cuaca, ekonomi, olahraga, dan ilmu pengetahuan. Proses prediksi bertujuan untuk memberikan gambaran mengenai kemungkinan masa depan, yang dapat membantu individu atau organisasi dalam mengambil keputusan yang lebih baik dan terinformasi[10]. Dengan Extreme Gradient Boosting (XGBoost) difungsikan sebagai alat bantu yang dapat memprediksi harga jual kelapa sawit tersebut. Selain itu, XGBoost mengoptimalkan pola akses cache, kompresi data, dan sharding untuk membentuk sistem penguatan pohon yang efisien dan skalabel.

Kombinasi fitur ini memungkinkan XGBoost memproses data berukuran besar dengan efisiensi sumber daya yang lebih tinggi dibandingkan sistem lainnya[11].



Gambar 2.1 Arsitektur XGBoost 1

Gambar 2.1 menjelaskan bahwa model XGBoost dibangun secara bertahap (*sekuensial*), di mana setiap pohon keputusan (*Tree*) hingga *Tree n* dilatih untuk memprediksi residual atau kesalahan prediksi dari model kumulatif sebelumnya. Artinya, setiap pohon baru berfungsi memperbaiki kesalahan dari gabungan pohon-pohon sebelumnya. Proses ini berulang secara iteratif hingga jumlah pohon yang ditentukan tercapai. Hasil akhir diperoleh dengan menjumlahkan prediksi semua pohon (*Sum of Results*), sehingga menghasilkan model prediktif yang lebih akurat. Pendekatan ini merupakan inti dari metode *boosting*, yaitu meningkatkan performa model secara progresif melalui perbaikan bertahap[12].

Berdasarkan arsitektur pada Gambar 2.1, proses kerja XGBoost dapat dijelaskan sebagai berikut.



1. Input Dataset

Tahap pertama adalah memasukkan dataset yang telah dikumpulkan. Dataset terdiri atas data historis harga Tandan Buah Segar (TBS) kelapa sawit beserta atribut yang digunakan sebagai variabel prediksi. Sebelum diproses, data telah melalui tahap preprocessing seperti pembersihan data (data cleaning), penanganan data kosong (missing value), serta normalisasi atau transformasi apabila diperlukan agar kualitas data menjadi lebih baik.

2. Inisialisasi Model

Setelah dataset siap digunakan, XGBoost membentuk model awal dengan memberikan nilai prediksi awal (base prediction). Nilai ini menjadi dasar sebelum proses pembentukan pohon keputusan dilakukan.

3. Pembangunan Decision Tree Pertama

Pada tahap ini algoritma membangun pohon keputusan pertama berdasarkan data pelatihan (training data). Pohon pertama menghasilkan prediksi awal terhadap harga sawit berdasarkan pola yang ditemukan pada data historis.

4. Perhitungan Error (Residual)

Hasil prediksi dari pohon pertama dibandingkan dengan nilai aktual sehingga diperoleh selisih atau residual. Residual menunjukkan besarnya kesalahan prediksi yang masih dimiliki model.

5. Pembentukan Decision Tree Berikutnya



Berdasarkan nilai residual tersebut, XGBoost membangun pohon keputusan baru yang bertujuan memperbaiki kesalahan pada pohon sebelumnya. Setiap pohon tidak mempelajari data dari awal, tetapi fokus mempelajari kesalahan (error) yang belum berhasil diprediksi.

6. Proses Gradient Boosting

Seluruh pohon keputusan dibangun secara bertahap menggunakan metode Gradient Boosting. Pada setiap iterasi, algoritma menghitung nilai gradient dari fungsi kerugian (loss function) untuk menentukan arah perbaikan model. Selain itu, XGBoost menerapkan teknik regularisasi (L1 dan L2) untuk mengurangi risiko overfitting, sehingga model mampu memberikan prediksi yang lebih stabil pada data baru.

7. Penggabungan Seluruh Pohon (Ensemble Model)

Setelah seluruh proses iterasi selesai, semua pohon keputusan digabungkan menjadi satu model ensemble. Prediksi akhir diperoleh dari akumulasi kontribusi seluruh pohon yang telah dibangun secara bertahap. Dengan mekanisme ini, setiap pohon saling melengkapi dalam meningkatkan ketepatan prediksi.

