

ANALISIS KOMPARASI ALGORITMA REGRESI MACHINE LEARNING DALAM ESTIMASI HARGA PENUTUPAN HARIAN EMAS PADA (DATA HISTORI) BROKER EXNESS TAHUN 2020-2024

Bagus Prabowo

Program Studi Teknik Informatika, FTI, Institut Teknologi Budi Utomo Jakarta,
bagus@itbu.ac.id

Abstrak

Pergerakan harga emas sangat dipengaruhi oleh faktor teknikal dan makroekonomi, sehingga memerlukan pendekatan prediksi yang lebih akurat dibanding metode konvensional. Penelitian ini bertujuan untuk membandingkan performa tiga algoritma regresi machine learning, yaitu Support Vector Regression (SVR), Random Forest, dan XGBoost dalam mengestimasi harga penutupan harian emas (XAU/USD) berdasarkan data historis dari broker Exness periode 2020–2024. Fitur yang digunakan mencakup harga open, high, low, inflation, dan interest rate. Penelitian ini menggunakan pendekatan CRISP-DM dan diimplementasikan dengan Python. Data dibagi menjadi data latih dan uji (80:20), serta dievaluasi menggunakan MAE, RMSE, dan R². Hasil penelitian menunjukkan bahwa algoritma Random Forest memberikan performa terbaik dalam hal akurasi prediksi.

Kata Kunci : Machine Learning, Harga Emas, Regresi, Random Forest, Estimasi.

1) PENDAHULUAN

Perdagangan emas telah lama menjadi salah satu instrumen investasi yang paling diminati oleh masyarakat di berbagai belahan dunia. Dalam kondisi perekonomian yang tidak stabil, emas sering dijadikan pilihan utama sebagai aset lindung nilai (hedging asset) karena nilainya cenderung stabil atau bahkan meningkat saat nilai mata uang dan pasar saham menurun. Keunggulan ini menjadikan emas sebagai alat diversifikasi portofolio yang efektif bagi investor, baik individu maupun institusi. Namun, harga emas juga mengalami fluktuasi signifikan akibat berbagai faktor seperti ketegangan geopolitik, inflasi, suku bunga, dan sentimen pasar global. Mekanisme harga emas pada pasangan XAU/USD dipengaruhi oleh sejumlah faktor fundamental dan eksternal. Studi oleh (Y. Zhang, 2021). Menunjukkan bahwa harga emas sangat dipengaruhi oleh fluktuasi nilai tukar dolar AS, tingkat inflasi, serta ketegangan geopolitik global yang memicu permintaan emas sebagai aset pelindung nilai. Selain itu, faktor permintaan fisik dari sektor perhiasan dan industri juga berkontribusi dalam dinamika harga emas. Harga emas di pasar

XAU/USD mengalami volatilitas yang mencerminkan respon cepat terhadap perubahan kondisi ekonomi dan geopolitik global, menjadikannya instrumen yang kompleks dan dinamis untuk diperdagangkan (Y. Zhang, 2021). Selain itu, dari sisi teknikal, harga pembukaan (open), harga tertinggi (high), dan harga terendah (low) dalam satu periode perdagangan juga memainkan peran penting dalam membentuk harga penutupan (closing price).

Estimasi harga penutupan emas sangat penting dalam konteks investasi dan manajemen risiko. Namun, proses estimasi ini bukanlah sesuatu yang sederhana. Kompleksitas hubungan antara faktor-faktor yang memengaruhi harga emas dan tingginya volatilitas pasar menjadikan metode estimasi konvensional, seperti regresi linear klasik, memiliki keterbatasan dalam menghasilkan estimasi yang akurat. Selain itu, pendekatan teknikal yang terlalu bergantung pada pola historis terkadang gagal menangkap dinamika pasar yang berubah cepat. Ketidaktepatan dalam memproyeksikan harga penutupan dapat menyebabkan kerugian finansial, terutama bagi investor jangka pendek yang

bergantung pada momentum pasar. Oleh karena itu, diperlukan pendekatan yang lebih canggih dan adaptif terhadap kondisi pasar yang dinamis dan kompleks.

Seiring dengan perkembangan teknologi informasi dan meningkatnya ketersediaan data, pendekatan berbasis machine learning. Machine Learning (ML) merupakan cabang dari kecerdasan buatan yang berfokus pada pengembangan algoritma yang memungkinkan komputer untuk belajar dari data dan melakukan prediksi tanpa perlu diprogram secara eksplisit. ML bekerja berdasarkan pendekatan statistik dan matematika untuk mengenali pola dalam data historis, sehingga dapat digunakan untuk menghasilkan prediksi atas variabel yang belum diketahui. Dalam praktiknya, ML terbagi dalam beberapa jenis, salah satunya adalah supervised learning yang paling banyak digunakan untuk prediksi harga aset keuangan karena proses pelatihannya berdasarkan data yang telah diberi label. Machine Learning (ML) mulai mendapat perhatian luas dalam dunia keuangan. Machine learning memungkinkan pemodelan hubungan non-linear dan interaksi multivariabel yang sulit ditangkap oleh metode statistik tradisional. Dalam dunia keuangan, Machine Learning telah dimanfaatkan secara luas untuk menganalisis dan memprediksi harga aset seperti saham, mata uang, maupun komoditas seperti emas. Salah satu penelitian yang relevan dilakukan oleh Ricky Lie Jaya (2022) yang menggunakan algoritma XGBoost, SVR, dan Random Forest dalam memprediksi return saham perusahaan yang tergabung dalam indeks LQ45.

Hasil studi tersebut menunjukkan bahwa model machine learning mampu mengenali pola kompleks dan memberikan prediksi yang kompetitif dibandingkan metode konvensional. Temuan ini memperkuat posisi algoritma-algoritma seperti Random Forest, SVR, dan XGBoost sebagai alat prediksi yang andal dan fleksibel dalam konteks keuangan.

Dalam konteks estimasi harga emas, model regresi machine learning seperti Random Forest, Support Vector Regression (SVR), dan XGBoost dinilai memiliki keunggulan dalam hal akurasi estimasi dan fleksibilitas terhadap beragam data input. Dengan memanfaatkan data historis harga pasar dan variabel ekonomi makro seperti inflasi dan suku bunga, model-model ini dapat digunakan untuk membangun sistem estimasi harga penutupan yang lebih presisi dan informatif.

