

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil

Penelitian ini bertujuan untuk menerapkan Metode *Decision Tree* menggunakan *Algorithma Classification and Regression Tree* untuk memprediksi harga minyak kelapa sawit mentah. Pada tahap ini perlu dilakukan mengklasifikasi harga minyak mentah, mulai dari proses mempersiapkan dataset yang akan digunakan dengan *range* 10 tahun, melakukan perhitungan *entropy* dan *gain*, kemudian pembentukan pohon keputusan.

4.1.1 Persiapan Data

Data historis Minyak Sawit Berjangka (FCPOc1). Dari *website* resmi *Investing.com* yang di unduh dai url: (<https://id.investing.com/commodities/palm-oil-historical-data>). Seperti pada tabel di bawah ini :

Tabel 4. 1 Persiapan Dataset

No	Nama	Record	Feature	Class
1	Data historis Minyak Sawit Berjangka (FCPOc1)	2541	6	1

Dataset ini merupakan dataset *public* dengan *record* yang cukup banyak dengan berbagai tipe. Karena dataset ini merupakan dataset *realtime* yang setiap harinya akan terus bertambah dengan data baru.

4.1.2 Implementasi Metode

Pembuatan model pohon keputusan dilakukan pada dataset yang terdiri dari 7 atribut yang merupakan atribut dari dataset histori minyak sawit berjangka. Algoritma yang digunakan adalah CART (*Classification and Regression Tree*). Metode ini bekerja dengan membuat pohon keputusan untuk memprediksi variabel target harga CPO, perhitungannya dilakukan sebagai berikut dengan menggunakan sampel 10 data dari dataset uji yang saya gunakan

- a. Pengambilan 10 sampel data di tahun 2025

Tabel 4. 2 Implementasi Metode

No	Tanggal	Harga Pembukaan	Harga Terakhir
1	02/01/2025	4.862,00	4.690,00
2	03/01/2025	4.720,00	4.723,00
3	06/01/2025	4.740,00	4.690,00
4	07/01/2024	4.659,00	4.713,00
5	08/01/2024	4.700,00	4.711,00
6	09/01/2024	4.695,00	4.635,00
7	10/01/2024	4.697,00	4.710,00
8	13/01/2024	4.800,00	4.815,00
9	14/01/2024	4.789,00	4.668,00
10	15/01/2024	4.637,00	4.675,00

Kita akan menggunakan harga pembukaan sebagai atribut prediktor (X) dan harga penutupan sebagai target (Y).

- b. Rata-rata harga terakhir

$$Y = \frac{\sum_{i=1}^n Y_i}{n} = \frac{\text{Total keseluruhan output}}{\text{Jumlah data output}}$$

$$Y = 4.690,00 + 4.723,00 + 4.690,00 + 4.713,00 + 4.711,00 + 4.635,00 + 4.710,00 + 4.815,00 + 4.668,00 + 4.675,00$$

$$Y = \frac{47.030}{10} = 4.703$$

c. Menghitung nilai varian

$$Variance(Y) = \frac{1}{N} \sum_{i=1}^n (Y_i - mean(Y))^2$$

$$Variance(Y) = \frac{1}{Jumlah\ Data} (Rata - rata\ Output - Output)^2$$

$$\begin{aligned} Variance(Y) &= \frac{1}{10} * ((4.703 - 4.690)^2 + (4.703 - 4.723)^2 + (4.703 - 4.690)^2 \\ &\quad + (4.703 - 4.713)^2 + (4.703 - 4.711)^2 + (4.703 - 4.635)^2 \\ &\quad + (4.703 - 4.710)^2 + (4.703 - 4.815)^2 + (4.703 - 4.668)^2 \\ &\quad + (4.703 - 4.675)^2 = 2.008 \end{aligned}$$

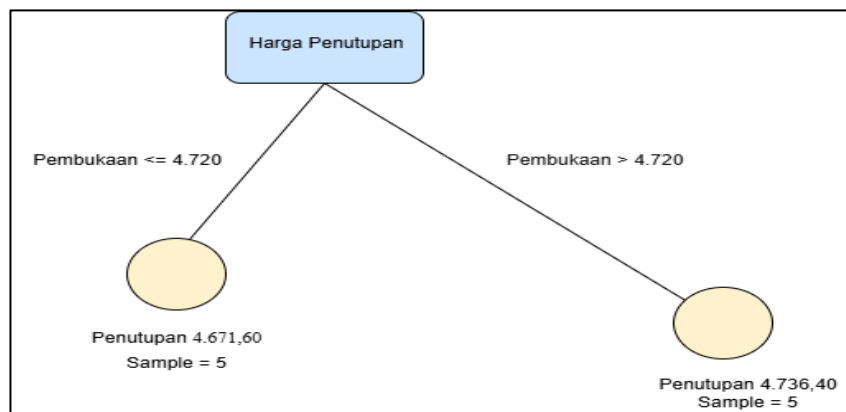
d. Menentukan atribut *splitting*

1. Nilai batas ≤ 4.720

$$Y = \frac{Jumlah\ Output}{Jumlah\ Data\ Output} = \frac{23.358}{5} = 4.671,60$$

2. Nilai batas > 4.720

$$Y = \frac{Jumlah\ Output}{Jumlah\ Data\ Output} = \frac{23.682}{5} = 4.736,40$$



Gambar 4. 1 Implementasi Metode

4.2 Pembahasan

4.2.1 Pengolahan Data

Dataset di tampilkan untuk mengetahui tipe data yang ada di dalam dataset tersebut.

Tabel 4. 3 Keterangan Dataset Awal

Nama Atribut	Tipe data
Tanggal	<i>Object</i>
Terakhir	<i>Object</i>
Pembukaan	<i>Object</i>
Tertinggi	<i>Object</i>
Terendah	<i>Object</i>
Vol.	<i>Object</i>
Perubahan%	<i>Object</i>

Dari dataset awal semua data memiliki tipe data *Object*. Setelah data diketahui maka data tersebut diolah dengan dikonversi ke dalam tipe data *numerik* agar sesuai dengan kebutuhan dari model pohon keputusan.

Tabel 4. 4 Konversi Tipe data

Nama Atribut	Tipe data	Nilai
Tanggal	<i>Datetime</i>	2014/08/18 – 2024/12/31
Terakhir	<i>Float</i>	1.759 – 8.163
Pembukaan	<i>Float</i>	1.751 – 8.200
Tertinggi	<i>Float</i>	1.759 – 8.757
Terendah	<i>Float</i>	1.750 – 7.780
<i>Vol.</i>	<i>Float</i>	0.000 – 632.000
Perubahan %	<i>Float</i>	0.000 – 1091.000

4.2.2 Eksplorasi Data

Pada tahap eksplorasi data ini dilakukan proses pembacaan data awal dari dataset dengan tujuan untuk mengetahui adanya data yang hilang serta mencari anomali-anomali pada dataset tersebut.

Tabel 4. 5 Explorasi Data

Index	Tanggal	Terakhir	Pembukaan	Tertinggi	Terendah	Vol.	Perubahan%
0	31/12/2024	4.861	4.912	4.800	4.800	1,31k	-1,30%
1	30/12/2024	4.925	4.956	4.993	4.920	0,96K	-1,01%
2	27/12/2024	4.975	4.905	4.980	4.878	1,66K	1,53%
3	26/12/2024	4.900	4.850	4.916	4.813	0,53K	0,10%
4	24/12/2024	4.895	4.885	4.923	4.846	0,76K	-0,12%
...	...	...	...	...	...	...	...
2536	22/08/2014	2.021	2.063	2.063	2.020	1,24K	-1,65%
2537	21/08/2014	2.055	2.070	2.070	2.044	0,92K	0,05%
2538	20/08/2014	2.054	2.078	2.078	2.051	0,56K	-0,96%
2539	19/08/2014	2.074	2.107	2.074	2.074	0,86K	-1,19%
2540	18/08/2014	2.099	2.205	2.096	2.099	1,04K	-3,09%

Pada dataset awal di atas kita dapat mengetahui terdapat 2541 baris dan 7 kolom atribut, dengan *Missing Value* terdapat pada kolom *Vol.*

4.2.3 Analisa Data

Analisis data merupakan serangkaian proses pengolahan data dan penelitian yang bertujuan untuk menemukan informasi penting guna membantu pengambilan keputusan dan pemecahan masalah. Proses ini dimulai dengan mengelompokkan data berdasarkan karakteristik, membersihkan data untuk mengatasi ketidak konsistenan dan nilai yang hilang, mentransformasikan data agar dapat dipersiapkan dengan baik untuk dianalisis, dan menyediakan pola pencarian yang relevan. Proses ini melibatkan beberapa fase, dimulai dengan membangun model yang memperoleh wawasan penting dari data.

