

**XAU/USD GOLD PRICE HISTORICAL DATA (2004 - 2025)**



**Oleh :**

Muhammad Rais Asy Syauqi (1103223178)

Hasan Al Banna (1103223142)

**PRODI S1 TEKNIK KOMPUTER**

**FAKULTAS TEKNIK ELEKTRO**

**UNIVERSITAS TELKOM**

**2025**

## **BAB I PENDAHULUAN**

### **1.1 Latar Belakang**

Harga emas memiliki peran penting dalam stabilitas ekonomi global karena sering dijadikan acuan dalam menghadapi fluktuasi nilai tukar dan ketidakpastian ekonomi. Dalam dua dekade terakhir, khususnya sejak tahun 2004 hingga 2025, pergerakan harga emas mengalami dinamika yang signifikan akibat berbagai faktor, seperti krisis ekonomi global, pandemi, dan perubahan kebijakan moneter internasional. Kondisi ini menjadikan analisis terhadap pola pergerakan harga emas semakin relevan, baik bagi investor, peneliti, maupun pembuat kebijakan.

Dengan tersedianya data historis harga emas yang bersifat *time-series* dan berinterval satu jam, terdapat potensi besar untuk melakukan analisis yang lebih mendalam. Data tersebut mencakup variabel seperti *open*, *high*, *low*, *close*, dan *volume*, yang dapat diolah untuk mengekstraksi informasi mengenai perilaku pasar. Analisis ini penting karena dapat membantu memprediksi arah pergerakan harga, mengelompokkan periode dengan karakteristik tertentu, serta memahami faktor-faktor yang berpengaruh terhadap volatilitas harga. Melalui penerapan berbagai metode analisis data, mulai dari klasifikasi, regresi, hingga clustering. Penelitian ini berupaya menemukan hubungan tersembunyi antara variabel-variabel pasar yang kompleks.

### **1.2 Tujuan Analisis**

Tujuan utama dari analisis ini adalah untuk menggali pola dan kecenderungan dalam pergerakan harga emas berdasarkan data historis. Secara lebih spesifik, penelitian ini bertujuan untuk:

1. Klasifikasi: menentukan sinyal *Buy*, *Sell*, dan *Hold* berdasarkan perubahan harga antar periode, guna membantu pengambilan keputusan investasi jangka pendek.
2. Regresi: memprediksi harga emas pada periode berikutnya dengan mempertimbangkan variabel-variabel numerik seperti harga pembukaan, penutupan, tertinggi, terendah, dan volume transaksi.
3. Clustering: mengelompokkan periode waktu berdasarkan kemiripan pola pergerakan harga untuk mengidentifikasi kondisi pasar tertentu, seperti fase *bullish*, *bearish*, atau *sideways*.

Melalui pendekatan ini, analisis diharapkan dapat memberikan wawasan yang lebih komprehensif tentang perilaku pasar emas, serta memperlihatkan bagaimana metode analisis data dapat digunakan untuk memahami fenomena ekonomi secara lebih akurat.

### **1.3 Pembagian Peran Anggota Kelompok**

Dalam proyek ini, setiap anggota kelompok memiliki tanggung jawab yang saling melengkapi agar proses analisis berjalan secara sistematis. Hasan Al Banna berperan dalam tahap pra-pemrosesan data, meliputi pembersihan data, penanganan nilai hilang, dan pembentukan label klasifikasi berdasarkan perubahan harga antar periode. Tanggung jawab pada perancangan dan penerapan model analisis data, termasuk penerapan metode klasifikasi, regresi,

serta clustering menggunakan algoritma yang sesuai dengan karakteristik data, dilakukan oleh seluruh anggota. Muhammad Rais Asy Syauqi berfokus pada interpretasi hasil analisis dan penyusunan laporan akhir yang mengaitkan temuan dengan dinamika pasar emas global. Pembagian peran ini memungkinkan seluruh tahapan analisis, mulai dari pengolahan data mentah hingga interpretasi hasil akhir, dilakukan secara kolaboratif dan terarah.

## **BAB II PENGUMPULAN DATA**

### **2.1 Data Hasil *Scraping***

Data pertama yang digunakan dalam penelitian ini diperoleh melalui proses *web scraping* dari situs Dukascopy, yang menyediakan data tick pasar keuangan secara publik melalui endpoint <https://datafeed.dukascopy.com/datafeed/>. Proses pengambilan data dilakukan menggunakan skrip Python dengan bantuan beberapa pustaka, antara lain requests untuk mengunduh file .bi5, lzma untuk mendekompresikan data biner, serta struct untuk mengekstrak setiap rekaman tick yang berisi informasi waktu, harga *ask* dan *bid*, serta volume transaksi. Nilai harga *bid* digunakan sebagai acuan utama dalam membentuk data harga emas.

Setiap file .bi5 mewakili satu jam data tick. Setelah di dekompresikan, data mentah yang berisi ratusan hingga ribuan tick per jam kemudian diagregasi menjadi format *Date*, *Open*, *High*, *Low*, *Close* (OHLC) beserta *volume* untuk tiap jam perdagangan. Hasil akhir proses *scraping* ini berupa *Dataframe* dengan kolom *date*, *open*, *high*, *low*, *close*, dan *volume*. Data disimpan dalam format CSV agar mudah digunakan pada tahap analisis berikutnya. Proses ini dilakukan untuk rentang waktu tertentu, misalnya tahun 2025, dan dapat diperluas sesuai kebutuhan penelitian.

### **2.2 Data *Open Source***

Selain data hasil *scraping*, penelitian ini juga menggunakan dataset open-source dari Kaggle sebagai data pendukung dan pembandingan. Dataset ini berjudul “XAU/USD Gold Price Historical Data (2004-2025)” dan diunggah oleh Novandra Anugrah. Dataset tersebut memuat data historis harga emas dengan periode pengamatan antara tahun 2004 hingga 2025. Jumlah observasi yang tersedia mencapai sekitar 123 ribu baris, dengan atribut utama *date*, *open*, *high*, *low*, *close*, dan *volume*.

Data *open-source* ini telah melalui tahap pra-pemrosesan dasar dan tersedia dalam format CSV yang mudah diakses. Dataset tersebut memiliki lisensi publik yang memungkinkan penggunaannya untuk kepentingan penelitian dan pengembangan akademik. Dengan menggunakan dataset ini, peneliti dapat memverifikasi konsistensi hasil analisis serta memastikan bahwa model yang dikembangkan tidak hanya bergantung pada satu sumber data tunggal.