Berdasarkan latar belakang tersebut, Penelitian ini bertujuan untuk membandingkan performa beberapa algoritma regresi dalam menghasilkan estimasi yang paling optimal.

2) METODOLOGI PENELITIAN

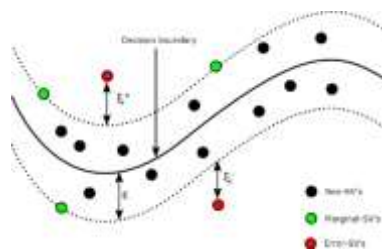
2.1. Jenis Penelitian

Penelitian ini termasuk dalam jenis kuantitatif prediktif, karena menggunakan data numerik historis untuk membangun model estimasi terhadap nilai masa depan, dalam hal ini adalah harga penutupan harian emas (XAU/USD). Model dibuat dengan pendekatan machine learning menggunakan tiga algoritma regresi, yaitu:

- 1) Support Vector Regression (SVR)
SVR dikenal mampu menangani hubungan yang kompleks antara variabel, baik yang bersifat linier maupun non-linier, melalui fungsi-fungsi tertentu yang disebut kernel. Selain itu, metode ini memungkinkan pengaturan sejauh mana model dapat mentoleransi kesalahan dan seberapa ketat model tersebut mengikuti data. Pengaturan ini dilakukan dengan menentukan nilai-nilai parameter tertentu yang memengaruhi fleksibilitas model terhadap data pelatihan. Dalam penelitian oleh Gananta, Purnama, dan Fredlina (2024), SVR digunakan untuk memprediksi harga emas harian berdasarkan data

historis tahun 2018 hingga 2021. Dengan menerapkan teknik pencarian kombinasi parameter terbaik, hasil penelitian menunjukkan bahwa SVR mampu memberikan prediksi harga emas yang akurat, terutama setelah dilakukan pengoptimalan, dengan tingkat kesalahan yang lebih rendah dibandingkan model SVR tanpa pengaturan parameter lanjutan [8]

(SVR)



Gambar 1 Support Vector Regression

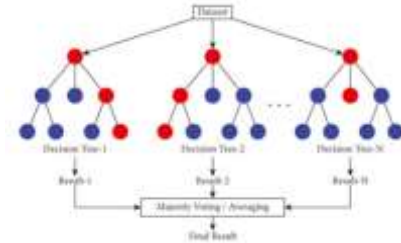
Sumber: https://medium-com.translate.goog/analytics-vidhya/support-vector-regression-svr-model-a-regression-based-machine-learning-approachf4641670c5bb?_x_tr_sl=en&_x_tr_tl=id&_x_tr_hl=id&_x_tr_pto=i mgs

2) Random Forest

Random Forest adalah salah satu metode dalam pembelajaran mesin yang termasuk dalam algoritma gabungan (ensemble). Algoritma ini bekerja dengan membangun sejumlah pohon keputusan secara bersamaan. Untuk masalah klasifikasi, hasil akhir diambil berdasarkan suara terbanyak, sedangkan untuk regresi, hasilnya diambil dari rata-rata semua prediksi yang dihasilkan oleh pohon-pohon tersebut.

Pendekatan ini membuat Random Forest menjadi model yang kuat dan stabil karena mampu menyeimbangkan kesalahan dari setiap pohon secara keseluruhan. Selain itu, metode ini juga dikenal mampu mengurangi risiko kelebihan pencocokan data

(overfitting), serta efektif dalam menangani data dengan jumlah fitur yang banyak. Random Forest dapat digunakan untuk menyelesaikan permasalahan baik dalam klasifikasi maupun prediksi nilai (regresi). [10]

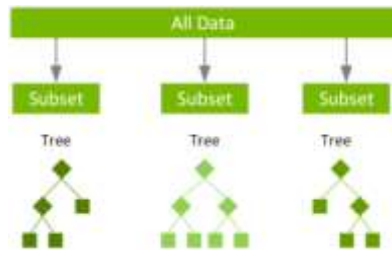


Gambar 2 Random Forest

Sumber : R2 score/I. Abdillah

3) XGBoost

Extreme Gradient Boosting (XGBoost) adalah salah satu algoritma machine learning yang termasuk dalam metode ensemble boosting dan dikembangkan untuk meningkatkan akurasi dan efisiensi komputasi dalam proses prediksi. Algoritma ini bekerja secara bertahap dengan memperbaiki kesalahan dari model sebelumnya dan menggunakan teknik regularisasi untuk mengurangi overfitting. Dalam konteks prediksi harga emas, karena harga emas sangat dipengaruhi oleh berbagai variabel ekonomi seperti inflasi, nilai tukar mata uang, dan suku bunga. Dengan kompleksitas hubungan antar variabel tersebut, XGBoost cocok digunakan karena mampu menangani data yang tidak linier dan mengelola variabel dalam jumlah besar. Kelebihan ini menjadikan XGBoost sebagai salah satu algoritma yang relevan dan potensial dalam membangun model prediksi harga emas yang akurat dan adaptif terhadap dinamika pasar.[8]

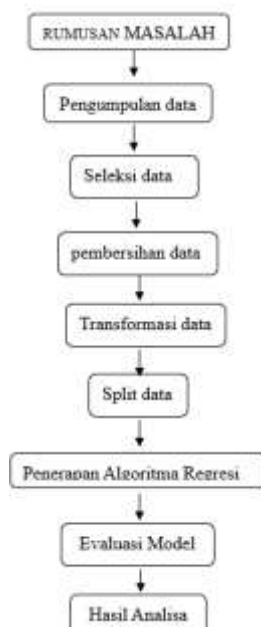


Gambar 3 Extreme Gradient Boosting (XGBoost)