Dalam analisa data pada dataset dilakukan pemisahan terkait dengan *class* atau label yang berfungsi sebagai target atau *output* dengan *atribut* atau *feature* yang berfungsi sebagai *input*.

Tabel 4. 6 Analisa Data

Kolom	Tipe Data
Tanggal	<i>Object</i>
Terakhir	<i>Object</i>
Pembukaan	<i>Object</i>
Tertinggi	<i>Object</i>
Vol.	<i>Object</i>
Perubahan%	<i>Object</i>

Pada Dataset ini, setiap kolom masih dalam tipe data objek yang Dimana nantinya harus di encode agar bisa di proses.

a. *Class* atau Label

Class atau label pada dataset ini adalah kolom Terakhir yang merupakan harga penutupan atau harga terakhir dari penjualan pada hari tersebut, yang akan berfungsi sebagai *output* dari model mesin *learning* yang di buat, dari analisa *price* atau terakhir dapat dilihat pada tabel berikut.

Tabel 4. 7 Analisa Data Terakhir

Keterangan	Nilai
Nilai Minimal	1.759
Nilai Maksimal	8.163
Jumlah Baris	3541 Baris Data
Keunikan Setiap Data	1534Baris Data

b. *Atribut* atau *Feature*

1. Tanggal

Tanggal adalah kolom yang merupakan tanggal dari penjualan minyak kelapa sawit mentah pada hari tersebut. Pada kolom tanggal ini terdapat 2541 baris data yang dimana setiap isi berbeda-beda sebanyak 2541 data. Tanggal yang di ambil dari data ini dimulai dari 18/08/2014 sampai 31/12/2024.

Tabel 4. 8 Analisa data Tanggal

Keterangan	Nilai
Jumlah Baris	2541 Baris Data
Keunikan setiap data	2541 Baris Data
Tanggal Awal	18/08/2014
Tanggal Akhir	31/12/2024

2. Pembukaan

Kolom pembukaan merupakan harga pembukaan/awal dari penjualan pada hari tersebut. Pada kolom pembukaan terdapat 2541 baris data dengan jumlah keunikan 1534 baris data, data yang paling banyak muncul yaitu 2.200,00 dengan frekuensi kemunculan sebanyak 11 kali.

Tabel 4. 9 Analisa data Pembukaan

Keterangan	Nilai
Jumlah Baris	2541 Baris Data
Keunikan setiap data	1534 Baris Data
Data yang sering muncul	2.200,00
Frekuensi kemunculan	11

3. Tertinggi

Kolom Tertinggi merupakan harga tertinggi dari penjualan pada hari tersebut. Pada kolom Tertinggi terdapat 2541 baris data dengan jumlah keunikan 1461 baris data, data yang paling banyak muncul yaitu 2.150,00 dengan frekuensi kemunculan sebanyak 13 kali.

Tabel 4. 10 Analisa data Tertinggi

Keterangan	Nilai
Jumlah Baris	2541 Baris Data
Keunikan setiap data	1461 Baris Data
Data yang sering muncul	2.150,00
Frekuensi kemunculan	13

4. Terendah

Kolom Terendah merupakan harga terendah dari penjualan pada hari tersebut. Pada kolom Terendah terdapat 2541 baris data dengan jumlah keunikan 1454 baris data, data yang paling banyak muncul yaitu 2.850,00 dengan frekuensi kemunculan sebanyak 4 kali.

Tabel 4. 11 Analisa data Terendah

Keterangan	Nilai
Jumlah Baris	2541 Baris Data
Keunikan setiap data	1454 Baris Data
Data yang sering muncul	3.850,00
Frekuensi kemunculan	11

5. *Vol.*

Kolom *Vol.* merupakan volume dari penjualan pada hari tersebut. Pada kolom *Vol.* terdapat 2392 baris data dengan jumlah keunikan 276 baris data, data yang paling banyak muncul yaitu 0,01k dengan frekuensi kemunculan sebanyak 87 kali.

Tabel 4. 12 Analisa Data Vol.

Keterangan	Nilai
Jumlah Baris	2392 Baris Data
Keunikan setiap data	276 Baris Data
Data yang sering muncul	0.01K
Frekuensi kemunculan	87

6. Perubahan%

Kolom perubahan merupakan harga perubahan dari penjualan kedalam bentuk persentase pada hari tersebut. Pada kolom Perubahan% terdapat 2541 baris data dengan jumlah keunikan 713 baris data, data yang paling banyak muncul yaitu 0,00% dengan frekuensi kemunculan sebanyak 80 kali.

Tabel 4. 13 Analisa data Perubahan%

Keterangan	Nilai
Jumlah Baris	2541
Keunikan setiap data	713
Data yang sering muncul	0,00%
Frekuensi kemunculan	80

4.2.4 Stastika Data

Pada tahap ini dilakukan proses pemanggilan informasi terkait dengan nilai-nilai statistik pada dataset, khususnya pada kolom-kolom yang bertipe *numerik* pada dataset ini, seperti tabel berikut ini.

Tabel 4. 14 Statistika Data

	Terakhir	Pembukaan	Tertinggi	Terendah	Vol.	Perubahan%
Count	2541.000	2541.000	2541.000	2541.000	2541.000	2541.000
Mean	3.207	3.204	3.234	3.178	57.609	131.288
Min	1.759	1.751	1.759	1.750	0.000	0,000
25%	2.278	2.274	2.290	2.263	9000	47.000
50%	2.298	2.799	2.811	2.779	32.000	100.000
75%	3.929	3.930	3.962	3.895	86.000	175.000
Max	8.163	8.200	8.757	7.780	632.000	1091.000
Std	1.137	1.132	1.155	1.114	67.409	125.888

Tabel 4. 15 Keterangan Statistik Data

Nama	Keterangan
Count	digunakan untuk menghitung jumlah data elemen dalam suatu array atau struktur data.
Mean	Nilai rata-rata dari sejumlah nilai atau data.
Minimal	Nilai terkecil dari data yang ada
25%	Nilai kuartil pertama.
50%	Nilai kuartil kedua.
75%	Nilai kuartil ketiga.
Maximal	Nilai tertinggi dari data yang ada.
Standar Deviasi	Nilai statistik yang di gunakan untuk mengukur sejumlah mana data tersebar dari rata-rata dalam satu himpunan data.

4.2.5 Visualisasi Data

Pada tahap ini akan dilakukan visualisasi data dengan menggunakan grafik atau diagram untuk mengetahui pola atau trend pada data.

a. *Class* atau label

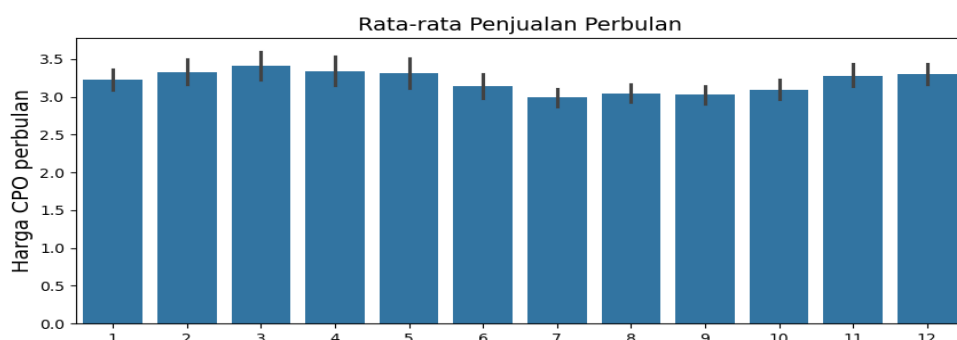
Visualisasi data pada *class* atau label dengan menggunakan kolom *price* atau harga penutupan penjualan minyak mentah sebagai berikut data yang tampilkan.