8. Output Prediksi

Tahap terakhir menghasilkan nilai prediksi harga kelapa sawit berdasarkan pola yang telah dipelajari dari data historis. Model yang telah dilatih kemudian digunakan untuk memprediksi harga sawit pada periode berikutnya, sehingga dapat membantu petani maupun pihak terkait dalam memperkirakan perubahan harga di masa mendatang.



Formulasi matematis XGBoost dijelaskan melalui komponen berikut:

$$obj(t) = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)}) + f_t(x_i) + \Omega(f_t) + constant$$

Keterangan:

- n : Jumlah total data latih (training samples)
- y_i : Label atau nilai aktual dari data ke- i
- $\hat{y}_i^{(t-1)}$: Hasil prediksi kumulatif hingga iterasi ke- $(t - 1)$
- $f_t(x_i)$: Output dari pohon ke- t yang ditambahkan di iterasi ke- t
- $l(y_i, \hat{y}_i)$: Fungsi kerugian (loss) antara nilai aktual dan prediksi
- $\Omega(f_t)$: Fungsi regularisasi yang menghitung kompleksitas pohon f_t

2.2 Kajian Literatur

Berikut penelitian yang telah dilakukan memiliki korelasi searah dengan penelitian yang akan dibahas dalam menyempurnakan perlu dilakukan kajian literatur diantaranya;

Tabel 2.1 Analisis Metode Berbagai Objek Penelitian Terdahulu

No	Kasus Penelitian	Model Algoritma	Sumber Dataset	Hasil Akurasi / Evaluasi
1	Prediksi harga cabai cayenne[13]	XGBoost Regression	Data historis harga cabai + faktor iklim & ekonomi	R^2 Train: 99% – R^2 Test: 92%, Cross-Validation 96%
2	Prediksi penjualan produk[4]	XGBoost Regression & XGBoost Linear	Data penjualan 24 produk SRC selama 3 bulan	Akurasi: 96.56%, 99.22%, 99.59% (variasi model XGBoost)
3	Prediksi harga saham Tesla[14]	XGBoost & LSTM	Data harga saham Tesla tahun 2025	XGBoost vs LSTM – LSTM lebih akurat; evaluasi RMSE/MAE/ R^2
4	Prediksi harga TBS sawit[15]	ARIMA (pembanding non-ML)	Data mingguan harga TBS sawit 2019–2024	ARIMA (1,1,1) → proyeksi trend naik





Tabel 2.1 Lanjutan

No	Kasus Penelitian	Model Algoritma	Sumber Dataset	Hasil Akurasi / Evaluasi
5	Prediksi harga saham menggunakan data historis[16]	XGBoost Regression	Data harga saham harian	OHLC, Volume, Trend
6	Perbandingan model regresi ML[6]	Random Forest, SVR, XGBoost	Dataset keuangan publik	Variabel regresi multi-fitur
7	Prediksi harga kripto (Ethereum)[17]	Random Forest & XGBoost	Data harga ETH	Open, High, Low, Close, Volume
8	Prediksi penjualan ritel berskala besar[18]	XGBoost vs LightGBM	6436 data retail Walmart	Store, CPI, Temperature, Fuel Price, Unemployment
9	Prediksi Harga Rumah Jabodetabek[11]	XGBoost, Random Forest	Data harga rumah Jabodetabek dari Kaggle	3.554 data, 27 variabel
10	Prediksi Harga Saham BCA Tbk[19]	XGBoost	Data harga saham Bank BCA Tbk (time series 2019–2023)	5 tahun, data harian

Tabel 2.1 memuat daftar penelitian terdahulu yang relevan dan memiliki keterkaitan kuat dengan topik penelitian mengenai prediksi harga sawit menggunakan algoritma XGBoost. Setiap penelitian memberikan gambaran mengenai penggunaan metode machine learning khususnya pada kasus prediksi harga, penjualan, atau data time-series.

