

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Penelitian yang dilakukan oleh A. Tholib et al. (2023) meneliti tentang perbandingan performa dua metode *deep learning*, yaitu *Long Short-Term Memory (LSTM)* dan *Gated Recurrent Unit (GRU)*, dalam memprediksi harga emas berdasarkan data historis dari tahun 2000 hingga 2022. Data yang digunakan telah melalui proses normalisasi dan dibagi menjadi data training dan data testing, dengan evaluasi performa menggunakan metrik *MAE*, *RMSE*, dan *MAPE*. Hasil penelitian menunjukkan bahwa model *LSTM* menghasilkan nilai *error* yang lebih rendah (*MAE* 0.0389, *RMSE* 0.0475, *MAPE* 5.2047%) dibandingkan *GRU*. Namun, prediksi *GRU* secara visual lebih mendekati data aktual, sehingga dapat disimpulkan bahwa *GRU* lebih efektif dan akurat dalam memprediksi harga emas berdasarkan data *time-series* [4].

Penelitian yang dilakukan oleh I. Aulia Maharani et al. (2024) menerapkan *ARIMA* untuk meramalkan harga emas di Indonesia berdasarkan data historis harian dari Januari 2014 hingga Desember 2023. Setelah dilakukan *preprocessing* dan uji stasioneritas dengan *ADF*, model terbaik yang diperoleh yaitu *ARIMA* (0,1,2) dengan nilai *AIC* terendah sebesar 71827.3. Evaluasi akurasi dilakukan menggunakan tiga skenario data uji (10%, 15%, dan 20%), dan skenario terbaik ditemukan pada data uji 15% dengan nilai *MAE* sebesar 6990,226, *RMSE* sebesar 8272,612, dan *MAPE* sebesar 0,62%. Hasil peramalan menunjukkan tren kenaikan harga emas yang konsisten selama periode prediksi, mencerminkan tingginya permintaan pasar dan indikasi inflasi [5].

Penelitian yang dilakukan oleh S. Indriyani et al. (2023) mengimplementasikan algoritma *Naive Bayes* dan *Particle Swarm Optimization (PSO)* untuk memprediksi harga emas dengan pendekatan klasifikasi naik dan turun. Menggunakan 120 data numerik dari situs Investing.com dan proses klasifikasi pada *RapidMiner*, penelitian ini menunjukkan bahwa algoritma *Naive Bayes* mampu menghasilkan akurasi sebesar 84,17%, *precision* 82,76%, dan *recall*

84,21%. Setelah dilakukan optimasi menggunakan *PSO*, akurasi meningkat menjadi 88,33%, *precision* menjadi 89,29%, dan *recall* mencapai 86,21%. Hasil ini menunjukkan bahwa kombinasi *Naive Bayes* dan *PSO* efektif dalam meningkatkan performa model prediksi harga emas berbasis data historis [6].

Penelitian yang dilakukan oleh M. Raafi Herlando et al. (2025) menganalisis dan membandingkan kinerja dua model *machine learning*, yaitu Regresi Linier dan *Support Vector Regression (SVR)*, untuk memprediksi harga minyak mentah Brent. Variabel prediktor yang digunakan meliputi data historis harga emas dan nilai tukar mata uang. Penelitian ini menerapkan tahapan pra-pemrosesan data, rekayasa fitur deret waktu, serta pengujian model pada empat skenario rentang waktu historis yang berbeda. Evaluasi kinerja menggunakan metrik *R-squared (R²)*, *Root Mean Squared Error (RMSE)*, dan *Mean Absolute Percentage Error (MAPE)*. Hasil penelitian menunjukkan bahwa Regresi Linier secara konsisten menunjukkan performa yang lebih unggul dibandingkan *SVR* pada seluruh skenario, meskipun secara umum kedua model belum mampu menghasilkan prediksi yang akurat. Hal ini mengindikasikan bahwa kompleksitas model tidak selalu menjamin peningkatan akurasi, khususnya jika variabel prediktor yang digunakan tidak memiliki daya prediksi yang kuat terhadap dinamika pasar yang kompleks [7].