Sumber: <https://www.nvidia.com/en->

2.2. Kerangka Penelitian

Kerangka pemikiran disusun sebagai panduan sistematis dalam menyelesaikan sebuah permasalahan penelitian. Kerangka ini menggambarkan tahapan-tahapan yang dilakukan mulai dari perumusan masalah, pengumpulan dan pengelolaan data, hingga penerapan metode yang relevan serta evaluasi hasil. Setiap tahap dirancang untuk mendukung proses analisis agar penelitian berjalan terarah dan menghasilkan kesimpulan yang sesuai tujuan. Diagram berikut menyajikan alur proses yang dilakukan dalam penelitian ini.:



Gambar 4. Kerangka Pemikiran
Sumber : Penelitian Mandiri 2025

Agar kegiatan penelitian dapat berlangsung secara runut, kerangka pemikiran dijelaskan sebagai berikut:

- 1) Rumusan masalah menjawab bagaimana pengaruh harga open, high, low, Inflation, dan Interest rate terhadap harga penutupan emas, serta algoritma mana yang memiliki performa prediksi terbaik..
- 2) Pengumpulan data historis harga emas harian diambil dari situs Exness.com, sedangkan data Inflation (inflasi) dan Interest rate (suku bunga) diperoleh dari investing.com
- 3) Pra-pemrosesan Data tahap ini meliputi pembersihan data (menghapus nilai kosong), transformasi (normalisasi/standarisasi), dan konversi format jika diperlukan.
- 4) Pembagian Dataset dibagi menjadi data latih dan data uji dengan rasio umum 80:20 untuk keperluan pelatihan dan validasi model.
- 5) Pemodelan dengan Algoritma Regresi yaitu Random Forest, SVR, XGBoost
- 6) Evaluasi Model yang digunakan adalah Mean Squared Error (MSE) Root Mean Squared Error (RMSE), dan R²-Squared (R²)
- 7) Analisis dan Interpretasi model dibandingkan berdasarkan nilai akurasi dan error, lalu ditarik kesimpulan terkait algoritma dengan performa terbaik.

2.3. Metode Penelitian

Penelitian ini menggunakan metode kuantitatif prediktif, dengan data yang dikumpulkan berdasarkan pengamatan langsung terhadap data historis harga emas dan variabel pendukung seperti open, high, low, inflasi, dan Interest rate. Data dianalisis secara numerik menggunakan algoritma regresi machine learning yaitu Random Forest, SVR, dan XGBoost.

Hasil prediksi dari masing-masing algoritma dievaluasi menggunakan metrik statistik seperti

MSE, RMSE, dan R^2 , untuk mengetahui algoritma dengan performa terbaik dalam estimasi harga penutupan emas.

2.4. Metode Analisis Data

Dalam penelitian ini, pendekatan yang digunakan untuk menganalisis data adalah Cross-Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM merupakan suatu standar metodologi yang luas digunakan dalam proyek-proyek data mining dan machine learning karena bersifat fleksibel, iteratif, dan dapat diterapkan dalam berbagai domain industri, termasuk dalam konteks prediksi harga. Pendekatan ini terdiri dari enam fase utama, yaitu:

1. Business Understanding
yaitu memahami tujuan penelitian dalam konteks bisnis atau praktik nyata, yaitu melakukan estimasi harga penutupan emas secara akurat.
2. Data Understanding
yaitu proses eksplorasi awal untuk memahami struktur, format, dan karakteristik data historis harga emas dari broker Exness.
3. Data Preparation
yang mencakup pembersihan data (data cleaning), transformasi data, dan pemisahan data untuk pelatihan dan pengujian model.
4. Modeling
yaitu proses membangun model prediksi menggunakan tiga algoritma regresi: Support Vector Regression (SVR), Random Forest, dan XGBoost.
5. Evaluation
yaitu mengevaluasi performa masing-masing model menggunakan metrik evaluasi seperti R-squared (R^2), Mean Absolute Error (MAE) dan Root Mean Square Error (RMSE) guna mengetahui model yang paling akurat.



Gambar 5 Diagram Metodologi
Sumber : Penelitian Mandiri 2025

3. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Tahap Business Understanding dalam pendekatan CRISP-DM bertujuan untuk memahami latar belakang dan konteks permasalahan bisnis yang ingin diselesaikan melalui proses analitik. Pada tahap ini, dilakukan identifikasi terhadap kebutuhan bisnis, tujuan yang ingin dicapai, serta perumusan masalah bisnis yang relevan. Pemahaman ini menjadi dasar dalam menentukan arah pendekatan teknis yang akan dilakukan pada tahap-tahap selanjutnya.

1. Latar Belakang Bisnis

Perdagangan emas global, khususnya pasangan XAU/USD, sangat dipengaruhi oleh dinamika pasar dan indikator makroekonomi seperti inflasi dan suku bunga. Pelaku pasar menghadapi tantangan besar dalam memprediksi harga penutupan emas karena volatilitas yang tinggi dan ketergantungan pada berbagai variabel yang saling berkaitan. Informasi seperti harga pembukaan, tertinggi, dan terendah harian juga memberikan sinyal teknikal yang penting namun sering kali sulit ditafsirkan secara manual. Dalam kondisi seperti ini, pendekatan analitik berbasis data historis dan pembelajaran mesin (*machine learning*) menjadi alternatif yang potensial untuk membantu

investor dan analis dalam mengambil keputusan yang lebih terukur. Selain memberikan estimasi harga penutupan, model prediksi juga dapat digunakan untuk memahami seberapa besar pengaruh variabel-variabel tertentu terhadap pergerakan harga.

Penelitian ini dilakukan dengan membangun model prediktif berbasis tiga algoritma regresi *machine learning* Support Vector Regression (SVR), Random Forest, dan XGBoost untuk mengevaluasi performa estimasi serta menilai pengaruh variabel input terhadap target. Dengan demikian, pendekatan ini diharapkan dapat memberikan wawasan bisnis yang lebih tajam dan mendukung strategi investasi berbasis data.

