

BAB III

METODE PENELITIAN

A. Jenis Penelitian

Penelitian ini tergolong penelitian kuantitatif dengan pendekatan machine learning, yang ditujukan untuk memprediksi harga emas harian berdasarkan data historis. Penelitian ini bersifat eksperimental karena dilakukan melalui serangkaian percobaan dengan menerapkan dan membandingkan algoritma Support Vector Regression (SVR) dan Random Forest Regression (RFR) pada data harga emas harian.

Eksperimen dilakukan dengan menggunakan beberapa skenario yang melibatkan variasi periode data historis, yaitu 1 tahun, 3 tahun, dan 5 tahun, serta penerapan kedua algoritma pada masing-masing periode tersebut. Setiap skenario melalui tahapan persiapan data, pembentukan sliding window, pembagian data latih dan data uji, pemodelan, serta evaluasi kinerja model menggunakan Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan koefisien determinasi (R^2).

Melalui serangkaian eksperimen tersebut, penelitian ini bertujuan untuk mengetahui kemampuan masing-masing algoritma dalam memprediksi harga emas harian serta membandingkan kinerja SVR dan RFR berdasarkan hasil evaluasi yang diperoleh. Dengan demikian, hasil penelitian dapat digunakan untuk menentukan algoritma dan periode data yang memberikan kinerja prediksi terbaik pada data harga emas yang digunakan.

B. Setting Penelitian

Penelitian dilaksanakan sepanjang tahun 2026. Pengambilan data dilakukan secara daring melalui platform Yahoo Finance yang menyediakan data historis harga emas. Selanjutnya, proses pengolahan data, pemodelan, dan evaluasi dijalankan menggunakan bahasa pemrograman Python di platform Google Colab.

C. Bahan Penelitian

Bahan penelitian yang dipakai berupa data sekunder bersumber dari Yahoo Finance, yaitu data historis harga emas harian (Gold Futures/XAU) periode 1 Januari 2021 sampai 31 Desember 2025 dalam bentuk time series aktual (real data) yang mencerminkan pergerakan harga emas pada rentang tersebut. Data ini bukan hasil prediksi maupun olahan dari model lain, sehingga layak dijadikan dasar analisis.

Data historis yang dipakai mencakup fitur Open (harga pembukaan), High (harga tertinggi), Low (harga terendah), Close (harga penutupan), dan Volume perdagangan. Kelima fitur ini menjadi variabel dalam membangun model prediksi harga emas harian menggunakan algoritma Support Vector Regression (SVR) dan Random Forest Regression (RFR).

D. Alat Penelitian

Perangkat keras (hardware) yang dipakai berupa satu unit komputer/laptop dengan spesifikasi prosesor Intel Core i3-5005U @ 2.00 GHz, RAM 12 GB, penyimpanan 224 GB SSD dan 466 GB HDD, kartu grafis Intel HD Graphics 5500, serta sistem operasi Windows 10 Home 64-bit. Perangkat ini mendukung seluruh rangkaian penelitian, mulai dari pengolahan data, pembangunan model machine learning, hingga implementasi aplikasi.

Adapun perangkat lunak (software) yang digunakan meliputi:

1. Sistem Operasi Windows 10
2. Bahasa pemrograman Python sebagai alat utama pengolahan data dan pemodelan machine learning
3. Google Colab sebagai lingkungan pengembangan berbasis cloud untuk menjalankan kode program
4. Library scikit-learn untuk implementasi algoritma SVR dan Random Forest Regression
5. Library pandas untuk pengolahan dan manipulasi data
6. Library matplotlib untuk visualisasi data

7. Microsoft Excel (opsional) untuk memeriksa data awal
8. Web browser untuk mengakses data dari Yahoo Finance

Python melalui Google Colab dipilih karena menawarkan fleksibilitas dalam pengolahan data serta kemudahan dalam mengimplementasikan berbagai algoritma machine learning. Library scikit-learn menyediakan fitur lengkap untuk membangun model SVR dan Random Forest Regression sekaligus mendukung proses evaluasi model secara efisien. Selain itu, penggunaan platform berbasis cloud memungkinkan komputasi dijalankan tanpa terkendala keterbatasan spesifikasi perangkat lokal.

