

## ***Analisis Feature Importance Random Forest untuk Faktor Penentu Harga Emas Antam***

**Annisa Hestiningtyas<sup>1</sup>, Sri Wahyuningsih<sup>2</sup>**

Universitas Islam Negeri Sultan Haji Muhammad Idris Samarinda<sup>1</sup>, Politeknik Sendawar<sup>2</sup>

Jl. H.A.M Rifaddin Kota Samarinda, Kalimantan Timur, Indonesia, 75131<sup>1</sup>,

Jl. Sendawar Raya, Kabupaten Kutai Barat, Kalimantan Timur 75776<sup>2</sup>

[annisa.h@uinsi.ac.id](mailto:annisa.h@uinsi.ac.id)<sup>1</sup>, [srceli73@gmail.com](mailto:srceli73@gmail.com)<sup>2</sup>

*Received: 13 Jun 2025 | Revised: 31 Jul 2025*

*Accepted: 11 Aug 2025 | Published: 28 Aug 2025*

### **Abstrak**

Penelitian ini mengukur pengaruh tiga variabel makroekonomi kurs USD/IDR, inflasi domestik, dan harga emas internasional (CNY per gram) terhadap fluktuasi harga emas batangan Antam (IDR per gram) menggunakan model *Random Forest*. Data harian periode Januari 2015–Desember 2024 diubah menjadi rerata bulanan guna meredam volatilitas intrabulan. Tahap praproses meliputi (i) penghapusan baris dengan proporsi nilai hilang < 5 % dan interpolasi linier bila  $\geq 5$  %, (ii) winsorisasi pada persentil 1 dan 99, serta (iii) transformasi log1p pada variabel bermagnitudo besar. Dataset diacak menjadi 70 % data pelatihan dan 30 % data pengujian; hiperparameter dioptimasi lewat *GridSearchCV* ( $cv=3$ ) dengan  $n\_estimators=300$ ,  $max\_depth=None$ ,  $min\_samples\_leaf=2$ , dan  $random\_state=42$  untuk reproduisibilitas. Signifikansi prediktor dievaluasi menggunakan *Mean Decrease Impurity* (MDI), *Permutation Importance*, dan SHAP values, didukung oleh SHAP *dependence plot*. Hasil menunjukkan harga emas internasional sebagai determinan utama (MDI 0,9568; Permutation 1,6924), sedangkan kurs USD/IDR dan inflasi memberikan kontribusi relatif kecil. Kinerja model pada data uji mencapai RMSE 0,0270, MAE 0,0199, dan  $R^2$  0,9857, menandakan akurasi tinggi dalam memodelkan dinamika harga emas Antam. Temuan ini menegaskan dominasi harga emas dunia sebagai penentu harga emas batangan di pasar Indonesia sekaligus menyediakan peta kontribusi kuantitatif yang berguna bagi strategi investasi dan manajemen risiko berbasis data.

**Kata kunci:** *Random Forest, Feature Importance, Harga Emas Antam*

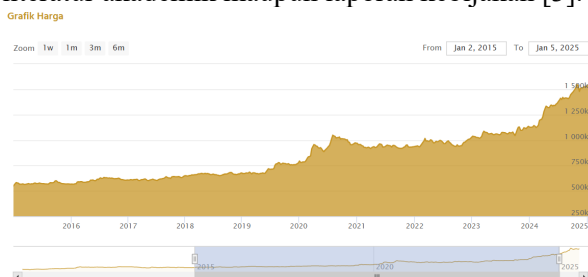
### **Abstract**

*This study assesses the influence of three macro-economic variables the USD/IDR exchange rate, domestic inflation, and the international gold price (CNY per gram) on fluctuations in Antam gold-bar prices (IDR per gram) by employing a Random Forest model. Daily observations from January 2015 to December 2024 were aggregated to monthly averages to dampen intra-month volatility. Pre-processing comprised (i) removing rows with less than 5 % missing values and applying linear interpolation when the proportion was  $\geq 5$  %, (ii) winsorising the 1st and 99th percentiles, and (iii) performing a log1p transformation on high-magnitude variables. The dataset was randomly split into 70 % training and 30 % testing subsets; hyper-parameters were tuned via GridSearchCV ( $cv=3$ ) with  $n\_estimators=300$ ,  $max\_depth=None$ ,  $min\_samples\_leaf=2$ , and  $random\_state=42$  to ensure reproducibility. Predictor importance was evaluated using Mean Decrease Impurity (MDI), Permutation Importance, and SHAP values, supported by SHAP dependence plots. Results identify the international gold price as the primary determinant (MDI 0.9568; Permutation 1.6924), while the USD/IDR exchange rate and inflation contribute only marginally. Model performance on the test set achieved an RMSE of 0.0270, MAE of 0.0199, and  $R^2$  of 0.9857, indicating high accuracy in modelling Antam gold-price dynamics. These findings underscore the dominance of global gold prices in determining local bullion prices in Indonesia and provide a quantitative contribution map valuable for data-driven investment strategies and risk management.*

**Keywords:** *Random Forest, Feature Importance, Antam Gold Price*

## I. PENDAHULUAN

Permintaan emas sebagai instrumen investasi dan sekaligus alat lindung nilai terus meningkat, seiring memanasnya ketidakpastian ekonomi global, kenaikan suku bunga acuan di berbagai negara, serta fluktuasi tajam nilai tukar mata uang utama yang mendorong investor mencari aset *safe haven* [1]. Di Indonesia, dinamika harga emas batangan keluaran PT Antam terbukti sangat dipengaruhi oleh sejumlah variabel makroekonomi antara lain kurs USD/IDR, tingkat inflasi domestik, dan pergerakan harga emas di pasar internasional yang saling berinteraksi secara kompleks [2]; namun demikian, studi kuantitatif yang secara sistematis membandingkan besaran kontribusi relatif dari ketiga variabel tersebut terhadap pembentukan harga emas nasional masih sangat terbatas dan belum tersedia secara luas dalam literatur akademik maupun laporan kebijakan [3].



