

PREDIKSI HARGA EMAS UNTUK PENGAMBILAN KEPUTUSAN INVESTASI MENGGUNAKAN ALGORITMA CART (CLASSIFICATION AND REGRESSION TREE)

Arie Ardiansyah¹, Sri Mulyati²

⁽¹⁾⁽²⁾ Program Studi Teknik Informatika, Universitas Pamulang, JL Puspitek Tangerang Selatan,
Indonesia, 15310
e-mail: ⁽¹⁾arieardiansyah1996@gmail.com

Abstract

The dynamic fluctuations in gold prices, influenced by various global economic factors, make gold price prediction an important topic in finance and investment. This study aims to analyze and predict gold prices using the Classification and Regression Tree (CART) algorithm, which is a decision tree-based machine learning method. Historical gold price data from 2013 to 2023 were used as training data, while data from 2024 to 2025 were used for testing. The research process includes preprocessing, data splitting, CART model construction, and performance evaluation using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2). The results indicate that the CART model is capable of providing reasonably accurate and effective predictions of gold prices, making it a viable tool for investment decision-making. With its simplicity and interpretability, CART helps uncover patterns in gold price data and offers valuable estimates for investors and market participants.

Keywords: Gold Price, Prediction, CART, Regression, Machine Learning.

Abstrak

Fluktuasi harga emas yang dinamis dan dipengaruhi oleh berbagai faktor ekonomi global menjadikan prediksi harga emas sebagai topik yang penting dalam bidang keuangan dan investasi. Penelitian ini bertujuan untuk menganalisis dan memprediksi harga emas dengan menggunakan algoritma *Classification and Regression Tree* (CART), yang merupakan salah satu metode dalam machine learning berbasis pohon keputusan. Data historis harga emas dari tahun 2013 hingga 2023 digunakan sebagai data latih, sementara data tahun 2024 hingga 2025 digunakan sebagai data uji. Proses penelitian meliputi tahap preprocessing, pembagian data, pembangunan model CART, dan evaluasi performa model menggunakan metrik *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan *R-squared* (R^2). Hasil penelitian menunjukkan bahwa model CART mampu memberikan prediksi harga emas dengan tingkat akurasi dan efektivitas yang cukup baik, sehingga dapat digunakan sebagai salah satu alternatif dalam pengambilan keputusan investasi. Dengan pendekatan yang sederhana dan interpretatif, CART dapat membantu dalam memahami pola-pola dalam data harga emas dan memberikan estimasi yang berguna bagi investor maupun pelaku pasar.

Kata Kunci: Harga Emas, Prediksi, CART, Regresi, *Machine Learning*.

1. PENDAHULUAN

Emas merupakan salah satu instrumen investasi yang penting karena nilainya cenderung stabil dalam jangka panjang dan berfungsi sebagai pelindung nilai (*hedging*) terhadap inflasi dan gejolak ekonomi. Namun, harga emas sangat fluktuatif dan dipengaruhi oleh berbagai faktor global seperti inflasi, suku bunga, nilai tukar mata uang, kondisi geopolitik, serta permintaan dan penawaran di pasar

internasional. Ketidakstabilan ini menyebabkan prediksi harga emas menjadi suatu tantangan yang kompleks bagi investor, pelaku usaha, maupun pemerintah dalam menyusun strategi investasi yang tepat [1].

Dalam menghadapi tantangan tersebut, teknologi prediksi menjadi solusi yang menjanjikan. Prediksi harga emas menggunakan data historis bertujuan untuk mengetahui peluang perubahan harga di masa depan agar dapat dijadikan dasar dalam pengambilan keputusan investasi [2]. Meskipun metode konvensional seperti regresi linier masih sering digunakan, pendekatan tersebut kurang efektif dalam menangkap pola data yang bersifat *non-linear*. Oleh karena itu, pendekatan berbasis *machine learning* mulai banyak dikembangkan karena memiliki kemampuan untuk menganalisis pola kompleks dalam data dan memberikan hasil prediksi yang lebih akurat [3, 4].

