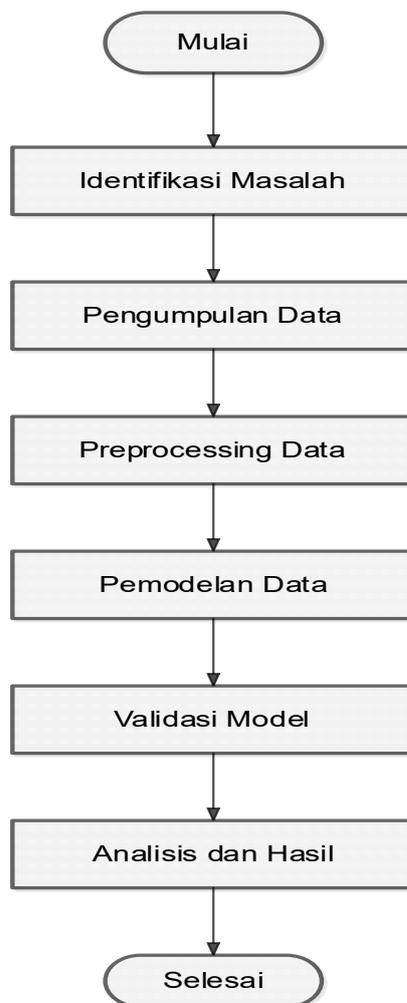


BAB III

METODOLOGI PENELITIAN

3.1 Kerangka Kerja Penelitian

Kerangka kerja penelitian adalah struktur konseptual yang dirancang untuk memberikan arahan sistematis dalam pelaksanaan penelitian. Kerangka ini mencakup tahapan yang mengintegrasikan teori, data, dan metode untuk mencapai tujuan penelitian, yakni membangun model prediksi hasil panen kelapa sawit yang akurat berdasarkan data historis PT. Sinar Pandawa. Gambaran kerangka kerja penelitian ini dapat dilihat pada gambar 3.1.



Gambar 3. 1 Kerangka Kerja

Kerangka kerja penelitian meliputi tahapan berikut:

1. Mulai: Penelitian dimulai dengan mengidentifikasi kebutuhan untuk memprediksi hasil panen kelapa sawit menggunakan data historis PT. Sinar Pandawa.
2. Identifikasi Masalah: Masalah utama yang diteliti adalah bagaimana pengelola perkebunan merencanakan jadwal panen dan alokasi sumber daya secara lebih efisien, untuk mendukung keberlanjutan operasional dan daya saing perusahaan menggunakan pendekatan berbasis data .
3. Pengumpulan Data: Data historis kelapa sawit dikumpulkan dari sumber terpercaya, yaitu laporan internal PT. Sinar Pandawa. Pengumpulan data yang dilakukan yaitu metode sekunder, melibatkan data yang sudah ada untuk menjawab pertanyaan penelitian atau mencapai tujuan penelitian.
4. Preprocessing Data: Data yang telah dikumpulkan diproses, dengan tahapan cleaning data, transformasi data dan normalisasi data
5. Pemodelan Data: Algoritma C4.5 digunakan untuk mengkategorikan data (tinggi, rendah dan sedang). Sementara regresi digunakan untuk memprediksi berdasarkan nilai numerik.
6. Validasi Model: Model yang dihasilkan divalidasi menggunakan teknik cross-validation untuk mengukur keakuratan prediksi.
7. Analisis dan Hasil: Data dianalisis untuk membandingkan performa kedua metode serta menghasilkan kesimpulan dan rekomendasi.
8. Selesai: Penelitian ditutup dengan penyusunan laporan yang mencakup hasil, analisis, dan implikasi praktis.

Kerangka kerja ini memberikan pedoman bagi peneliti untuk mengidentifikasi, mengorganisasi, dan melaksanakan langkah-langkah penelitian secara terstruktur. Dengan mengacu pada kerangka kerja ini, peneliti dapat memastikan bahwa pendekatan yang digunakan konsisten dengan pertanyaan penelitian, hipotesis, dan tujuan yang telah ditentukan.

3.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari laporan internal PT. Sinar Pandawa, yaitu data historis hasil panen yang diakumulasikan secara bulanan dalam kurun waktu satu tahun.

