

PREDIKSI HARGA MINYAK MENTAH DENGAN MENGGUNAKAN ALGORITMA CART (*CLASSIFICATION AND REGRESSION TREE*)

Muhammad Reynaldi¹, Santi Rahayu², Teti Desyani³

¹Teknik Informatika, Universitas Pamulang, Jl. Raya Puspitek, Buaran, Kec. Pamulang, Kota Tangerang Selatan, Banten 15310
e-mail: ¹reynaldi.rr80@gmail.com

Abstract

Crude oil is an important energy source that experiences significant price fluctuations due to various factors such as geopolitics and global market dynamics. Research conducted by Muhammad Reynaldi in 2024 using the title “Crude Oil Price Prediction Using the CART (Classification and Regression Tree) Algorithm”, aims to disseminate examples of crude oil price predictions carefully and easily interpreted. The problem faced in this study is the difficulty of predicting crude oil prices due to fluctuations determined by many external factors. The purpose of this study is to create a more accurate prediction method with traditional methods, and to assist decision making in the energy sector. The results of the study show that the CART solving procedure for predicting crude oil prices with an MSE value of 0.892, an MAE value of 0.700, an RSME value of 0.944, and an R-square value of 0.957. This study places significance on overcoming the challenges of crude oil price fluctuations and provides simple solutions for governments, companies, and energy industry players.

Keywords: Crude oil, CART algorithm, price prediction, decision tree, machine learning.

Abstrak

Minyak mentah adalah asal tenaga penting yg mengalami fluktuasi harga secara signifikan dampak banyak sekali faktor misalnya geopolitik & dinamika pasar global. Penelitian yang dilakukan Muhammad Reynaldi dalam tahun 2024 menggunakan judul “Prediksi Harga Minyak Mentah Menggunakan Algoritma CART (Classification and Regression Tree)“, bertujuan buat menyebarkan contoh prediksi harga minyak mentah dengan seksama & gampang diinterpretasikan. Masalah yang dihadapi pada penelitian ini merupakan sulitnya memprediksi harga minyak mentah dampak fluktuasi yang ditentukan banyak sekali faktor eksternal. Tujuan penelitian ini merupakan membuat metode prediksi yang lebih seksama dengan metode tradisional, dan membantu pengambilan keputusan pada sektor energi. Hasil penelitian memperlihatkan bahwa prosedur pemecahan CART untuk memprediksi harga minyak mentah dengan nilai MSE 0.892, nilai MAE 0.700, nilai RSME 0.944, dan nilai *R-square* 0.957. Penelitian ini menaruh signifikansi pada mengatasi tantangan fluktuasi harga minyak mentah & memperlihatkan solusi simpel bagi pemerintah, perusahaan, dan pelaku industri tenaga.

Kata Kunci: Minyak mentah, algoritma CART, prediksi harga, pohon keputusan, pembelajaran mesin.

1. PENDAHULUAN

Minyak mentah merupakan campuran kompleks *hidrokarbon* yang terbentuk secara alami di bumi. Minyak mentah dapat dipisahkan menjadi beberapa fraksi melalui penyulingan untuk memperoleh turunan yang paling bermanfaat bagi

masyarakat, seperti bahan bakar dan *petrokimia* (Silva, 2023). Perkiraan harga minyak sangat membantu manajemen untuk mengambil keputusan operasional yang lebih baik, mengurangi biaya operasional, meningkatkan keuntungan dan meningkatkan keunggulan kompetitif

(Bollapragada, Mankude, & Udayabhanu, 2021). Minyak mentah adalah salah satu sumber energi terpenting di bumi, Sejauh ini, Indonesia tetap menjadi bahan bakar terkemuka di dunia, dengan hampir sepertiga dari konsumsi energi global (Gao & Lei, 2019), banyak dampak yang dapat disebabkan oleh harga minyak mentah, sehingga pemerintah harus mampu mengantisipasi dan mempersiapkan hal tersebut.

Salah satu cara untuk mengatasi masalah tersebut adalah dengan melakukan peramalan menggunakan analisis deret waktu. Analisis deret waktu melibatkan pengamatan berdasarkan urutan waktu yang berkorelasi antara pengamatan yang berdekatan. Dalam deret waktu, setiap observasi dari variabel terkait dengan variabel itu sendiri pada waktu sebelumnya. (Rahma, et al., 2024). Harga minyak mentah dunia terus berubah. Perubahan ini muncul karena konflik geopolitik Rusia-Ukraina pada tahun 2022. Konflik tersebut menyebabkan harga minyak mentah dunia naik di atas *US\$100* per-barel sejak 2014 (Aliffia, Sediono, Suliyanto, Mardianto, & Amelia, 2023). Penurunan harga minyak mentah dunia di paruh kedua tahun 2022 hingga paruh pertama tahun 2023, membuat beberapa pihak memprediksikan harga minyak mentah *Brent* berada di bawah *US\$80* per-barel. Namun, pada paruh kedua tahun 2023 proyeksi tersebut berbalik, harga minyak mentah dunia tahun 2024 diprediksi mengalami kenaikan (Paramita, Chasmir, Fakhreza, Lestari, & Ramiayu, 2023).

Dalam memprediksi harga minyak mentah ada beberapa metode yang bisa dipilih yaitu menggunakan metode pembelajaran mesin seperti *Decision Tree* (Chen & He, 2019), *ARIMA* (Lestari, Sumargo, & Ladayya, 2022), *Artificial Neural Network* (Rosiani, et al., 2023), *Support Vector Regression* (Redha & Yousif, 2023), Metode-metode ini melatih dan menguji data kemudian dinilai dengan metrik RMSE, MAE, dan R^2 .

Metode *ARIMA (Autoregressive Integrated Moving Average)*, yaitu suatu metode yang didasarkan pada nilai-nilai suatu peubah yang telah terjadi pada waktu lampau, lalu digunakan untuk menentukan pola historis data, dan kemudian digunakan untuk mengekstrapolasikan pola tersebut pada masa yang akan datang (Wulandari & Gernowo, 2019). Tetapi kelemahan dari model

ARIMA bisa saja gagal menangkap pola kompleks yang timbul dari spekulasi pasar dan peristiwa geopolitik (Gasper & Mbwambo, 2023).