1. Pola harga penutupan pertahun

**Gambar 4. 2 Pola Harga Penutupan Pertahun**

Grafik ini menunjukkan pergerakan harga penutupan CPO dari tahun 2014 hingga akhir 2024. Pada awal periode, harga berada pada level terendah dan kemudian mengalami kenaikan bertahap dengan beberapa fluktuasi kecil hingga sekitar tahun 2017. Dari tahun 2017 sampai 2020, harga mengalami penurunan dan fluktuasi kecil. Setelah itu, harga relatif stabil sampai tahun 2020. Pada periode 2021 hingga awal 2022, terjadi lonjakan harga yang sangat tajam, diikuti oleh penurunan drastis sepanjang tahun 2022. Mulai pertengahan 2022 hingga akhir 2024, harga kembali stabil dan menunjukkan tren kenaikan perlahan. Kesimpulannya, harga CPO mengalami beberapa fase penting: kenaikan bertahap, penurunan dan fluktuasi kecil, lonjakan tajam, penurunan cepat, dan stabilisasi dengan kenaikan perlahan.

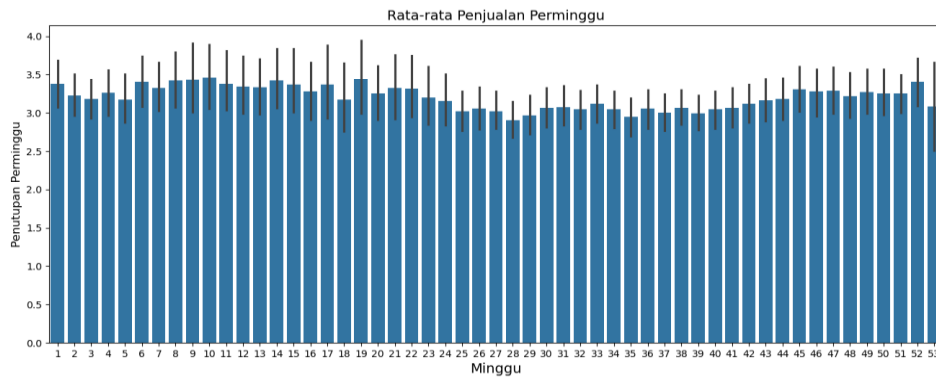
2. Pola harga penutupan perbulan



Gambar 4. 3 Pola Harga Penutupan Perbulan

Grafik ini menunjukkan harga rata-rata CPO setiap bulan. Harga tertinggi terjadi pada bulan ke-3 dan ke-11, yaitu sekitar 3,4, sedangkan harga terendah terjadi pada bulan ke-7, sekitar 3,0. Sedangkan garis hitam di atas batang (error bar) menunjukkan variasi atau ketidakpastian harga, yaitu seberapa besar harga bisa naik atau turun dari nilai rata-rata.

3. Pola harga penutupan perminggu



Gambar 4. 4 Pola Harga Penutupan Perminggu

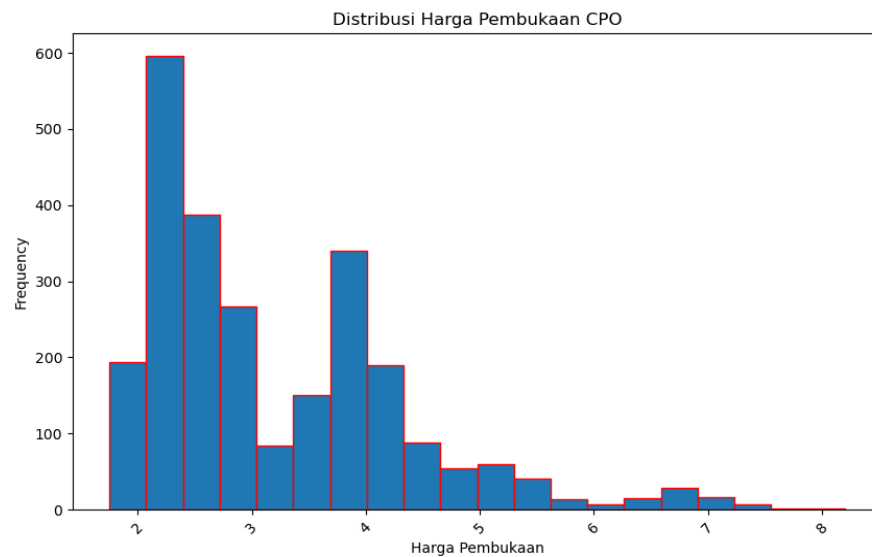
Grafik ini menunjukkan rata-rata penjualan setiap minggu selama 53 minggu. Penjualan tertinggi terjadi sekitar minggu ke-9 dan minggu ke-19, dengan nilai sekitar 3,5. Sedangkan penjualan terendah terjadi sekitar minggu ke-28, dengan nilai sekitar 3,0. Setiap batang menunjukkan rata-rata penjualan di minggu itu, dan garis hitam di atasnya menunjukkan seberapa besar perbedaan penjualan dalam minggu tersebut. Secara umum, penjualan mingguan cukup stabil dan tidak banyak berubah dari minggu ke minggu.

b. *Atribut* atau *Feature*

1. Pembukaan

Harga pembukaan atau kolom *Pembukaan* pada dataset mengacu pada harga CPO yang di perdagangan pada priode tertentu.

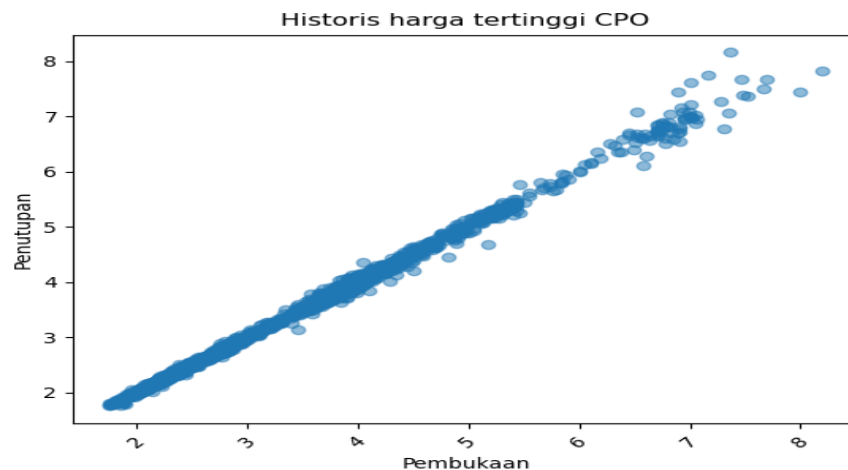
a) Distribusi harga pembukaan



Gambar 4. 5 Distribusi Harga Pembukaan

Grafik ini menunjukkan sebaran harga pembukaan CPO. Sebagian besar harga pembukaan berada di kisaran 2 sampai 4, dengan puncak frekuensi tertinggi sekitar 2 sampai 2.5, mencapai hampir 600 kejadian. Setelah harga 4, jumlah kejadian menurun tajam, dan harga di atas 6 sangat jarang muncul. Harga pembukaan CPO umumnya rendah hingga sedang, berkisar antara 2 sampai 4. Harga yang jauh lebih tinggi sangat jarang terjadi.

b) Distribusi harga pembukaan dengan harga penutupan



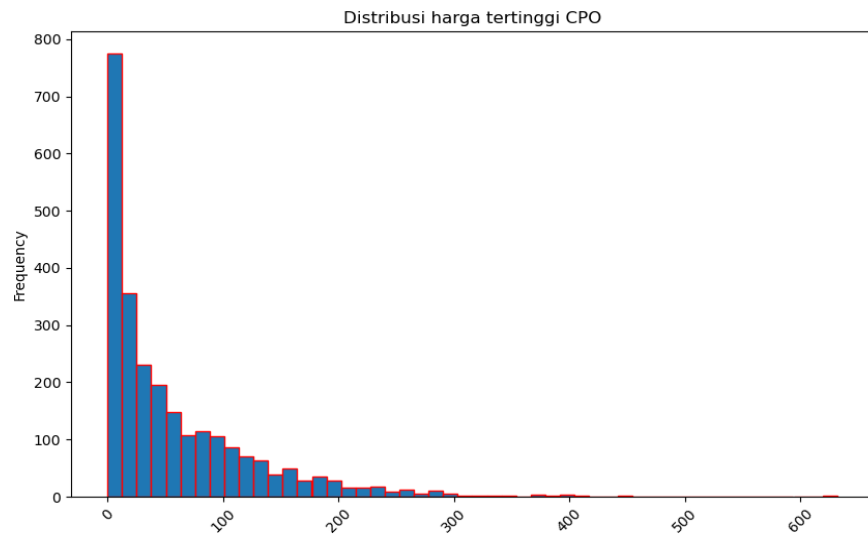
Gambar 4. 6 Harga Pembukaan Dengan Harga Penutupan

Harga pembukaan dan penutupan CPO biasanya sangat dekat dan stabil. Setiap titik mewakili satu data harga pembukaan dan harga penutupan. Titik-titik ini membentuk pola yang hampir seperti garis lurus miring ke atas, yang berarti harga pembukaan dan penutupan biasanya hampir sama. Namun, ada beberapa titik yang menyebar atau jauh dari garis utama, yang berarti harga penutupan berbeda cukup jauh dari harga pembukaan. Titik-titik yang menyebar ini menunjukkan adanya fluktuasi harga yang lebih besar, artinya harga CPO berubah lebih banyak pada beberapa data.