Salah satu metode *machine learning* yang dapat digunakan untuk menangani data *non-linear* adalah algoritma CART (*Classification and Regression Tree*). CART bekerja dengan membagi *dataset* menjadi subset berdasarkan nilai atribut tertentu dan menghasilkan model pohon keputusan yang mudah diinterpretasikan. Algoritma ini juga memiliki keunggulan dalam menangani *outlier* serta tidak memerlukan asumsi distribusi tertentu pada data. Beberapa studi menunjukkan bahwa penggunaan CART dalam prediksi harga komoditas memberikan hasil yang baik dan akurat, khususnya saat dikombinasikan dengan analisis deret waktu [5, 6]. Namun, meskipun potensial, masih relatif sedikit penelitian yang secara spesifik mengkaji penerapan algoritma CART dalam konteks prediksi harga emas dibandingkan metode lain seperti *Random Forest* atau *Support Vector Regression* (SVR), sehingga terdapat celah penelitian yang perlu diisi [7, 8]. Keterbatasan kajian ini menjadi peluang untuk mengeksplorasi efektivitas CART secara lebih mendalam, khususnya dalam konteks fluktuasi harga emas yang kompleks dan dinamis.

Melalui penelitian ini, penulis mengimplementasikan algoritma CART untuk membangun model prediksi harga emas berdasarkan data historis. Dengan pendekatan ini, diharapkan model prediksi yang dihasilkan mampu membantu investor dalam membuat keputusan yang lebih rasional serta mengurangi risiko akibat fluktuasi harga. Selain itu, penelitian ini juga bertujuan untuk mengevaluasi kinerja algoritma CART dalam konteks prediksi harga emas menggunakan metrik evaluasi seperti MAE, RMSE, dan R^2 , sehingga dapat memberikan kontribusi nyata dalam pengembangan metode prediksi berbasis *machine learning* [9, 10].

2. METODE

Data Mining

Data mining adalah proses menemukan pola, tren, dan informasi berguna dari kumpulan data besar menggunakan teknik statistik, algoritma, dan pembelajaran mesin [11].

Machine Learning

Machine Learning (ML) adalah studi ilmiah tentang algoritma dan model statistik yang digunakan sistem komputer untuk melakukan tugas tertentu tanpa diprogram secara eksplisit yang bertujuan untuk mengajarkan mesin mengenai bagaimana cara mengelola data secara lebih efisien [12].

Decision Tree

Decision Trees merupakan metode pembelajaran yang serbaguna dan nonparametrik yang diawasi yang digunakan secara luas dalam pembelajaran mesin untuk tugas klasifikasi dan regresi. Keahlian mereka terletak pada pembuatan model prediktif yang didasarkan pada aturan keputusan yang tidak rumit yang diekstrak dari data atribut, dengan demikian memastikan interpretasi yang jelas dan meniadakan kebutuhan untuk penskalaan fitur. *Decision tree* dengan mahir mengkategorikan data dan memperkirakan dalam segmen yang berbeda, membuktikan keampuhannya dalam berbagai skenario [13].

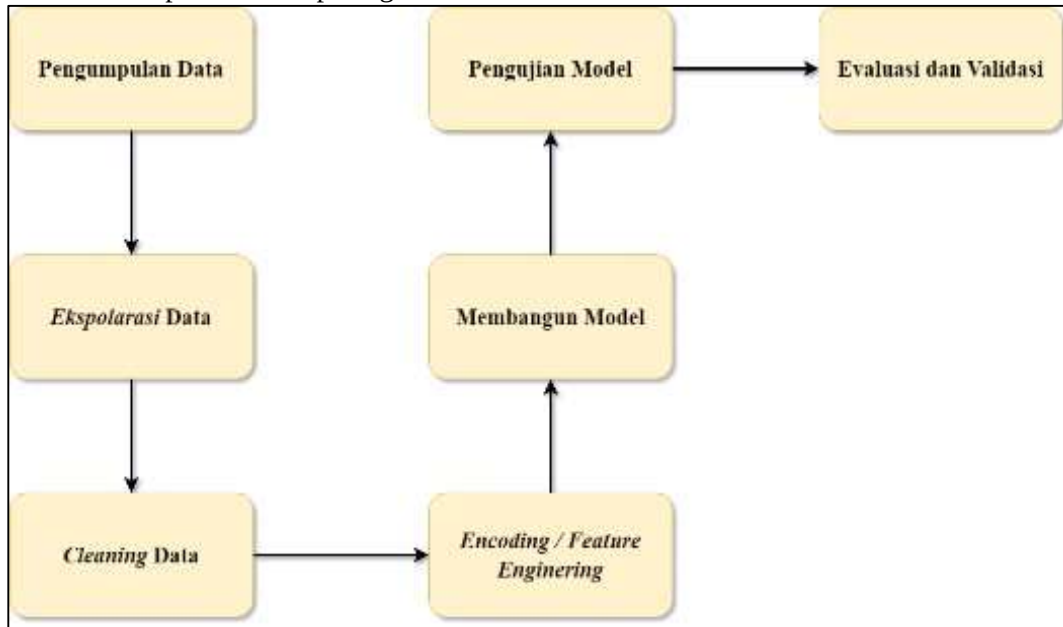
CART (Classification and Regression Tree)