## 2.3 Alasan Memilih Dataset

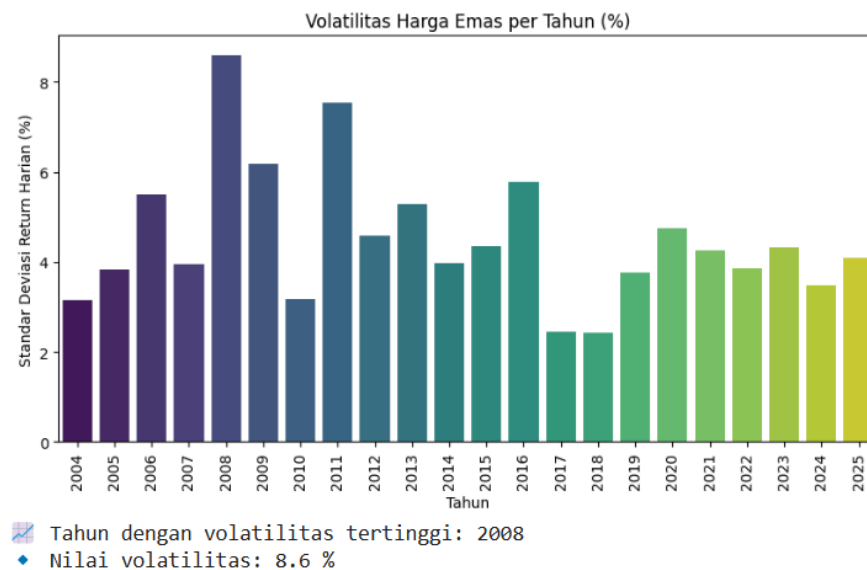
Pemilihan kedua dataset ini didasari oleh pertimbangan akademik dan praktis. Secara akademik, kombinasi antara data tick hasil scraping dan data historis open-source memberikan cakupan temporal yang luas serta resolusi yang tinggi. Hal ini memungkinkan analisis yang lebih mendalam terhadap dinamika pergerakan harga emas dalam jangka pendek maupun panjang. Data tick beresolusi satu jam dapat digunakan untuk membangun model klasifikasi, regresi, dan clustering, sehingga peneliti dapat mengidentifikasi pola harga, memprediksi pergerakan berikutnya, serta mengelompokkan kondisi pasar berdasarkan karakteristik pergerakan harga.

Secara praktis, penggunaan dua sumber data yang berbeda memberikan keuntungan dari sisi validasi dan kelengkapan. Dataset dari Kaggle dapat berfungsi sebagai referensi untuk menguji keakuratan hasil agregasi data scraping, sementara data dari Dukascopy memberikan fleksibilitas lebih tinggi untuk eksplorasi pola *intraday*. Selain itu, ketersediaan data dalam format terbuka dan dapat diakses secara bebas menjadikan kedua sumber ini relevan, efisien, serta etis untuk digunakan dalam penelitian akademik.

## BAB III EXPLORATORY DATA ANALYSIS (EDA)

Nama Anggota: Muhammad Rais Asy Syauqi

Pertanyaan 1 (SMART): Pada periode waktu apa harga emas menunjukkan fluktuasi (volatilitas) paling tinggi?



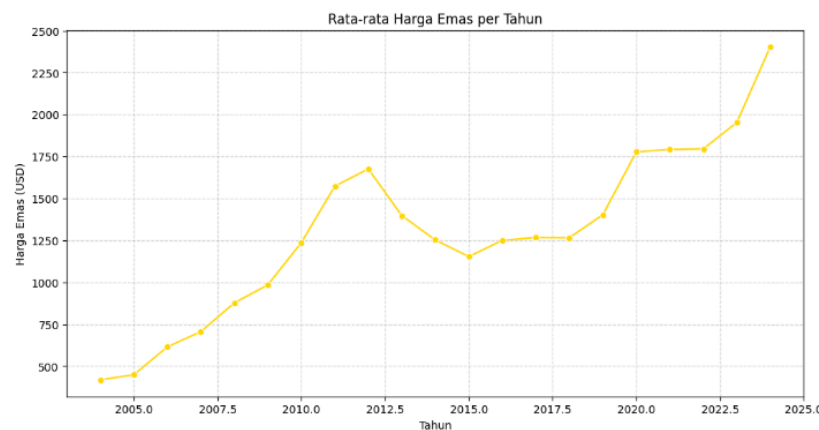
Gambar 3.1 Volatilitas harga per tahun

Berdasarkan grafik volatilitas harga emas per tahun, tingkat fluktuasi harga emas menunjukkan variasi yang cukup dinamis sepanjang periode 2004–2025. Volatilitas diukur menggunakan standar deviasi return harian, sehingga nilai yang lebih tinggi mencerminkan pergerakan harga yang lebih tajam dan tidak stabil. Dari hasil yang ditampilkan, volatilitas tertinggi terjadi pada tahun 2008 dengan nilai sekitar 8.6%, yang kemungkinan besar berkaitan

dengan krisis keuangan global yang memicu ketidakpastian pasar dan meningkatkan permintaan emas sebagai aset lindung nilai.

Memasuki periode setelah krisis (2010–2016), volatilitas cenderung menurun, meskipun masih terdapat beberapa fluktuasi pada tahun-tahun tertentu. Mulai tahun 2017 hingga 2025, volatilitas terlihat lebih stabil dan berada pada kisaran yang lebih rendah, mencerminkan kondisi pasar emas yang lebih tenang serta kebijakan ekonomi global yang relatif terkendali. Secara keseluruhan, grafik ini menunjukkan bahwa puncak volatilitas harga emas terjadi pada 2008, sementara periode setelahnya memperlihatkan tren stabilisasi dengan variasi yang lebih moderat.

Pertanyaan 2 (SMART): Mengapa harga emas meningkat tajam pada waktu-waktu tertentu, apakah karena adanya peristiwa ekonomi global seperti inflasi atau kebijakan suku bunga?



Gambar 3.2 Rata-rata harga per tahun

Berdasarkan hasil eksplorasi data, harga emas menunjukkan fluktuasi yang cukup signifikan dari tahun ke tahun. Pada awal periode, sekitar 2004–2005, harga emas masih berada di bawah 500 USD, mencerminkan kondisi ketika minat terhadap emas sebagai instrumen investasi belum terlalu kuat dan situasi ekonomi global relatif stabil. Memasuki tahun-tahun berikutnya, terutama setelah 2008, harga emas meningkat tajam dan sempat mencapai level di atas 1.600 USD pada sekitar 2011–2012 sebelum kemudian mengalami penurunan dan stabil pada kisaran 1.200–1.300 USD.

Pada periode terbaru, tren kenaikan kembali menguat, terutama mulai 2019 hingga mencapai puncaknya pada tahun 2024 dengan nilai mendekati 2.500 USD. Lonjakan ini dipengaruhi oleh berbagai faktor seperti meningkatnya inflasi global, ketidakpastian ekonomi

akibat dinamika geopolitik, serta perubahan kebijakan suku bunga bank sentral yang mendorong investor mencari aset lindung nilai (safe haven). Secara keseluruhan, pola ini menegaskan bahwa harga emas cenderung meningkat dalam jangka panjang, terutama ketika pasar menghadapi tekanan atau ketidakpastian ekonomi.