Tabel 3. 1 Pengumpulan Data

No	Bulan	Curah hujan (mm)	Pupuk (kg)	Hasil panen (ton)
1	Januari	690	27000	1500
2	Februari	150	83000	2000
3	Maret	289	59000	1900
4	April	227	73000	1900
5	Mei	785	25000	1300
6	Juni	343	51000	1700
7	Juli	333	57000	1800
8	Agustus	130	87000	2000
9	September	143	80000	2000
10	Oktober	530	25000	1300
11	November	152	85000	2000
12	Desember	380	45000	1500

3.3 Pra-pemrosesan Data

Pra-pemrosesan dilakukan untuk memastikan data dianalisis menggunakan metode yang diteliti. Data di klasifikasi sesuai dengan algoritma agar mempermudah langkah selanjutnya. Pada penelitian ini, terdapat tiga variabel, yaitu

curah hujan, pupuk, dan hasil panen. Klasifikasi dilakukan dengan menentukan nilai max dan min berdasarkan ketentuan rumus excel. Selanjutnya, perhitungan dilakukan dengan formula excel menggunakan fungsi IF. Berikut tabel klasifikasi dibawah.

Tabel 3. 2 Klasifikasi Data

No	Bulan	Kategori Curah Hujan	Kategori Pupuk	Kategori Hasil Panen
1	Januari	Tinggi	Sedikit	Sedang
2	Februari	Rendah	Banyak	Tinggi
3	Maret	Rendah	Sedang	Tinggi
4	April	Rendah	Banyak	Tinggi
5	Mei	Tinggi	Sedikit	Rendah
6	Juni	Sedang	Sedang	Sedang
7	Juli	Sedang	Sedang	Sedang
8	Agustus	Rendah	Banyak	Tinggi
9	September	Rendah	Banyak	Tinggi
10	Oktober	Sedang	Sedikit	Rendah
11	November	Rendah	Banyak	Tinggi
12	Desember	Sedang	Sedang	Sedang

3.4 Analisa dengan Algoritma C4.5

Algoritma C4.5 dapat digunakan untuk analisis data, pengambilan keputusan, dan sistem pakar. Ini adalah teknik klasifikasi yang efektif dan mudah dipahami yang menggunakan gain informasi untuk memilih atribut terbaik dalam membagi data dan menggunakan pengukuran kinerja seperti RMSE, SE, dan AE untuk mengevaluasi model.

1. Menghitung nilai entropy dan information gain dari masing-masing atribut terhadap kelas hasil panen.

Tabel 3. 3 Nilai Entropy Curah Hujan

Kategori Curah Hujan	Jumlah Data	Entropy Subset	Proporsi ($ S_v / S $)	Weighted Entropy
Rendah	4	0.9183	0.3333	0.3061
Sedang	5	1.585	0.4167	0.6604
Tinggi	3	1.585	0.25	0.3962

Tabel 3. 4 Nilai Entropy Total Dan Gain Curah Hujan

Entropy Total	Information Gain (Curah Hujan)
1.4591	0.825

Tabel 3. 5 Nilai Entropy Pupuk

Kategori Pupuk	Jumlah Data	Entropy Subset	Proporsi ($ S_v / S $)	Weighted Entropy
Banyak	4	0.9183	0.3333	0.3061
Sedang	4	0.9183	0.3333	0.3061
Sedikit	4	0.9183	0.3333	0.3061