Classification and regression trees adalah strategi klasifikasi berbasis data yang menggunakan data historis untuk membuat *Decision Trees* [14]. CART (*Classification and Regression Trees*) adalah teknik pembelajaran mesin yang digunakan untuk membangun pohon keputusan baik untuk klasifikasi maupun regresi. Tahapan metode CART meliputi pemilihan atribut dan kriteria pemisahan, pembangunan pohon keputusan, pemangkasan pohon untuk menghindari *overfitting*, dan evaluasi serta validasi hasil. Proses ini

memastikan bahwa pohon keputusan yang dihasilkan efektif dalam klasifikasi atau regresi dengan performa yang baik pada data yang tidak terlihat sebelumnya [15].

Dalam penelitian ini metode yang digunakan adalah menerapkan algoritma Decision Tree untuk memprediksi harga emas dengan menggunakan algoritma CART dimana data yang diambil merupakan *dataset* data emas berjangka yang diambil dari website *investing.com*. Dengan alamat url : (<https://id.investing.com/commodities/Gold-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 2. 1 Perancangan Penelitian

- Pengolahan Data**
Data minyak sawit yang diolah yaitu di ambil dari *website Investing.com* dengan rentan waktu 10 tahun dari tahun 2014 -2024 data tersebut dengan format berbentuk CSV
- Eksplorasi Data**
Eksplorasi data untuk memahami pola, tren, dan hubungan antar variabel. Pada tahap ini visualisasi data seperti grafik *time-series* atau *scatter plot* sering digunakan untuk melihat bagaimana variabel-variabel berubah seiring waktu dan apakah terdapat *outlier* atau anomali dalam data.
- Cleaning Data**
Cleaning data atau pembersihan data untuk mengatasi masalah seperti data *missing values* dan *outlier*. Data yang kosong dapat diisi dengan nilai rata-rata (*imputasi*) atau dihapus jika tidak relevan.
- Encoding (Feature Enginering)**
Feature engineering yaitu menciptakan dan memilih variabel yang paling relevan untuk meningkatkan performa model. Pada proses ini dilakukan untuk mengubah data kategorikal menjadi bentuk *numerik* yang dapat di proses oleh algoritma pembelajaran atau model *statistic*.
- Membangun Model**
Membangun model dengan algoritma *Decision Tree* diimplementasikan dengan menggunakan *library scikit-learn*. *Scikit-Learn* atau *Sklearn* adalah *library* berbasis *Python* untuk membangun model pembelajaran mesin. Dan menyediakan banyak algoritma pembelajaran untuk regresi, pengelompokan, dan klasifikasi. *Sklearn* kompatibel dengan *NumPy* dan *SciPy*.
- Pengujian Model**

Pengujian model menggunakan data uji untuk melihat akurasi dan kesalahan prediksi. Beberapa metrik evaluasi yang umum digunakan adalah *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R-squared* (R^2).

g. Evaluasi dan Validasi

Evaluasi dan validasi untuk memastikan model bekerja dengan baik pada data baru dan tidak hanya pada data latih. Jika hasilnya belum optimal, tuning parameter atau metode *ensemble* dapat dicoba untuk meningkatkan akurasi.

3. HASIL

Membangun Model

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

Tabel 3.1 Membangun Model

No	Dataset yang Digunakan	Dimensi
1	Data Historis Emas Berjangka	2811 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 3. 2 *Splitting* Data

Dataset	Jumlah
Data Historis	2811 Baris Data & 7 Kolom
Data Uji	346 Baris Data & 7 Kolom

a. *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 3. 1 *Training Model decision Tree*