2. Tertinggi

Harga tertinggi atau kolom Tertinggi pada dataset mengacu pada harga CPO yang di perdagangan pada priode perdagangan tertentu.

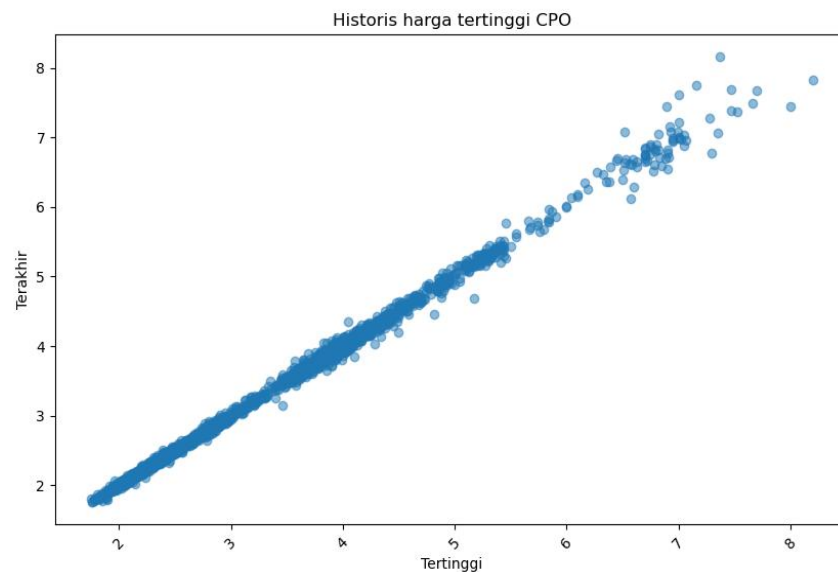
a) Distribusi harga tertinggi



Gambar 4. 7 Distribusi Harga Tertinggi

Gambar ini adalah grafik batang yang menunjukkan seberapa sering harga tertinggi CPO (Minyak Sawit Mentah) muncul dalam berbagai rentang harga. Batang yang lebih tinggi berarti harga tersebut lebih sering muncul. Dari grafik terlihat bahwa harga tertinggi CPO biasanya berada di angka yang rendah, dan semakin tinggi harganya, semakin jarang muncul.

b) Distribusi harga tertinggi dengan harga penutupan



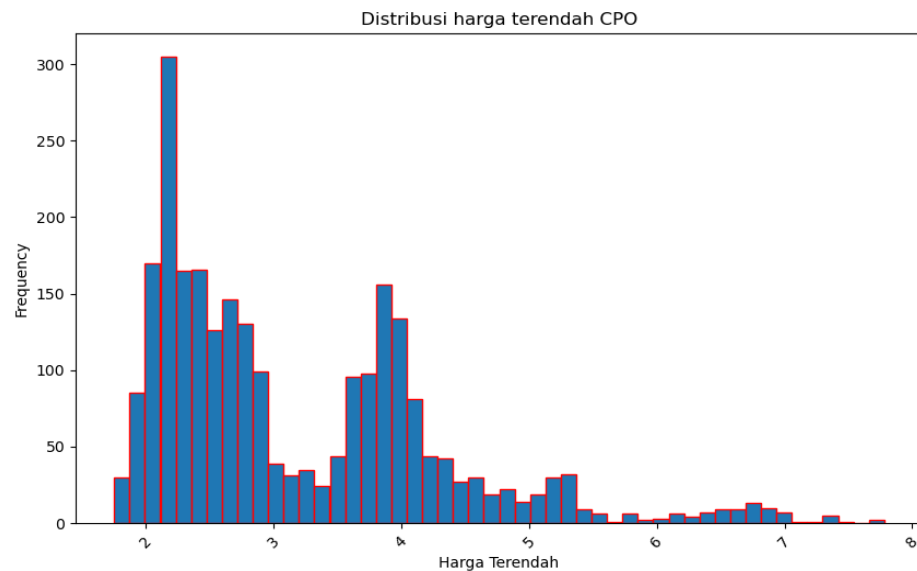
Gambar 4. 8 Distribusi Harga Tertinggi Dengan Harga Penutupan

Grafik ini menunjukkan hubungan antara harga tertinggi dan harga terakhir CPO (Crude Palm Oil) dari waktu ke waktu. – Sumbu horizontal (Tertinggi) menunjukkan harga tertinggi yang pernah dicapai. – Sumbu vertikal (Terakhir) menunjukkan harga terakhir yang tercatat. Titik-titik pada grafik sebagian besar membentuk garis lurus dari kiri bawah ke kanan atas, yang berarti ketika harga tertinggi naik, harga terakhir juga ikut naik. Namun, ada beberapa titik yang menyebar atau menyimpang dari garis lurus ini, menunjukkan bahwa pada beberapa waktu harga terakhir tidak selalu tepat sama dengan harga tertinggi, ada variasi kecil di antara keduanya. Jadi, meskipun harga tertinggi dan harga terakhir biasanya bergerak bersama, terkadang ada perbedaan kecil di antara keduanya.

3. Terendah

Harga terendah atau kolom Terendah pada dataset mengacu pada harga CPO yang di perdagangan pada priode tertentu.

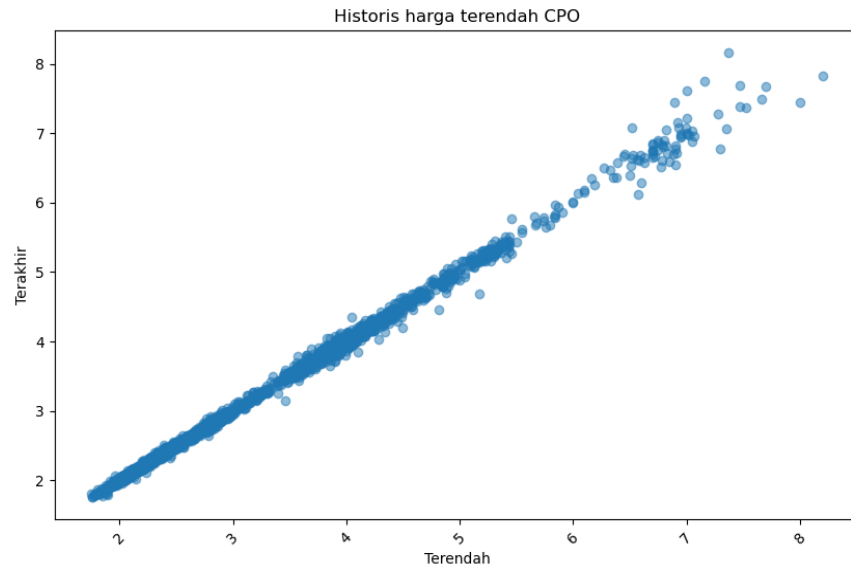
a) Distribusi harga terendah.