E. Prosedur Penelitian

Prosedur penelitian mengacu pada kerangka kerja Cross-Industry Standard Process for Data Mining (CRISP-DM) yang telah diuraikan pada Bab II, digunakan untuk memastikan proses analisis data berjalan sistematis dan terstruktur. Tahapan penelitian berdasarkan CRISP-DM akan dijelaskan sebagai berikut:

1. Business Understanding

Tahap ini menjadi langkah awal untuk memahami permasalahan penelitian sekaligus tujuan yang hendak dicapai. Permasalahan yang diangkat dalam penelitian ini adalah fluktuasi harga emas harian yang sulit diprediksi secara akurat lewat metode konvensional, sehingga dibutuhkan pendekatan machine learning untuk memodelkan pola pergerakannya agar prediksi lebih akurat. Tujuan penelitian ini mencakup: membangun model prediksi harga emas harian menggunakan SVR dan RFR, membandingkan performa kedua algoritma, serta menganalisis pengaruh penggunaan periode data historis 1 tahun, 3 tahun, dan 5 tahun terhadap hasil prediksi harga emas harian.

2. Data Understanding

Tahap ini bertujuan memahami karakteristik data sebelum masuk ke tahap pemodelan. Data yang dipakai adalah data historis

harga emas harian Gold Futures (GC=F) dari Yahoo Finance, terdiri atas fitur Open, High, Low, Close, dan Volume (OHLCV) sebagai fitur masukan, dengan target prediksi berupa harga penutupan (Close) pada hari berikutnya.

Pada tahap ini dilakukan identifikasi fitur data, analisis awal distribusi data, serta pengamatan pola pergerakan harga emas sebagai data time series. Pemeriksaan kualitas data juga dilakukan untuk mengecek keberadaan missing value, data duplikat, dan konsistensi format data. Hasil Data Understanding menjadi dasar sebelum masuk ke tahap Data Preparation.

3. Data Preparation

Tahap ini mempersiapkan data sebelum pemodelan. Hasil pemeriksaan kualitas data menunjukkan tidak ada missing value maupun data duplikat, sehingga proses data cleaning tidak diperlukan. Selanjutnya, variabel target dibentuk dari nilai harga Close hari berikutnya. Data time series kemudian disusun menggunakan metode Sliding Window dengan ukuran lag 30, sehingga setiap sampel memanfaatkan informasi fitur OHLCV selama 30 hari sebelumnya untuk memprediksi harga Close hari berikutnya.

Ukuran lag 30 hari dipilih karena dinilai mampu merepresentasikan pola pergerakan harga emas dalam jangka pendek, yaitu sekitar satu bulan perdagangan, sekaligus menjaga keseimbangan antara jumlah informasi historis yang digunakan sebagai input model dan jumlah data yang tersedia untuk pelatihan. Pendekatan jendela pengamatan 30 hari juga digunakan untuk menangkap pola pergerakan harga dalam jangka pendek tanpa mengurangi jumlah data pelatihan secara signifikan.

Setelah data terbentuk, validasi dilakukan dengan metode TimeSeriesSplit skema 5-Fold Cross Validation. Metode ini membagi data secara kronologis menjadi lima fold, sehingga pada setiap iterasi data pelatihan selalu berasal dari periode sebelum data pengujian.

Pendekatan ini dipilih karena mempertahankan karakteristik data time series sekaligus menghindari penggunaan data dari periode mendatang dalam proses pelatihan, sehingga hasil evaluasi menjadi lebih representatif.