Artificial Neural Network (ANN) atau Jaringan Saraf Tiruan merupakan salah satu pemodelan kompleks yang dapat memprediksi bagaimana ekosistem merespon perubahan variabel lingkungan dengan terinspirasi oleh cara kerja sistem saraf biologis, khususnya pada sel otak manusia dalam memproses informasi (Anisa, 2022). Tetapi kelemahan *Artificial Neural Network* yaitu Hasil dari model *Artificial Neural Network* seringkali sulit untuk diinterpretasikan, sehingga membuatnya kurang transparan dibandingkan metode tradisional (Veri, Surmayati, & Guslendra, 2022).

SVM (Support Vector Regression) adalah sebuah metode yang digunakan untuk kasus klasifikasi, namun prinsip metode tersebut dapat dikembangkan ke dalam regresi dan metode peramalan *Time Series* (Isnaeni, Sudarmin, & Rais, 2022). Kelemahan dari SVR yaitu Menentukan parameter seperti nilai C dan epsilon dapat menjadi tantangan, dan sering memerlukan teknik pencarian seperti Grid Search untuk mendapatkan hasil yang optimal (Suryani & Fadhillah, 2024).

Decision Tree merupakan salah satu cara data *processing* dalam memprediksi masa depan dengan cara membangun klasifikasi atau regresi model dalam bentuk struktur pohon (Ramadhan, 2022). Pohon keputusan memiliki keuntungan karena mudah dipahami, dan biasanya berkinerja baik dengan data pelatihan. Akan tetapi, kelemahan utama mereka yang dimaksudkan untuk dipahami sebagai *overfitting* adalah bahwa mereka memiliki kemampuan terbatas untuk generalisasi dan berkinerja buruk saat menangani data di luar sampel (Sofianos, Zaganidis, Papadimitriou, & Gogas, 2024).

Metode yang digunakan dalam penelitian ini adalah Pohon Keputusan menunjukkan kinerja metode *Classification and Regression Tree (CART)* dalam prediksi harga minyak mentah. Khususnya dalam lingkungan perdagangan waktu nyata dengan sumber daya komputasi terbatas. Dalam penelitian ini, data historis harga minyak diolah menggunakan *CART* untuk menghasilkan model peramalan yang dapat memprediksi harga minyak di masa depan dengan akurasi tinggi dan mudah dipahami.

2. PENELITIAN YANG TERKAIT

Berdasarkan penelitian yang dilakukan oleh Engu Chen, Xin James He yang berjudul *Crude Oil Price Prediction with Decision Tree Based Regression Approach*, metode yang digunakan dalam penelitian ini yaitu menggunakan Classification and Regression Tree (CART), Random Forest, ARIMA. Sedangkan masalah yang dihadapi peneliti adalah kesulitan dalam Pemilihan metode terbaik dalam memprediksi harga minyak mentah. Penelitian ini bertujuan untuk memprediksi harga minyak mentah menggunakan berbagai metode pohon keputusan. Hasil penelitian menunjukkan bahwa Classification and Regression Tree (CART) dengan MAE 4,8093, RMSE 6,9553, R^2 0,9688 sedangkan ARIMA dengan MAE 2,9444, RMSE 3,9452, R^2 0,9799 sedangkan Random Forest terbukti sebagai model paling akurat, dengan MAE 1,2481 dan RMSE 1,8613, meskipun sedikit kalah dalam nilai R^2 0,9966 (Chen & He, 2019).

Berdasarkan penelitian yang dilakukan oleh Indah Lestari, Bagus Sumargo, dan Faroh Ladayya yang berjudul "Pemodelan ARIMA Intervensi untuk Meramalkan Harga Minyak Mentah Dunia", metode yang digunakan dalam penelitian ini ARIMA. Sedangkan masalah yang dihadapi adalah ketidakpastian harga minyak yang disebabkan oleh fluktuasi permintaan akibat lockdown dan pembatasan mobilitas juga Penurunan harga minyak mentah dunia yang signifikan akibat intervensi dari pandemi COVID-19. Adapun tujuan dari penelitian ini adalah Memberikan informasi mengenai fluktuasi harga minyak mentah yang dipengaruhi oleh intervensi (pandemi COVID-19). Hasil penelitian model intervensi terbaik yang ditemukan adalah ARIMA nilai model intervensi (0,2,2), MAPE sebesar 9,29% (Lestari, Sumargo, & Ladayya, 2022).

Berdasarkan penelitian yang dilakukan oleh Donghua Wang dan Tianhui Fang yang berjudul *Forecasting Crude Oil Prices with a WT-FNN Model*, metode yang digunakan dalam penelitian ini Feedforward Neural Network (FNN), dan pengembangannya yaitu Weight-Feed-Forward Neural Network (WT-FNN). Sedangkan masalah yang dihadapi adalah peneliti Metode pembelajaran mesin, termasuk FNN, rentan terhadap overfitting, yang mengakibatkan model tidak dapat generalisasi dengan baik pada data

baru. Penelitian ini bertujuan agar dapat memberikan solusi yang lebih baik untuk prediksi harga minyak mentah dengan mempertimbangkan karakteristik data yang dinamis dan relevan. Hasil penelitian menunjukkan akurasi prediksi FNN menunjukkan akurasi yang lebih rendah dibandingkan dengan WT-FNN, dimana FNN memiliki nilai MAE 2.7134 RMSE 3.6593 sedangkan WT-FNN memiliki nilai yang lebih baik yaitu MAE 2.7052 RMSE 3.6236 (Wang & Fang, 2022).