Gambar 4. 9 Distribusi Harga Terendah

Harga terendah CPO paling sering berada di kisaran 2.1 sampai 2.3 dengan frekuensi tertinggi sekitar 300 kali. Harga terendah paling rendah sekitar 1.8 dengan frekuensi sekitar 30, dan harga tertinggi hampir mencapai 7.8 tetapi sangat jarang terjadi dengan frekuensi hampir nol. Jadi, sebagian besar harga terendah CPO berkisar di angka rendah antara 1.8 sampai 4.0.

b) Distribusi harga teringgi dengan harga penutupan



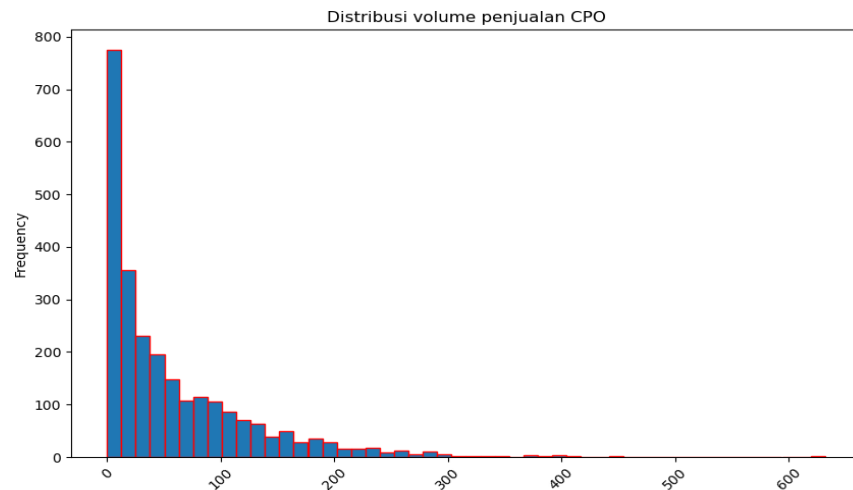
Gambar 4. 10 Distribusi Harga Terendah Dengan Harga Penutupan

Grafik titik yang menunjukkan hubungan antara harga terendah dan harga terakhir CPO (minyak sawit mentah). Setiap titik mewakili data harga pada waktu tertentu. Grafik ini memperlihatkan bahwa ketika harga terendah naik, harga terakhir juga biasanya naik, menunjukkan hubungan yang kuat dan positif antara kedua harga tersebut. Titik-titik pada grafik sebagian besar tersebar rapat membentuk pola diagonal dari kiri bawah ke kanan atas, yang menandakan konsistensi dan keterkaitan yang erat antara harga terendah dan harga terakhir. Namun, ada beberapa titik yang tersebar agak jauh dari pola utama, menunjukkan adanya variasi atau fluktuasi harga pada beberapa waktu tertentu.

4. *Vol.*

Pada atribut volume atau kolom *Vol.* pada dataset mengacu pada volume perdagangan CPO yang di perdagangan pada priode tertentu.

a) Distribusi *Vol.*

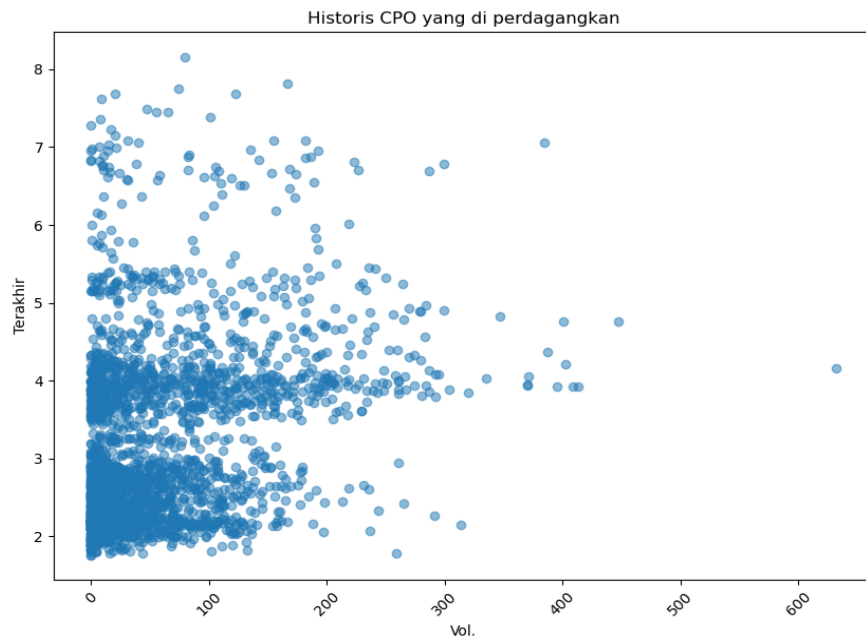


Gambar 4. 11 Distribusi *Vol.*

Grafik ini menunjukkan seberapa sering volume penjualan CPO (minyak sawit mentah) terjadi dalam berbagai jumlah. Volume penjualan sangat kecil (dekat 0) terjadi paling sering, dengan frekuensi sekitar 775 kali. – Volume sekitar 20 terjadi sekitar 350 kali. Volume sekitar 40 terjadi sekitar 230 kali. Volume sekitar 60 terjadi sekitar 150 kali. – Volume sekitar 80 terjadi sekitar 110 kali. Volume sekitar 100 terjadi sekitar 100 kali. Volume sekitar 150 terjadi sekitar 65 kali. Volume sekitar 200 terjadi sekitar 30 kali. Volume sekitar 250 terjadi sekitar 15 kali. Volume sekitar 300 terjadi sekitar 10 kali. Volume di atas 400 sangat jarang terjadi, frekuensinya kurang dari 5 kali. Jadi,

sebagian besar penjualan CPO terjadi dalam jumlah kecil hingga sedang, sedangkan penjualan dalam jumlah besar sangat jarang terjadi.

b) Distribusi *Vol.* dengan harga penutupan



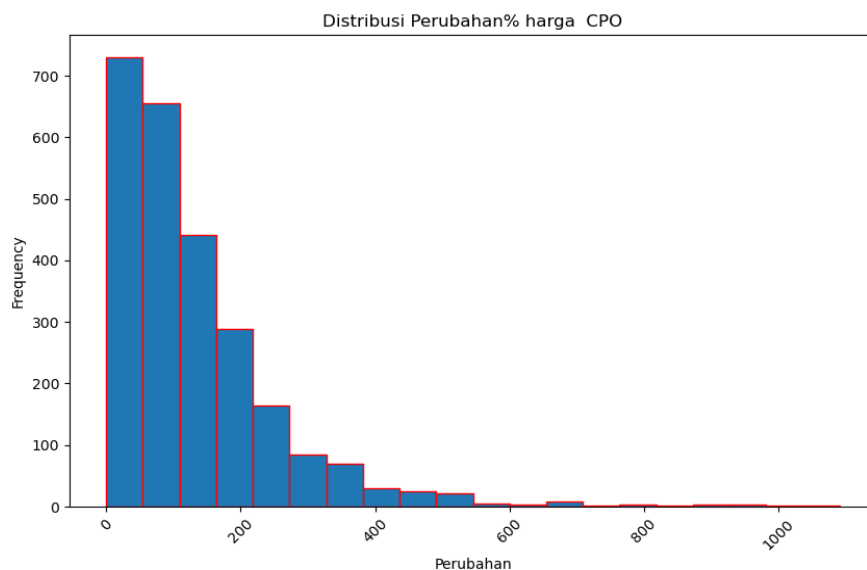
Gambar 4. 12 Distribusi Vol. Dengan Harga Penutupan

Sebagian besar transaksi CPO terjadi dengan volume di bawah 50, dengan harga terakhir yang bervariasi antara sekitar 2 hingga 7. Volume perdagangan yang sangat besar, mencapai lebih dari 600, sangat jarang terjadi dan tidak menunjukkan pola yang jelas terhadap harga terakhir, yang berkisar antara 2 hingga 8.

5. Perubahan%

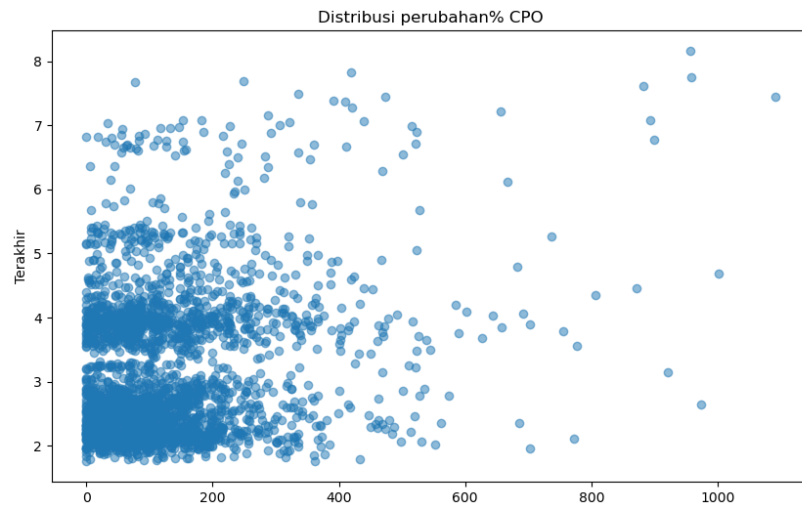
Pada atribut Perubahan atau *kolom* Perubahan% pada dataset mengacu pada perubahan harga perdagangan CPO yang di perdagangan pada periode tertentu.

a) Distribusi Perubahan%

**Gambar 4.13 Distribusi Perubahan%**

Grafik ini menunjukkan sebaran perubahan harga CPO. Sebagian besar perubahan harga berada di kisaran 0 hingga 200 dengan frekuensi tertinggi sekitar 730 kali. Perubahan antara 200 hingga 400 terjadi sekitar 150-300 kali, sedangkan perubahan di atas 400 sangat jarang, dengan frekuensi kurang dari 50 kali. Perubahan harga tertinggi mencapai sekitar 1000 dengan frekuensi sangat rendah, hampir nol. Ini menunjukkan harga CPO biasanya stabil dengan fluktuasi kecil, dan perubahan besar hanya terjadi sesekali.

b) Distribusi Perubahan% dengan harga penutupan.