2. Tujuan Bisnis

Tujuan utama dari penelitian ini adalah untuk membangun sistem prediksi harga penutupan harian emas (XAU/USD) yang akurat dan dapat diandalkan berdasarkan data historis. Sistem ini bertujuan untuk membantu investor, analis pasar, dan pengambil keputusan dalam menentukan strategi yang lebih tepat dalam aktivitas perdagangan emas. Secara lebih rinci, tujuan bisnis dari penelitian ini meliputi:

1) Mengidentifikasi faktor-faktor yang berpengaruh terhadap harga penutupan emas, seperti harga pembukaan, harga tertinggi, harga terendah, inflasi, dan suku bunga. Dengan mengetahui hubungan antara variabel-variabel ini, pengguna dapat memahami

karakteristik dasar dari pergerakan harga emas.

- 2) Membangun model prediksi berbasis algoritma regresi *machine learning* yang dapat memperkirakan harga penutupan emas menggunakan data historis. Model ini diharapkan mampu menangkap hubungan non-linier antar variabel dan memberikan hasil estimasi yang mendekati nilai aktual.
- 3) Membandingkan performa tiga *algoritma regresi machine learning*, yaitu SVR, *Random Forest*, dan XGBoost dalam konteks prediksi harga penutupan emas. Hasil komparasi ini akan menunjukkan model mana yang paling akurat dan sesuai untuk diterapkan pada data serupa.
- 4) Memberikan dasar analisis yang lebih objektif dan berbasis data kepada pelaku pasar untuk mendukung pengambilan keputusan investasi yang lebih baik dan lebih terinformasi.

Dengan tercapainya tujuan-tujuan tersebut, diharapkan model prediksi yang dikembangkan tidak hanya berguna dalam konteks akademik, tetapi juga dapat diterapkan dalam praktik nyata di sektor keuangan dan investasi, khususnya pada perdagangan komoditas emas.

3. Rumusan Masalah Bisnis

Tujuan utama dari penelitian ini adalah untuk membantu pelaku pasar dalam memprediksi harga penutupan harian emas (XAU/USD) secara akurat melalui model berbasis *machine learning*, sekaligus mengetahui pengaruh masing-masing variabel input

terhadap harga tersebut. Informasi ini penting bagi investor dalam merumuskan strategi perdagangan, manajemen risiko, dan pengambilan keputusan berdasarkan data.

Secara khusus, penelitian ini bertujuan untuk:

- 1) Menganalisis pengaruh harga open, high, low, Inflation, dan Interest rate terhadap harga penutupan emas.
- 2) Membangun model prediksi harga penutupan emas menggunakan algoritma SVR, Random Forest, dan XGBoost.
- 3) Membandingkan kinerja ketiga model menggunakan metrik evaluasi MAE, RMSE, dan R^2 , untuk menentukan algoritma yang paling optimal.

Melalui pencapaian tujuan ini, hasil penelitian diharapkan dapat memberikan nilai tambah bagi pelaku pasar dan menjadi referensi dalam penerapan teknologi prediktif di bidang keuangan.

3.2 Data Understanding

Tahap Data Understanding dalam pendekatan CRISP-DM bertujuan untuk memahami struktur, isi, dan karakteristik awal dari data yang akan digunakan dalam proses pemodelan. Fokus utamanya adalah melakukan eksplorasi terhadap sumber data, jenis fitur, jumlah data, serta identifikasi potensi masalah pada data mentah, seperti tipe data tidak sesuai atau nilai yang hilang, tanpa melakukan perubahan atau transformasi pada tahap ini.

1. Sumber dan Struktur Data

Data yang digunakan dalam penelitian ini terdiri atas dua jenis, yaitu data teknikal dan data makroekonomi. Data teknikal berupa harga historis emas terhadap

dolar Amerika Serikat (XAU/USD) diperoleh dari broker Exness melalui platform MetaTrader 4 (MT4). Data ini mencakup informasi harga pembukaan (open), harga tertinggi (high), harga terendah (low), dan harga penutupan (close) per hari.

Sementara itu, data makroekonomi yang meliputi tingkat inflasi dan suku bunga diperoleh dari situs Investing.com. Kedua jenis data tersebut kemudian digabungkan berdasarkan tanggal yang sama untuk membentuk satu dataset terpadu yang memuat data harga dan kondisi ekonomi makro dalam satu format.

Dataset mencakup rentang waktu dari Januari 2020 hingga Maret 2024, dengan total sebanyak 1.639 baris data harian. Setiap baris merepresentasikan satu hari perdagangan. Dataset ini memiliki tujuh kolom utama:

- 1) Date: tanggal pencatatan data
- 2) Open: harga pembukaan
- 3) High: harga tertinggi harian
- 4) Low: harga terendah harian
- 5) inflation: tingkat inflasi saat itu
- 6) Interest Rate: suku bunga acuan
- 7) Close : harga penutupan (sebagai target prediksi)

Hasil pemeriksaan awal menunjukkan bahwa seluruh kolom numerik pada dataset awal masih bertipe object karena penggunaan tanda koma (,) sebagai pemisah desimal. Selain itu, terdapat satu nilai kosong (missing value) pada kolom inflation, sedangkan kolom lainnya tidak memiliki nilai hilang. Permasalahan ini akan dijelaskan lebih lanjut pada tahap Data Preparation.

Tabel 1 Struktur Data Awal

No	Nama Kolom	Tipe Data	Jumlah Non-Null	Deskripsi
1	Date	object	1.639	Tanggal pencatatan harga emas
2	Open	object	1.639	Harga pembukaan pasar emas pada hari tersebut
3	High	object	1.639	Harga tertinggi yang dicapai oleh emas pada hari tersebut
4	Low	object	1.639	Harga terendah yang dicapai oleh emas pada hari tersebut
5	Inflation	object	1.638	Nilai inflasi pada periode bersangkutan
6	Interest rate	object	1.639	Tingkat suku bunga pada periode tersebut
7	Close	object	1.639	Harga penutupan pasar emas pada hari tersebut (harga penutup)

Sumber : Penelitian Mandiri 2025

2. Tampilan Data Awal

Setelah dilakukan penggabungan antara data teknis harga emas dan data makroekonomi, langkah berikutnya adalah meninjau isi dataset melalui tampilan awal data. Peninjauan ini bertujuan untuk memperoleh gambaran awal mengenai struktur dan nilai-nilai yang terdapat dalam dataset, sebelum dilakukan proses pembersihan dan transformasi data lebih lanjut.