Berdasarkan penelitian yang dilakukan oleh Redha A. Redha, Suhad A. Yousif pada tahun 2023 yang berjudul *Crude Oil Price Forecasts Using Support Vector Regression and Technical Indicators*, metode yang digunakan dalam penelitian ini Support Vector Regression. Sedangkan masalah yang dihadapi peneliti adalah prediksi harga minyak sangat dipengaruhi oleh faktor eksternal yang sulit diprediksi, seperti pasokan, permintaan, dan peristiwa politik. Adapun tujuan dari penelitian ini adalah untuk mengembangkan metode yang lebih baik dan lebih akurat dalam memprediksi harga minyak mentah, yang penting bagi stabilitas ekonomi dan perencanaan strategis. Hasil penelitian penggunaan SVR dengan indikator teknis dapat menghasilkan prakiraan harga minyak mentah yang lebih akurat dengan RMSE 1.5456, MAE 1.3219, MAPE 1.9173 (Redha & Yousif, 2023).

Berdasarkan penelitian yang dilakukan oleh Diyah Rosiani, Muhamad Gibril Walay, Pradini Rahalintar, Arya Dwi Candra, Akhmad Sofyan, Yesaya Arison Haratua yang berjudul *Application of Artificial Intelligence in Predicting Oil Production Based on Water Injection Rate*, metode yang digunakan dalam penelitian ini Multiple Linear Regression, Polynomial Regression, Artificial Neural Networks. Sedangkan masalah yang dihadapi adalah metode simulasi reservoir tradisional memerlukan data yang komprehensif, waktu yang lama, dan biaya tinggi. Adapun tujuan dari penelitian ini adalah untuk menerapkan teknik AI dalam memprediksi laju produksi minyak berdasarkan laju injeksi air dari dua sumur injeksi. Hasil penelitian Model ANN menunjukkan performa terbaik dengan nilai Root Mean Square Error (RMSE) terendah sebesar 0.142 dan nilai R^2 tertinggi sebesar 16.2% (Rosiani, et al., 2023).

Berdasarkan penelitian yang dilakukan oleh Hualing Lin, dan Qiubi Sun yang berjudul *Crude Oil Prices Forecasting: An Approach of Using CEEMDAN-Based Multi-Layer Gated Recurrent Unit Networks*, metode yang digunakan dalam penelitian ini yaitu menggunakan Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN). Sedangkan masalah yang dihadapi peneliti adalah Data harga minyak mentah sering kali menunjukkan karakteristik non-linier dan non-stasioner, yang menyulitkan model tradisional untuk membuat prediksi yang akurat. Penelitian ini bertujuan untuk memberikan kontribusi yang signifikan terhadap pemahaman dan prediksi harga minyak mentah, yang sangat penting bagi stabilitas ekonomi global. Hasil penelitian ini berhasil menunjukkan bahwa kombinasi metode CEEMDAN dapat meningkatkan akurasi prediksi harga minyak mentah, yang penting untuk stabilitas ekonomi dan keamanan finansial dengan nilai RMSE 0.9276, MAE 0.9134, MAPE 0.0094 (Lin & Sun, 2020).

3. METODE PENELITIAN

3.1 Analisa Kebutuhan

Dalam penelitian ini pengujian data akan menggunakan aplikasi *Anaconda* yang support dengan *Jupyter notebook* dengan bahasa pemrograman yaitu *Python* dengan tujuan yaitu dapat meningkatkan hasil prediksi harga minyak mentah dengan menggunakan metode *decision tree*.

Untuk menunjang penelitian ini dibutuhkan spesifikasi kebutuhan tabel spesifikasi kebutuhan dibawah ini :

Tabel 3. 1 Analisa Kebutuhan

Perangkat Keras	Spesifikasi
Processor	Core i3 atau setara
Hard Disk Drive	Minimum 240 Gb
Memory	Minimum 8 Gb

Tabel 3. 2 Kebutuhan Perangkat Lunak

Perangkat Lunak	Spesifikasi
Sistem Operasi	Windows 10 atau lebih tinggi

Aplikasi Simulator	<i>Pemrograman Python 3.x</i>
Aplikasi Pengolah Kata	<i>Minimum Microsoft Excel 2013</i>
<i>Library Python</i>	- <i>Pandas</i> - <i>Numpy</i> - <i>Matplotlib</i> - <i>Scikit-Lern</i>

3.2 Teknik Analisis

3.2.1 Pengumpulan Data (Dataset)

Pada Penelitian ini data yang digunakan untuk melakukan penelitian adalah dataset Data historis Minyak Mentah WTI yang berasal dari *website Investing.com* dengan Alamat url: (<https://www.investing.com/commodities/crude-oil-historical-data>) yang memiliki atribut dengan keterangan sebagai berikut :

Tabel 3. 3 Pengumpulan Data

Input/Atribute	Keterangan
<i>Date</i>	Tanggal data harga minyak mentah WTI (<i>West Texas Intermediate</i>) diambil dan tercatat.
<i>Price</i>	Harga penutupan indeks pada hari perdagangan tersebut.
<i>Open</i>	Harga pembukaan indeks pada hari perdagangan tersebut.
<i>High</i>	Harga tertinggi indeks pada hari perdagangan tersebut.
<i>Low</i>	Harga terendah indeks pada hari perdagangan tersebut.
<i>Vol.</i>	Jumlah minyak mentah yang di perdagangan pada hari perdagangan tersebut.
<i>Change %</i>	Persentase perubahan harga indeks dibandingkan dengan hari perdagangan sebelumnya.