Tabel 3. 6 Nilai Entropy Total Dan Gain Pupuk

Entropy Total	Information Gain (Pupuk)
1.4591	0.7925

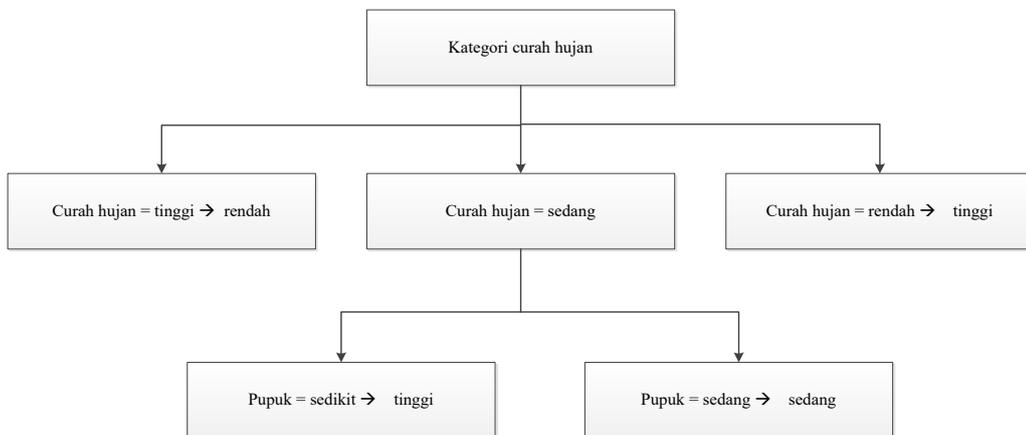
2. Pembangunan Pohon Keputusan C4.5
 - pembentukan pohon dilakukan secara rekursif berdasarkan nilai Information Gain tertinggi. Setiap node pohon mewakili atribut tertentu, sedangkan setiap cabang merepresentasikan kategori nilai atribut tersebut. Daun pohon menyatakan keputusan klasifikasi hasil panen: *Rendah*, *Sedang*, atau *Tinggi*. Gain tertinggi pada atribut yaitu Curah hujan 0,825. Oleh karena itu curah hujan digunakan sebagai akar.

- Pembentukan Cabang Berdasarkan Kategori Atribut Akar
Data dibagi menjadi cabang-cabang berdasarkan kategori dari atribut akar, yaitu: Curah Hujan = Rendah, Sedang, dan Tinggi. Untuk setiap cabang, distribusi kelas hasil panen dianalisis.

Akar Pohon = Kategori Curah Hujan

rediksi: Rendah

- Rendah → prediksi: Tinggi
- Sedang → butuh eksplorasi lanjut (karena campuran kelas Sedang dan Tinggi)



Gambar 3. 2 Pohon Keputusan

3. Melakukan prediksi berdasarkan pohon

Tabel 3. 7 Prediksi Berdasarkan Pohon

No	Curah Hujan (mm)	Kategori Curah Hujan	Prediksi Kategori Panen
1	690	Tinggi	Rendah
2	150	Rendah	Tinggi
3	289	Sedang	Sedang
4	227	Sedang	Sedang
5	785	Tinggi	Rendah
6	343	Sedang	Sedang
7	333	Sedang	Sedang
8	130	Rendah	Tinggi
9	143	Rendah	Tinggi

No	Curah Hujan (mm)	Kategori Curah Hujan	Prediksi Kategori Panen
10	530	Tinggi	Rendah
11	152	Rendah	Tinggi
12	380	Sedang	Sedang

4. Konversi Kelas Hasil Klasifikasi ke Nilai Numerik

Agar hasil klasifikasi dapat dievaluasi secara numerik, setiap kelas dikonversi ke nilai representatif:

- *Rendah* → 1300 ton,
- *Sedang* → 1600 ton,
- *Tinggi* → 2000 ton.

Nilai-nilai ini dapat ditentukan berdasarkan rata-rata hasil panen aktual dalam masing-masing kategori.

Tabel 3. 8 Klasifikasi ke Nilai Numerik

No	Curah Hujan (mm)	Kategori Curah Hujan	Prediksi Kategori Panen	Prediksi Ton
1	690	Tinggi	Rendah	1600
2	150	Rendah	Tinggi	2000
3	289	Sedang	Sedang	1600
4	227	Sedang	Sedang	1600
5	785	Tinggi	Rendah	1300
6	343	Sedang	Sedang	1600
7	333	Sedang	Sedang	1600
8	130	Rendah	Tinggi	2000
9	143	Rendah	Tinggi	2000
10	530	Tinggi	Rendah	1300
11	152	Rendah	Tinggi	2000
12	380	Sedang	Sedang	1600

5. Menghitung nilai AE (absoult error), SE(squad error), MAE(mean absoult error) dan RMSE (root mean square error). MAE adalah rata-rata nilai error, RMSE adalah rata-rata error tapi lebih menekankan pada error lebih besar.