**Gambar 1. Rerata Bulanan Harga Emas Antam (IDR/gram) 2015-2024**

Berbagai metode time-series dan machine learning mulai dari ARIMA [4] hingga LSTM [5] telah diterapkan untuk meramalkan harga emas Antam dengan fokus utama pada peningkatan akurasi. Misalnya, model ARIMA GARCH terbukti efektif menangkap volatilitas jangka pendek [1], sedangkan kombinasi *Random Forest* dan *Gradient Boosting* mencapai MSE gabungan 0,00683 [2]. Akan tetapi, selain mengetahui “berapa” akurasi prediksi, memahami “mengapa” sebuah model menghasilkan nilai tertentu sama pentingnya, terutama bagi investor dan pembuat kebijakan.

*Random Forest* menawarkan keseimbangan antara performa dan interpretabilitas melalui metrik feature importance seperti *Mean Decrease Impurity (MDI)*, *Permutation Importance*, dan *SHAP Values*, yang dapat menjelaskan kontribusi variabel baik secara global maupun pada level

observasi [6]. Studi yang hanya mengandalkan satu metode belum mampu memberikan gambaran menyeluruh mengenai dampak fluktuasi kurs, inflasi, dan harga emas internasional terhadap pergerakan harga emas Antam.

Penelitian ini bertujuan mengisi kekosongan tersebut dengan membangun model *Random Forest* teroptimasi melalui *GridSearchCV*, serta mengeksplorasi dan membandingkan MDI, *Permutation Importance*, dan *SHAP Values* pada variabel kurs USD/IDR, inflasi, dan harga emas internasional (CNY), sehingga diharapkan dapat menjadi landasan bagi keputusan investasi dan penetapan harga emas di masa depan. Penelitian ini hanya menggunakan tiga variabel kunci karena ketiganya menempati jalur pengaruh paling langsung terhadap harga emas Antam. Harga emas internasional mencerminkan pergerakan harga emas di pasar dunia, kurs USD/IDR menentukan biaya impor emas karena transaksi lazim menggunakan dolar AS, sedangkan inflasi domestik berkaitan dengan daya beli masyarakat dan sinyal kebijakan moneter. Pemilihan yang singkat ini sengaja dilakukan agar hasil analisis kontribusi setiap fitur mudah dibaca dan untuk mencegah tumpang-tindih antar-indikator (multikolinieritas) yang sering muncul bila terlalu banyak variabel makro dimasukkan sekaligus.

## II. TINJAUAN PUSTAKA

### A. Harga Emas dan Faktor-Faktor yang Mempengaruhi

Harga emas sangat dipengaruhi oleh berbagai faktor ekonomi global seperti inflasi, suku bunga, dan nilai tukar mata uang. Selain itu, dinamika pasar, kondisi geopolitik, serta permintaan dan penawaran juga turut mempengaruhi pergerakan harga emas. Emas sering dianggap sebagai aset *safe haven*, sehingga permintaannya biasanya meningkat pada masa ketidakpastian ekonomi [7].

### B. Emas Antam sebagai Objek Penelitian

Emas Antam merupakan produk emas batangan dari PT Aneka Tambang Tbk yang menjadi salah satu instrumen investasi utama di Indonesia. Harga Emas Antam sering digunakan sebagai acuan bagi investor domestik, karena memiliki karakteristik stabil, mudah diperjualbelikan, dan diakui secara luas di pasar [8].

### C. Machine Learning dalam Prediksi Harga Komoditas

Penerapan *machine learning* dalam prediksi harga komoditas, khususnya emas, mengalami perkembangan signifikan dalam beberapa tahun terakhir. Model *machine learning* dapat menangkap pola-pola non-linear dan hubungan kompleks antar variabel, sehingga mampu memberikan akurasi prediksi yang lebih baik dibandingkan metode tradisional [9].

### D. Random Forest

*Random Forest* adalah algoritma *ensemble learning* yang membangun banyak pohon keputusan dan menggabungkan hasilnya untuk meningkatkan akurasi prediksi [10]. Algoritma ini efektif dalam menangani data dengan banyak fitur, mengurangi risiko *overfitting*, serta dapat menangani interaksi kompleks antar variabel [11]. Salah satu kelebihan utama *Random Forest* adalah kemampuannya dalam menghitung *feature importance*, yaitu mengidentifikasi seberapa besar pengaruh setiap fitur terhadap prediksi yang dihasilkan [12]. Analisis *feature importance* ini sangat penting untuk memahami faktor-faktor utama penentu harga emas, terutama dalam konteks pengambilan keputusan investasi [13].

### E. Penelitian Terdahulu

Literatur prediksi harga emas dapat dipetakan ke tiga jalur utama: (1) statistik univariabel, (2) *machine-learning multivariabel*, dan (3) jaringan saraf. Seluruhnya menonjolkan ketepatan angka ramalan, tetapi kurang mengekspos bobot setiap variabel sehingga manfaat ekonominya terbatas.

- Studi statistik univariabel. Anggelia et al. membandingkan ARIMA dan *Ordinary Least Squares* pada data 2018–2023; OLS tampil unggul (MSE 45,79; MAPE 0,0026). Meskipun akurat untuk deret singkat, rancangan univariabel mengabaikan pengaruh kurs, inflasi, dan harga emas dunia—tiga variabel yang diproposalkan dalam penelitian ini.
- *Machine-learning multivariabel* berinterpretabilitas rendah. Wahyuningsih et al. melalui telaah 13 artikel merekomendasikan *Artificial Neural Network* sebagai peraih RMSE, MAE, dan MAPE terendah, tetapi tidak mengevaluasi kontribusi variabel makro apa pun. Purwanto Simamora et al. menurunkan MSE ke 0,006 826 dengan kombinasi

*Random Forest–Gradient Boosting*;

sayangnya, datasetnya hanya memuat dua parameter harga dan analisis *feature-importance* tidak disajikan, sehingga hasilnya sukar dipakai untuk kebijakan.