Tabel 2 Lima Data Teratas

Date	Open	High	Low	Inflation	Interest rate	Close
2020-01-02	1217.180	1522.340	1212.450	2.29	1.75	1319.430
2020-01-03	1318.460	1574.100	1315.270	2.29	1.75	1323.840
2020-01-06	1333.330	1588.110	1324.250	2.29	1.75	1327.780
2020-01-07	1372.560	1591.360	1280.270	2.29	1.75	1372.380
2020-01-08	1372.200	1604.900	1351.290	2.29	1.75	1356.820

Sumber : Penelitian Mandiri 2025

Tabel di atas menunjukkan bahwa format angka pada kolom harga dan variabel makroekonomi masih menggunakan tanda koma sebagai pemisah desimal. Hal ini menunjukkan bahwa data masih dalam format mentah dan belum siap langsung digunakan dalam pemodelan machine learning. Penyesuaian tipe data dan format angka akan dijelaskan lebih lanjut pada tahap Data Preparation.

3. Pemeriksaan Data Hilang (Missing Value)

Salah satu aspek penting dalam tahap Data Understanding adalah pemeriksaan terhadap nilai yang hilang (missing values). Nilai hilang merupakan kondisi di mana suatu entri dalam dataset tidak memiliki data pada kolom tertentu.

Keberadaan nilai hilang dapat memengaruhi hasil analisis serta mengganggu kinerja model machine learning jika tidak ditangani dengan tepat. Untuk itu, dilakukan analisis guna mengidentifikasi jumlah nilai yang hilang pada setiap kolom dalam dataset. Hasil pemeriksaan ditampilkan pada gambar berikut:

Tabel 3 Jumlah Missing Value pada Setiap Kolom Dataset

Nama Kolom	Jumlah Nilai Hilang
Date	0
Open	0
High	0
Low	0
Inflation	1
Interest rate	0
Close	0

Sumber : Penelitian Mandiri 2025

Berdasarkan hasil tersebut, seluruh kolom pada dataset tidak memiliki nilai yang hilang, kecuali pada kolom Inflation yang tercatat memiliki satu data kosong. Jumlah ini sangat kecil, yaitu hanya 1 dari total 1.639 baris data (sekitar 0,06%).

Meskipun proporsinya rendah, nilai kosong tersebut tetap perlu diperhatikan, karena dapat menyebabkan gangguan dalam proses pelatihan model. Oleh karena itu, pada tahap selanjutnya, nilai hilang ini akan ditangani dengan pendekatan yang sesuai agar tidak memengaruhi integritas data secara keseluruhan.

4. Statistik Deskriptif Fitur Numerik

Untuk memahami karakteristik awal data numerik dalam dataset, dilakukan analisis statistik deskriptif terhadap enam fitur numerik utama, yaitu Open, High, Low, Inflation, Interest rate, dan Close. Statistik yang ditampilkan meliputi jumlah data (count), nilai rata-rata (mean), standar deviasi

(std), nilai minimum (min), kuartil pertama (25%), median (50%), kuartil ketiga (75%), dan nilai maksimum (max). Hasilnya dapat dilihat pada Tabel 4.x.

Secara umum, seluruh fitur memiliki jumlah entri yang sama, yaitu sebanyak 1.638 data, setelah satu baris dengan nilai kosong dihapus sebelumnya. Harga pembukaan (Open) memiliki rata-rata sebesar 1.991,71 dan harga penutupan (Close) sebesar 1.992,69, menunjukkan bahwa pergerakan harga emas selama periode pengamatan cenderung stabil secara keseluruhan. Fitur High dan Low masing-masing mencerminkan harga tertinggi dan terendah harian, dengan nilai maksimum mencapai 3.245,44 dan minimum hingga 1.451,28, mencerminkan adanya fluktuasi yang signifikan.

Variabel makroekonomi seperti Inflation memiliki nilai rata-rata sebesar 0,363 dengan standar deviasi 0,355, menandakan variasi yang relatif kecil dan nilai yang mendekati nol, serta berkisar dari -0,8 hingga 1,3. Sedangkan Interest rate memiliki nilai rata-rata 2,212 dan deviasi standar sebesar 2,301, yang menunjukkan adanya sebaran data yang lebih besar, dengan nilai maksimum 5,33 dan minimum 0,05. Analisis ini memberikan gambaran awal yang penting mengenai distribusi dan skala nilai masing-masing fitur. Informasi ini juga menjadi dasar pertimbangan dalam tahap selanjutnya, seperti normalisasi dan pemilihan parameter model, agar proses pelatihan algoritma regresi dapat berjalan secara optimal dan tidak bias terhadap fitur yang memiliki rentang nilai jauh lebih besar.

Tabel 4 Statistik Deskriptif Fitur Numerik Dataset Harga Emas

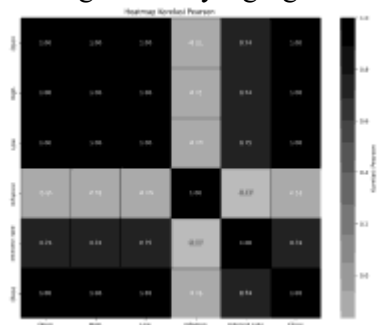
Statistik	Open	High	Low	Inflation	Interest rate	Close
Count	1.638.000	1.638.000	1.638.000	1.638.000	1.638.000	1.638.000
Mean	1.991.709	2.901.989	1.979.961	0.363	2.212	1.992.686
Std	373.218	377.893	333.874	0.355	2.301	396.473
Min	1.451.250	3.245.444	1.451.280	-0.800	0.050	3.431.587
25%	1.783.841	2.794.853	1.774.274	0.108	0.680	1.784.188
50% (Median)	1.888.107	2.898.437	1.874.484	0.400	0.930	1.888.433
75%	2.074.607	2.937.742	2.011.897	0.600	1.880	2.077.468
Max	3.188.993	3.245.436	3.184.793	1.300	5.330	3.237.739

Sumber : Penelitian Mandiri 2025

5. Analisis Korelasi Antar Fitur

Untuk memperoleh pemahaman yang lebih mendalam mengenai hubungan antar variabel dalam dataset, dilakukan analisis korelasi antar fitur menggunakan metode Pearson. Korelasi Pearson merupakan ukuran statistik yang digunakan untuk mengukur kekuatan dan arah hubungan linear antar dua variabel numerik. Nilai koefisien berkisar antara -1 hingga 1, dengan interpretasi:

- 1) Nilai mendekati 1 menunjukkan hubungan linear positif yang sangat kuat.
- 2) Nilai mendekati -1 menunjukkan hubungan linear negatif yang sangat kuat.
- 3) Nilai mendekati 0 mengindikasikan tidak adanya hubungan linear yang signifikan.