Gambar 4. 14 Distribusi Perubahan% Dengan Harga Penutupan

Gambar ini menunjukkan sebaran perubahan persentase CPO dan nilai terakhirnya dalam bentuk titik-titik biru. Sebagian besar titik terkumpul pada perubahan persentase antara 0 hingga sekitar 400, dengan nilai terakhir berkisar antara 2 sampai 5. Ada beberapa titik yang menunjukkan perubahan sangat besar, hingga lebih dari 1000, dan nilai terakhir mencapai sekitar 8.

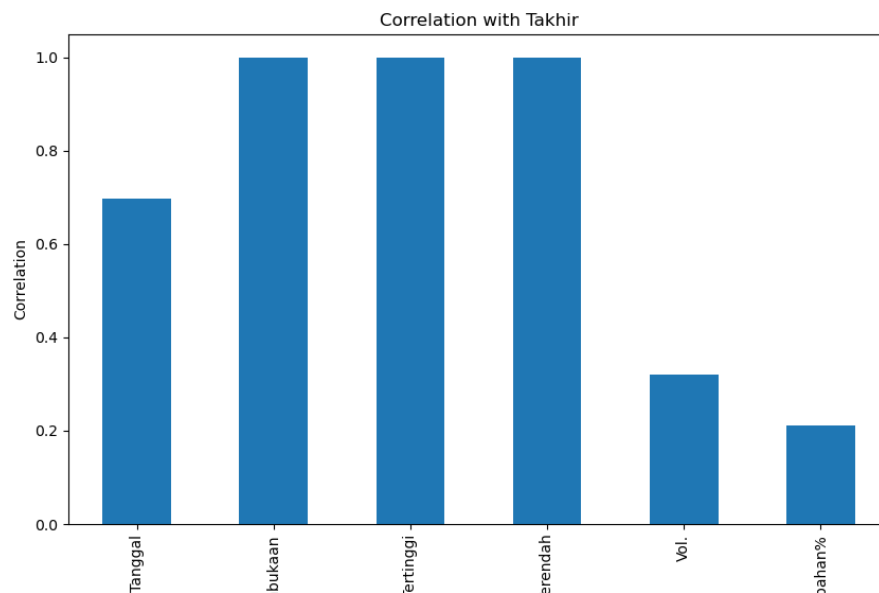
4.2.6 Korelasi Atribut (Input) Terhadap Label (Output)

Korelasi atribut (*input*) dan label (*output*) dalam analisis data bertujuan untuk mengetahui seberapa kuat hubungan antara variabel independen (*input*) dan variabel dependen (*output*). Dalam konteks peramalan harga CPO menggunakan algoritma CART, korelasi ini membantu menentukan atribut mana yang memiliki dampak terbesar terhadap harga CPO.

Tabel 4. 16 Korelasi Atribut (*Input*) Dan Label (*Output*)

<i>Input</i>	Korelasi	<i>Output</i>
Tanggal	0.696594	Terakhir
Pembukaan	0.998489	Terakhir
Tertinggi	0.999191	Terakhir
Terendah	0.999264	Terakhir
Vol.	0.319968	Terakhir
Perubahan %	0.212197	Terakhir

Berdasarkan tabel diatas dapat dijelaskan terdapat korelasi hubungan antara atribut dengan label. Dari data tabel diatas maka dapat dilakukan visualisasi dalam bentuk grafik batang, sebagai berikut.

**Gambar 4. 15 Kolerasi Atibut (*Input*) Dan *Label* (*Output*)**

Dalam grafik ini, semua nilai korelasi bernilai positif, yang berarti semakin mendekati 1, semakin kuat hubungan tersebut. Jika angka korelasi mendekati 1 maka semakin berpengaruh terhadap harga penutupan tersebut. Nilai korelasi pada

grafik menunjukkan bahwa atribut Tanggal, Pembukaan, Tertinggi, dan Terendah memiliki korelasi positif yang mendekati nilai 1, yang berarti hubungan sangat kuat. Atribut *Vol.* dan *Perubahan%* memiliki korelasi positif yang lemah karena nilai kolerasinya rendah.

4.2.7 Pembersihan Data

Pembersihan data merupakan proses penting dalam analisis data yang bertujuan untuk mengidentifikasi dan memperbaiki atau menghapus data yang tidak akurat, tidak lengkap, atau tidak konsisten. Pada tahap pemberisihan data ini dilakukan penghapusan data dengan nilai yang kosong atau yang memiliki data duplikat ataupun data dengan nilai *outlier*.

a. *Missing Value*

Missing Value atau nilai yang hilang merujuk pada data yang seharusnya ada tetapi tidak terisi dalam dataset. Kehadiran nilai yang hilang dapat menjadi tantangan dalam analisis data dan pengembangan model, karena dapat mempengaruhi hasil analisis dan kinerja model. Setelah mengidentifikasi terdapat *Missing Value* data sebanyak 149 pada kolom *Vol.* sebagai berikut.

Tabel 4. 17 Missing Value

Kolom	Nilai <i>Missing value</i>
Tanggal	0
Terakhir	0
Pembukaan	0
Tertinggi	0
Terendah	0
<i>Vol.</i>	149
<i>Perubahan%</i>	0

b. Data Duplikat

Data duplikat merujuk pada entri yang sama dalam dataset yang dapat mengganggu analisis dan mempengaruhi kinerja model. Menghapus data duplikat penting untuk memastikan keakuratan dan konsistensi hasil analisis serta meningkatkan efektivitas model yang dibangun. Setelah diidentifikasi, entri yang duplikat harus dihapus untuk memastikan dataset bersih.

Tabel 4. 18 Data Duplikat

Kolom	Data Duplikat
Tanggal	0
Terakhir	0
Pembukaan	0
Tertinggi	0
Terendah	0
Vol.	0
Perubahan%	0

c. Data *Outlier*

Data *Outlier* adalah nilai atau entri dalam dataset yang secara signifikan berbeda dari nilai-nilai lainnya. *Outlier* dapat muncul karena berbagai alasan, termasuk kesalahan pengukuran, variasi alami dalam data, atau fenomena yang tidak biasa. Dalam konteks machine learning dan analisis data, outlier dapat mempengaruhi hasil analisis dan kinerja model.. Pada dataset ini tidak memiliki *Outlier*.

Tabel 4. 19 Data Outlier

Kolom	Data Outlier
Tanggal	0
Terakhir	0
Pembukaan	0
Tertinggi	0
Terendah	0
Vol.	0
Perubahan%	0

Setelah diketahui data yang memiliki *missing value*, data duplikat, dan data *outlier* maka data tersebut akan dibersihkan dengan menghapus data *outlier* dan mengubah baris yang kosong menjadi nilai 0 pada baris *missing value* dari dataset tersebut.

4.2.8 Encoding (Feature Engineering)

Encoding merupakan proses konversi tipe data untuk pembuatan *machine learning*, sehingga dapat menganalisis lebih lanjut dan pengambilan keputusan yang akurat, seperti pada tabel berikut :

Tabel 4. 20 Encoding (Feature Engineering)

Kolom	Tipe Data Awal	Tipe Data Konversi
Tanggal	<i>Object</i>	<i>Datetime</i>
Terakhir	<i>Object</i>	<i>Float</i>
Pembukaan	<i>Object</i>	<i>Float</i>
Tertinggi	<i>Object</i>	<i>Float</i>
Terendah	<i>Object</i>	<i>Float</i>
Vol.	<i>Object</i>	<i>Float</i>
Perubahan%	<i>Object</i>	<i>Float</i>

Dari tabel diatas dapat dilihat terjadi konversi dalam tipe data, dataset tersebut yang awalnya terdapat tipe data *object* di konversi menjadi *Float*, dan kolom

Tanggal menjadi *Datetime* untuk pembuatan model *machine learning* yang bersifat *numerik*.