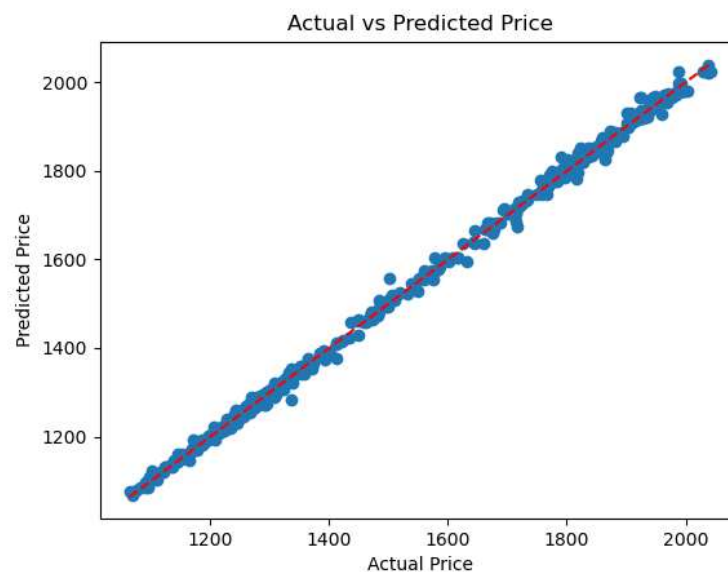
Pengujian Model

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 Errorr* (MSE), *Mean Absolute Errorr* (MAE), *R-Square* digunakan untuk mengukur sejumlah model dapat mengklasifikasikan data dengan benar.

Tabel 3. 3 *Pengujian Model*

Model	Hasil
<i>Mean Squared Errorr (MSE)</i>	111.2
<i>Mean Absolute Errorr (MAE)</i>	0.998
<i>R-Square</i>	7.082

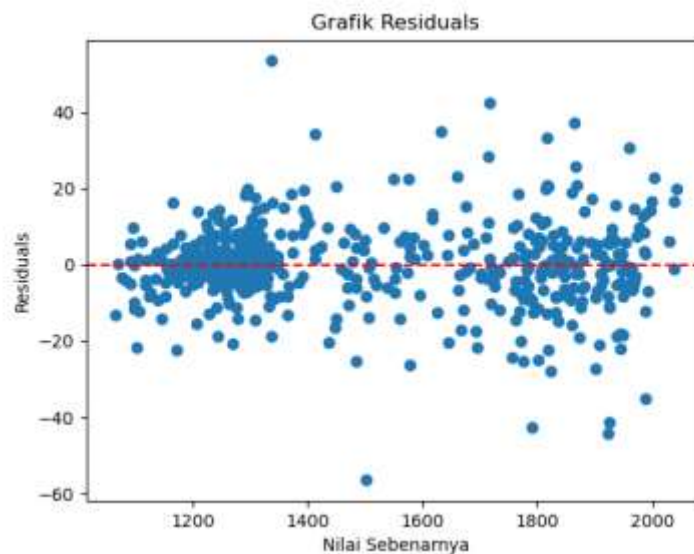
Berdasarkan tabel diatas pengujian model dari pembuatan model tersebut mendapatkan nilai akurasi *R-Square* 7.082, nilai MSE sebesar 111.2, dan nilai MAE sebesar 0.998. Hasil pengujian model mesin learning dapat dilihat dalam bentuk grafik atau *kurva* dibawah ini.



Gambar 3. 2 *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 3. 3 Grafik *Residual Model Machine Learning*

Dari Gambar grafik *residual model machine learning* tersebut cukup baik karena semakin banyak persebaran data yang mendekati nilai 0 maka data tersebut semakin baik dengan mengikuti garis linieritas, walaupun ada beberapa data yang menjauh dari garis linier tetapi persebaran yang mendekati nilai 0 cukup banyak.

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 2024 - Maret 2025 sebelumnya, adapun *dataset* yang digunakan dapat dilihat pada tabel berikut.