Gambar 6 Heatmap Korelasi Pearson Antar Fitur

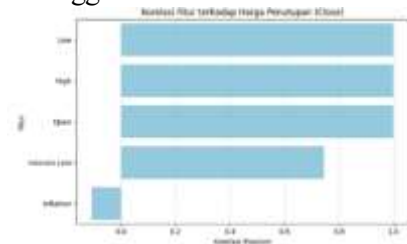
Sumber : Penelitian Mandiri 2025

Berdasarkan hasil tersebut, Open, High, dan Low memiliki korelasi sangat kuat (1.00) satu sama lain, yang menunjukkan hubungan teknikal yang erat dalam satu hari perdagangan. Interest rate menunjukkan korelasi sedang terhadap fitur

teknikal (sekitar 0.74), sedangkan Inflation memiliki korelasi sangat lemah terhadap semua fitur. Hasil ini mengindikasikan bahwa fitur teknikal saling berhubungan erat, dan perlu diperhatikan kemungkinan multikolinearitas dalam pemodelan.

6. Korelasi Fitur terhadap Harga Penutupan (Close)

Analisis ini bertujuan untuk mengetahui seberapa besar hubungan linier antara masing-masing fitur dengan target Close. Korelasi dihitung menggunakan metode Pearson, yang menghasilkan nilai dalam rentang -1 hingga 1.



Gambar 7 Korelasi Fitur terhadap Harga Penutupan (Close)

Sumber : Penelitian Mandiri 2025

Berdasarkan gambar di atas, fitur Open, High, dan Low memiliki nilai korelasi sangat tinggi terhadap harga penutupan, yaitu mendekati 1.00. Hal ini menunjukkan bahwa fitur-fitur teknikal tersebut berkontribusi besar dalam membentuk harga akhir emas harian. Interest rate juga menunjukkan hubungan yang cukup kuat dengan nilai korelasi sekitar 0.74. Sementara itu, Inflation hanya menunjukkan korelasi sangat lemah, sehingga pengaruhnya terhadap harga penutupan kurang signifikan dalam konteks linear.

3.3 Data Preparation

Tahap Data Preparation dalam pendekatan CRISP-DM merupakan

proses penting yang bertujuan untuk membersihkan dan menyiapkan data mentah agar siap digunakan dalam pemodelan. Pada tahap ini dilakukan sejumlah langkah seperti konversi tipe data, penanganan nilai kosong, serta penyusunan dataset akhir yang akan diproses oleh algoritma regresi machine learning.

Langkah-langkah pada tahap ini dilakukan berdasarkan temuan pada proses Data Understanding, serta menyesuaikan dengan format data yang diperoleh dari sumber asli. Persiapan data yang baik akan meningkatkan akurasi model dan menghindari kesalahan pada tahap pemodelan.

1. Konversi Format Desimal dan Tipe Data

Berdasarkan eksplorasi pada tahap sebelumnya, seluruh fitur numerik dalam dataset awal, termasuk Open, High, Low, Inflation, Interest rate, dan Close, masih bertipe object. Hal ini disebabkan oleh penggunaan tanda koma (,) sebagai pemisah desimal, yang merupakan format standar lokal pada file CSV yang digunakan.

Format angka seperti ini menyebabkan Python membaca nilai numerik sebagai string (object), sehingga tidak dapat langsung digunakan dalam perhitungan matematis maupun proses pemodelan. Jika tidak dilakukan konversi, maka proses training model machine learning akan gagal karena tipe data yang tidak sesuai dengan kebutuhan algoritma regresi.

Untuk mengatasi hal ini, dilakukan dua langkah utama dalam proses konversi:

- 1) Penggantian tanda koma menjadi titik (.) sebagai pemisah desimal menggunakan fungsi `.str.replace(',', '.')`.
- 2) Konversi isi kolom menjadi tipe numerik (float) menggunakan

fungsi `pd.to_numeric()` dengan parameter `errors='coerce'` untuk memastikan nilai yang tidak valid dikonversi menjadi NaN. Selain itu, karakter-karakter tidak valid seperti spasi dan titik di akhir angka juga dihapus menggunakan metode `str.strip()` dan `str.rstrip('.')`. Seluruh proses ini dilakukan terhadap kolom-kolom numerik yang relevan secara iteratif. Dengan langkah-langkah tersebut, seluruh fitur numerik berhasil dikonversi ke tipe data `float64` dan telah siap digunakan dalam proses pemodelan regresi tanpa kendala kompatibilitas tipe data.

Tabel 5 Tipe Data Kolom Dataset Setelah Konversi

Nama Kolom	Tipe Data
Date	object
Open	float64
High	float64
Low	float64
Inflation	float64
Interest rate	float64
Close	float64

Sumber : Penelitian Mandiri 2025

2. Penanganan Nilai Kosong (Missing Value)

Pemeriksaan terhadap nilai kosong dilakukan menggunakan fungsi `df.isnull().sum()`. Hasil pemeriksaan menunjukkan bahwa hanya kolom Inflation yang memiliki satu entri kosong, sementara kolom lainnya terisi lengkap tanpa nilai yang hilang. Karena nilai kosong tersebut hanya terjadi pada satu baris dari total 1.639 baris data (sekitar 0,06%), maka diputuskan untuk menghapus baris tersebut menggunakan fungsi `df.dropna()`.

Keputusan penghapusan diambil karena pengaruhnya terhadap distribusi data sangat kecil dan tidak berdampak signifikan terhadap kinerja model. Setelah proses penghapusan dilakukan, dilakukan verifikasi ulang untuk

memastikan tidak terdapat lagi nilai kosong atau entri tidak valid yang tersisa dalam dataset. Dengan demikian, data telah bersih dan siap digunakan pada tahap pemodelan.