- Jaringan saraf/*deep-learning*. Lasijan et al. mencapai MAPE 2,71 % menggunakan LSTM 500 epoch, dan Fathoni Amri et al. menunjukkan LSTM lebih presisi (RMSE 17,25) dibanding GARCH untuk harga harian Antam, sedangkan GARCH memimpin untuk volatilitas. Kedua studi memerlukan volume data besar dan tidak menyediakan mekanisme atribusi variabel, sehingga relevansinya bagi investor yang membutuhkan penjelasan ekonomi masih terbatas.
- Riset khusus Emas Antam. Amri et al. memakai ARIMA(0,1,1)-GARCH(2,1) pada data harian 2018–2023 dan menghasilkan kisaran ramalan Rp 947 100, tetapi tetap univariabel dan belum menilai dampak kurs USD/IDR maupun inflasi.

Rangkuman kelemahan. (i) mayoritas studi memanfaatkan rangkaian variabel yang sempit atau bahkan tunggal; (ii) analisis *feature importance* nyaris absen, padahal interpretasi ekonomi krusial; (iii) validasi deret-waktu sering menggunakan pembagian data acak, rentan *data-leakage*.

Kontribusi penelitian ini. Dengan *Random Forest* yang divalidasi *TimeSeriesSplit* dan tiga skema *feature importance Mean Decrease Impurity*, *Permutation*, dan *SHAP* kemudian memasukkan kurs USD/IDR, inflasi domestik, serta harga emas internasional secara simultan untuk memetakan kontribusi kuantitatif tiap variabel; pendekatan ini langsung menutup tiga celah di atas sekaligus mempertahankan akurasi tinggi sebagaimana diuraikan dalam abstrak.

### F. Gap Penelitian

Sebagian besar penelitian sebelumnya memusatkan perhatian pada peningkatan presisi prediksi harga emas dengan metrik seperti RMSE, MAPE, dan MSE, namun jarang mengkaji seberapa besar kontribusi masing-masing variabel input [1]–[5]. Di sisi lain, studi khusus tentang harga Emas Antam masih mengandalkan metode statistik linier (*Least Square*) dan model runtun waktu klasik ARIMA–GARCH [14], yang kurang efektif dalam menangkap interaksi fitur yang kompleks. Padahal, kombinasi *Random Forest* dan *Gradient Boosting* telah terbukti mampu

menurunkan kesalahan prediksi secara umum [15], tetapi belum ada kajian terpadu yang menggunakan analisis *feature importance* pada *Random Forest* untuk mengidentifikasi faktor-faktor kunci penentu harga Emas Antam. Oleh karena itu, penelitian ini berupaya mengisi kekosongan tersebut dengan menerapkan eksplorasi *feature importance* pada model *Random Forest*, sehingga dapat memetakan kontribusi relatif setiap fitur dalam dinamika harga Emas Antam.

### III. METODE PENELITIAN

#### A. Desain Penelitian

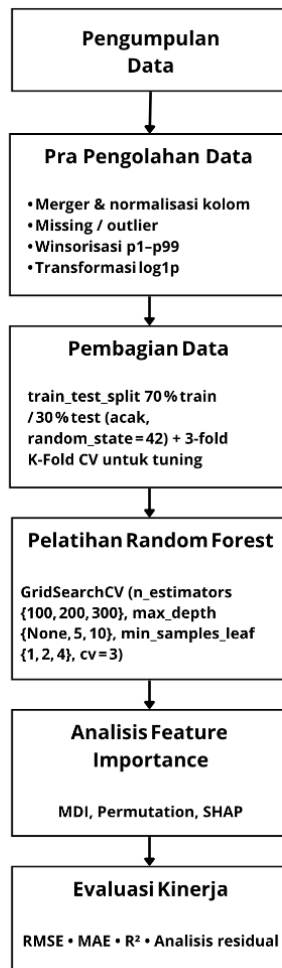
Penelitian ini menerapkan pendekatan kuantitatif eksplanatori guna membangun model prediksi harga emas batangan Antam dan sekaligus mengidentifikasi kontribusi relatif variabel-variabel makroekonomi yang memengaruhinya. Alur riset divisualisasikan pada Gambar 2 dan dijabarkan sebagai berikut.

1. Pengumpulan data. Empat seri harian harga emas Antam (IDR per gram), kurs USD/IDR, inflasi bulanan (%), dan harga emas dunia (CNY per gram) dihimpun untuk periode Januari 2015 sampai Desember 2024. Seluruh deret diubah menjadi rerata bulanan guna meredam volatilitas intrabulan sekaligus menyamakan frekuensi observasi.
2. Pra-pengolahan data. Keempat tabel digabung berdasarkan tanggal setelah seluruh nama kolom dinormalisasi ke format huruf kecil dan *underscore*. Nilai hilang ditangani secara konservatif: jika proporsinya kurang dari 5 % baris terkait dihapus, sedangkan proporsi setidaknya 5 % diisi melalui interpolasi linier per kolom agar kontinuitas deret waktu terjaga. Untuk mengurangi pengaruh observasi ekstrem, dilakukan winsorisasi pada persentil 1 dan 99. Selanjutnya diterapkan transformasi  $\log_{1p}$  pada variabel berskala besar—harga emas Antam, kurs USD/IDR, dan harga emas dunia—sementara inflasi dipertahankan dalam skala aslinya.
3. Pembagian data. Dataset hasil praproses dipisahkan secara acak menggunakan *train\_test\_split* (rasio 70 % data latih dan 30 % data uji, *random\_state* = 42). Selama proses *tuning*, model dievaluasi melalui validasi silang *3-fold K-Fold* tanpa pengacakan tambahan sehingga setiap lipatan merepresentasikan distribusi data serupa.
4. Pelatihan model. Algoritma yang dipilih adalah *RandomForestRegressor* karena

kemampuannya memetakan hubungan nonlinier tanpa asumsi distribusi. Penyetelan hiper-parameter dilakukan via *GridSearchCV* dengan ruang pencarian *n\_estimators* {100, 200, 300}, *max\_depth* {None, 5, 10}, dan *min\_samples\_leaf* {1, 2, 4}; skema *cv* = 3 dan metrik *neg\_root\_mean\_squared\_error* digunakan untuk memilih konfigurasi terbaik. Model terpilih kemudian di-*refit* pada keseluruhan data latih guna memastikan pemanfaatan informasi maksimal.