3.2.2 Pengolahan Data Awal (Data Training)

Pengelolaan awal ini dilakukan pemilihan atribut atau tahap *preprocessing* yang akan digunakan dalam penelitian ini yang sebelumnya sebanyak 2891 data, yang digunakan sebagai acuan penelitian dalam menganalisis dengan 7 atribut menjadi 2890 data dengan 7 atribut dikarenakan terdapat *Missing Value* pada atribut *Vol.* seperti tabel di bawah ini:

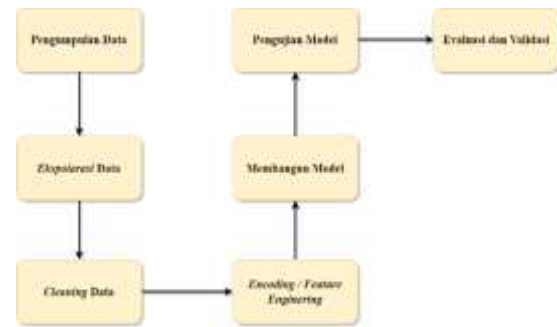
Tabel 3. 4 Pengolahan Data Awal

Nama Atribut	Nilai	Tipe Data
<i>Date</i>	Awal : 2013-01-02 Akhir : 2023-12-30	<i>Object</i>
<i>Price</i>	-37.630 - 123.700	<i>Float</i>
<i>Open</i>	-14.000 - 124.660	<i>Float</i>
<i>High</i>	13.690 - 130.500	<i>Float</i>
<i>Low</i>	-40.320 - 120.790	<i>Float</i>
<i>Vol.</i>	1.000 - 993.240	<i>Object</i>
<i>Change %</i>	0.00 - 305.97	<i>Object</i>

3.3 Perancangan Penelitian

Dalam penelitian ini metode yang digunakan adalah menerapkan algoritma *Decision Tree* untuk memprediksi harga Minyak Mentah dengan menggunakan algoritma CART dimana data yang diambil merupakan dataset data historis emas berjangka yang diambil langsung dari *website investing.com*. Dengan alamat url : (<https://www.investing.com/commodities/crude-oil-historical-data>).

Ada beberapa tahapan yang dibutuhkan dalam penelitian ini yaitu: Pengolahan data, Eksplorasi data, *Cleaning data*, *Feature engineering*, membangun model, kemudian Pengujian Model dan terakhir evaluasi dan validasi hasil seperti terlihat pada gambar di bawah ini.



Gambar 3. 1 Perancangan Penelitian

4. HASIL DAN PEMBAHASAN

4.1 Hasil

Penelitian ini bertujuan untuk menerapkan Algoritma *Classification and Regression Tree* untuk memprediksi harga minyak 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.

4.2 Pembahasan

4.2.1 Pengolahan Data

Pada tahapan pengolahan data, dataset yang ada di tampilkan untuk mengetahui apa saja data yang ada di dalam dataset tersebut.

Tabel 4. 1 Keterangan Dataset Awal

Nama Atribut	Tipe data
<i>Date</i>	<i>Object</i>
<i>Price</i>	<i>Float</i>
<i>Open</i>	<i>Float</i>
<i>High</i>	<i>Float</i>
<i>Low</i>	<i>Float</i>
<i>Vol.</i>	<i>Object</i>
<i>Change %</i>	<i>Object</i>

Dari dataset awal diatas dapat kita lihat memiliki 3 tipe data *Object* dan 4 tipe data *Float*. Setelah data diketahui maka data tersebut diolah dengan dikonversi tipe data tersebut ke dalam tipe data *numerik* agar sesuai dengan kebutuhan dari model pohon keputusan.

Tabel 4. 2 Konversi Tipe data

Nama Atribut	Tipe data	Nilai
Tanggal	<i>Date</i>	2013/01/02 - 2023/12/30
Terakhir	<i>Float</i>	12.34 - 123.70

Pembukaan	Float	12.96 - 124.66
Tertinggi	Float	13.69 - 130.50
Terendah	Float	10.07 - 120.79
Vol.	Float	1.00 - 993.24
Perubahan %	Float	0.00 - 37.66

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. 3 Eksplorasi Data

Index	Date	Price	Open	High	Low	Vol.	Change %
0	12/29/2023	71.65	71.99	72.62	71.25	214.49K	-0.17%
1	12/28/2023	71.77	73.80	74.40	71.72	262.75K	-3.16%
2	12/27/2023	74.11	75.32	75.66	73.77	253.32K	-1.93%
3	12/26/2023	75.57	73.56	76.18	73.13	208.72K	2.41%
4	12/25/2023	73.79	73.49	73.94	73.48	NaN	0.31%
...
2886	01/08/2023	93.15	93.32	93.80	92.67	195.87K	-0.04%
2887	01/07/2023	93.19	93.21	93.35	92.42	166.29K	0.11%
2888	01/04/2023	93.09	92.86	93.21	91.52	210.75K	0.18%
2889	01/03/2023	92.92	92.91	93.30	92.49	189.81K	-0.21%
2890	01/02/2023	93.12	91.78	93.87	91.56	203.87K	1.42%

Pada dataset awal di atas kita dapat mengetahui terdapat 2891 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. 4 Analisa Data

Kolom	Type Data
Date	Object
Price	Object
Open	Object
High	Object
Low	Object
Vol.	Object
Change %	Object

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 *price* atau terakhir yang merupakan harga penutupan atau harga terakhir dari penjualan minyak mentah 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. 5 Analisa data Price

Keterangan	Nilai
Nilai Minimal	-37.630
Nilai Maksimal	123.700
Jumlah Baris	2891 Baris Data
Keunikan Setiap Data	2379 Baris Data

b. *Atribut* atau *Feature*

1. *Date* (Tanggal)

Date adalah kolom tanggal yang merupakan tanggal dari penjualan minyak mentah pada hari tersebut. Pada kolom tanggal ini terdapat 2891 baris data yang dimana setiap isi berbeda-beda sebanyak 2379 data. Tanggal yang di ambil dari data ini dimulai dari 01/02/2013 sampai 12/29/2023.

Tabel 4. 6 Analisa data Date

Keterangan	Nilai
------------	-------

Jumlah Baris	2891 Baris Data
Keunikan setiap data	2379 Baris Data
Tanggal Awal	01/02/2013
Tanggal Akhir	12/29/2023

2. *Open* (Pembukaan)

Open adalah data pembukaan yang merupakan harga pembukaan/awal dari penjualan minyak mentah pada hari tersebut.

Pada kolom pembukaan terdapat 2891 baris data dengan jumlah keunikan 2349 baris data, dimana data yang paling banyak muncul yaitu 52.45, 61.96, 47.72, 61.55 dengan frekuensi kemunculan sebanyak 4 kali.