Penelitian yang dilakukan oleh A. Hurifiani et al. (2024) penerapan algoritma regresi linier untuk memprediksi penjualan Alat Tulis Kantor (ATK) di BUMDes Desa Cintarasa. Penelitian ini menggunakan pendekatan *Knowledge Discovery in Database (KDD)* dengan meliputi tahapan *selection*, *preprocessing*, *transformation*, data mining, *evaluation*, dan *knowledge*. Data penjualan ATK yang digunakan mencakup periode Mei hingga September 2023, dengan atribut utama meliputi tanggal, harga, total penjualan, dan total pembelian. Model regresi linier diterapkan untuk menghasilkan prediksi penjualan harian yang kemudian dievaluasi menggunakan metrik *Root Mean Squared Error (RMSE)*, *Absolute Error*, dan *Relative Error*. Hasil penelitian menunjukkan bahwa nilai *RMSE* yang diperoleh sebesar 7.802,802, *Absolute Error* sebesar 6.894,646, dan *Relative Error* sebesar 21,05%. Berdasarkan nilai evaluasi tersebut, model prediksi dapat

dikategorikan cukup baik dan layak digunakan untuk memperkirakan penjualan ATK pada periode berikutnya [8].

2.2 Landasan Teori

2.2.1 Harga Emas (XAU/IDR)

Harga emas (XAU/IDR) menjadi fokus utama, dengan sumber data berasal dari situs resmi Investing.com. Dataset yang digunakan mencakup periode 1 Januari 2022 hingga 31 Desember 2024 dengan frekuensi harian. Atribut yang tersedia dalam data meliputi tanggal transaksi (*Date*), harga pembukaan (*Open*), harga tertinggi (*High*), harga terendah (*Low*), harga penutupan (*Price*), dan persentase perubahan harga harian (*Change %*).

Penelitian ini memfokuskan analisis pada harga penutupan (*Price*) sebagai variabel dependen yang diprediksi, sementara dimensi waktu (*Date*) diperlakukan sebagai variabel independen. Harga penutupan dipilih karena merepresentasikan nilai pasar terakhir dalam satu hari perdagangan, sehingga menjadi indikator penting dalam analisis tren dan prediksi harga [9].

Pemilihan Investing.com sebagai sumber data didasarkan pada reputasinya sebagai salah satu *platform* penyedia data keuangan yang kredibel, akurat, dan banyak digunakan dalam penelitian akademik maupun industri. Rentang waktu tiga tahun dipilih untuk memastikan ketersediaan data yang cukup beragam, sehingga memungkinkan model prediksi mempelajari pola perubahan harga emas secara lebih representatif.

Melalui analisis terhadap data ini, penelitian bertujuan membangun model prediksi berbasis algoritma regresi linier dan ARIMA yang mampu mengestimasi harga emas di masa depan secara sederhana namun dengan tingkat akurasi yang memadai.

2.2.2 Regresi Linier

Regresi linier merupakan salah satu metode statistik yang digunakan untuk memodelkan hubungan antara satu variabel independen dengan satu variabel dependen dalam bentuk garis lurus. Algoritma ini bertujuan untuk menemukan persamaan linier yang paling sesuai (*best fit*) terhadap data historis, sehingga dapat digunakan untuk melakukan prediksi nilai pada masa mendatang [10].

Secara matematis, regresi linier sederhana dapat dinyatakan dalam bentuk persamaan:

$$Y = \alpha + bX$$

Dimana Y merupakan variabel dependen (target yang diprediksi), X adalah variabel independen (fitur input), α merupakan konstanta (titik potong dengan sumbu Y), b adalah koefisien regresi (kemiringan garis regresi). Dalam konteks prediksi harga emas, variabel independen direpresentasikan oleh waktu (*Date*), sedangkan harga penutupan emas berfungsi sebagai variabel dependen.

Rumus mencari konstanta a :

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

Dimana $\sum y$ adalah total variabel dependen, $(\sum x^2)$ adalah hasil pangkat dua variabel independen, $\sum x$ adalah total variabel independen, $\sum xy$ adalah total hasil perkalian antara variabel dependen dan variabel independen, n adalah jumlah data.

Rumus mencari koefisien b :

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

Dimana $\sum y$ adalah total variabel dependen, $(\sum x^2)$ adalah hasil pangkat dua variabel independen, $\sum x$ adalah total variabel independen, $\sum xy$ adalah total hasil perkalian antara variabel dependen dan variabel independen, n adalah jumlah data. Dengan memanfaatkan regresi linier, diharapkan dapat diperoleh model yang mampu memetakan hubungan tren historis harga emas terhadap waktu, sehingga dapat digunakan untuk melakukan estimasi harga di masa depan. Kelebihan regresi linier terletak pada kesederhanaan, interpretabilitas, serta kemampuannya untuk menghasilkan model yang efisien dengan kompleksitas rendah. Oleh karena itu, metode ini dipilih untuk penelitian ini sebagai pendekatan awal dalam membangun model prediksi harga emas. Algoritma regresi linier diimplementasikan menggunakan pustaka *Scikit-learn* dalam bahasa pemrograman *Python*. Proses manipulasi dan pembersihan data dilakukan dengan *Pandas* dan *NumPy*, sedangkan visualisasi data dan hasil prediksi dilakukan menggunakan *Matplotlib*. Seluruh proses dijalankan pada platform *Google Colab*.