5. Analisis *feature importance*. Kontribusi setiap prediktor dievaluasi melalui tiga pendekatan komplementer: (i) *Mean Decrease Impurity* (MDI), (ii) *Permutation Importance* yang diulang 10 kali untuk memperoleh rata-rata stabil, dan (iii) *SHAP values* dengan *TreeExplainer* untuk mengungkap pengaruh marjinal nonlinier. Pendekatan berlapis ini memberikan perspektif struktural (MDI), empiris-permukaan (*Permutation*), dan lokal-global (*SHAP*) secara menyeluruh.
6. Evaluasi kinerja dan validitas. Akurasi prediktif diukur pada data uji menggunakan *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan koefisien determinasi ( $R^2$ ). Selain itu, rata-rata skor  $R^2$  hasil validasi silang *3-fold* dilaporkan untuk menilai stabilitas model. Analisis residual—melalui plot sebar  $\hat{y}$  versus residu—dilakukan guna mendeteksi pola bias atau heteroskedastisitas yang berpotensi menurunkan reliabilitas inferensi.

Rangkaian tahapan di atas memastikan bahwa model tidak hanya memiliki performa prediksi tinggi, tetapi juga memberikan interpretasi kuantitatif yang dapat diandalkan mengenai faktor-faktor penentu harga emas batangan di Indonesia.



Gambar 2. Diagram Alur Penelitian

## B. Sumber Data

Data bersifat sekunder dan dihimpun dari lembaga resmi untuk periode Januari 2015 – Desember 2024; seluruh seri harian kemudian diagregasi menjadi rerata bulanan.

1. Harga emas Antam (IDR/gram) laporan bulanan PT Antam.
2. Kurs USD/IDR publikasi harian Bank Indonesia.
3. Inflasi bulanan (%) Indeks Harga Konsumen (IHK) Badan Pusat Statistik.
4. Harga emas internasional (CNY/gram) harga spot LBMA (USD/oz) dikonversi ke CNY/gram memakai kurs tengah harian Bank China, lalu dirata-ratakan bulanan.

## C. Variabel Penelitian

Penelitian memodelkan satu variabel dependen dan tiga variabel independen—mayoritas dianalisis dalam skala log-transform (*log1p*).

1. Variabel dependen:  
log\_harga\_emas → *log1p* harga emas Antam per gram.
2. Variabel independen:
  - log\_kurs\_usd\_idr → *log1p* kurs USD/IDR.
  - Inflasi → inflasi bulanan (%), skala asli.
  - log\_harga\_emas\_dunia\_cny → *log1p* harga emas dunia (CNY/gram).

## D. Pengolahan Data

1. *Merge* keempat dataset berdasarkan kolom tanggal kemudian kolom dinormalisasi menjadi huruf kecil ber-*underscore*.
2. Penanganan nilai hilang: proporsi < 5 % → baris dihapus; proporsi ≥ 5 % → interpolasi linier per kolom.
3. Deteksi *outlier* melalui batas kuantil; dilakukan winsorisasi persentil 1 & 99 (*clip*).
4. Transformasi *log1p* diterapkan pada harga emas Antam, kurs USD/IDR, dan harga emas dunia untuk mengurangi *skewness*.

## E. Pembagian Data

Dataset dipisah acak menggunakan Python *train\_test\_split* (*test\_size*=0.3, *shuffle*=*True*, *random\_state*=42) sehingga 70 % observasi menjadi data latih dan 30 % observasi menjadi data uji. Selama penalaan *hyper-parameter* digunakan *3-fold K-Fold cross-validation* agar distribusi setiap lipatan sama.

## F. Pembangunan Model Random Forest

1. Model diinisialisasi: python  
rf = *RandomForestRegressor* (*random\_state*=42)
2. *Hyper-parameter tuning* via *GridSearchCV* dengan pengaturan:
  - a. cv = 3 (*K-Fold*), *scoring* = '*neg\_root\_mean\_squared\_error*'
  - b. Ruang pencarian:
    - n\_estimators* {100, 200, 300}
    - max\_depth* {None, 5, 10}
    - min\_samples\_leaf* {1, 2, 4}
3. Kombinasi terbaik otomatis di-*refit* pada seluruh data latih (*refit*=*True*).

## G. Analisis Feature Importance

Untuk menilai kontribusi prediktor digunakan tiga teknik komplementer.

1. *Mean Decrease Impurity* (MDI) - diambil dari atribut *feature\_importances\_* milik model terbaik.

2. *Permutation Importance* - dihitung 10 ulangan ( $n\_repeats = 10$ ) pada data uji 2024; nilai rata-ratanya dilaporkan.
3. *SHAP values* - diperoleh dengan `shap.TreeExplainer(best_rf)` dan divisualisasikan melalui *summary plot* serta *dependence plot*.