Sebelum proses pemodelan, dilakukan normalisasi data menggunakan metode Min-Max Normalization pada kedua algoritma, yaitu Support Vector Regression (SVR) dan Random Forest Regression (RFR). Metode Min-Max Normalization dipilih karena mampu mengubah nilai setiap fitur ke dalam rentang skala yang seragam tanpa mengubah pola atau hubungan relatif antar nilai dalam data. Hal ini penting karena fitur yang digunakan memiliki rentang nilai yang berbeda, khususnya Volume yang memiliki skala nilai lebih besar dibandingkan fitur harga seperti Open, High, Low, dan Close. Dengan melakukan normalisasi, perbedaan skala antarfitur dapat dikurangi sehingga proses pemodelan tidak didominasi oleh fitur yang memiliki nilai numerik lebih besar. Pemilihan Min-Max Normalization juga sesuai untuk SVR karena algoritma tersebut sensitif terhadap perbedaan skala fitur, terutama dalam perhitungan jarak dan penerapan kernel RBF. Oleh karena itu, penskalaan diperlukan agar setiap fitur dapat berkontribusi secara lebih seimbang dalam proses pembentukan model. Pada RFR, normalisasi tidak menjadi persyaratan utama karena algoritma berbasis pohon relatif tidak sensitif terhadap skala fitur. Namun, normalisasi tetap diterapkan pada RFR untuk menjaga konsistensi perlakuan data pada kedua algoritma dan memastikan hasil eksperimen dapat dibandingkan secara adil.

Selanjutnya, dilakukan proses hyperparameter tuning untuk memperoleh kombinasi nilai parameter yang sesuai dengan karakteristik data penelitian. Hyperparameter tuning merupakan proses pencarian nilai hyperparameter yang digunakan sebelum proses pelatihan model. Proses ini dilakukan karena pemilihan nilai hyperparameter dapat memengaruhi kemampuan model dalam

mempelajari pola data dan menghasilkan prediksi. Penggunaan hyperparameter tuning diharapkan dapat menghasilkan konfigurasi model yang lebih sesuai dibandingkan penggunaan parameter tanpa proses pencarian. Dalam penelitian ini, metode yang digunakan untuk melakukan hyperparameter tuning disesuaikan dengan masing-masing algoritma, yaitu GridSearchCV pada SVR dan RandomizedSearchCV pada RFR.

Pada SVR, GridSearchCV digunakan untuk melakukan pencarian secara sistematis terhadap kombinasi nilai hyperparameter yang telah ditentukan dalam parameter grid. Sementara itu, pada RFR digunakan RandomizedSearchCV yang melakukan pencarian dengan memilih sejumlah kombinasi parameter secara acak dari ruang parameter yang telah ditentukan. Kedua metode tersebut menggunakan TimeSeriesSplit 5-Fold dalam proses validasinya, sehingga setiap kombinasi parameter dievaluasi berdasarkan pembagian data yang tetap mempertahankan urutan kronologis. Kombinasi hyperparameter yang menghasilkan kinerja terbaik berdasarkan proses cross-validation kemudian digunakan dalam pembentukan model.

Pada setiap fold, parameter normalisasi hanya dihitung berdasarkan data pelatihan (fit), kemudian digunakan untuk mentransformasi data pengujian (transform). Dengan demikian, informasi dari data pengujian tidak digunakan dalam proses pembentukan parameter normalisasi sehingga risiko terjadinya data leakage dapat dihindari.

4. Modeling

Tahap ini membangun model prediksi harga emas harian menggunakan SVR dan RFR, memakai data yang telah disiapkan pada tahap Data Preparation yaitu data historis yang sudah disusun dengan Sliding Window lag 30. Pelatihan dan pengujian model dilakukan dengan TimeSeriesSplit 5-Fold Cross Validation, sehingga pada setiap

iterasi data training selalu berasal dari periode sebelum data testing, sesuai karakteristik data time series.

Pada SVR, optimasi parameter dilakukan dengan GridSearchCV memakai skema TimeSeriesSplit 5-Fold Cross Validation. Parameter yang dioptimasi meliputi kernel, C, gamma, dan epsilon guna memperoleh kombinasi terbaik sebelum model dibangun dan dievaluasi.

Pada RFR, optimasi parameter dilakukan dengan RandomizedSearchCV memakai skema yang sama. Parameter yang dioptimasi meliputi `n_estimators`, `max_depth`, `min_samples_split`, `min_samples_leaf`, dan `max_features`. Kombinasi parameter terbaik yang diperoleh kemudian dipakai untuk membangun model.