Tabel 6 Jumlah Missing Values Setelah Pembersihan

Nama Kolom	Jumlah Nilai Hilang
Date	0
Open	0
High	0
Low	0
Inflation	0
Interest rate	0
Close	0

Sumber : Penelitian Mandiri 2025

3. Pemisahan Fitur dan Target serta Standarisasi Fitur

Pada tahap ini, dilakukan pemisahan antara variabel fitur dan target. Variabel target yang digunakan adalah Close, yaitu harga penutupan harian emas, sementara variabel fitur terdiri dari Open, High, Low, Inflation, Interest rate. Kolom Date diabaikan karena bersifat non-numerik dan tidak memberikan kontribusi langsung terhadap prediksi numerik dalam model regresi.

Berikut adalah potongan kode pemisahan fitur dan target:

```
# Pisahkan fitur dan target
X = df.drop(columns=['Date', 'Close'])
y = df['Close']
```

Gambar 8 Kode Pemisahan Fitur dan Target

Sumber : Penelitian Mandiri 2025
Setelah fitur dan target dipisahkan, dilakukan proses standarisasi fitur numerik dengan menggunakan `StandardScaler` dari pustaka `scikit-learn`. Standarisasi ini bertujuan untuk mengubah skala seluruh fitur numerik agar memiliki rata-rata (mean) nol dan standar deviasi satu. Langkah ini khusus diterapkan pada model `Support Vector Regression (SVR)` karena SVR sensitif terhadap skala data. Sementara itu, model lain seperti `Random Forest` dan

XGBoost tidak memerlukan proses standardisasi karena bersifat tree-based dan tidak terpengaruh oleh skala fitur. Berikut adalah potongan kode standardisasi:

```
# Scaling fitur numerik khusus untuk SVR
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

Gambar 9 Standardisasi Fitur untuk SVR

Sumber : Penelitian Mandiri 2025

Proses ini menghasilkan data fitur terstandardisasi yang akan digunakan dalam pelatihan model SVR. Dengan adanya transformasi ini, model diharapkan dapat mempelajari pola dengan lebih akurat dan stabil.

4. Pembagian Data Training dan Testing

Setelah data fitur (X) dan target (y) berhasil dipisahkan dan dilakukan proses standardisasi (khusus untuk SVR), tahap berikutnya adalah membagi data ke dalam dua subset: data pelatihan (training) dan data pengujian (testing). Tujuan dari pembagian ini adalah untuk mengevaluasi performa model secara objektif pada data yang belum pernah dilihat sebelumnya.

Pembagian dilakukan dengan menggunakan fungsi `train_test_split` dari library `sklearn.model_selection`. Proporsi yang digunakan

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Gambar 10 Pembagian Data Latih dan Data Uji

Sumber : Penelitian Mandiri 2025

yaitu 80% data untuk pelatihan dan 20% untuk pengujian. Parameter `random_state` disetel untuk menjaga konsistensi hasil pembagian di setiap eksekusi kode.

3.4 Modelling

Pada tahap ini dilakukan pemodelan regresi menggunakan tiga algoritma berbeda, yaitu Random Forest, Support Vector Regressor (SVR), dan XGBoost. Sebelumnya, data fitur telah dibagi menjadi data latih dan data uji. Khusus untuk model SVR, data fitur perlu dilakukan standardization menggunakan metode `StandardScaler` karena model ini sensitif terhadap skala fitur.

Setiap model diinisialisasi dengan parameter dasar, di mana jumlah estimator untuk Random Forest dan XGBoost ditetapkan sebanyak 100 pohon untuk menjaga keseimbangan antara akurasi dan efisiensi waktu komputasi. Proses pelatihan dilakukan menggunakan data latih, sedangkan evaluasi dilakukan pada data uji.

```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.svm import SVR
from xgboost import XGBRegressor
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