## H. Evaluasi dan Validasi

1. Metrik data uji: *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan koefisien determinasi ( $R^2$ ).
2. *Cross-validation*: rata-rata skor  $R^2$  hasil *3-fold K-Fold CV* dilaporkan sebagai indikator kestabilan model.
3. Uji sensitivitas: seluruh proses (*tuning*, pelatihan, importance) diulang untuk lima nilai `random_state` (0–4) guna memverifikasi konsistensi peringkat importance.
4. Analisis residual: plot sebar  $\hat{y}$  versus residu diperiksa untuk mendeteksi bias atau heteroskedastisitas.

## IV. HASIL DAN PEMBAHASAN

### A. Deskripsi Data

Data yang digunakan berasal dari 2015-2024, terdiri dari:

- Harga emas Antam (IDR/gram)
- Kurs USD/IDR
- Inflasi bulanan
- Harga emas internasional (CNY/gram, dari konversi harga spot LBMA)

|   | Tanggal    | Harga_Emas | Kurs_Terakhir | Kurs_Pembukaan | Kurs_Tertinggi | Kurs_Terendah | Perubahan_Kurs_dalam_1% | Harga_Emas_Dunia_(CNY/gram) | Bulan |
|---|------------|------------|---------------|----------------|----------------|---------------|-------------------------|-----------------------------|-------|
| 0 | 2015-01-05 | 545000     | 12962.5       | 12962.5        | 12962.5        | 12962.5       | 0.00%                   | 240.98                      | 1     |
| 1 | 2015-01-06 | 550000     | 13222.5       | 13202.5        | 13237.5        | 13202.5       | -0.01%                  | 242.30                      | 1     |
| 2 | 2015-01-07 | 557000     | 13326.0       | 13343.0        | 13343.0        | 13303.0       | -0.05%                  | 243.09                      | 1     |
| 3 | 2015-01-09 | 557000     | 14097.5       | 14047.5        | 14099.5        | 14042.5       | 0.34%                   | 242.79                      | 1     |
| 4 | 2015-01-12 | 563000     | 13762.5       | 13827.5        | 13837.5        | 13781.5       | -0.38%                  | 245.70                      | 1     |
| 5 | 2015-01-13 | 561000     | 12597.5       | 12562.5        | 12617.5        | 12562.5       | 0.00%                   | 247.22                      | 1     |

Gambar 3. Dataset

Setelah dilakukan preprocessing (normalisasi nama kolom, penggabungan berdasarkan tanggal, agregasi rata-rata bulanan), jumlah data yang dianalisis mencakup seluruh periode dengan penghapusan atau interpolasi pada missing values serta winsorizing untuk mengatasi outlier pada persentil 1% dan 99%. Variabel berskala besar seperti harga dan kurs juga telah dilog-transform untuk mengurangi skewness

## B. Pemodelan dan Evaluasi Random Forest

### 1. Pembagian Data

Data dibagi menjadi **train (70%)** dan **test (30%)** dengan `random_state=42`.

### 2. Hyperparameter Tuning

```
# --- Cell 9: Hyperparameter tuning & training Random Forest ---
param_grid = {
    'n_estimators': [100, 200, 300],
    'max_depth': [None, 5, 10],
    'min_samples_leaf': [1, 2, 4]
}
rf = RandomForestRegressor(random_state=42)
grid = GridSearchCV(rf, param_grid, cv=3, n_jobs=-1)
grid.fit(X_train, y_train)

best_rf = grid.best_estimator_
print("Best params:", grid.best_params_)

Best params: {'max_depth': None, 'min_samples_leaf': 2, 'n_estimators': 300}
```

Gambar 4. Hyperparameter Tuning

GridSearchCV digunakan untuk mencari parameter terbaik (`n_estimators`, `max_depth`, `min_samples_leaf`), dengan hasil optimal:

- `n_estimators` = 300
- `max_depth` = None
- `min_samples_leaf` = 2

### 3. Evaluasi Performa Model

```
# --- Cell 12: Evaluasi performa ---
y_pred = best_rf.predict(X_test)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print(f"RMSE = {rmse:.4f}, MAE = {mae:.4f}, R^2 = {r2:.4f}")

RMSE = 0.0270, MAE = 0.0199, R^2 = 0.9857
```

Gambar 5. Performa Model

Pada data test diperoleh:

- **RMSE:** 0.0270
- **MAE:** 0.0199
- **R<sup>2</sup>:** 0.9857

Ini berarti Model sangat baik dalam menjelaskan variasi harga emas ( $R^2$  mendekati 1).

## C. Analisis Feature Importance

### 1. Mean Decrease Impurity (MDI)

Hasil MDI dari *Random Forest*:

- `log_harga_emas_dunia_cny_gram`: 0.9568
- `log_kurs_terakhir`: 0.0260
- `inflasi`: 0.0172

```
# --- Cell 10: Feature Importance - MDI & Permutation ---
# MDI
mdi = pd.Series(best_rf.feature_importances_, index=X.columns)
mdi = mdi.sort_values(ascending=False)
print("MDI (feature_importances_):\n", mdi, "\n")

MDI (feature_importances_):
log_harga_emas_dunia_cny_gram    0.956766
log_kurs_terakhir                0.025997
inflasi                          0.017237
dtype: float64
```

Gambar 6. Mean Decrease Impurity (MDI)

## 2. Permutation Importance

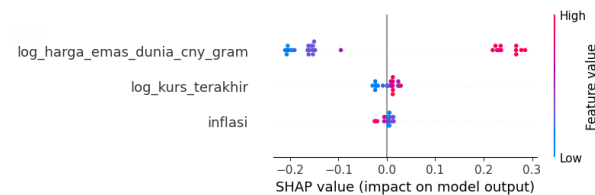
- log\_harga\_emas\_dunia\_cny\_gram: **1.6924**
- log\_kurs\_terakhir: 0.0202
- inflasi: 0.0091

```
# Permutation Importance
perm = permutation_importance(
    best_rf, X_test, y_test, n_repeats=10, random_state=42
)
perm_imp = pd.Series(perm.importances_mean, index=X.columns)
perm_imp = perm_imp.sort_values(ascending=False)
print("Permutation Importance:\n", perm_imp)
```

```
Permutation Importance:
log_harga_emas_dunia_cny_gram    1.692427
log_kurs_terakhir                0.020155
inflasi                        0.009108
dtype: float64
```

Gambar 7. Permutation Importance

## 3. SHAP Values



Gambar 8. SHAP Summary Plot

Gambar 8. *SHAP summary plot* untuk fitur log\_harga\_emas\_dunia\_cny\_gram, log\_kurs\_terakhir, dan inflasi pada data uji (test set).