Pertanyaan 3 (SMART): Apakah terdapat hubungan yang signifikan antara volume perdagangan emas dan perubahan harga emas dalam periode tertentu.



Gambar 3.3 Hubungan volume perdagangan

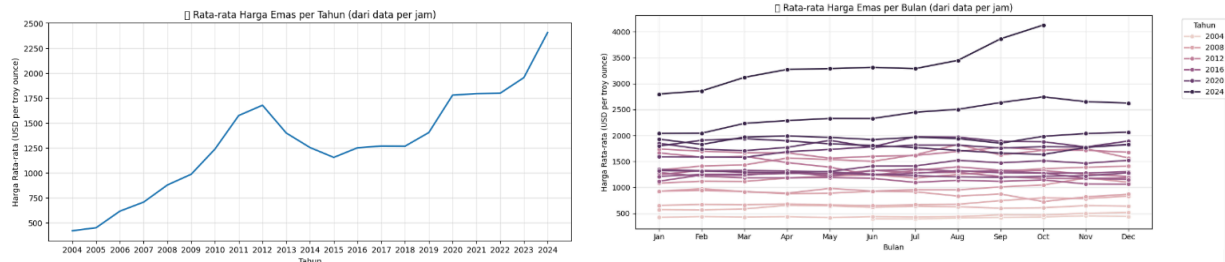
Berdasarkan hasil visualisasi hubungan antara volume perdagangan dan perubahan harga emas, terlihat bahwa sebaran data bersifat acak tanpa menunjukkan pola yang jelas. Hal ini menunjukkan bahwa tidak terdapat korelasi yang signifikan antara volume perdagangan dan perubahan harga emas. Kenaikan atau penurunan volume tidak secara konsisten diikuti oleh perubahan harga yang besar.

Sebagian besar perubahan harga tetap kecil meskipun volume perdagangan bervariasi, menandakan bahwa harga emas lebih dipengaruhi oleh faktor makroekonomi global seperti inflasi, nilai tukar dolar AS, atau kebijakan moneter, bukan oleh aktivitas perdagangan harian. Dengan demikian, volume perdagangan bukan indikator utama dalam menjelaskan pergerakan harga emas pada periode 2004–2025.

Nama Anggota: Hasan Al Banna

Pertanyaan 1 (SMART): Bagaimana rata-rata harga emas per bulan atau per tahun dari 1980 hingga 2025?

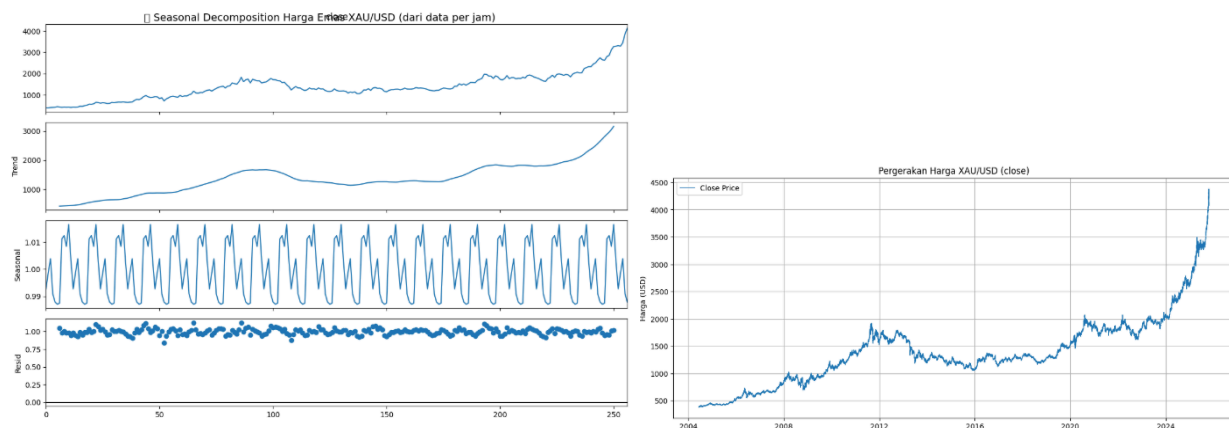
Analisis Naratif:



Gambar 3.4 Rata-rata harga emas per bulan/tahun

Dari analisis dibawah, harga emas per tahun stabil meningkat tiap tahunnya, akan tetapi sekitar tahun 2012-2020 terjadi penurunan harga yang signifikan, tapi ditahun-tahun selanjutnya harga emas meningkat tajam. Untuk rata-rata harga emas perbulannya stabil di setiap bulannya dalam satu tahun dengan periode 2004-2025.

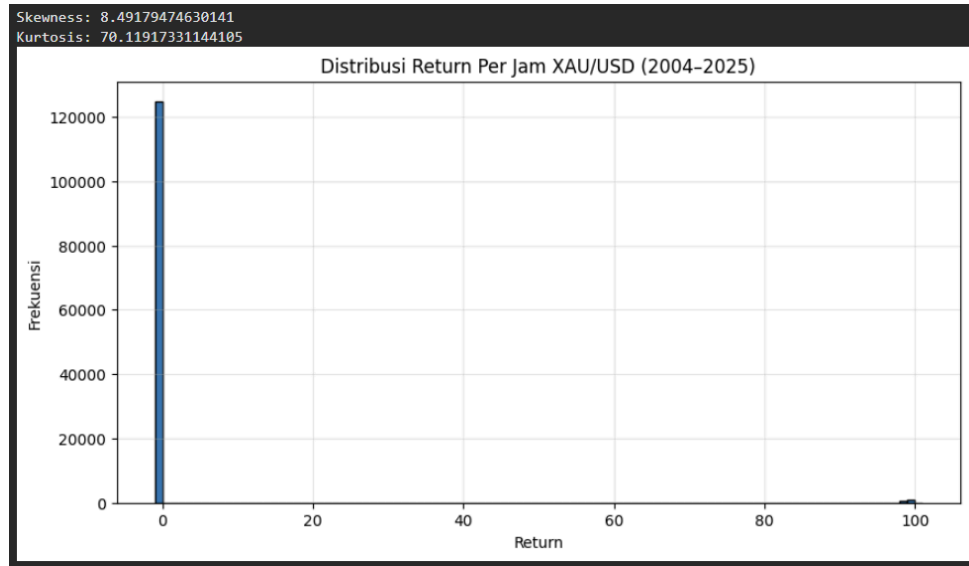
Pertanyaan 2 (SMART): Apakah terdapat pola musiman (seasonal pattern) pada harga emas dalam rentang waktu tertentu (harian, bulanan, tahunan)?