Tabel 4. 7 Analisa data *Open*

Keterangan	Nilai
Jumlah Baris	2891 Baris Data
Keunikan setiap data	2349 Baris Data
Data yang sering muncul	52.45, 61.96, 47.72, 61.55
Frekuensi kemunculan	4

3. *High* (Tertinggi)

High adalah kolom tertinggi yang merupakan harga tertinggi dari penjualan minyak mentah pada hari tersebut.

Pada kolom tertinggi terdapat 2891 baris data dengan jumlah keunikan 2337 baris data, dimana data yang paling banyak muncul yaitu 46.53 dengan frekuensi kemunculan sebanyak 6 kali.

Tabel 4. 8 Analisa data *High*

Keterangan	Nilai
Jumlah Baris	2891 Baris Data
Keunikan setiap data	2337 Baris Data
Data yang sering muncul	46.53
Frekuensi kemunculan	6

4. *Low* (Terendah)

Low adalah kolom terendah yang merupakan harga terendah dari penjualan minyak mentah pada hari tersebut.

Pada kolom tertinggi terdapat 2891 baris data dengan jumlah keunikan 2358 baris data, dimana data yang paling banyak muncul yaitu 42.36, 45.61, 53.05, 102.1, 53.35, 92.86 dengan frekuensi kemunculan sebanyak 4 kali.

Tabel 4. 9 Analisa data *Low*

Keterangan	Nilai
Jumlah Baris	2891 Baris Data
Keunikan setiap data	2358 Baris Data
Data yang sering muncul	42.36, 45.61, 53.05, 102.1, 53.35, 92.86
Frekuensi kemunculan	4

5. *Vol.*

Vol. adalah kolom *Vol.* yang merupakan volume dari penjualan minyak mentah pada hari tersebut.

Pada kolom tertinggi terdapat 2891 baris data dengan jumlah keunikan 2709 baris data, dimana data yang paling banyak muncul yaitu 1.01M dengan frekuensi kemunculan sebanyak 7 kali.

Tabel 4. 10 Analisa Data *Vol.*

Keterangan	Nilai
Jumlah Baris	2891 Baris Data
Keunikan setiap data	2709 Baris Data
Data yang sering muncul	1.01M
Frekuensi kemunculan	7

6. *Change %* (Perubahan%)

Change % adalah kolom perubahan yang merupakan harga perubahan dari penjualan minyak mentah kedalah bentuk persentase pada hari tersebut.

Pada kolom tertinggi terdapat 2891 baris data dengan jumlah keunikan

909 baris data, dimana data yang paling banyak muncul yaitu 0.13% dengan frekuensi kemunculan sebanyak 14 kali.

Tabel 4. 11 Analisa data *Change %*

Keterangan	Nilai
Jumlah Baris	2891
Keunikan setiap data	909
Data yang sering muncul	0.13%
Frekuensi kemunculan	14

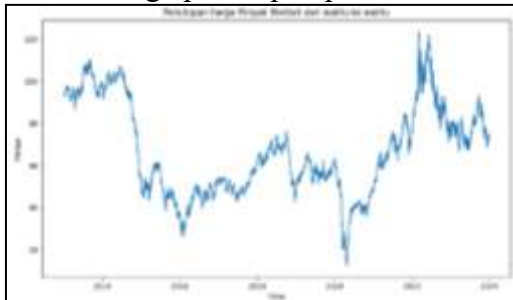
4.2.4 Visualisasi Data

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 ditampilkan.

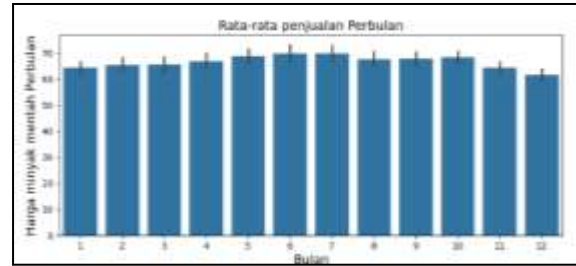
1. Pola harga penutupan pertahun



Gambar 4. 1 Pola Harga Penutupan Pertahun

Pada gambar diatas dapat kita lihat terlihat grafik ketidak setabilan fluktuasi harga yang sangat signifikan dari harga penjualan minyak mentah, dimulai penurunan terjadi pada tahun 2015 lalu stabil hingga sampai 2020 harga minyak mentah mengalami penurunan harga yang sangat signifikan di karenakan *covid-19*, lalu pada tahun 2022 harga minyak meningkat kembali hingga saat ini mulai stabil kembali.

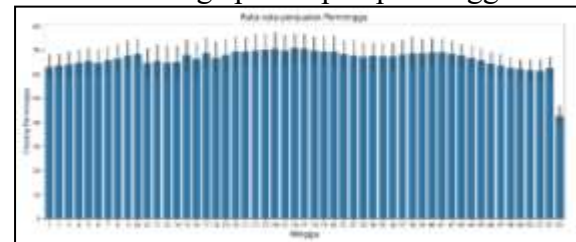
2. Pola harga penutupan perbulan



Gambar 4. 2 Pola Harga Penutupan Perbulan

Pada gambar diagram diatas dapat kita lihat harga penutupan minyak mentah, rata-rata penjualan perbulan tertinggi pada pertengahan tahun atau pada bulan ke 6 dan 7.