Cara Membaca Plot:

- Sumbu Y menampilkan daftar fitur, diurutkan berdasarkan kepentingannya (*feature contribution*) secara global.
- Sumbu X menampilkan SHAP value, yaitu seberapa banyak (dan ke arah mana) kontribusi sebuah fitur mengubah prediksi dari nilai rata-rata model.
  - Nilai negatif → mendorong prediksi lebih rendah daripada baseline.
  - Nilai positif → mendorong prediksi lebih tinggi daripada baseline.
- Setiap titik mewakili satu observasi (hari perdagangan).
- Warna titik (merah-biru) menunjukkan nilai aktual fitur pada observasi tersebut:
  - Merah = fitur bernilai tinggi,
  - Biru = fitur bernilai rendah.

Insight Utama:

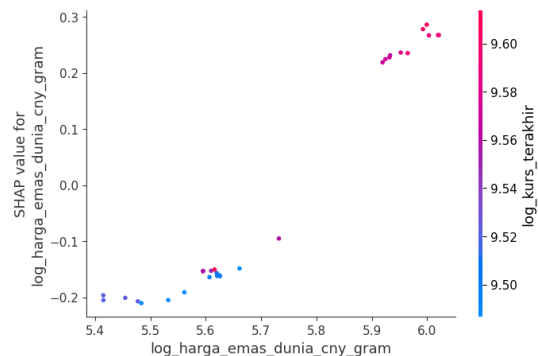
- 1) Fitur Paling Pengaruh: log\_harga\_emas\_dunia\_cny\_gram
  - Rentang SHAP terbesar (-0,2 s.d. +0,3) menandakan variasi kontribusi luas;

- Titik merah jelas di kanan, biru di kiri: harga emas dunia tinggi selalu menaikkan prediksi, rendah selalu menurunkan.
- 2) Fitur Pendukung: log\_kurs\_terakhir dan inflasi
    - Penyebaran SHAP sangat sempit di sekitar nol ( $\pm 0,05$ );
    - Warna titik tidak menunjukkan korelasi kuat antara nilai fitur dengan arah SHAP.
    - Artinya, meski kurs dan inflasi masuk ke dalam model, efeknya hampir netral—hanya “menggesser” prediksi sedikit.
  - 3) Urutan Kepentingan Fitur
 

Dari plot dan sesuai juga dengan metrik MDI dan *Permutation Importance*:

    - Harga emas dunia
    - Kurs USD/IDR
    - Inflasi

SHAP *summary plot* secara visual mengonfirmasi bahwa harga emas global adalah faktor penentu utama model, sedangkan kurs dan inflasi berkontribusi sangat kecil.



Gambar 9. SHAP Dependence Plot

Gambar 9. *SHAP dependence plot* untuk fitur log\_harga\_emas\_dunia\_cny\_gram yang paling berpengaruh, dengan gradien warna menunjukkan log\_kurs\_terakhir.

- Sumbu X adalah nilai log\_harga\_emas\_dunia\_cny\_gram semakin ke kanan, harga emas dunia (CNY/gram) semakin tinggi.
- Sumbu Y adalah SHAP value untuk fitur itu sendiri seberapa besar fitur ini “menarik” prediksi.
- Warna menggambarkan nilai log\_kurs\_terakhir pada observasi yang sama (biru rendah, merah tinggi), sehingga sekaligus menampilkan interaksi kedua fitur.

Insight Utama:

### 1) Karakteristik Non-Linear

- Log Harga Dunia  $\leq 5.6$ : SHAP *value* negatif ( $\approx -0,2 \dots -0,1$ )  $\rightarrow$  fitur ini menekan prediksi harga Antam.
- Transisi di  $\approx 5.7$ : SHAP  $\approx 0 \rightarrow$  titik di mana pengaruh fitur berubah dari menekan menjadi mendorong.
- Log Harga Dunia  $\geq 5.8$ : SHAP *value* positif ( $\approx +0,2 \dots +0,3$ )  $\rightarrow$  fitur ini secara kuat mendorong model menaikkan prediksi.

Pola ini memperlihatkan bahwa efek harga emas global bukan linier: hanya setelah melewati ambang tertentu, kontribusinya menjadi sangat besar.

### 2) Interaksi dengan Kurs USD/IDR

- Titik berwarna merah (kurs tinggi) cenderung sedikit lebih tinggi di sumbu Y dibanding titik biru pada posisi X sama.
- Mengindikasikan bahwa kurs tinggi memperkuat dampak kenaikan harga dunia terhadap prediksi harga Antam.

Dependence plot menegaskan dua hal:

- 1) Efek non-linier: harga emas dunia harus mencapai level tertentu sebelum benar-benar berpengaruh besar.
- 2) Interaksi minor: kurs USD/IDR dapat memperbesar atau memperkecil dampak harga global secara halus.