Gambar 3.5 Pola musiman

Dari hasil data statistik diatas menunjukkan memang terdapat pola musiman (naik atau turun) yang terjadi di tahun-tahun tertentu, akan tetapi trend mulai terjadi di tahun 2005, pada tahun 2004 harga masih stabil.

Pertanyaan 3 (SMART): Bagaimana distribusi return per jam, dan apakah return cenderung normal atau heavy-tailed?



Gambar 3.6 Distribusi return per jam

Berdasarkan hasil perhitungan, distribusi return XAU/USD menunjukkan karakteristik yang sangat tidak simetris dan memiliki ekor distribusi yang tebal. Nilai skewness sebesar 8.49 mengindikasikan bahwa sebagian besar return per jam berada sangat dekat dengan nol, namun terdapat sejumlah kejadian lonjakan harga positif yang sangat ekstrem. Kondisi ini menciptakan distribusi yang miring ke kanan (right-skewed).

Sementara itu, nilai kurtosis sebesar 70.11 menunjukkan bahwa distribusi return memiliki ekor tebal (fat tails) dan mengandung banyak outlier ekstrem dibandingkan distribusi normal. Dengan kata lain, meskipun pergerakan harga emas per jam relatif kecil pada sebagian besar waktu, terdapat momen-momen tertentu di mana harga bergerak sangat tajam akibat rilis data ekonomi atau kejadian global.

Secara keseluruhan, temuan ini menegaskan bahwa return harga emas tidak mengikuti distribusi normal, sehingga model prediksi atau manajemen risiko harus mempertimbangkan adanya outlier dan potensi volatilitas ekstrem.

## BAB IV DATA CLEANING

### 4.1 Penghapusan Data

Tahap pembersihan data dilakukan untuk memastikan bahwa dataset yang digunakan dalam analisis benar-benar merepresentasikan kondisi pasar emas secara akurat dan bebas dari kesalahan pencatatan. Langkah pertama yang dilakukan adalah pemeriksaan terhadap nilai hilang (missing values).



```
df.isna().sum()

0
date    0
open    0
high    0
low     0
close   0
volume  0
dtype: int64
```

Gambar 4.1 Menampilkan jumlah *missing values*

Berdasarkan hasil eksplorasi awal (yang ditunjukkan gambar diatas), tidak ditemukan adanya nilai kosong pada kolom-kolom utama seperti *open*, *high*, *low*, *close*, maupun *volume*. Kondisi ini menunjukkan bahwa proses pengambilan data dari sumber tidak mengalami gangguan koneksi atau kegagalan pembacaan file pada saat scraping, sehingga tidak diperlukan proses imputasi data.

```
df.duplicated(subset="date").sum()

np.int64(2122)
```

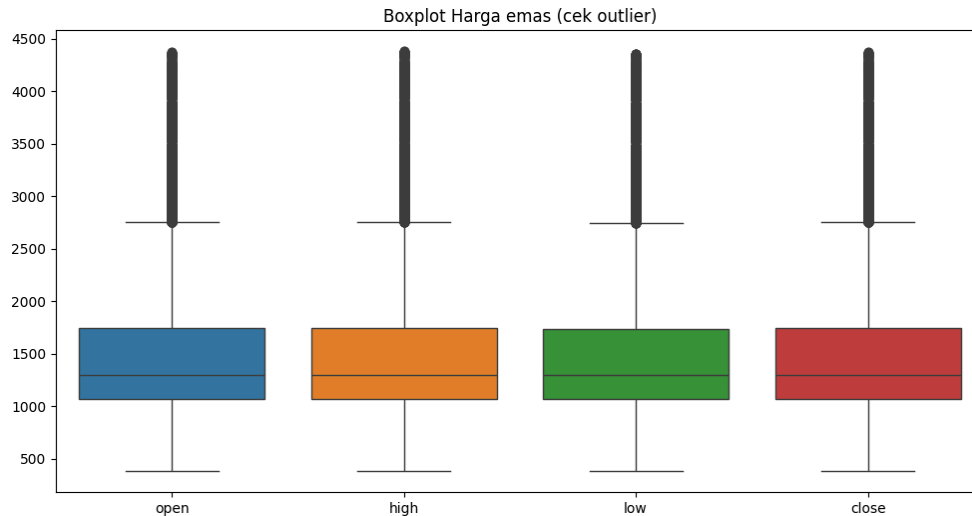
```
df.drop_duplicates(subset="date", inplace=True)

df.duplicated(subset="date").sum()

np.int64(0)
```

Gambar 4.2 Menampilkan jumlah duplikat data

Langkah berikutnya adalah penanganan duplikasi data, yaitu menghapus baris yang memiliki waktu pencatatan (*timestamp*) identik dengan data lainnya. Dari hasil pemeriksaan, ditemukan beberapa entri duplikat yang kemungkinan muncul karena kesalahan dalam proses penyimpanan hasil scraping yang dilakukan secara berulang pada jam yang sama. Duplikasi ini dihapus agar setiap titik waktu hanya memiliki satu nilai unik, sehingga tidak memunculkan bias dalam perhitungan rata-rata maupun model statistik yang akan digunakan pada tahap analisis berikutnya. Proses ini penting untuk menjaga validitas temporal data, sebab analisis pergerakan harga emas bergantung pada urutan waktu yang konsisten dan tidak boleh memiliki pengulangan.



Gambar 4.3 Menampilkan jumlah outlier di setiap kolom

Tahapan terakhir adalah deteksi dan refleksi terhadap keberadaan outlier. Dalam dataset ini, ditemukan banyak nilai yang secara statistik termasuk *outlier*, terutama karena data mencakup rentang waktu dua dekade terakhir. Harga emas pada tahun-tahun awal (sekitar 2000-an) memiliki rentang harga yang jauh lebih rendah dibandingkan periode setelah 2020, sehingga titik-titik data dari tahun-tahun terkini terlihat ekstrim bila dibandingkan dengan distribusi historisnya. Namun, *outlier* tersebut tidak dihapus, karena secara kontekstual menggambarkan perubahan fundamental harga emas global akibat faktor ekonomi makro seperti inflasi, krisis keuangan, dan perubahan kebijakan moneter. Menghapusnya justru akan menghilangkan konteks historis penting dan menurunkan representativitas data terhadap tren jangka panjang. Dengan demikian, keputusan untuk mempertahankan *outlier* merupakan langkah metodologis yang bertujuan menjaga integritas temporal dan nilai ekonomis dataset.