2.2.3 ARIMA

ARIMA (*AutoRegressive Integrated Moving Average*) merupakan salah satu metode statistik yang banyak digunakan dalam analisis dan peramalan data *time-series*, khususnya ketika data bersifat non-stasioner atau memiliki tren dan pola musiman yang tidak konstan. ARIMA menggabungkan tiga komponen utama, yaitu komponen *autoregressive (AR)*, *differencing* atau *integrated (I)*, dan *moving average (MA)*, yang secara bersamaan membentuk kerangka kerja prediktif yang fleksibel terhadap data deret waktu. Komponen *AR* merepresentasikan ketergantungan nilai saat ini terhadap nilai-nilai sebelumnya, dengan asumsi bahwa nilai masa depan dapat dijelaskan sebagai fungsi linear dari observasi masa lalu. Komponen *Integrated (I)* digunakan untuk membuat data menjadi stasioner melalui proses *differencing*, yaitu mengurangi nilai saat ini dengan nilai sebelumnya satu kali atau lebih, hingga tren atau fluktuasi jangka panjang dihilangkan. Sementara itu, komponen *MA* menangkap hubungan antara nilai saat ini dengan kesalahan (*error*) dari prediksi masa lalu. Secara umum, ARIMA dilambangkan dengan notasi ARIMA (p, d, q), dimana p menunjukkan banyaknya lag pada komponen *AR*, d menunjukkan derajat *differencing* untuk mencapai stasioneritas, q menunjukkan banyaknya lag pada komponen *MA*. Digunakan ARIMA (0,1,1), yang berarti tidak terdapat komponen *AR* atau $p = 0$, dilakukan *differencing* satu kali untuk menjadikan data stasioner ($d = 1$), dan terdapat satu komponen *MA* atau $q = 1$ [11].

Secara matematis, ARIMA (0,1,1) dapat dinyatakan dengan persamaan berikut:

$$Y_t = Y_{t-1} + \theta_1 \times \epsilon_{t-1}$$

Dimana Y_t adalah nilai prediksi saat ini, Y_{t-1} adalah nilai aktual sebelumnya, θ_1 adalah parameter dari komponen *MA*, dan ϵ_{t-1} merupakan nilai error pada waktu sebelumnya. Sebelum membangun ARIMA, dilakukan tahap identifikasi awal yang mencakup pengujian stasioneritas data menggunakan uji *Augmented Dickey-Fuller (ADF)*, serta penentuan parameter *AR* (p) dan *MA* (q) berdasarkan grafik *Autocorrelation Function (ACF)* dan *Partial Autocorrelation Function (PACF)*. Validasi performa model dilakukan menggunakan metrik evaluasi yang relevan. Dalam konteks penelitian ini, ARIMA digunakan untuk memodelkan dan

memprediksi harga emas berdasarkan data historis XAU/IDR. Diharapkan model ini mampu menangkap pola pergerakan harga yang memiliki karakteristik musiman maupun fluktuatif, serta memberikan hasil prediksi yang akurat dalam mendukung proses pengambilan keputusan keuangan berbasis data deret waktu. Proses pemodelan dilakukan menggunakan pustaka *Statsmodels* dalam bahasa pemrograman *Python*, dengan visualisasi grafik *ACF* dan *PACF* menggunakan *Matplotlib* dan *Seaborn*. Seluruh proses analisis dan pemodelan dijalankan pada lingkungan *Google Colab* [12].

2.2.4 Metrik Evaluasi

Untuk mengukur performa model prediksi yang dibangun, penelitian ini menggunakan dua metrik evaluasi utama, yaitu *MAE* dan *RMSE*. Kedua metrik ini dipilih karena mampu memberikan gambaran yang jelas mengenai tingkat kesalahan antara nilai prediksi dan nilai aktual [13].

1. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) mengukur rata-rata nilai absolut dari selisih antara nilai yang diprediksi dengan nilai aktual. Rumus *MAE* adalah sebagai berikut:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Dimana n merupakan jumlah dataset, y_i adalah nilai aktual ke- i , \hat{y}_i adalah nilai prediksi ke- i , *MAE* memberikan gambaran seberapa besar rata-rata kesalahan prediksi dalam satuan aslinya tanpa mengkuadratkan selisihnya, sehingga tidak terlalu sensitif terhadap *outlier*. Semakin kecil nilai *MAE*, semakin baik kinerja model [14].

2. Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) merupakan akar kuadrat dari *Mean Squared Error (MSE)*, yang mengembalikan nilai kesalahan ke dalam satuan yang sama dengan variabel target. Rumus *RMSE* adalah sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Dimana n merupakan jumlah dataset, y_i adalah nilai aktual ke- i , \hat{y}_i adalah nilai prediksi ke- i , $RMSE$ memudahkan interpretasi hasil karena memiliki satuan yang sama dengan data aslinya (dalam hal ini, harga emas). Seperti MSE , semakin kecil nilai $RMSE$, maka semakin baik akurasi model prediksi. Penggunaan MSE dan $RMSE$ secara bersamaan pada penelitian ini bertujuan untuk memberikan gambaran yang lebih komprehensif mengenai tingkat kesalahan prediksi dari regresi linier dan ARIMA [15].

2.2.5 Data Time Series

Data *time series* atau data deret waktu merupakan sekumpulan data yang dikumpulkan dan dicatat secara berurutan berdasarkan waktu. Karakteristik utama dari data *time series* yaitu adanya ketergantungan antara nilai data saat ini dengan nilai data pada waktu sebelumnya, sehingga pola historis dapat dimanfaatkan untuk memprediksi kondisi di masa mendatang. Dalam konteks penelitian ini, harga emas harian merupakan contoh dari data *time series* karena nilainya berubah dari waktu ke waktu dan dapat dianalisis berdasarkan tren, musiman (*seasonality*), serta fluktuasi jangka pendek. Analisis data *time series* umumnya memerlukan perlakuan khusus seperti uji stasioneritas dan *differencing* agar dapat dimodelkan secara akurat, salah satunya dengan ARIMA yang dirancang khusus untuk menangani karakteristik data semacam ini [16].

2.2.6 Python

Python merupakan bahasa pemrograman yang banyak digunakan dalam bidang sains data, statistik, dan pengembangan model *machine learning* karena sintaksnya yang sederhana, fleksibel, serta didukung oleh berbagai pustaka (*library*) yang kuat. Penelitian ini, *Python* digunakan sebagai alat bantu utama dalam proses pra-pemrosesan data, implementasi algoritma prediksi regresi linier dan ARIMA, serta visualisasi hasil. Beberapa pustaka yang digunakan antara lain *pandas* untuk manipulasi data, *scikit-learn* untuk regresi linier, *statsmodels* untuk penerapan ARIMA, serta *matplotlib* dan *seaborn* untuk visualisasi grafik.

Pemilihan *Python* didasarkan pada kemudahan integrasi antar *library*, dukungan komunitas yang luas, serta kemampuannya dalam menangani data dalam skala besar secara efisien [17].

2.2.7 Prediksi

Prediksi merupakan proses memperkirakan nilai atau kejadian yang akan terjadi di masa mendatang berdasarkan pola historis yang terbentuk dari data sebelumnya. Dalam konteks penelitian dan analisis data, prediksi memiliki peran penting dalam mendukung proses pengambilan keputusan yang bersifat strategis, baik dalam jangka pendek maupun jangka panjang. Model prediktif yang akurat dapat membantu dalam mengidentifikasi tren, memperkirakan perubahan, serta mengoptimalkan sumber daya yang tersedia. Salah satu metode statistik yang sering diterapkan dalam proses prediksi adalah regresi linier, yang dapat menggambarkan hubungan antara variabel independen dan variabel dependen dalam bentuk persamaan matematis. Selain itu, ARIMA juga banyak digunakan untuk menganalisis data deret waktu, khususnya data yang bersifat non-stasioner. ARIMA menggabungkan tiga komponen utama, yaitu *AR*, *differencing (I)*, dan *MA*, yang secara bersama membentuk model yang fleksibel dalam menangkap pola historis dan memproyeksikan nilai masa depan. Prediksi dilakukan terhadap harga emas berdasarkan data historis XAU/IDR menggunakan pendekatan regresi linier dan ARIMA untuk menghasilkan estimasi yang akurat dan dapat digunakan untuk mendukung pengambilan keputusan berbasis data [18] [19].