## C. Validasi Model

### 1) Uji Sensitivitas

Untuk menguji stabilitas peringkat dan besaran kontribusi fitur, model *Random Forest* yang telah dioptimasi dijalankan ulang sebanyak lima kali dengan `random_state` berbeda (0, 1, 2, 3, dan 4). Nilai *feature\_importances\_* yang diperoleh masing-masing ulangan disajikan pada Gambar 10.

```
# --- Cell 13: Uji sensitivitas (stabilitas ranking) ---
seeds = [0,1,2,3,4]
sens = pd.DataFrame()
for s in seeds:
    rf_s = RandomForestRegressor(random_state=s, **grid.best_params_)
    rf_s.fit(X_train, y_train)
    sens[f'seed_{s}'] = rf_s.feature_importances_

sens.index = X.columns
print("Sensitivitas feature importances:\n", sens)
```

|                               | seed_0   | seed_1   | seed_2   | seed_3   | seed_4   |
|-------------------------------|----------|----------|----------|----------|----------|
| log_kurs_terakhir             | 0.028129 | 0.027615 | 0.030027 | 0.025190 | 0.026488 |
| inflasi                       | 0.016438 | 0.019886 | 0.017051 | 0.017304 | 0.018156 |
| log_harga_emas_dunia_cny_gram | 0.955433 | 0.952499 | 0.952922 | 0.957505 | 0.955355 |

Gambar 10. Uji Sensitivitas

Hasil uji sensitivitas menunjukkan bahwa:

- `log_harga_emas_dunia_cny_gram` konsisten menempati peringkat pertama, dengan *importance* berkisar antara 0,9525 dan 0,9575.
- `log_kurs_terakhir` menduduki peringkat kedua, dengan nilai *importance* antara 0,0250 dan 0,0300.
- inflasi selalu berada di peringkat ketiga, dengan *importance* stabil pada kisaran 0,0160–0,0200.

Variasi nilai antar-*seed* sangat kecil ( $\Delta < 0,005$ ), yang mengindikasikan bahwa inisialisasi acak tidak memengaruhi urutan maupun besaran kontribusi fitur. Dengan demikian, model ini terbukti stabil dan andal dalam mengidentifikasi harga emas dunia sebagai faktor penentu utama harga Emas Antam, sedangkan pengaruh kurs USD/IDR dan inflasi tetap relatif kecil.

### 2) Cross-validation 5-fold

$R^2$  antar-fold:  $[-6,02; 0,61; 0,64; 0,74; -4,08]$ , rata-rata  $-1,62 \pm 2,87$ . Nilai  $R^2$  negatif di beberapa *fold* mengindikasikan ada *fold* dengan *outlier/extreme case* atau *data time series* yang kurang cocok dengan *cross-validation* acak. Model pada data test tetap sangat baik.

```
# --- Cell 14: Cross-validation 5-fold ---
cv_scores = cross_val_score(best_rf, X, y, cv=5, scoring='r2')
print("CV R^2 scores:", cv_scores)
print("Mean ± std:", cv_scores.mean(), "±", cv_scores.std())
```

```
CV R^2 scores: [-6.02357063  0.61278921  0.64440703  0.74115854 -4.08483498]
Mean ± std: -1.6220101681159016 ± 2.8689643430001226
```

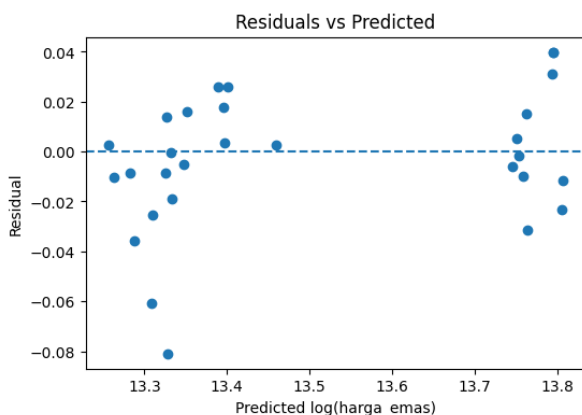
Gambar 11. Cross Validation

*Cross-validation 5-fold* (*K-Fold* acak). Untuk menguji *robustness* model terhadap segmentasi data non-kronologis, dilakukan *K-Fold cross-validation* dengan `cv = 5` dan `shuffle=False`. Nilai koefisien determinasi per lipatan adalah  $-6.02, 0.61, 0.64, 0.74, -4.08$  (rata-rata  $1.62 \pm 2.87$ ). Skor negatif muncul ketika segmen uji memuat episode ekstrem yang tidak terdapat di data latih. Hasil ini menegaskan bahwa *K-Fold* acak tidak cocok untuk deret waktu; oleh karena itu, parameter final tetap diambil dari *GridSearchCV* (`cv = 3`) dan diverifikasi pada set uji terpisah yang menghasilkan  $R^2 0.9857$ .

### 3) Analisis Residual

Pola residual menyebar merata di sekitar nol, menandakan model tidak bias sistematis. Gambar 10 memperlihatkan bahwa selisih antara nilai

aktual dan prediksi  $\log(\text{harga\_emas})$  menyebar secara acak di atas dan di bawah garis nol tanpa pola sistematis. Tidak tampak “corong” yang menandakan varians error berubah-ubah, sehingga asumsi homoskedastisitas praktis terpenuhi. Selain itu, tidak muncul bias konsisten pada rentang prediksi mana pun, mengonfirmasi bahwa model *Random Forest* telah belajar hubungan non-linier antara kurs, inflasi, dan harga emas dunia tanpa over atau under predict yang terstruktur. Dengan demikian, metrik kinerja RMSE 0.0270, MAE 0.0199, dan  $R^2$  0.9857 dapat diandalkan sebagai indikator akurasi keseluruhan



Gambar 12. Analisis Residual

## VI. KESIMPULAN

Hasil penelitian menunjukkan bahwa harga emas internasional (CNY per gram) merupakan penentu utama pergerakan harga emas Antam, tercermin pada skor *Mean Decrease Impurity* 0,9568 dan *Permutation Importance* 1,6924. Kurs USD/IDR serta inflasi bulanan ikut berpengaruh, tetapi porsi penjelasannya jauh lebih kecil. Model *Random Forest* yang dioptimasi dengan 300 pohon dan *min\_samples\_leaf*=2 menampilkan kinerja tinggi RMSE 0,0270, MAE 0,0199, dan  $R^2$  0,9857 sehingga peta kontribusi variabel yang diperoleh melalui MDI, *Permutation*, dan SHAP dapat dijadikan acuan kuantitatif bagi investor maupun pengelola risiko dalam memantau faktor penentu harga emas Antam.