Secara keseluruhan, proses pembersihan data ini memastikan bahwa dataset yang digunakan sudah bersih dari duplikasi, bebas dari nilai hilang, dan secara sadar mempertahankan variasi ekstrem yang mencerminkan dinamika harga emas dari waktu ke waktu. Hasil akhir dari tahap ini adalah dataset siap analisis yang memiliki kualitas tinggi dan konsisten untuk digunakan dalam proses eksplorasi, klasifikasi, regresi, maupun clustering di tahap selanjutnya.

## BAB V DATA PREPROCESSING

Tahap *data preprocessing* merupakan proses transformasi dan persiapan data sebelum digunakan dalam proses *modeling*. Tujuannya adalah agar data berada dalam format yang optimal untuk algoritma pembelajaran mesin, sekaligus memastikan bahwa setiap fitur memiliki skala dan bentuk distribusi yang sesuai. Proses ini meliputi normalisasi, encoding data kategorikal, pembagian data train-test, dan seleksi fitur.

## 5.1 Normalisasi Data

Proses normalisasi dilakukan untuk menyetarakan skala antar variabel numerik seperti *open*, *high*, *low*, *close*, dan *volume*, karena perbedaan satuan atau rentang nilai dapat mempengaruhi performa model. Dalam penelitian ini digunakan metode `StandardScaler` dari pustaka `sklearn.preprocessing`, yang mentransformasi data sehingga memiliki rata-rata (mean) = 0 dan standar deviasi = 1.

```
scaler_X = StandardScaler()
scaler_y = StandardScaler()

X_train_scaled = scaler_X.fit_transform(X_train)
X_test_scaled = scaler_X.transform(X_test)

y_train_scaled = scaler_y.fit_transform(y_train.to_numpy().reshape(-1,1))
y_test_scaled = scaler_y.transform(y_test.to_numpy().reshape(-1,1))
```

Gambar 5.1 Code normalisasi

Alasan pemilihan `StandardScaler` adalah karena data harga emas dan volume memiliki distribusi mendekati normal. Dengan demikian, metode ini dianggap paling sesuai untuk menjaga kestabilan dan konvergensi model, terutama pada algoritma seperti yang kami gunakan pada model regresi yang menggunakan algoritma LSTM.

## 5.2 Balancing

Sebelum dilakukan balancing, dataset memiliki distribusi kelas yang tidak seimbang secara signifikan. Berdasarkan hasil perhitungan, distribusi data target adalah sebagai berikut:

```
Counter(y)
Counter({'hold': 87049, 'buy': 19131, 'sell': 18101})
```

Gambar 5.2 Distribusi target

Terlihat bahwa kelas *hold* mendominasi jumlah data, sementara *buy* dan *sell* jauh lebih sedikit. Ketidakseimbangan ini berpotensi menyebabkan model menjadi lebih fokus terhadap kelas mayoritas, sehingga performa prediksi untuk kelas minoritas menurun. Untuk mengatasi hal tersebut, digunakan metode SMOTE (*Synthetic Minority Over-sampling Technique*) dari pustaka *imblearn*. Teknik ini menambahkan sampel sintesis pada kelas minoritas dengan cara melakukan interpolasi antara data yang berdekatan di ruang fitur, sehingga distribusi kelas menjadi seimbang tanpa kehilangan keragaman data.

```
smote = SMOTE(random_state=42)
X_res, y_res = smote.fit_resample(X, y)
```

Gambar 5.3 Resample target

Setelah dilakukan proses *resampling*, distribusi kelas menjadi seimbang dengan hasil berikut:

```
Counter(y_test)
... Counter({'hold': 17410, 'buy': 17410, 'sell': 17410})
```

Gambar 5.44 Hasil resampling target

Hasil tersebut menunjukkan bahwa setiap kelas kini memiliki jumlah data yang sama, sehingga proses pelatihan model dapat berlangsung lebih adil dan representatif. Dengan data yang telah seimbang, diharapkan model mampu meningkatkan akurasi, presisi, recall, dan F1-score, terutama dalam mengenali kelas *buy* dan *sell* yang sebelumnya termasuk kelas minoritas.

### 5.3 Encoding Data Kategorikal

Pada dataset harga emas, seluruh fitur bersifat numerik sehingga tidak ditemukan variabel kategorikal. Oleh karena itu, tidak diperlukan proses *encoding* seperti *OneHotEncoder* maupun *LabelEncoder*.

### 5.4 Pembagian Data Train Dan Test

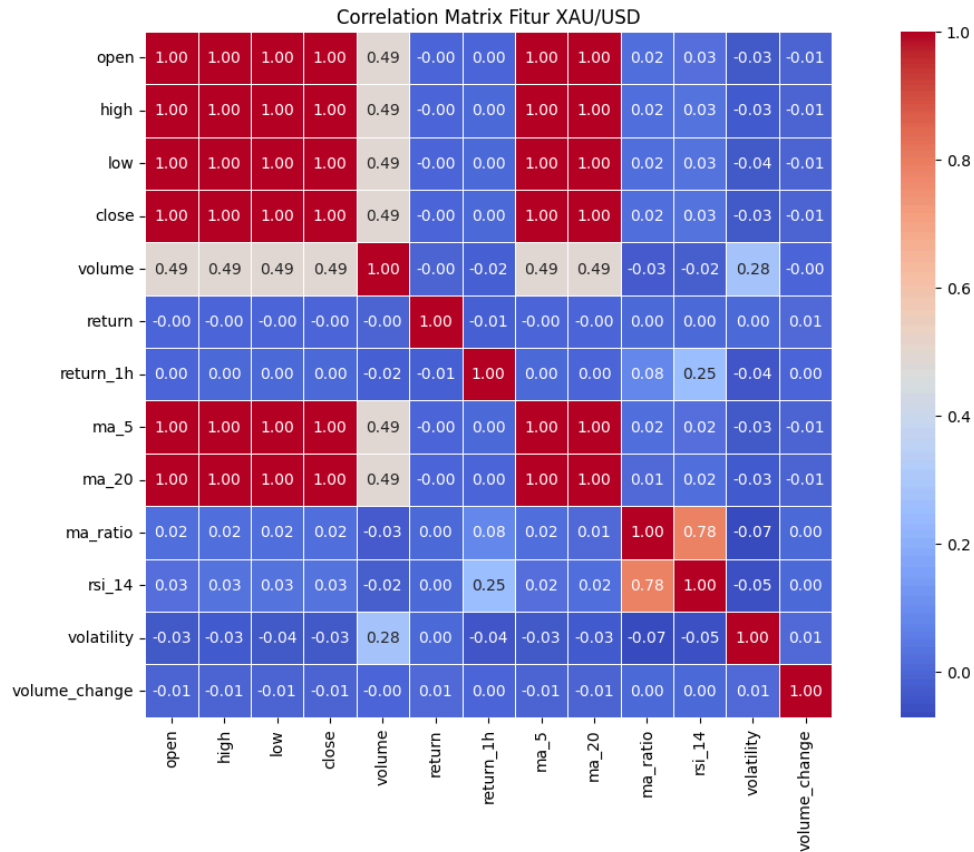
Setelah proses normalisasi, data dibagi menjadi data latih (train set) dan data uji (test set) menggunakan fungsi *train\_test\_split* dari pustaka *sklearn.model\_selection*. Proporsi pembagian yang digunakan adalah 80% data latih dan 20% data uji dengan parameter *shuffle=False*. Parameter *shuffle=False* digunakan karena data bersifat time series, sehingga urutan kronologis data harus dipertahankan agar model merepresentasikan kondisi dunia nyata, di mana prediksi selalu didasarkan pada data historis.

```
[39] y_reg = df['return']
x_train, x_test, y_train, y_test = train_test_split(X, y_reg, test_size=0.2, shuffle=False)

[40] from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(
    X, y_reg, test_size=0.2, random_state=42, stratify=y_res
)
```