Tabel 3 .4 Evaluasi Data

No	<i>Dataset</i> yang Digunakan	Dimensi
1	Data Historis Emas Berjangka	346 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: 7.855963089867732
Mean Squared Error: 61.71615606936417
Mean Absolute Error: 2.777456647398906
R-Square: 0.9994831258433321
Accuracy: 99.94831258433321 %
```

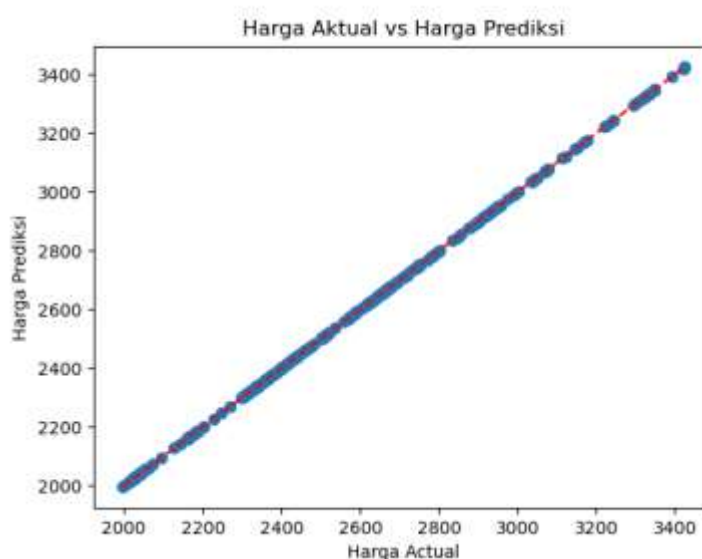
Gambar 3. 4 Validasi Hasil

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

Tabel 3 .5 Validasi Hasil

Model	Hasil
Root Mean Squared Errorr (RMSE)	7.855
Mean Squared Errorr (MSE)	61.71
Mean Absolute Errorr (MAE)	2.777
R-Square	0.999

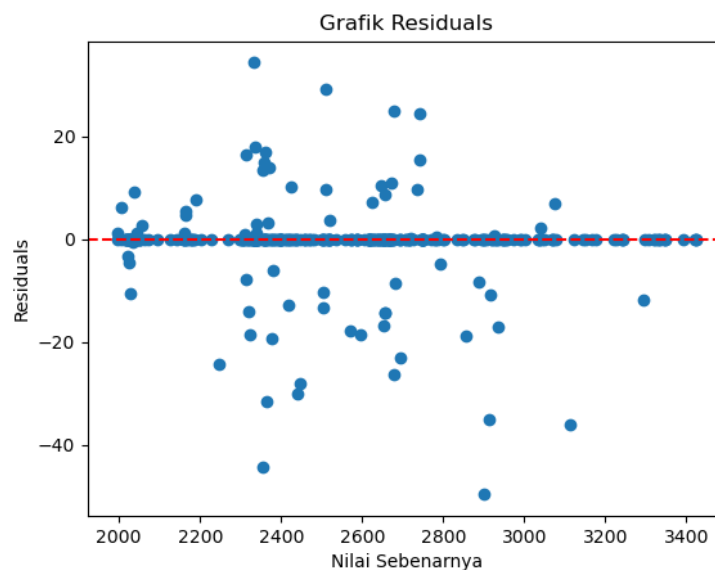
Hasil dari model memiliki tingkat akurasi yang dengan nilai *R-Square* 0.999, nilai RMSE sebesar 7.855, nilai MSE sebesar 61.71, dan nilai MAE sebesar 2.777, Sedangkan grafik harga aktual dengan harga prediksi sebagai berikut :



Gambar 3. 5 Hasil Actual Dengan Prediksi

Pada gambar di atas menunjukkan bahwa semakin dekat titik data dengan garis linier, maka semakin baik performa model dalam memprediksi harga emas. Kurva regresi memberikan gambaran *visual* tentang

hubungan antara variabel *input* dan *output* dalam model regresi, misalnya jika titik data berada di atas kurva regresi itu menunjukkan nilai prediksi lebih tinggi di dibandingkan dengan nilai aktual dan sebaliknya. Sedangkan grafik *residual* dari data uji tersebut sebagai berikut :



Gambar 3. 6 Grafik Residual Data

Dari gambar di atas dapat dilihat bahwa hasil prediksi harga emas memiliki tingkatan *residual* yang sangat baik karena mengikuti garis linier pada grafik tersebut dan persebaran data banyak yang mendekati dengan nilai 0, karena semakin banyak nilai data yang mendekati garis linier atau nilai 0 maka data prediksi tersebut semakin baik. Persebaran data tersebut di ambil dari harga aktual dikurangi harga prediksi maka data tersebut merupakan jarak selisih dari harga aktual dengan harga prediksi.

4. PEMBAHASAN

Persiapan Data

Data historis emas berjangka. Dari website resmi *Investing.com* yang di unduh dai url: (<https://id.investing.com/commodities/Gold1-historical-data>). Seperti pada tabel di bawah ini :

Tabel 4. 1 Persiapan Dataset

No	Nama	Record	Feature	Class
1	Data Historis Emas Berjangka	2811	6	1

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

Tabel 4. 2 Atribut Dataset

Atrribute	Keterangan
<i>Date</i>	Tanggal data harga emas diambil dan tercatat.
<i>Price</i>	Harga penutupan indeks pada hari perdagangan tersebut di website investing.com.
<i>Open</i>	Harga pembukaan indeks pada hari perdagangan tersebut di website investing.com
<i>High</i>	Harga tertinggi indeks pada hari perdagangan tersebut website investing.com

<i>Low</i>	Harga terendah indeks pada hari perdagangan tersebut di website investing.com.
<i>Vol.</i>	Jumlah emas yang di perdagangan pada hari perdagangan tersebut di website investing.com.
<i>Change %</i>	Persentase perubahan harga indeks dibandingkan dengan hari perdagangan sebelumnya di website investing.com

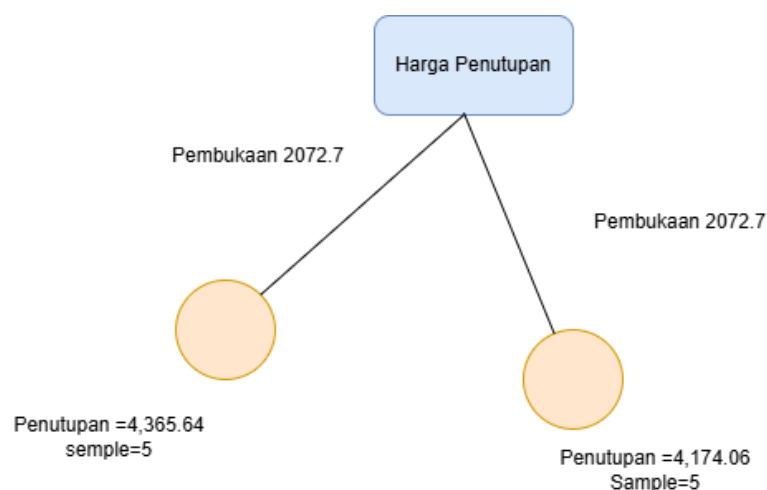