3. Pola harga penutupan perminggu



Gambar 4. 3 Pola Harga Penutupan Perminggu

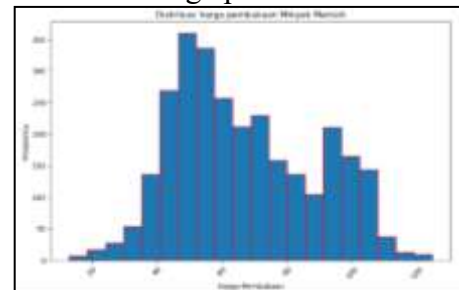
Pada gambar digram diatas distribusi harga minyak mentah sangat *fluktuatif*, dengan harga yang idak setabil pada tiap minggunya dan terjadi penurunan di minggu terakhir menjelang akhir tahun.

b. *Atribut* atau *Feature*

1. Pembukaan (Harga Pembukaan)

Harga pembukaan atau kolom *Open* pada dataset mengacu pada harga minyak mentah yang di perdagangan pada priode perdagangan tertentu.

a) Distribusi harga pembukaan

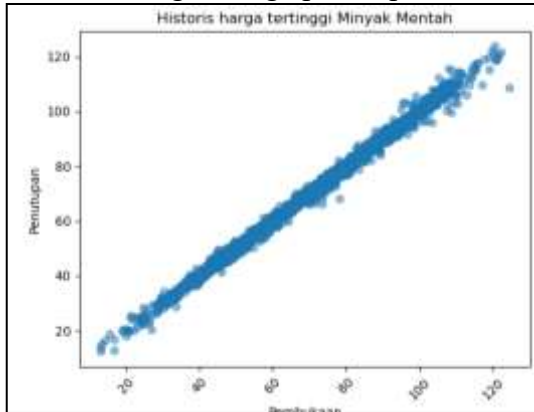


Gambar 4. 4 Distribusi Harga Pembukaan

Dari visualisasi gambar diatas dapat kita lihat untuk kolom *Open*

memiliki frekuensi tertinggi dalam pembukaan harga 45-50 memiliki frekuensi tertinggi.

- b) Distribusi harga pembukaan dengan harga penutupan



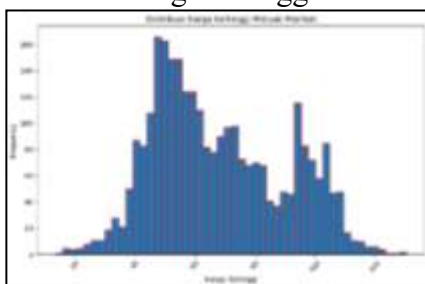
Gambar 4. 5 Distribusi Harga Pembukaan Dengan Penutupan

Pada visualisasi gambar diatas adalah hubungan antara atribut pembukaan dengan penutupan yang memperlihatkan bahwa kemunculan data pada harga pembukaan dan harga penutupan perbedaan yang tidak terlalu signifikan atau tidak jauh berbeda, data nilai diatas sama-sama berkumpul di tengah.

2. *High* (Harga Tertinggi)

Harga tertinggi atau kolom *High* pada dataset mengacu pada harga minyak mentah yang di perdagangan pada priode perdagangan tertentu.

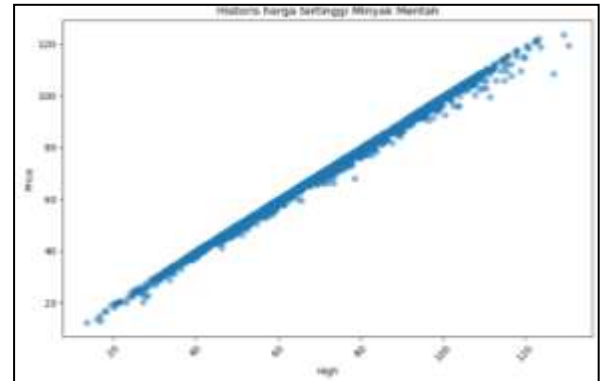
- a) Distribusi harga tertinggi



Gambar 4. 6 Distribusi Harga Tertinggi Dari visualisasi gambar diatas dapat kita lihat untuk kolom *High* memiliki frekuensi tertinggi dalam

Tertinggi harga 45-50 memiliki frekuensi tertinggi mencapai 160 data lebih.

- b) Distribusi harga tertinggi dengan harga penutupan



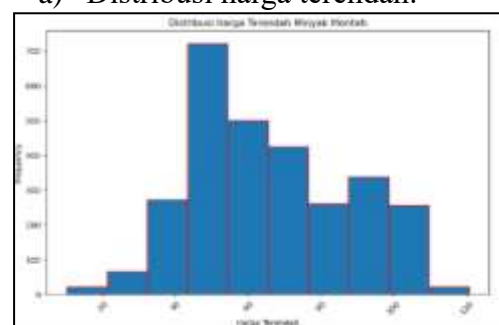
Gambar 4. 7 Distribusi Harga Tertinggi Dengan Penutupan

Pada visualisasi gambar diatas adalah hubungan antara atribut harga tertinggi dengan penutupan yang memperlihatkan bahwa kemunculan data pada harga tertinggi dan harga penutupan perbedaan yang tidak terlalu signifikan atau cukup sedikit perbedaan, dimana data yang berkumpul berada di Tengah.

3. *Low* (Harga Terendah)

Harga terendah atau kolom *Low* pada dataset mengacu pada harga minyak mentah yang di perdagangan pada priode perdagangan tertentu.

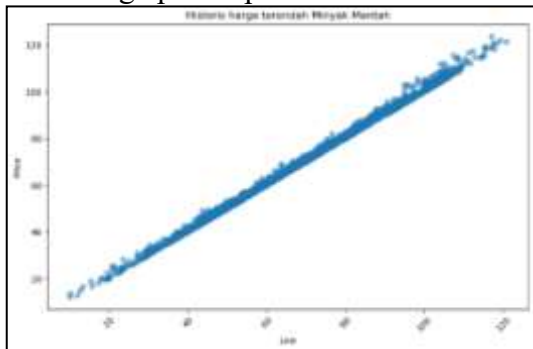
- a) Distribusi harga terendah.



Gambar 4. 8 Distribusi Harga Terendah

Dari visualisasi gambar diatas dapat kita lihat untuk kolom *Low* memiliki frekuensi tertinggi dalam terendah harga 45-50 memiliki frekuensi tertinggi mencapai 700 data lebih.