4.2.9 Membangun Model

Membangun model mesin learning merupakan sebuah langkah untuk membuat prediksi, dataset yang saya gunakan berjangka 10 tahun sebagai berikut.

Tabel 4. 21 Membangun Model

No	Dataset yang Digunakan	Dimensi
1	Data Historis Minyak Sawit Berjangka (FCPOc1)	2541 Baris Data & 11 Kolom

a. *Splitting* Data

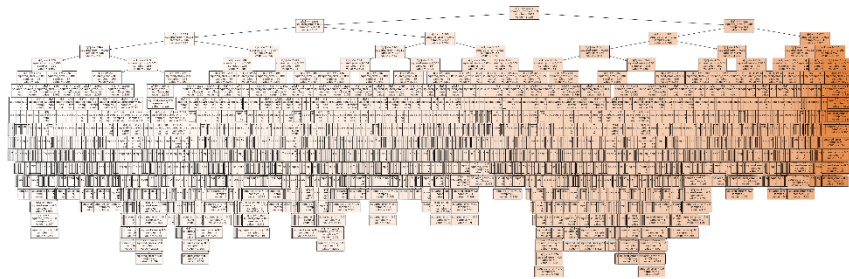
Data *Splitting* atau pemisahan data adalah metode membagi data menjadi dua bagian atau lebih yang membentuk sub himpunan data. Umumnya *splitting* data yaitu membagi data menjadi 2 bagian, data historis yaitu menguji coba data historis 10 tahun kebelakang dan data uji untuk menguji data latih selama setahun kebelakang.

Tabel 4. 22 *Splitting* Data

Dataset	Jumlah
Data Historis	1779 Baris Data & 7 Kolom
Data Uji	762 Baris Data & 7 Kolom

b. *Training Model*

Training Model atau Pelatihan model adalah proses memasukkan data rekayasa ke algoritma pembelajaran mesin *learning* untuk menghasilkan model dengan parameter optimal yang dapat dilatih dan di pelajari . Pada penelitian saya ini menggunakan model *Decision Tree* untuk menganalisis dan mengambil Keputusan berdasarkan data yang ada, model *decision tree* adalah metode pengambilan Keputusan yang menggambarkan alur logika dalam bentuk pohon. Seperti gambar berikut ini gambar pohon keputusan :



Gambar 4. 16 Training Model Decision Tree

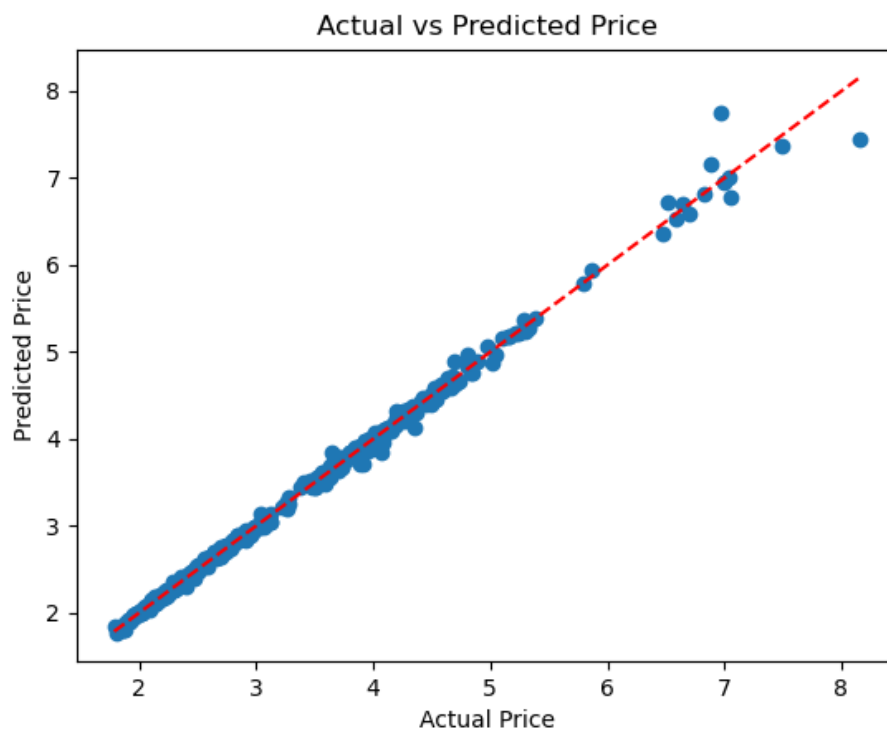
4.2.10 Pengujian Model

Pengujian model ini dilakukan untuk memastikan bahwa model dapat memberikan prediksi yang akurat dan dapat diandalkan dalam mengklasifikasi dengan data test. Dalam pengujian ini, beberapa metrik evaluasi seperti *Mean Squared Error* (MSE), *Mean Absolute Error* (MAE), *R-Square* digunakan untuk mengukur sejumlah model dapat mengklasifikasikan data dengan benar.

Tabel 4. 23 Pengujian Model

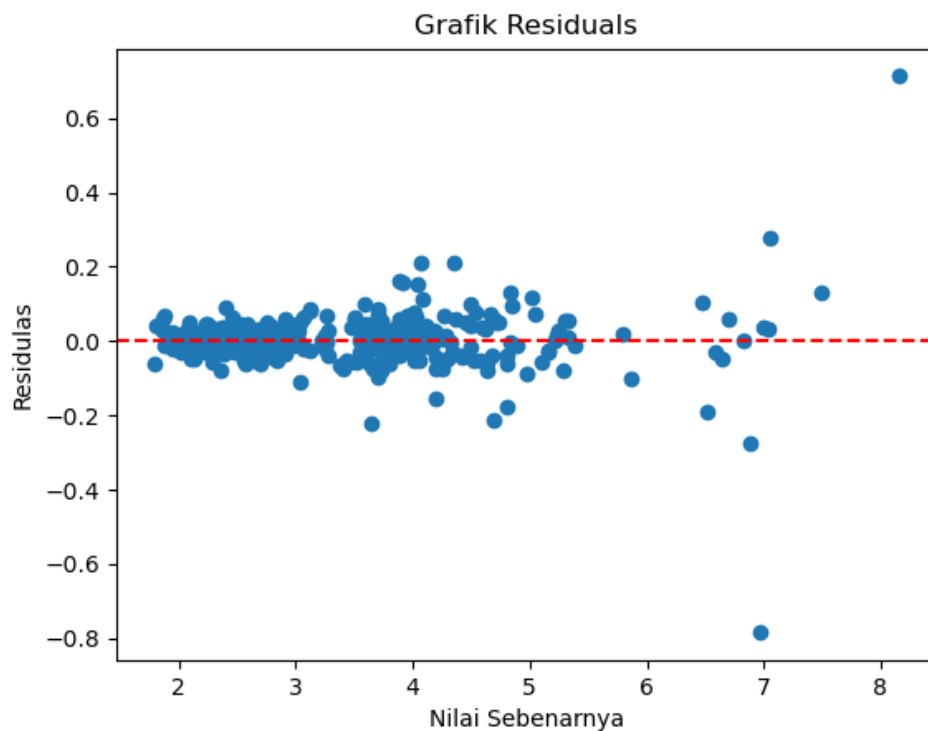
Model	Hasil
<i>Mean Squared Error (MSE)</i>	0.004
<i>Mean Absolute Error (MAE)</i>	0.032
<i>R-Square</i>	0.996

Berdasarkan tabel diatas pengujian model dari pembuatan model tersebut mendapatkan nilai akurasi *R-Square* 0.996, nilai MSE sebesar 0.032, dan nilai MAE sebesar 0.004. Hasil pengujian model mesin learning dapat dilihat dalam bentuk grafik atau *kurva* dibawah ini.