Penelitian ini menggunakan algoritma XGBoost Regression untuk memprediksi harga cabai cayenne berdasarkan data historis dan faktor-faktor eksternal seperti cuaca dan ekonomi. XGBoost dipilih karena kemampuannya mengatasi pola non-linear dan menangani fitur dalam jumlah banyak. Hasil penelitian menunjukkan bahwa XGBoost mampu memberikan akurasi yang

sangat tinggi, sehingga relevan sebagai model untuk prediksi harga sawit yang juga memiliki pola fluktuatif serupa[13]

Sementara itu penerapan XGBoost Regression dan XGBoost Linear untuk memprediksi penjualan 24 produk di salah satu toko SRC. Fokus penelitian adalah membandingkan performa model XGBoost pada data penjualan retail yang bersifat time-series dan dipengaruhi banyak variabel. Penelitian ini memperkuat bahwa XGBoost efektif pada kasus bisnis dan dapat diadaptasi pada prediksi harga sawit[4]

Membandingkan XGBoost dengan LSTM dalam memprediksi harga saham Tesla. Studi ini menyoroti bahwa meskipun XGBoost memiliki performa baik, model deep learning seperti LSTM lebih unggul dalam menangkap pola jangka panjang. Temuan ini penting sebagai pembanding dalam prediksi harga sawit yang juga bergerak secara fluktuatif[14]

Penelitian ini digunakan sebagai pembanding karena menerapkan ARIMA (1,1,1) untuk memprediksi harga TBS sawit. Meskipun bukan model machine learning, penelitian ini menunjukkan pola harga sawit yang dapat diproyeksikan menggunakan metode time-series klasik. Hal ini relevan untuk menilai kelebihan XGBoost dibanding ARIMA[15].

Menggunakan XGBoost Regression untuk memprediksi harga saham berdasarkan data historis seperti OHLC, volume, dan tren. Penelitian ini menguatkan bukti bahwa XGBoost mampu mempelajari pola harga yang dinamis dan tidak stabil seperti harga sawit[16].





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.
3. Penggunaan tanpa izin untuk kepentingan komersial atau pelanggaran hak cipta 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.

Membandingkan beberapa model regresi seperti Random Forest, SVR, dan XGBoost. Hasilnya menunjukkan XGBoost memberikan performa kompetitif dan stabil. Ini menunjukkan bahwa XGBoost merupakan model yang kuat untuk prediksi berbasis regresi[6]

Selanjutnya penelitian ini membandingkan Random Forest dan XGBoost dalam memprediksi harga Ethereum. XGBoost menghasilkan performa lebih baik, membuktikan keunggulannya dalam menangani data harga kripto yang sangat fluktuatif – mirip dengan harga sawit[17].

Penelitian ini membandingkan XGBoost dan LightGBM untuk prediksi penjualan Walmart. Hasilnya menunjukkan LightGBM sedikit lebih unggul, namun XGBoost tetap memberikan performa tinggi. Penelitian ini memperkuat posisi XGBoost sebagai model yang andal dan fleksibel dalam regresi time-series[18].

Menggunakan XGBoost dan Random Forest untuk prediksi harga rumah. Meskipun XGBoost tidak selalu paling unggul pada dataset tertentu, penelitian ini memberikan gambaran mengenai kebutuhan tuning agar performanya optimal[11].

Selanjutnya penelitian ini menggunakan XGBoost untuk memprediksi harga saham BCA dan menghasilkan akurasi sangat tinggi (MAPE 2,48%). Hal ini menunjukkan XGBoost cocok diterapkan pada pola harga yang kompleks seperti harga sawit[19].

Seluruh penelitian ini menunjukkan bahwa dataset time-series dapat dipelajari dengan baik menggunakan XGBoost, sehingga mendukung pemilihan model untuk prediksi harga sawit.