Pembuatan model pohon keputusan dilakukan pada *dataset* yang terdiri dari 7 atribut yang merupakan atribut dari *dataset* histori emas berjangka. Algoritma yang digunakan adalah CART (*Classification and Regression Tree*). Metode ini bekerja dengan membuat pohon keputusan untuk memprediksi *variabel*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

a. Pengambilan 10 sampel data dari tahun 2024-2025

Tabel 4. 3 Implementasi Metode

No	Tanggal	Harga Pembukaan	Harga Terakhir
1	09/05/2025	3310.2	3344.0
2	08/05/2025	3373.1	3306.0
3	07/05/2025	3448.1	3391.9
4	06/05/2025	3345.7	3422.8
5	05/05/2025	3247.1	3322.3
6	08/01/2024	2052.6	2033.5
7	05/01/2024	2051.4	2049.8
8	04/01/2024	2049.3	2050.0
9	03/01/2024	2067.9	2042.8
10	02/01/2024	2072.7	2073.4

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



Gambar 4. 1 Implementasi Metode

Pengolahan Data

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

Tabel 4. 4 Keterangan Dataset Awal

Nama Atribut	Tipe data
<i>Date</i>	<i>Object</i>
<i>Price</i>	<i>Object</i>
<i>Open</i>	<i>Object</i>
<i>High</i>	<i>Object</i>
<i>Low</i>	<i>Object</i>
<i>Vol.</i>	<i>Object</i>
<i>Change %</i>	<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. 5 Konversi Tipe data

Nama Atribut	Tipe data	Nilai
Tanggal	<i>Datetime</i>	2013/01/02 - 2023/12/29
Terakhir	<i>Float</i>	1049.60 – 2083.50
Pembukaan	<i>Float</i>	1051.50 – 2089.50
Tertinggi	<i>Float</i>	1062.70 – 2140.70
Terendah	<i>Float</i>	1045.40 – 2074.60
<i>Vol.</i>	<i>Float</i>	0.12 – 816.53
Perubahan %	<i>Float</i>	0.00-9.34

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. Pada *dataset* awal di atas kita dapat mengetahui terdapat 2811 baris dan 7 kolom atribut, dengan *Missing Value* terdapat pada kolom *Vol.*

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
<i>Date</i>	<i>Object</i>
<i>Price</i>	<i>Object</i>
<i>Open</i>	<i>Object</i>
<i>High</i>	<i>Object</i>

<i>Vol.</i>	<i>Object</i>
<i>Change %</i>	<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. 7Analisa Data Terakhir

Keterangan	Nilai
Nilai Minimal	1049.60
Nilai Maksimal	2083.50
Jumlah Baris	2811 Baris Data
Keunikan Setiap Data	2265 Baris Data

Atribut atau Feature

1. Date

Date adalah kolom yang merupakan tanggal dari penjualan emas pada hari tersebut. Pada kolom *Date* ini terdapat 2811 baris data yang dimana setiap isi berbeda-beda sebanyak 2811 data. Tanggal yang di ambil dari data ini dimulai dari 01/02/2013 sampai 12/29/2023.

Tabel 4. 8 Analisa Data Date

Keterangan	Nilai
Jumlah Baris	2811 Baris Data
Keunikan setiap data	2811 Baris Data
Tanggal Awal	01/02/2013
Tanggal Akhir	12/29/2023

2. Open Kolom pembukaan merupakan harga pembukaan/awal dari penjualan pada hari tersebut. Pada kolom pembukaan terdapat 2811 baris data dengan jumlah keunikan 2219 baris data, data yang paling banyak muncul yaitu 1280.7 dengan frekuensi kemunculan sebanyak 6 kali.

Tabel 4. 9 Analisa data Pembukaan

Keterangan	Nilai
Jumlah Baris	2811 Baris Data
Keunikan setiap data	2219 Baris Data
Data yang sering muncul	1280.7
Frekuensi kemunculan	6

3. High