**Gambar 4. 17 Hasil Pengujian Model *Machine Learning***

Berdasarkan grafik perbandingan antara harga aktual dan harga prediksi, terlihat bahwa sebagian besar titik data berada sangat dekat dengan garis merah. Hal ini menunjukkan bahwa model memiliki tingkat akurasi yang tinggi, karena prediksi harga yang dihasilkan hampir sama dengan nilai sebenarnya. Namun,

terdapat beberapa penyimpangan kecil pada nilai harga yang lebih tinggi, di mana titik-titik sedikit menjauhi garis. Meskipun demikian, secara keseluruhan model telah menunjukkan performa yang sangat baik dan mampu memprediksi dengan akurasi yang tinggi pada sebagian besar data. Sedangkan grafik *Residual* dari dataset tersebut sebagai berikut :



Gambar 4. 18 Grafik Residual Model *Machine Learning*

Grafik residual menunjukkan bahwa sebagian besar titik residual berada cukup dekat dengan garis nol, khususnya pada rentang nilai sebenarnya antara 2 hingga 5. Hal ini mengindikasikan bahwa model memiliki performa prediksi yang baik dan kesalahan prediksi relatif kecil pada rentang nilai tersebut. Grafik *residual* tersebut cukup baik karena semakin banyak persebaran data yang mendekati nilai 0 maka data tersebut semakin baik dengan mengikuti garis *linieritas*.

4.3 Evaluasi dan Validasi Hasil

Pada tahap ini dilakukan prosen pengujian model mesin *learning* yang telah dibuat dengan menggunakan pola data atau dataset uji yang telah disiapkan mulai dari bulan januari – Mei 2024 sebelumnya, adapun dataset yang digunakan dapat dilihat pada tabel berikut.

Tabel 4. 24 Evaluasi Data

No	Dataset yang Digunakan	Dimensi
1	Data Historis Minyak Sawit Berjangka (FCPOc1)	69 Baris Data & 7 Kolom

Pada tabel diatas merupakan dataset yang akan digunakan untuk pengujian evaluasi dan validasi hasil, Dimana data yang diperoleh dari data uji sebanyak 208 data dengan 7 kolom.

```

rmse= mean_squared_error(y_uji, y_pred2, squared=False)
print("Root Mean Squared Error:", rmse)

mse = mean_squared_error(y_uji, y_pred2)
print('Mean Squared Error:', mse)

mae = mean_absolute_error(y_uji, y_pred2)
print('Mean Absolute Error:', mae)

r2 = r2_score(y_uji, y_pred2)
print('R-Square:', r2)

r2=r2 *100
print('Accuracy:',r2,'%')

```

Root Mean Squared Error: 0.06209598831410828
 Mean Squared Error: 0.0038559117647058716
 Mean Absolute Error: 0.04858823529411761
 R-Square: 0.9310121400582015
 Accuracy: 93.10121400582015 %

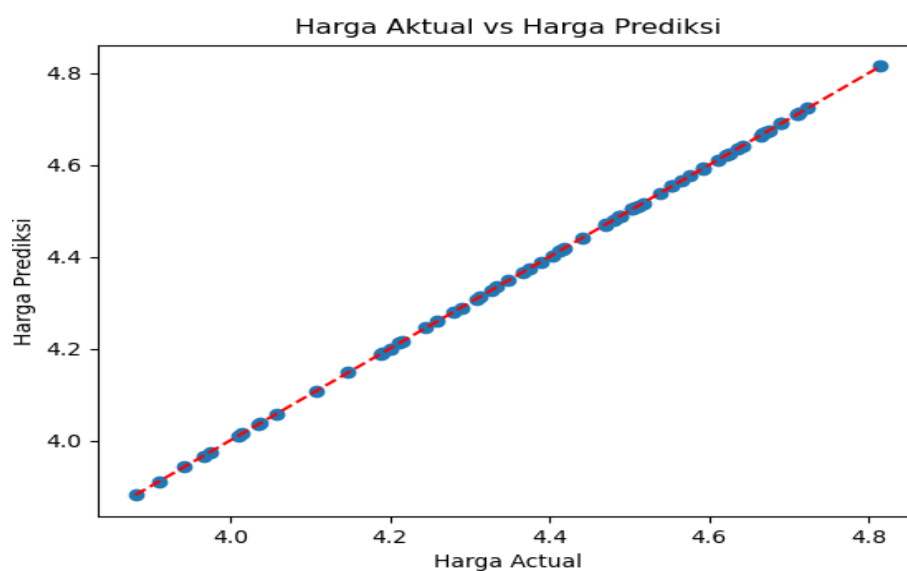
Gambar 4. 19 Validasi Hasil

Gambar diatas merupakan perhitungan dari pengujian model dan validasi hasil menggunakan data uji yang digunakan untuk menghitung parameter model uji dalam penelitian ini dan didapat kan hasil sebagai berikut :

Tabel 4. 25 Validasi Hasil

Model	Hasil
<i>Root Mean Squared Errorr (RMSE)</i>	0.062
<i>Mean Squared Errorr (MSE)</i>	0.003
<i>Mean Absolute Errorr (MAE)</i>	0.046
<i>R-Square</i>	0.937

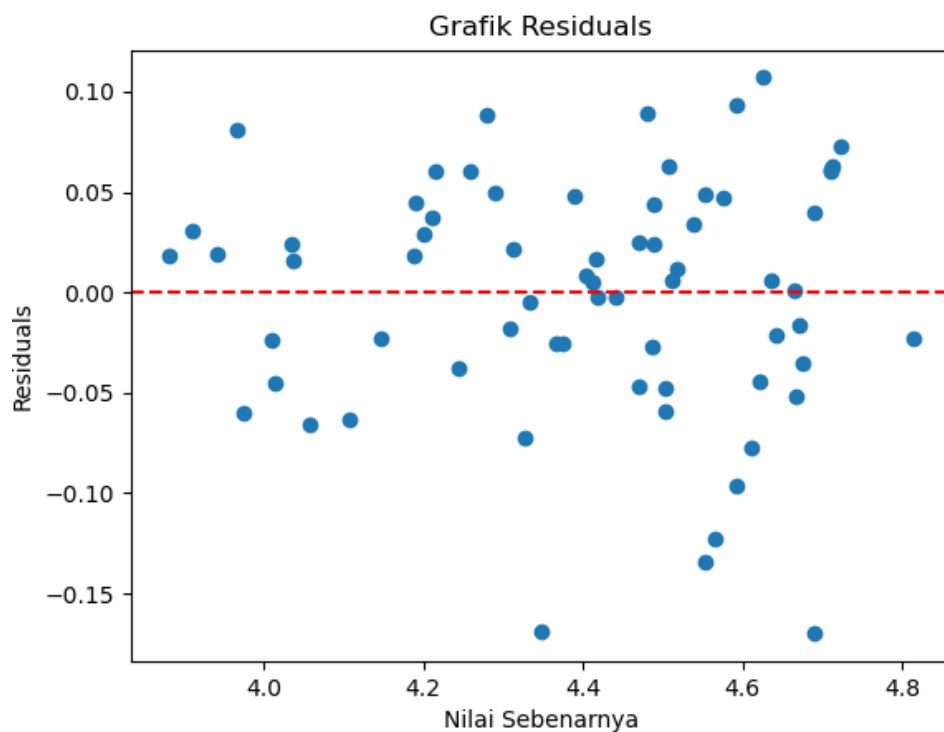
Hasil dari model memiliki tingkat akurasi yang dengan nilai *R-Square* 0.937, nilai RMSE sebesar 0.062, nilai MSE sebesar 0.003, dan nilai MAE sebesar 0.046, Sedangkan grafik harga aktual dengan harga prediksi sebagai berikut :



Gambar 4. 20 Hasil Aktual Dengan Prediksi

Pada gambar diatas mmenunjukkan bahwa semakin dekat titik data dengan garis linier, maka semakin baik performa model dalam memprediksi harga CPO.

Titik-titik biru yang merepresentasikan harga aktual vs harga prediksi tersebar sangat rapat di sekitar garis merah putus-putus, yang menunjukkan bahwa model prediksi memiliki akurasi yang sangat tinggi. Sedangkan grafik residual dari data uji tersebut sebagai berikut :



Gambar 4. 21 Grafik Residual Data Uji

Berdasarkan grafik residual yang dihasilkan, terlihat bahwa sebagian besar titik residual tersebar secara acak dan berada sangat dekat dengan garis nol. Jika model bagus, titik-titik residual akan menyebar secara acak di sekitar nol. Jika membentuk pola tertentu, berarti model belum tepat. Hal ini menunjukkan bahwa selisih antara nilai prediksi dan nilai aktual relatif kecil, yang mengindikasikan bahwa model memiliki tingkat akurasi yang tinggi.