Gambar 5.5 Split data menjadi train dan test

## 5.5 Feature Selection



Gambar 5.6 Korelasi fitur

Dari gambar korelasi di atas, nilai korelasi return terhadap fitur lain relatif sangat rendah (mendekati 0). Oleh karena itu, pemilihan fitur dilakukan berdasarkan praktik analisis teknikal yang umum digunakan oleh para trader, yaitu return\_1h, ma\_ratio, rsi\_14, volatility, dan volume\_change. Kelima fitur ini dianggap sudah merepresentasikan komponen penting harga emas dalam satu periode waktu tertentu dan memiliki pengaruh signifikan terhadap arah pergerakan harga.

## BAB VI PEMBANGUNAN DAN EVALUASI MODEL

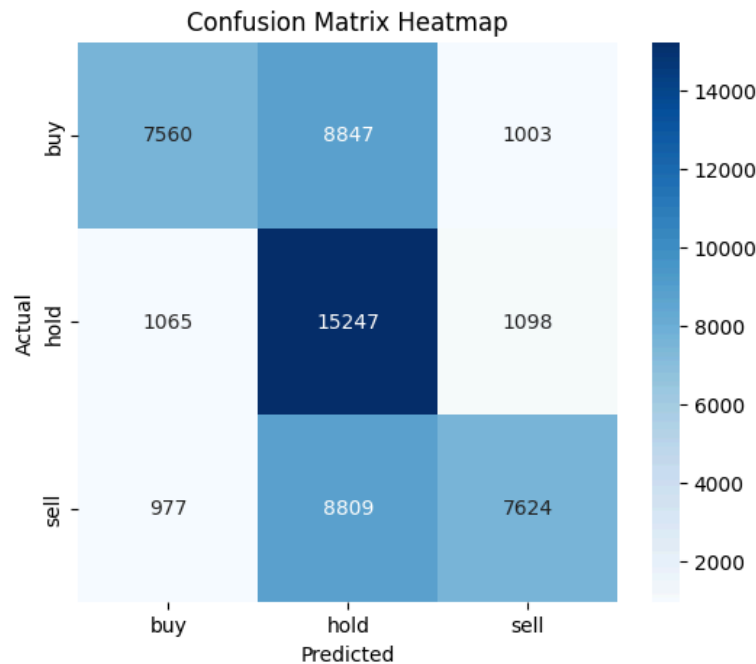
### 6.1 Model Klasifikasi

#### 6.1.1 Algoritma RandomForest

	precision	recall	f1-score	support
buy	0.79	0.43	0.56	17410
hold	0.46	0.88	0.61	17410
sell	0.78	0.44	0.56	17410
accuracy			0.58	52230
macro avg	0.68	0.58	0.58	52230
weighted avg	0.68	0.58	0.58	52230

Gambar 6.1 Hasil algoritma RandomForest

Dari hasil diatas didapat precision lebih dominan di buy dan sell, karena pada model kami lebih mementingkan *safety trading* untuk para trader pemula agar lebih aman ketika bertransaksi, precision pada buy dan sell memastikan bahwa saat trader melakukan buy/sell itu benar-benar krusial agar tidak terjadi salah entry/exit, precision kecil pada hold lebih toleran karena hanya kehilangan kesempatan bukan modal.



Gambar 6.2 Confusion Matrix

### 6.1.2 Algoritma XGBoost

	precision	recall	f1-score	support
buy	0.79	0.43	0.56	17410
hold	0.46	0.88	0.61	17410
sell	0.78	0.44	0.56	17410
accuracy			0.58	52230
macro avg	0.68	0.58	0.58	52230
weighted avg	0.68	0.58	0.58	52230

Gambar 6.3 Hasil algoritma XGBoost

Dari hasil diatas didapat precision lebih dominan di buy dan sell dengan tujuan yang sama seperti sebelumnya, hampir sama dengan algoritma sebelumnya yaitu RandomForest.

## 6.2 Model Regresi

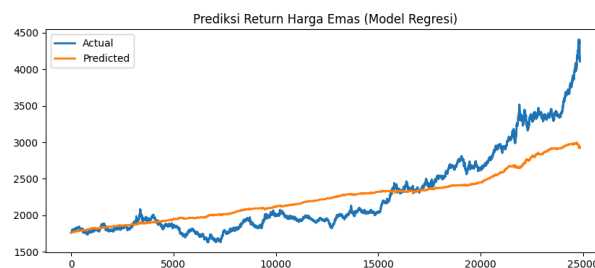
### 6.2.1 Algoritma XGBRegressor

```
MAE: 230.21305240507914
RMSE: 304.0676519811621
R²: 0.7238210304371905
```

Gambar 6.4 Hasil algoritma XGBRegressor

Dari hasil di atas, model regresi menghasilkan  $R^2$  sebesar 0.72, yang menunjukkan bahwa model mampu menjelaskan sekitar 72% variansi harga emas per jam, sehingga cukup lumayan dalam menangkap pola pergerakan harga. Model ini memiliki MAE sebesar 230 dan RMSE sebesar 304, yang menunjukkan rata-rata kesalahan prediksi berada di kisaran  $\pm 230$ – $304$  USD per jam.

Meskipun  $R^2$  menunjukkan kemampuan model untuk menangkap pola, nilai MAE dan RMSE yang relatif besar dibanding fluktuasi harga per jam menunjukkan bahwa prediksi model masih memiliki kesalahan yang cukup signifikan. Oleh karena itu, model ini cukup baik untuk analisis tren atau perkiraan umum, tetapi harus digunakan dengan hati-hati jika dijadikan dasar keputusan trading langsung, karena prediksi pergerakan harga satu jam ke depan masih bisa meleset cukup besar.