Tabel 2.2 Dataset dan Akurasi Penelitian Terdahulu

No	Dataset	Metode	Akurasi / Evaluasi	Ref
1	Harga cabai historis	XGBoost	$R^2 = 92\%$	[13]
2	Penjualan retail	XGBoost	Akurasi 96–99%	[4]
3	Harga TBS Sawit	ARIMA	Trend naik	[15]
4	Saham BCA	XGBoost	MAPE 2,48%	[19]
5	Harga TBS Sawit 2021–2025	XGBoost	Dievaluasi RMSE, MAE, R^2	Penelitian ini

Tabel 2.2 menggambarkan sumber dataset, karakteristik data, serta tingkat akurasi yang diperoleh masing-masing penelitian. Secara umum, seluruh penelitian menggunakan dataset berbasis time-series yang dipengaruhi banyak faktor, sehingga relevan untuk prediksi harga sawit.

Menunjukkan ringkasan penelitian terdahulu yang menggunakan berbagai dataset dan metode dalam prediksi harga dan penjualan. Sebagian besar penelitian membuktikan bahwa algoritma XGBoost mampu memberikan tingkat akurasi yang tinggi dibandingkan metode konvensional. Namun, penelitian yang secara khusus memprediksi harga sawit masih terbatas dan umumnya menggunakan metode statistik seperti ARIMA. Hal ini menunjukkan adanya peluang penelitian untuk mengimplementasikan XGBoost pada prediksi harga sawit dengan memanfaatkan data historis yang lebih panjang dan evaluasi model yang lebih komprehensif.





Tabel 2.3 Perbandingan Metode Prediksi Harga

Metode	Jenis Data	Kemampuan Non-Linear	Akurasi
ARIMA	Time series	Rendah	Sedang
KNN	Numerik	Sedang	Sedang
SVR	Numerik	Tinggi	Tinggi
Random Forest	Multi-fitur	Tinggi	Tinggi
XGBoost	Multi-fitur	Sangat Tinggi	Sangat Tinggi

Tabel 2.3 Membandingkan kemampuan beberapa metode prediksi harga berdasarkan jenis data, kemampuan menangani hubungan non-linear, dan tingkat akurasi. Metode statistik seperti ARIMA memiliki keterbatasan dalam memodelkan data non-linear, sedangkan metode machine learning seperti Random Forest dan XGBoost memiliki kemampuan yang lebih baik. XGBoost menunjukkan performa yang unggul karena mampu menggabungkan beberapa model weak learner sehingga menghasilkan prediksi yang lebih stabil dan akurat, terutama pada data harga sawit yang memiliki pola tidak linear dan dipengaruhi oleh berbagai faktor.

Tabel 2.4 Variabel Penelitian Prediksi Harga Sawit

No	Variabel	Jenis Variabel	Keterangan
1	Harga TBS	Dependen	Harga sawit per kg
2	Bulan	Independen	Pola musiman
3	Periode	Independen	Mingguan/bulanan
4	Tahun	Independen	Tren jangka panjang
5	Umur Sawit	Independen	Pengaruh produksi

Tabel 2.4 menjelaskan variabel yang digunakan dalam penelitian prediksi harga sawit. Harga TBS berperan sebagai variabel dependen yang dipengaruhi oleh beberapa variabel independen seperti periode waktu, bulan, dan tahun. Pemilihan variabel ini didasarkan pada karakteristik data harga sawit yang

memiliki pola musiman dan tren jangka panjang. Dengan menggunakan variabel-variabel tersebut, model XGBoost diharapkan mampu mempelajari pola historis harga sawit secara lebih akurat.

Tabel 2.5 Akurasi Model Prediksi Harga

Metode	RMSE	MAE	R ²
ARIMA	Tinggi	Tinggi	Sedang
Random Forest	Sedang	Sedang	Tinggi
XGBoost	Rendah	Rendah	Tinggi

Tabel 2.5 menyajikan perbandingan performa model prediksi harga berdasarkan nilai RMSE, MAE, dan koefisien determinasi (R²). Metode XGBoost menunjukkan nilai kesalahan yang lebih rendah dibandingkan metode lainnya, yang menandakan tingkat prediksi yang lebih presisi. Hal ini menguatkan alasan pemilihan XGBoost sebagai metode utama dalam penelitian ini, karena mampu menghasilkan prediksi yang lebih mendekati nilai aktual harga sawit.