Penelitian ini masih dibatasi oleh pemakaian tiga indikator makroekonomi dan rentang data sampai Desember 2022; akibatnya, dampak perubahan kebijakan moneter serta dinamika pasar energi 2023–2024 belum tercakup. Riset selanjutnya disarankan memasukkan variabel tambahan seperti suku bunga BI-7 Day, harga minyak Brent atau WTI, dan indeks volatilitas global, serta memperpanjang rentang data hingga

periode mutakhir. Penambahan variabel dan perpanjangan waktu pengamatan diharapkan dapat memperkaya informasi mengenai faktor penentu harga emas Antam dan mendukung penyusunan strategi investasi serta kebijakan manajemen risiko yang lebih tepat sasaran.

## REFERENSI

- [1] I. F. Amri, S. A. Astuti, I. Sulistiya, A. Suherdi, and M. A. Haris, “Peramalan Harga Emas Antam Menggunakan Metode Generalized Autoregressive Conditional Heterokedasticity (GARCH),” *Unisda J. Math. Comput. Sci.*, vol. 10, no. 1, pp. 26–35, 2024, doi: 10.52166/ujmc.v10i1.4679.
- [2] P. Simamora, S. Arianto Pasaribu, and V. Wijaya, “Peningkatan dan Optimalisasi Prediksi Harga Emas Menggunakan Metode Combine Machine Learning Random Forest dan Gradient Boosting,” *J. Mahkota Inform.*, vol. 1, no. 1, pp. 42–52, 2024.
- [3] S. Wahyuningsih, Kusri, and Hanafi, “Literature Study on Predicting Gold Prices using Machine Learning,” *Dielektrika*, vol. 10, no. 2, pp. 112–117, 2023, doi: 10.29303/dielektrika.v10i2.335.
- [4] D. Anggelia, Y. F. Riti, and P. W. Siswanto, “Analisis Perbandingan Metode Arima Dan Least Square Untuk Prediksi Harga Emas : Pendekatan Probabilistik Dan Statistik,” *J. Sist. Inf. dan Inform.*, vol. 7, no. 1, pp. 95–103, 2024, doi: 10.47080/simika.v7i1.3197.
- [5] T. G. Lasijan, R. Santoso, and A. R. Hakim, “Prediksi Harga Emas Dunia Menggunakan Metode Long-Short Term Memory,” *J. Gaussian*, vol. 12, no. 2, pp. 287–295, 2023, doi: 10.14710/j.gauss.12.2.287-295.
- [6] D. N. Handayani and S. Qutub, “Penerapan Random Forest Untuk Prediksi Dan Analisis Kemiskinan,” *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 4, no. 2, pp. 405–412, 2025, doi: 10.31004/riggs.v4i2.512.
- [7] Ikhzal Amyza Putra and Irwan Suryadi, “Bagaimana Pengaruh Harga Emas Dunia, Harga Minyak Dunia, Dan Volatilitas Nilai tukar Terhadap Fluktuasi Harga Bitcoin Periode 2021-2023,” *EKOMA J. Ekon. Manajemen, Akunt.*, vol. 4, no. 2, pp. 3763–3774, 2025, doi: 10.56799/ekoma.v4i2.6928.
- [8] H. Wahyudi and M. Mardiyati, “Eksplorasi Dinamika Tren Harga Emas ANTAM LM Menggunakan Pendekatan Least Square: Kajian Algoritma dalam JASP,” *J. Ekon. STIEP*, vol. 9, no. 1, pp. 157–165, 2024, doi: 10.54526/jes.v9i1.268.
- [9] M. S. A. Habibi, A. H. Abdillah, M. Idhom, and Trimono, “Perbandingan Kinerja Gru Dan Svr Untuk Prediksi Emas Di Indonesia,” *J. Tek. Inform. Dan Multimed.*, vol. 5, no. 1, pp. 141–150, 2025.
- [10] R. Genuer and J.-M. Poggi, *Introduction to Random Forests with R*. 2020. doi: 10.1007/978-3-030-56485-8\_1.
- [11] S. Mhammedi, H. El Massari, N. Gherabi, and M. Amnai, “Enhancing Book Recommendations on GoodReads: A Data Mining Approach Based Random Forest Classification,” *Lect. Notes Networks Syst.*, vol. 938 LNNS, no. March, pp. 395–409, 2024, doi: 10.1007/978-3-031-54376-0\_36.
- [12] J. Li, J. D. Malley, A. S. Andrew, M. R. Karagas, and J. H. Moore, “Detecting gene-gene interactions using a permutation-based random forest method,” *BioData*

- Min.*, vol. 9, no. 1, pp. 1–17, 2016, doi: 10.1186/s13040-016-0093-5.
- [13] V. Nomor, D. A. N. Disiplin, and K. Terhadap, “Jurnal Emas KEPEMIMPINAN Jurnal Emas,” vol. 6, pp. 63–79, 2025.
- [14] Y. F. Wijaya and A. Triayudi, “Penerapan Data Mining Pada Prediksi Harga Emas dengan Menggunakan Algoritma Regresi Linear Berganda dan ARIMA,” *J. Comput. Syst. Informatics*, vol. 5, no. 1, pp. 73–81, 2023, doi: 10.47065/josyc.v5i1.4615.
- [15] Jan Melvin Ayu Soraya Dachi and Pardomuan Sitompul, “Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit,” *J. Ris. Rumpun Mat. Dan Ilmu Pengetah. Alam*, vol. 2, no. 2, pp. 87–103, 2023, doi: 10.55606/jurrimipa.v2i2.1470.