Tabel 3. 9 Nilai AE, SE, MAE dan RMSE

Kategori Curah Hujan	Prediksi Kategori	Prediksi Ton	AE	SE
Rendah	1300	25,700	660,490,000	40000
Tinggi	2000	81,000	6,561,000,000	0
Sedang	1600	57,400	3,294,760,000	90000
Sedang	1600	71,400	5,097,960,000	90000
Rendah	1300	23,700	561,690,000	0
Sedang	1600	49,400	2,440,360,000	10000
Sedang	1600	55,400	3,069,160,000	40000
Tinggi	2000	85,000	7,225,000,000	0
Tinggi	2000	78,000	6,084,000,000	0
Rendah	1300	23,700	561,690,000	0
Tinggi	2000	83,000	6,889,000,000	0
Sedang	1600	43,400	1,883,560,000	10000

MAE : 56,425

RMSE : 60,779

3.5 Analisa Regresi Linear

Analisis regresi linear adalah untuk membangun model matematis yang dapat memprediksi hasil panen kelapa sawit berdasarkan dua variabel prediktor, yaitu curah hujan dan jumlah pupuk. Model ini memungkinkan estimasi kuantitatif hasil panen tanpa perlu pengelompokan ke dalam kelas-kelas kategori.

Berikut langkah-langkah regresi linear:

Langkah Analisis Regresi Linear

1. Identifikasi Variabel

Pada tahap ini, di tentukan variabel-variabel yang digunakan dalam model prediksi:

- Variabel Independen (X):
X1: Curah hujan (mm)
X2: Pupuk (kg)
- Variabel dependen (Y):
Y : Hasil panen (ton)

2. Membangun model regresi

Setelah variabel diidentifikasi, dilakukan proses pelatihan model regresi linear untuk mencari hubungan matematis antara variabel input dan output

Tabel 3. 10 Model Regresi

Curah Hujan (X1)	Pupuk (X2)	Hasil Panen (Y)	X1^2	X2^2	X1*X2
690	27,000	1,500	476,100	729,000,000	18,630,000
150	83,000	2,000	22,500	6,889,000,000	12,450,000
289	59,000	1,900	83,521	3,481,000,000	17,051,000
227	73,000	1,900	51,529	5,329,000,000	16,571,000
785	25,000	1,300	616,225	625,000,000	19,625,000
343	51,000	1,700	117,649	2,601,000,000	17,493,000
333	57,000	1,800	110,889	3,249,000,000	18,981,000
130	87,000	2,000	16,900	7,569,000,000	11,310,000
143	80,000	2,000	20,449	6,400,000,000	11,440,000
530	25,000	1,300	280,900	625,000,000	13,250,000
152	85,000	2,000	23,104	7,225,000,000	12,920,000
380	45,000	1,500	144,400	2,025,000,000	17,100,000
4,152	697,000	20,900	1,964,166	46,747,000,000	186,821,000

Tabel 3. 11 Model Regresi

X1*Y	X2*Y
1,035,000	40,500,000
300,000	166,000,000
549,100	112,100,000
431,300	138,700,000
1,020,500	32,500,000
583,100	86,700,000
599,400	102,600,000
260,000	174,000,000
286,000	160,000,000
689,000	32,500,000
304,000	170,000,000
570,000	67,500,000

6,627,400 1,283,100,000

Kemudian tentukan matriks normal equation dengan sistem persamaan:

Yang menghasilkan:

$$\begin{bmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1 X_2 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum X_1 Y \end{bmatrix}$$

$$[\sum X_2 \quad \sum X_1 X_2 \quad \sum X_2^2] [b_2] = [\sum X_2 Y]$$

$$b_0 = 1159.98$$

$$b_1 = -0.0702$$

$$b_2 = 0.0104$$

Persamaan regresi linear yang digunakan

$$Y = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2$$

$$Y = 1159.98 - 0.0702 \cdot X_1 + 0.0104 \cdot X_2$$

Tabel 3. 12 Model Regresi

X1 (Curah Hujan)	X2 (Pupuk)	Y (Hasil Panen)	\hat{Y} (Prediksi Y)
690	27,000	1,500	1392.342
150	83,000	2,000	2012.65
289	59,000	1,900	1753.2922
227	73,000	1,900	1903.2446
785	25,000	1,300	1364.873
343	51,000	1,700	1666.3014
333	57,000	1,800	1729.4034
130	87,000	2,000	2055.654
143	80,000	2,000	1981.9414
530	25,000	1,300	1382.774
152	85,000	2,000	2033.3096
380	45,000	1,500	1601.304

Interpretasi:

- Nilai **intersep (b₀)** sebesar 1159,98 menunjukkan hasil panen saat curah hujan dan pupuk bernilai nol (nilai dasar).
- Koefisien **b₁ = -0,0702** menunjukkan bahwa setiap peningkatan curah hujan sebesar 1 mm akan menurunkan hasil panen sebesar 0,0702 ton, dengan asumsi variabel lain tetap.
- Koefisien **b₂ = 0,0104** menunjukkan bahwa setiap penambahan pupuk sebanyak 1 kg akan meningkatkan hasil panen sebesar 0,0104 ton.