Tabel 2.6 Penggunaan XGBoost pada Penelitian Terdahulu

No	Objek Penelitian	Metode	Hasil
1	Harga Cabai	XGBoost	Akurasi tinggi
2	Penjualan Retail	XGBoost	Stabil & presisi
3	Saham	XGBoost	Error rendah
4	Harga Sawit	ARIMA (pembanding)	Akurasi terbatas
5	Harga Sawit (Penelitian ini)	XGBoost	-

Tabel 2.6 menunjukkan penerapan algoritma XGBoost pada berbagai bidang penelitian, seperti prediksi harga cabai, penjualan retail, dan pasar saham. Hasil penelitian terdahulu menunjukkan bahwa XGBoost konsisten memberikan performa yang baik dengan tingkat error yang rendah. Namun, penerapan



XGBoost dalam prediksi harga sawit masih sangat terbatas. Oleh karena itu, penelitian ini bertujuan untuk mengisi kekosongan tersebut dengan menerapkan XGBoost secara khusus pada data harga sawit.

Tabel 2.7 Posisi Penelitian yang Dilakukan Berbagai Objek

Metode	Objek	Kelemahan	Ref
ARIMA	Harga Sawit	Tidak tangani non-linear	[15]
XGBoost	Cabai	Bukan sawit	[13]
XGBoost	Saham	Faktor pasar berbeda	[14]
XGBoost	Harga Sawit	Fokus dataset historis	Penelitian ini

Tabel 2.7 menggambarkan posisi penelitian yang dilakukan dibandingkan dengan penelitian terdahulu. Penelitian sebelumnya umumnya menggunakan metode statistik atau menerapkan XGBoost pada objek yang berbeda seperti cabai dan saham. Penelitian ini menempatkan XGBoost sebagai metode utama untuk memprediksi harga sawit, sehingga memiliki kebaruan dalam objek penelitian dan pendekatan yang digunakan. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem prediksi harga sawit berbasis machine learning.

Beberapa penelitian menunjukkan bahwa XGBoost unggul dari sisi akurasi, stabilitas, dan kemampuan menangani fitur kompleks. Namun beberapa studi juga menyoroti kelemahan seperti sensitivitas terhadap hyperparameter dan kesulitan menangkap pola jangka panjang, sehingga perlu tuning atau dikombinasikan dengan model lain seperti LSTM.





Pada penelitian retail dan saham, XGBoost terbukti kuat dalam menangani pola volatil dan data besar. Sementara pada dataset terbatas, performa model dapat menurun sehingga rekomendasinya adalah menambah data historis.

Penelitian ARIMA menunjukkan pentingnya membandingkan XGBoost dengan metode time-series klasik agar penelitian lebih komprehensif. Penelitian rumah dan saham juga menyoroti pentingnya feature engineering dan penambahan variabel eksternal agar model lebih robust.

Secara keseluruhan, rekomendasi dari penelitian sebelumnya mengarah pada:

- a. Perlu tuning hyperparameter XGBoost.
- b. Dataset yang lebih panjang meningkatkan akurasi.
- c. Pertimbangan penggunaan variabel eksternal.
- d. Kombinasi model dapat meningkatkan performa.

Hal ini menjadi dasar kuat mengapa algoritma XGBoost dipilih pada penelitian harga sawit.

2.3 Rangkuman

Berdasarkan kajian literatur, dapat disimpulkan bahwa XGBoost merupakan algoritma yang sangat efektif dalam memprediksi harga komoditas, harga saham, dan data penjualan. Hampir seluruh penelitian terdahulu menunjukkan bahwa XGBoost memiliki akurasi tinggi dan mampu menangani data yang kompleks serta non-linear.

Penelitian terkait harga cabai, penjualan retail, saham, kripto, hingga harga rumah membuktikan bahwa XGBoost unggul dalam berbagai konteks prediksi. Sementara penelitian ARIMA memberikan pembandingan metode tradisional untuk prediksi harga sawit.

Temuan-temuan ini memperkuat alasan pemilihan XGBoost sebagai model utama dalam penelitian prediksi harga sawit. Dengan dataset historis sawit dan metode preprocessing yang tepat, model XGBoost diharapkan dapat memberikan tingkat akurasi tinggi serta membantu memecahkan masalah fluktuasi harga sawit yang selama ini sulit diprediksi.