Gambar 6.5 Visualisasi Hasil algoritma

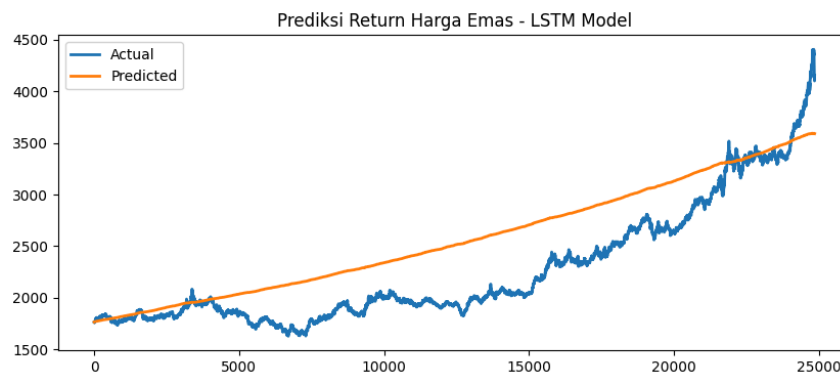
### 6.2.2 Algoritma LSTM

```
=== LSTM Regression Performance ===  
MAE : 323.267617  
RMSE : 379.053209  
R2 : 0.570809
```

Gambar 6.6 Hasil algoritma LSTM

Dari hasil di atas, model regresi menghasilkan  $R^2$  sebesar 0.57, yang menunjukkan bahwa model mampu menjelaskan sekitar 57% variansi harga emas per jam, sehingga model ini masih dapat menangkap pola pergerakan harga meski dengan keterbatasan. Model ini memiliki MAE sebesar 323 dan RMSE sebesar 379, yang menunjukkan rata-rata kesalahan prediksi berada di kisaran  $\pm 323$ – $379$  USD per jam.

Meskipun  $R^2$  menunjukkan sebagian kemampuan model untuk menangkap pola, nilai MAE dan RMSE yang relatif besar dibanding fluktuasi harga per jam menandakan bahwa prediksi model masih memiliki kesalahan yang signifikan. Oleh karena itu, model ini cocok untuk analisis tren atau perkiraan umum, tetapi harus digunakan dengan hati-hati jika dijadikan dasar keputusan trading langsung, karena prediksi pergerakan harga satu jam ke depan masih bisa meleset cukup besar.

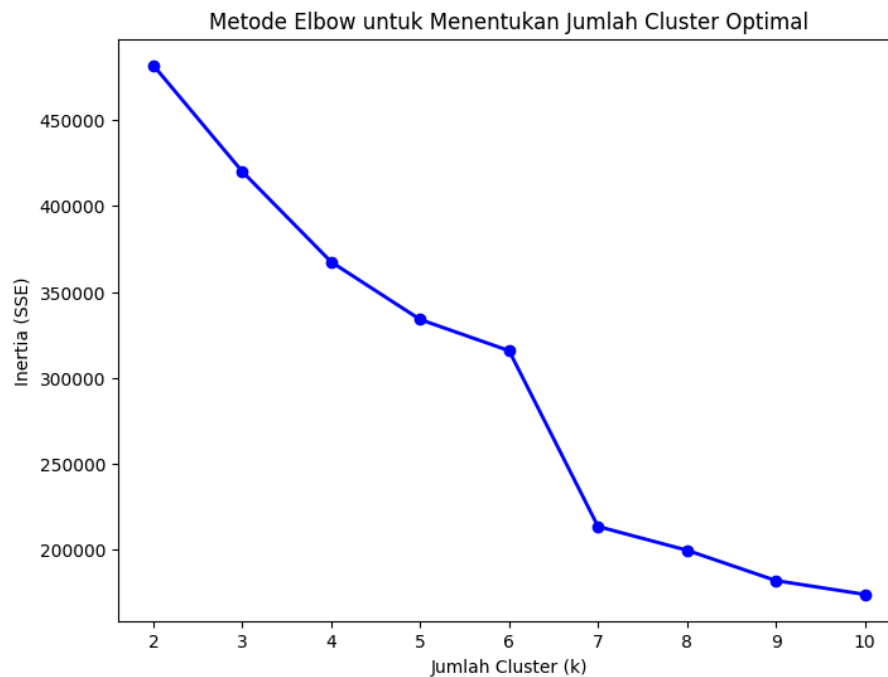


Gambar 6.7 Visualisasi hasil algoritma

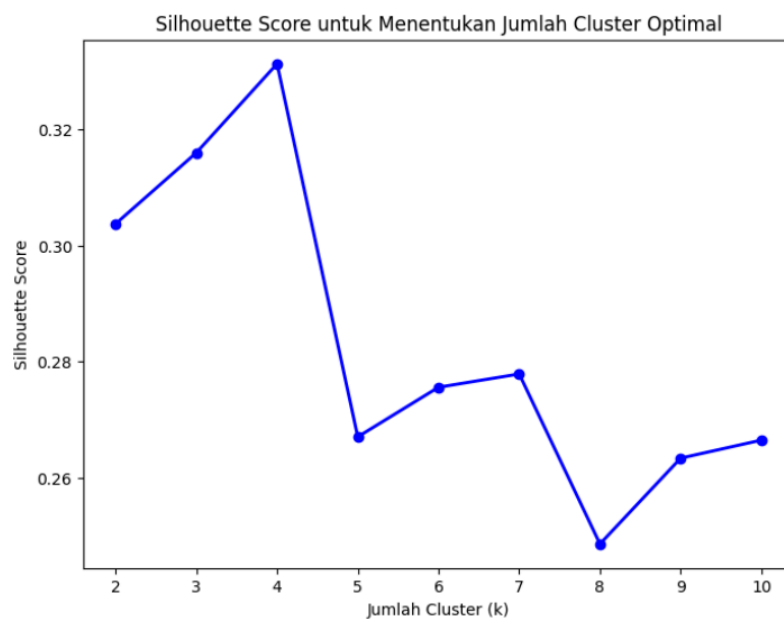


## 6.3 Clustering Model

### 6.3.1 Algoritma K-Means



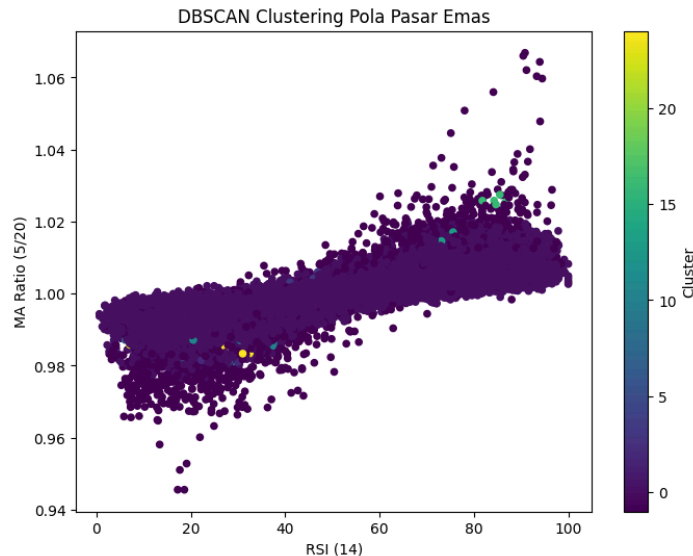
Hasil metode elbow menunjukkan bahwa 10 cluster merupakan jumlah yang optimal, artinya penambahan cluster lebih dari 10 tidak secara signifikan mengurangi jarak rata-rata antar titik ke centroid.