Kolom *High* merupakan harga tertinggi dari penjualan pada hari tersebut. Pada kolom tertinggi terdapat 2811 baris data dengan jumlah keunikan 2239 baris data, data yang paling banyak muncul yaitu 1283.9 dengan frekuensi kemunculan sebanyak 6 kali.

Tabel 4. 10 Analisa data *High*

Keterangan	Nilai
Jumlah Baris	2811 Baris Data
Keunikan setiap data	2239 Baris Data
Data yang sering muncul	1283.9
Frekuensi kemunculan	6

4. *Low*

Kolom *Low* merupakan harga terendah dari penjualan pada hari tersebut. Pada kolom terendah terdapat 2811 baris data dengan jumlah keunikan 2266 baris data, data yang paling banyak muncul yaitu 1320.8 dengan frekuensi kemunculan sebanyak 6 kali.

Tabel 4. 11 Analisa data *Low*

Keterangan	Nilai
Jumlah Baris	2811 Baris Data
Keunikan setiap data	2266 Baris Data
Data yang sering muncul	1320.8
Frekuensi kemunculan	6

5. *Vol.*

Kolom *Vol.* merupakan *Volume* dari penjualan pada hari tersebut. Pada kolom *Vol.* terdapat 2811 baris data dengan jumlah keunikan 2140 baris data, data yang paling banyak muncul yaitu 1.01 M dengan frekuensi kemunculan sebanyak 17 kali.

Tabel 4. 12 Analisa Data *Vol.*

Keterangan	Nilai
Jumlah Baris	2811 Baris Data
Keunikan setiap data	2140 Baris Data
Data yang sering muncul	1.01M
Frekuensi kemunculan	17

6. *Change %*

Kolom perubahan merupakan harga perubahan dari penjualan kedalam bentuk persentase pada hari tersebut. Pada kolom *Change %* terdapat 2811 baris data dengan jumlah keunikan 482 baris data, data yang paling banyak muncul yaitu 0,00% dengan frekuensi kemunculan sebanyak 31 kali.

Tabel 4. 13 Analisa data *Change %*

Keterangan	Nilai
Jumlah Baris	2811
Keunikan setiap data	482

Data yang sering muncul	0.00%
Frequensi kemunculan	31

Statistika 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 %
<i>Count</i>	2811.00	2811.00	2811.00	2811.00	2811.00	2811.00
<i>Mean</i>	1486.73	1487.06	1497.15	1476.35	135.31	0.67
<i>Std</i>	285.58	285.72	287.94	283.19	122.22	0.69
<i>Min</i>	1049.60	1052.50	1062.70	1045.40	0.12	0.00
25%	1253.00	1252.95	1260.35	1244.25	1.69	0.21
50%	1336.70	1336.90	1345.00	1327.00	139.26	0.47
75%	1781.80	1783.10	1795.15	1772.10	212.82	0.90
<i>MAX</i>	2083.50	2089.50	2140.70	2074.60	816.53	9.34

Tabel 4. 15 Keterangan Statistika Data

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

5. KESIMPULAN

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

- Hasil dari penelitian ini tersebut terbukti bahwa alghoritma *Classification And Regression tree* sangat baik diterapkan untuk memprediksi harga emas sehingga model *machine learning* ini dan dapat mempermudah pemerintah atau perusahaan untuk mengambil keputusan yang lebih efektif dalam sektor investasi.
- Bahwa metode *decision tree* dengan menggunakan alghoritma *Classification And Regression* atau CART dapat diterapkan kedalam model *Machine Learning* untuk memprediksi harga emas berjangka dibandingkan dengan metode linear dengan nilai MSE 61.71, nilai MAE 2.777, nilai RSME 7.885, dan nilai R-square 0.999, dan serta tingkat akurasi sebesar 99.95%.

6. UCAPAN TERIMA KASIH