- b) Distribusi harga terendah dengan harga penutupan



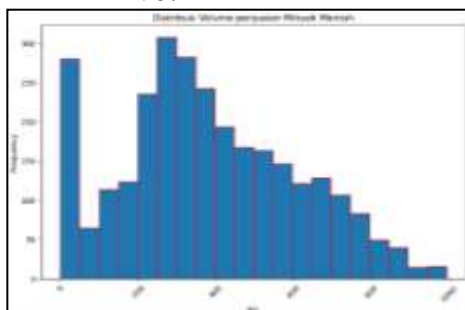
Gambar 4. 9 Distribusi Harga Terendah Dengan Penutupan

Pada visualisasi gambar diatas adalah hubungan antara atribut harga terendah dengan penutupan yang memperlihatkan bahwa kemunculan data pada harga terendah dan harga penutupan perbedaan yang tidak terlalu signifikan atau cukup sedikit perbedaan, dimana data yang berkumpul berada di Tengah.

4. *Vol.*

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

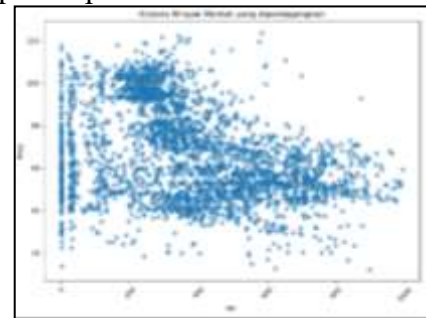
- a) Distribusi *Vol.*



Gambar 4. 10 Distribusi *Vol.*

Dari visualisasi gambar diatas dapat kita lihat untuk kolom *Vol.* memiliki frekuensi tertinggi dalam volume penjualan dengan nilai 250-300 yang memiliki frekuensi tertinggi mencapai 300 data.

- b) Distribusi *Vol.* dengan harga penutupan



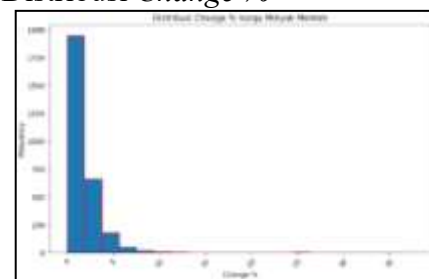
Gambar 4. 11 Distribusi *Vol.* Dengan Harga Penutupan

Pada visualisasi gambar diatas adalah hubungan antara atribut volume dengan penutupan yang memperlihatkan bahwa kemunculan data pada volume penjualan dan harga penutupan memiliki nilai yang bervariasi tetapi banyak berkumpul diangka 250-300 pada volume.

5. *Change %* (Perubahan)

Pada atribut Perubahan atau kolom *Change %* pada dataset mengacu pada perubahan harga perdagangan minyak mentah yang di perdagangan pada priode perdagangan tertentu.

- a) Distribusi *Change %*

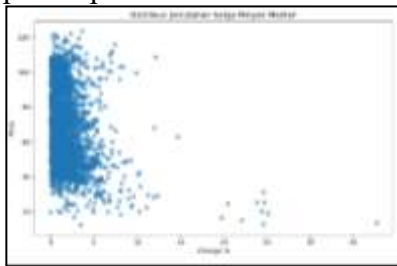


Gambar 4. 12 Distribusi *Change %*

Dari visualisasi gambar diatas dapat kita lihat untuk kolom

Change % memiliki presentase frekuensi tertinggi dalam perubahan harga penjualan dengan nilai 0-2,5% yang memiliki frekuensi tertinggi hamper mencapai 2000 data.

- b) Distribusi *Change %* dengan harga penutupan.



Gambar 4. 13 Distribusi *Change %* Dengan Harga Penutupan

Pada visualisasi gambar diatas adalah hubungan antara atribut perubahan dengan penutupan yang memperlihatkan bahwa kemunculan data pada perubahan memiliki persentase 0-2,5% dan harga penutupan memiliki nilai yang bervariasi tetapi banyak berkumpul diangka 40-110 pada *Price*-nya

5. KESIMPULAN

Berdasarkan Implementasi dan pengujian data yang telah di lakukan pada bab sebelumnya dapat diketahui Kesimpulan dari penelitian tersebut yaitu :

- Bahwa metode *decision tree* dengan menggunakan alghoritma *Classification And Regression* atau CART dapat diterapkan kedalam model *Machine Learning* untuk memprediksi Minyak Mentah WTI berjangka dengan nilai MSE 0.892, nilai MAE 0.700, nilai RSME 0.944, dan nilai *R-square* 0.957.
- Hasil dari penelitian ini tersebut terbukti bahwa alghoritma *Classification And Regression tree* sangat baik diterapkan untuk memprediksi harga minyak mentah sehingga model *machine learning* ini dan

dapat mempermudah Pemerintah atau Perusahaan untuk mengambil keputusan yang lebih efektif dalam sektor energi.

DAFTAR PUSTAKA

- Aliffia, N., Sediono, Suliyanto, Mardianto, M. F., & Amelia, D. (2023). Modeling Of World Crude Oil Price Based On Pulse Function Intervention Analysis Approach. *Media Statistika*, 136-147.
- Alpaydin, E. (2021). Machine Learning: The New AI. *MIT Press Journal*, 9-27.
- Anderson, & Williams. (2023). Python Libraries for Machine Learning: A Comprehensive Review. *Journal of Machine Learning Research (JMLR)*, 1-45.
- Anisa, N. (2022, Febuary 14). *Mengenal Artificial Neural Network*. Retrieved from BINUS University School of Information Systems: <https://sis.binus.ac.id/2022/02/14/mengenal-artificial-neural-network/>
- AWS. (2024). Amazon Forecast Panduan Developer. *Amazon Forecast*, 105-110.
- Bollapragada, R., Mankude, A., & Udayabhanu, V. (2021). Forecasting The Price Of Crude Oil. *DECISION*, 207-231.
- Chen, C. Y., & Chang, Y.-W. (2024). Missing Data Imputation Using Classification And Regression Trees. *PeerJ Comput*, 72-103.
- Chen, E., & He, X. J. (2019). Crude Oil Price Prediction with Decision Tree Based Regression Approach. *Journal of International Technology and Information Management*, 1-16.
- Delgado, M. F., & Cernadas, E. (2023). Do We Need Hundreds Of Classifiers To Solve Real World Classification Problems Machine Learning. *Machine Learning*, 1381-1407.
- Dombrowski, Q., Gniady, T., & Kloster, D. (2019). Introduction to Jupyter Notebooks. *Programming Historian* 8, 67-88. doi:10.46430/phen0087
- Gao, S., & Lei, Y. (2019). A New Approach For Crude Oil Price Prediction Based On Stream. *Geoscience Frontiers* 8, 183-187.
- Gaspar, L., & Mbwambo, H. (2023). Forecasting Crude Oil Prices By Using ARIMA Model: Evidence From Tanzania. *Journal of Accounting, Finance and Auditing Studies*, 158-175.
- Gholizadeh, & Samira. (2022). Top Popular Python Libraries in Research. *Journal of Robotics and Automation Research*, 142-145.