Setelah parameter terbaik ditetapkan, masing-masing model dievaluasi ulang menggunakan TimeSeriesSplit 5-Fold Cross Validation. Hasil prediksi pada tiap fold digabungkan (out-of-fold prediction) menjadi satu kumpulan hasil prediksi, yang selanjutnya dievaluasi memakai metrik Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan Coefficient of Determination (R^2).

5. Evaluation

Tahap ini mengukur kinerja model yang telah dibangun memakai TimeSeriesSplit 5-Fold Cross Validation. Pada tiap fold, model dilatih dengan data training dan diuji dengan data testing yang tersusun kronologis, sehingga karakteristik time series tetap terjaga dan data leakage dapat dihindari. Hasil prediksi tiap fold kemudian dibandingkan dengan data aktual untuk memperoleh nilai evaluasi.

Evaluasi memakai tiga metrik: MAE untuk mengukur rata-rata kesalahan prediksi, RMSE untuk mengukur kesalahan prediksi dengan penalti lebih besar pada kesalahan tinggi, dan R^2 untuk mengukur kemampuan model menjelaskan variasi data aktual.

Penelitian ini menerapkan dua pendekatan evaluasi. Pendekatan pertama menghitung nilai MAE, RMSE, dan R^2 pada tiap fold lalu

merata-ratakannya untuk melihat kestabilan performa model. Pendekatan kedua menghitung nilai evaluasi dari gabungan seluruh hasil prediksi lima fold (out-of-fold prediction), yang dijadikan hasil evaluasi utama dalam penelitian ini.

Hasil evaluasi tersebut menjadi dasar untuk membandingkan kinerja SVR dan RFR sekaligus menentukan model dengan performa prediksi terbaik.

6. Deployment

Tahap terakhir CRISP-DM ini bertujuan mengimplementasikan model prediksi ke dalam aplikasi yang dapat dipakai pengguna. Model terbaik yang diperoleh dari proses pemodelan SVR dan RFR diimplementasikan ke dalam aplikasi berbasis web menggunakan Streamlit.

Aplikasi yang dibangun mampu menampilkan data historis harga emas, melakukan prediksi berdasarkan model terlatih, serta menampilkan hasil prediksinya. Aplikasi juga menyajikan informasi performa model berdasarkan hasil evaluasi, sehingga pengguna dapat melihat hasil prediksi dengan mudah lewat antarmuka berbasis web.

Streamlit dipilih karena mampu mengintegrasikan model machine learning yang telah dibangun di Python ke dalam aplikasi web secara sederhana, interaktif, dan mudah diakses tanpa memerlukan proses pengembangan aplikasi yang rumit.

F. Teknik Analisis Data

Teknik analisis data pada penelitian ini dilakukan melalui beberapa tahapan. Diawali dengan pengumpulan data historis harga emas harian dari Yahoo Finance dengan ticker GC=F, mencakup fitur OHLCV, dengan tiga periode pengambilan data: 1 tahun, 3 tahun, dan 5 tahun.

Selanjutnya, data dipersiapkan dengan membentuk variabel target berdasarkan fitur Close, lalu disusun menjadi data time series menggunakan metode Sliding Window dengan lag 30 setiap sampel

memanfaatkan informasi harga emas 30 hari sebelumnya untuk memprediksi nilai Close.

Model prediksi kemudian dibangun menggunakan SVR dan RFR. Masing-masing algoritma dilatih dengan data yang sudah disiapkan, dilanjutkan pencarian parameter terbaik sesuai metode masing-masing algoritma sebelum dievaluasi.

Kinerja model dievaluasi memakai MAE, RMSE, dan R^2 . MAE dan RMSE mengukur tingkat kesalahan prediksi, sedangkan R^2 mengukur kemampuan model menjelaskan variasi data aktual. Hasil evaluasi kedua algoritma dibandingkan untuk menentukan model dengan performa prediksi terbaik.

Seluruh proses analisis data diimplementasikan menggunakan Python di Google Colab, memanfaatkan beberapa library: pandas untuk pengolahan data, NumPy untuk komputasi numerik, scikit-learn untuk implementasi algoritma machine learning dan evaluasi model, matplotlib untuk visualisasi data, serta yfinance untuk mengambil data historis harga emas dari Yahoo Finance.