Penulis menyadari bahwa penyusunan skripsi ini tidak lepas dari bantuan dan dukungan dari berbagai pihak. Oleh karena itu penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Ibu Sri Mulyati S.kom., M.kom selaku dosen pembimbing yang telah memberikan arahan menuntun saya hingga proses penelitian ini berjalan dengan baik.
2. Kepada Kedua orang tua dan teman-teman yang selalu memberikan semangat dan doa serta dukungan kepada penulis agar penulis dapat menyelesaikan penelitian ini.
3. Kepada semua pihak yang tidak disebutkan namanya yang telah membantu dalam penyusunan penelitian ini saya ingin mengucapkan banyak terima kasih.

DAFTAR PUSTAKA

- [1] D. Sari and I. Utami, "Analisa Prediksi Harga Emas Menggunakan Metode Machine Learning," *Jurnal Teknologi Dan Sistem Informasi*, pp. 55-62, 2020.
- [2] Nadir and Sukmana, "Sistem Prediksi Harga Emas Berdasarkan Data Time Series Menggunakan Metode Artificial Neural Network (ANN)," *ITScience*, pp. 426-432, 2023.
- [3] Fikri and Hafid Akbar, "Prediksi Harga Emas Dengan Algoritma," *ejurnal*, pp. 182-189, 2023.
- [4] V. Kumar and R. Singh, "Data Mining Techniques in Marketing: A Review," *International Journal of Business Analytics*, pp. 45-60, 2022.
- [5] L. Sari, H. Rahman and Y. Nugroho, "Hybrid Approaches in Gold Price Forecasting: CART and Time Series Analysis," *Journal of Commodity Markets*, pp. 87-102, 2021.
- [6] Lee and David, "A Comparative Study of Time Series and Machine Learning Models for Gold Price Prediction," *Journal of Computational Finance*, pp. 70-99, 2022.
- [7] S. R. Nayak, "A comprehensive review on stock market prediction using hybrid models," *Journal of King Saud University - Computer and Information Sciences*, pp. 1045-1062, 2021.
- [8] Y. Zhang, X. Chen and J. Li, "Forecasting gold prices using ARIMA model.," *Journal of Commodity Markets*, pp. 100-110, 2020.
- [9] M. Rahman, A. Iskandar and T. Suryanto, "Economic Factors in Gold Price Prediction Using CART," *Journal of Economic Research*, pp. 200-215, 2023.
- [10] G. Kim, "Challenges and Opportunities in Predicting Commodity Prices: Focus on Gold," *Resources Policy*, pp. 100-104, 2024.
- [11] C. Puspitasari, "Binus University," Senin April 2022. [Online]. Available: <https://binus.ac.id/malang/2022/05/sekilas-tentang-data-science-data-mining-dan-machine-learning/>.
- [12] F. Baharuddin and . A. Tjahyanto, "Peningkatan Performa Klasifikasi Machine Learning Melalui Perbandingan Metode Machine Learning dan Meningkatkan *Dataset*," *Jurnal SISFOKOM (Sistem Informasi dan Komputer)*, pp. 25-31, 2022.
- [13] R. W. Romadhonia, . A. Sofro, D. Ariyanto, D. A. Maulana and J. . B. Prihanto, "Application of Decision Trees in Athlete Selection: A Cart Algorithm Approach," *Journal Of Data Science*, pp. 1-9, 2023.
- [14] H. . A. Salman , . A. Kalakech and A. Steiti, "Random Forest Algorithm Overview," *Babylonian Journal of Machine Learning*, pp. 69-79, 2024.
- [15] R. Agustin , S. Defit and Sumijan , "Komparasi Algoritma CART dan C 4.5 pada Citra Tandan Buah Sawit untuk Mengetahui Tingkat Kematangan dalam Penentuan Harga," *Jurnal KomtekInfo*, pp. 263-273, 2024.