# Scaling khusus untuk SVR
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# Inisialisasi model
models = {
    'Random Forest': RandomForestRegressor(n_estimators=100, random_state=42),
    'SVR': SVR(),
    'XGBoost': XGBRegressor(n_estimators=100, random_state=42, verbosity=0)
}
```

Gambar 11 Inisialisasi Model Regresi

Sumber : Penelitian Mandiri 2025

Gambar ini menunjukkan proses inisialisasi model regresi yang digunakan dalam penelitian ini, yaitu Random Forest, SVR, dan XGBoost. Sebelum menerapkan SVR, data fitur terlebih dahulu diskalakan menggunakan `StandardScaler` agar model dapat menghasilkan performa yang optimal.

3.5 Evaluation

Evaluasi model bertujuan untuk mengukur seberapa baik model prediktif dalam memperkirakan harga penutupan emas berdasarkan fitur-fitur input. Pada penelitian ini, evaluasi dilakukan terhadap tiga algoritma

regresi, yaitu SVR, Random Forest, dan XGBoost.

Pengukuran performa dilakukan menggunakan tiga metrik evaluasi regresi:

- 1) MAE: Mengukur rata-rata selisih absolut antara nilai aktual dan hasil prediksi. Semakin kecil nilainya, semakin baik model.
- 2) RMSE: Mengukur kesalahan prediksi dengan penalti yang lebih besar untuk prediksi yang jauh meleset.
- 3) R²: Mengukur seberapa besar variasi pada variabel target yang dapat dijelaskan oleh model. Nilai ideal mendekati 1.

1. Hasil Evaluasi Model

Setelah proses pelatihan dan pengujian model dilakukan, langkah selanjutnya adalah mengevaluasi performa masing-masing algoritma regresi. Evaluasi ini dilakukan dengan menggunakan tiga metrik utama, yaitu Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan R-Squared (R²). Berdasarkan hasil prediksi terhadap data uji, Tabel 4.7 menunjukkan nilai evaluasi dari ketiga algoritma regresi:

Tabel 7 Hasil Evaluasi Kinerja Model

Regresi

Model	MAE	RMSE	R ²
Random Forest	7.5065	11.3120	0.9987
SVR	110.2023	215.4136	0.5266
XGBoost	7.9251	12.2479	0.9985

Sumber : Penelitian Mandiri 2025

Tabel ini menampilkan hasil evaluasi tiga algoritma regresi. Random Forest menunjukkan performa terbaik dengan error terendah dan nilai R² mendekati 1, sementara SVR memiliki performa terendah dengan error tinggi dan R² rendah.

2. Analisis Perbandingan Model

Dari hasil evaluasi yang diperoleh, terlihat bahwa setiap algoritma memberikan performa yang bervariasi terhadap data uji. Berikut adalah ringkasan analisis dari ketiga model:

- a. SVR menunjukkan hasil prediksi yang cukup baik, namun sensitif terhadap parameter dan skala data. Performa SVR sangat bergantung pada proses normalisasi dan pemilihan kernel.
- b. Random Forest memberikan hasil yang stabil dan cenderung memiliki nilai error yang rendah. Hal ini sejalan dengan sifat algoritma yang mampu menangani data kompleks dan menangkap non-linearitas tanpa banyak pengaturan parameter.
- c. XGBoost memiliki keunggulan dalam hal efisiensi dan akurasi, terutama pada dataset yang memiliki pola kompleks. Namun, performanya tetap bergantung pada penyetelan parameter lebih lanjut (tuning), yang dalam penelitian ini belum dilakukan secara mendalam.

3. Model Terbaik

Evaluasi terhadap tiga algoritma regresi machine learning Random Forest, XGBoost, dan SVR dilakukan dengan menggunakan tiga metrik evaluasi utama, yaitu MAE, RMSE, dan R². Hasil evaluasi menunjukkan perbedaan yang cukup signifikan dalam performa ketiganya.

Random Forest menunjukkan performa paling optimal dibandingkan dua model lainnya. Model ini menghasilkan nilai MAE sebesar 7.5065, RMSE sebesar 11.3120, dan R² sebesar 0.9987. Nilai R² tersebut mengindikasikan bahwa model ini mampu menjelaskan 99,87% variasi dari harga penutupan emas yang diamati, menjadikannya sangat andal untuk prediksi data time-series. Selain itu, nilai error yang rendah memperlihatkan bahwa prediksi yang dihasilkan sangat dekat dengan nilai aktual.

XGBoost tampil sebagai model kedua terbaik, dengan performa yang sangat kompetitif terhadap Random Forest.

Model ini menghasilkan MAE sebesar 7.9251, RMSE sebesar 12.2479, dan R^2 sebesar 0.9985. Meskipun nilai MAE dan RMSE sedikit lebih tinggi dibandingkan Random Forest, perbedaannya sangat kecil dan tetap berada dalam kategori sangat baik. XGBoost juga memiliki keunggulan dalam hal efisiensi pemrosesan dan kemampuan tuning hyperparameter yang fleksibel, menjadikannya pilihan yang layak untuk skenario yang memerlukan penyesuaian parameter secara mendalam.

SVR merupakan model dengan performa terendah dalam penelitian ini. Model ini mencatat MAE sebesar 110.2023, RMSE sebesar 215.4136, dan R^2 sebesar 0.5266. Nilai R^2 yang rendah menunjukkan bahwa model ini hanya mampu menjelaskan sekitar 52,66% dari variasi harga penutupan emas, yang berarti kualitas prediksinya kurang dapat diandalkan. Kinerja yang rendah ini kemungkinan disebabkan oleh ketidaksesuaian algoritma SVR terhadap kompleksitas distribusi data yang digunakan, serta parameter default yang tidak optimal tanpa tuning lanjutan.

Secara keseluruhan, Random Forest terbukti sebagai model paling akurat dan stabil dalam estimasi harga penutupan harian emas berdasarkan data historis dari broker Exness tahun 2020–2024.

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan terhadap estimasi harga penutupan harian emas (XAU/USD) menggunakan algoritma regresi machine learning dengan pendekatan CRISP-DM, maka diperoleh beberapa kesimpulan sebagai berikut:

1. Berdasarkan hasil analisis korelasi, variabel harga tertinggi (High) dan harga terendah (Low) memiliki hubungan paling kuat terhadap harga

penutupan emas. Variabel harga pembukaan (Open), inflasi, dan suku bunga juga memberikan kontribusi, meskipun pengaruhnya relatif lebih rendah dibanding faktor teknikal utama.

2. Ketiga algoritma regresi yang diuji, yaitu Support Vector Regression (SVR), Random Forest, dan XGBoost, mampu membentuk model prediksi harga penutupan emas dengan akurasi yang bervariasi. Evaluasi menggunakan MAE, RMSE, dan R^2 menunjukkan perbedaan karakteristik performa pada masing-masing algoritma.
3. Algoritma Random Forest menunjukkan performa paling unggul dengan nilai MAE dan RMSE terendah serta R^2 tertinggi (0,9988). Hal ini mengindikasikan bahwa Random Forest memiliki kemampuan yang lebih konsisten dan akurat dalam mengestimasi harga penutupan emas dibandingkan SVR dan XGBoost.

DAFTAR PUSTAKA

- Y. Zhang, J. Wang, dan L. Li, "Structural analysis and forecast of gold price returns," *Journal of Management Science and Engineering*, vol. 6, no. 2, pp. 135–145, Jun. 2021. [Online]. Tersedia: <https://doi.org/10.1016/j.jmse.2021.02.011>
- XGboost/ A. A. Saputra, B. N. Sari, dan C. Rozikin, "Penerapan Algoritma Extreme Gradient Boosting (XGBoost) untuk Analisis Risiko Kredit," *Jurnal Ilmiah Wahana Pendidikan*, vol. 10, no. 7, pp. 27–36, Apr. 2024.
- R2 score/I. Abdillah, *Prediksi Harga Menggunakan Ensemble Feature Selection dan Machine Learning*, Skripsi, Fakultas Sains dan Teknologi, UIN Syarif Hidayatullah Jakarta, 2024. [Online]. Available: <https://repository.uinjkt.ac.id/dspace/bitstream/123456789/76169/1/IRFAN%20ABDILLAH%20-%20FAH.pdf>