- [14] González-Carrillo, C. D., Restrepo-Calle, F., Ramírez-Echeverry, J. J., & González, F. A. (2021). Automatic Grading Tool for Jupyter Notebooks in Artificial Intelligence Courses. *Sustainability*, 33-39.
- [15] Hunter, & D., J. (2021). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 90-95.
- [16] Irmanita, R., Prasetyowati, S. S., & Sibaroni, Y. (2021). Classification of Malaria Complication Using CART (Classification and Regression Tree) and Naïve Bayes. *RESTI Journal System Engineering and Information Technology*, 10–16.
- [17] Isnaeni, Sudarmin, & Rais, Z. (2022). Analisis Support Vector Regression (Svr) Dengan Kernel Radial Basis Function (Rbf) Untuk Memprediksi Laju Inflasi Di Indonesia. *Journal of Statistics and Its Application on Teaching and Research*, 30-38.
- [18] James, G., Witten, D., & Tibshirani, T. H. (2020). *An Introduction to Statistical Learning with Applications in R*. New York Heidelberg Dordrecht London: Springer. doi:10.1007/978-1-4614-7138-7
- [19] Johnson, P. L., & Zhang, H. (2022). Analysis of Python Standard Library and Third-Party Libraries: Impact on Modern Software Development. *IEEE Transactions on Software Engineering*, 567-589.
- [20] Johnson, R., & Smith, K. (2023). Machine Learning Applications Using Scikit-learn. *Journal of Machine Learning Research*, 125-147.
- [21] Kumar, A., & Singh, R. (2023). Scalable decision tree algorithms for big data: A comprehensive review. *Journal of Big Data*, 1-25.
- [22] Kusumantara, P. M., Putrawanto, D. I., Mukaromah, S., & Agussalim. (2023). Analisis Model Peramalan Harga Minyak Mentah (Crude Oil) Menggunakan Weighted Moving Average. *Jurnal Nasional Komputasi dan Teknologi Informasi*, 395-401.
- [23] Lestari, I., Sumargo, B., & Ladayya, F. (2022). Pemodelan ARIMA Intervensi untuk Meramalkan Harga Minyak Mentah Dunia. *Statistika*, 133 – 146.
- [24] Li, X., Wang, Y., & Zhou, Z.-H. (2021). Adaptive Decision Tree Learning In Dynamic Environments. *Jurnal: IEEE Transactions on Knowledge and Data Engineering*, 3174-3186.
- [25] Lin, H., & Sun, Q. (2020). Crude Oil Prices Forecasting: An Approach of Using CEEMDAN-Based Multi-Layer Gated Recurrent Unit Networks. *Energies*, 1-21.
- [26] Martinez, A., & Johnson, B. (2024). TensorFlow: Evolution and Applications in Modern Deep Learning. *Journal of Artificial Intelligence Research*, 178-195.
- [27] Martinez, L., & Chen, W. (2021). Python in the Era of Digital Transformation: A Systematic Review. *ACM Computing Surveys*, 67-77.
- [28] McKinney. (2022). Pandas: Powerful Python Data Analysis Toolkit. *Journal of Open Source Software*, 6-23.
- [29] Muddana, A. L., & Vinayakam, S. (2024). *Basic Python*. India: Springer Nature Switzerland. doi:10.1007/978-3-031-52473-8_1
- [30] Nuha, H. H. (2023). Mean Squared Error (MSE) dan Penggunaannya. *SSRN*, 1-2.
- [31] Ogunleye, & Olufemi, J. (2021). The Concept of Data Mining. *Artificial Intelligence*, 1-18.
- [32] Paramita, R., Chasmir, D., Fakhreza, T. H., Lestari, V. P., & Ramiayu, D. D. (2023, February 10). *Lonjakan Harga Minyak Dunia Dan Mitigasi Risikonya*. Retrieved from Badan Keahlian DPR RI: <https://berkas.dpr.go.id/pa3kn/videografis/public-file/videografis-public-24.pdf>
- [33] Rahma, A. K., Abidin, Q. Z., Prasetyo, J. K., Larasati, B., Amelia, D., & Chamidah, N. (2024). Prediction Of Crude Oil Prices In Indonesia Using Fourier Series Estimator And Arima Method. *BAREKENG*, 1673–1682.
- [34] Ramadhan, A. S. (2022, January 21). *Decision Tree Algoritma Pada Data Mining*. Retrieved from BINUS University School of Information Systems: <https://sis.binus.ac.id/2022/01/21/decision-tree-algoritma-beserta-contohnya-pada-data-mining/>
- [35] Redha, R. A., & Yousif, S. A. (2023). Crude Oil Price Forecasts Using Support Vector Regression and Technical Indicators. *Iraqi Journal of Science*, 2512-2520.
- [36] Richardson, & Thompson. (2023). Comprehensive Review of Machine Learning: From Theory to Application. *IEEE Transactions on Artificial Intelligence*, 156-172.