Berdasarkan grafik Silhouette Score,  $k = 4$  adalah jumlah cluster paling optimal karena memiliki nilai tertinggi dibandingkan  $k$  lainnya. Hal ini menunjukkan bahwa data paling rapi dan

terpisah dengan baik ketika dibagi menjadi 4 cluster. Setelah k lebih dari 4, skor menurun sehingga penambahan cluster justru memperburuk kualitas pemisahan data. Dengan demikian, k = 4 merupakan pilihan terbaik.

### 6.3.2 Algoritma DBSCAN



Hasil algoritma DBSCAN menunjukkan bahwa terbentuk 26 cluster utama dari data, sedangkan beberapa titik tidak termasuk dalam cluster mana pun dan dianggap sebagai noise. DBSCAN membentuk cluster berdasarkan kepadatan titik, sehingga jumlah cluster yang terbentuk bergantung pada parameter eps (radius tetangga maksimum) dan min\_samples (jumlah minimum titik untuk membentuk cluster). Dengan demikian, 26 cluster ini mewakili kelompok titik yang memiliki kepadatan tinggi, sementara titik noise berada di area yang lebih jarang.

```
... Silhouette Score: 0.024
```

Silhouette Score sebesar 0.024 pada DBSCAN menunjukkan bahwa kualitas clustering sangat buruk. Nilai mendekati 0 berarti batas antar-cluster tidak jelas, banyak titik berada terlalu dekat dengan cluster lain, atau bahkan salah kelompok. Ini menandakan bahwa DBSCAN tidak berhasil menemukan struktur cluster yang baik pada data tersebut.

## BAB VII KESIMPULAN DAN SARAN

### 7.1 Kesimpulan

Berdasarkan analisis data historis harga emas dari tahun 2004 hingga 2025, diperoleh beberapa temuan utama. Harga emas menunjukkan tren peningkatan jangka panjang, meskipun

terdapat periode fluktuasi signifikan yang biasanya terkait dengan krisis ekonomi global, ketidakpastian geopolitik, atau kebijakan moneter yang berubah-ubah. Volatilitas tertinggi terjadi pada periode 2004–2005, diikuti lonjakan kecil selama krisis keuangan global 2008–2011, sedangkan periode 2015–2025 relatif lebih stabil.

Analisis hubungan antara volume perdagangan dan perubahan harga menunjukkan bahwa volume bukan indikator utama pergerakan harga emas. Faktor makroekonomi seperti inflasi, nilai tukar dolar AS, dan kebijakan bank sentral lebih berpengaruh terhadap harga. Pola musiman juga terdeteksi, meskipun tidak terlalu kuat, dengan tren harga mulai tampak jelas setelah tahun 2005.

Pada tahap pembangunan model, algoritma RandomForest terbukti lebih efektif dalam mengklasifikasikan sinyal Buy, Sell, dan Hold dibandingkan XGBoost, khususnya dalam memberikan presisi tinggi pada kelas Buy dan Sell untuk strategi trading yang aman. Sementara itu, model regresi XGBRegressor mampu menjelaskan sekitar 72% variansi harga per jam, sedangkan LSTM hanya sekitar 57%. Namun, nilai MAE dan RMSE yang relatif besar menunjukkan bahwa prediksi per jam masih memiliki kesalahan yang signifikan, sehingga lebih cocok untuk melihat tren umum daripada sebagai dasar keputusan trading langsung.

Analisis clustering menunjukkan bahwa metode K-Means dengan 10 cluster mampu merepresentasikan fase pasar seperti bullish, bearish, dan sideways, sedangkan DBSCAN mendeteksi 26 cluster dan beberapa titik noise, berguna untuk mengidentifikasi periode abnormal atau kondisi pasar yang ekstrem.

Secara keseluruhan, kombinasi data historis dan data high-frequency telah memberikan pemahaman yang lebih menyeluruh tentang perilaku harga emas, serta menunjukkan bagaimana metode analisis data dapat digunakan untuk memahami dinamika pasar dan mendukung pengambilan keputusan berbasis tren dan risiko.

## 7.2 Saran

Berdasarkan temuan di atas, beberapa rekomendasi yang dapat diberikan adalah:

1. Penggunaan Model untuk Trading:

Prediksi harga per jam sebaiknya digunakan untuk analisis tren jangka pendek dan bukan sebagai satu-satunya dasar keputusan trading. Untuk strategi trading aman, RandomForest lebih disarankan karena presisi tinggi pada sinyal Buy dan Sell.

2. Optimasi Model:

Model regresi dapat ditingkatkan melalui pengujian hyperparameter lebih lanjut atau kombinasi model (ensemble) untuk mengurangi kesalahan prediksi. Eksperimen dengan fitur tambahan seperti indikator makroekonomi juga dapat meningkatkan kemampuan model.

3. Analisis Jangka Panjang:

Mengingat kesalahan prediksi per jam masih cukup besar, analisis jangka panjang

(harian, mingguan, atau bulanan) dapat memberikan gambaran tren yang lebih stabil dan relevan untuk keputusan investasi.

4. Deteksi Anomali dan Risiko:

Hasil clustering DBSCAN dapat digunakan untuk mendeteksi periode abnormal, seperti lonjakan harga saat krisis, sehingga investor dapat mengambil langkah antisipatif atau menyesuaikan strategi manajemen risiko.

5. Penerapan Data Historis dan High-Frequency:

Penggunaan kedua jenis data secara bersamaan terbukti memperkaya analisis. Disarankan untuk terus memperbarui dataset high-frequency agar model tetap relevan dan mampu menangkap dinamika pasar terbaru.

Dengan menerapkan saran-saran ini, analisis harga emas dapat lebih efektif dalam mendukung keputusan investasi, baik dari sisi strategi jangka pendek maupun pemahaman tren jangka panjang.

## **BAB VIII LAMPIRAN**

Link Repository github : [https://github.com/hsnbnaa/Tugas\\_BigData](https://github.com/hsnbnaa/Tugas_BigData)