3. Evaluasi model

Evaluasi dilakukan untuk mengevaluasi performa model, dilakukan perhitungan terhadap nilai prediksi (\hat{Y}), error, dan kuadrat error.

Tabel 3. 13 Evaluasi Model

X1 (Curah Hujan)	X2 (Pupuk)	Y (Hasil Panen)	\hat{Y} (Prediksi Y)	Error (Y - \hat{Y})	Error²
690	27,000	1,500	1392.342	107.658	11590.245
150	83,000	2,000	2012.65	-12.65	160.0225
289	59,000	1,900	1753.2922	146.7078	21523.1786
227	73,000	1,900	1903.2446	-3.2446	10.5274292
785	25,000	1,300	1364.873	-64.873	4208.50613
343	51,000	1,700	1666.3014	33.6986	1135.59564
333	57,000	1,800	1729.4034	70.5966	4983.87993
130	87,000	2,000	2055.654	-55.654	3097.36772
143	80,000	2,000	1981.9414	18.0586	326.113034
530	25,000	1,300	1382.774	-82.774	6851.53508
152	85,000	2,000	2033.3096	-33.3096	1109.52945
380	45,000	1,500	1601.304	-101.304	10262.5004

Kemudian, dilakukan perhitungan Mean Squared Error (MSE) dan Root Mean Squared Error (RMSE):

$$MSE = \frac{1}{n} \sum (Y - \hat{Y})^2$$

$$RMSE = \sqrt{MSE}$$

Hasilnya

RMSE	3545.98834
MAE	1039.4784
R-Squared (R ²)	0.51871419

MAE (Mean Absolute Error) = rata-rata selisih absolut prediksi \approx 1.039 ton.

RMSE = rata-rata selisih kuadrat prediksi \approx 3.546 ton.

R² = 0,519 \rightarrow sekitar 51,9% variasi hasil panen bisa dijelaskan oleh model ini (cukup moderat).

3.6 Perbandingan metode

Berdasarkan hasil evaluasi model, kedua metode menunjukkan kemampuan dalam memprediksi hasil panen kelapa sawit, namun dengan pendekatan yang

berbeda. Metode regresi linear memberikan hasil prediksi dalam bentuk numerik langsung, dan menghasilkan RMSE sebesar 3.545,99, MAE sebesar 1.039,48, serta R^2 sebesar 0,519, yang menunjukkan bahwa 51,9% variasi hasil panen dapat dijelaskan oleh model. Sementara itu, algoritma C4.5 mengklasifikasikan hasil panen ke dalam kategori (rendah, sedang, tinggi), yang kemudian dikonversi ke nilai numerik untuk evaluasi. Metode ini menghasilkan RMSE sebesar 60.779 dan MAE sebesar 56.425. Secara numerik, regresi linear memberikan akurasi yang lebih tinggi karena nilai RMSE dan MAE jauh lebih rendah dibandingkan C4.5. Namun, C4.5 tetap relevan karena interpretasinya sederhana dan dapat digunakan untuk pengambilan keputusan berbasis aturan logika, seperti dalam sistem pakar.

Tabel 3. 14 Perbandingan Metode

Kriteria Evaluasi	Regresi Linear	Algoritma C4.5
Tipe Output	Numerik (kontinu)	Kategori (tinggi, sedang, rendah)
Metode Evaluasi	RMSE, MAE, R^2	AE, SE, MAE, RMSE
RMSE	3.545,99	60.779
MAE	1.039,48	56.425
R-Squared (R^2)	0,519	—
Interpretasi Output	Prediksi hasil panen dalam ton	Prediksi kelas hasil panen yang dikonversi ke ton
Kelebihan	Memberikan estimasi langsung dan kuantitatif	Mudah dipahami, berbentuk aturan logika pohon
Kekurangan	Sensitif terhadap outlier	Kehilangan informasi numerik saat dikategorikan