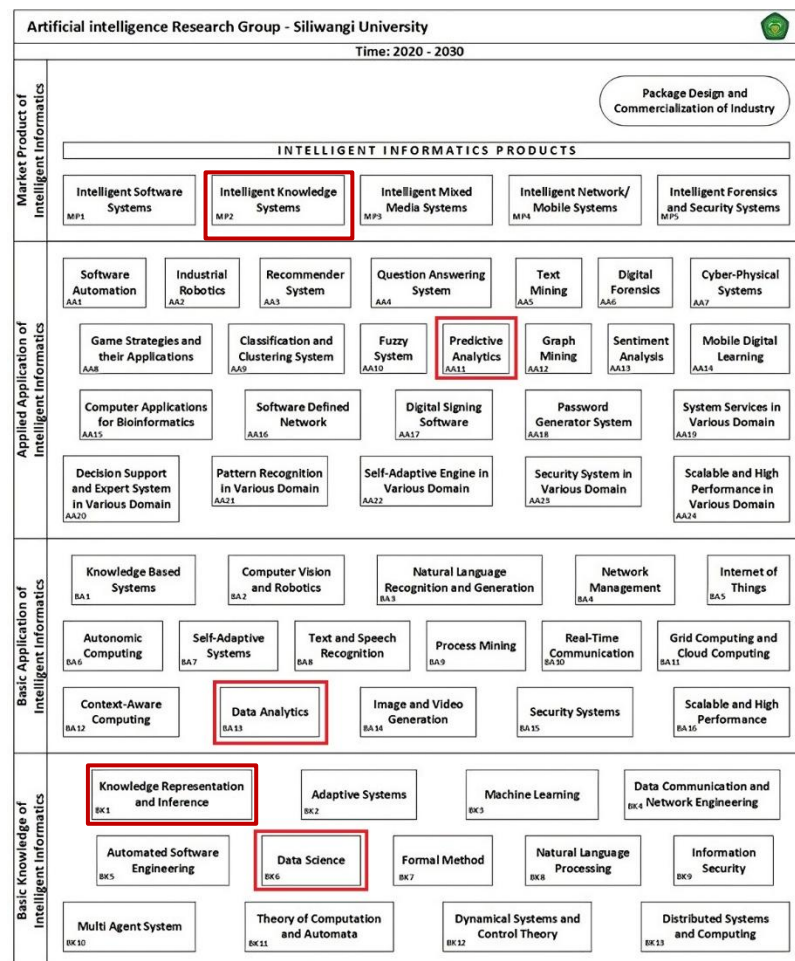


BAB III

METODOLOGI PENELITIAN

3.1 Roadmap Penelitian

Secara keseluruhan, rencana penelitian ini sejalan dengan *roadmap* Kelompok Keahlian Informatika dan Sistem Inteligen (ISI). *Roadmap* penelitian dapat dilihat pada Gambar 3.1.



Gambar 3. 1 *Roadmap* Penelitian (Artificial Intelligence Siliwangi, 2020)

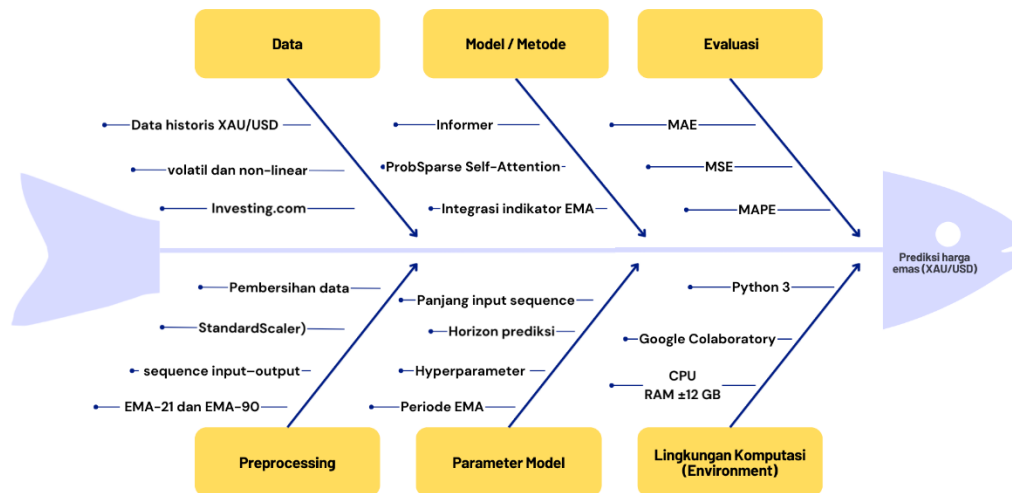
Roadmap penelitian ini berfokus pada pengembangan *Intelligent Informatics Products* dengan penekanan pada *Intelligent Knowledge Systems* seperti ditunjukkan pada Gambar 3.1. Posisi penelitian berada pada ranah *Data*

Science dan *Data Analytics*, yang ditunjukkan melalui pemanfaatan data historis harga emas (*XAU/USD*) sebagai data deret waktu yang diproses melalui tahapan pembersihan data, normalisasi, *pembentukan sequence input-output*, serta penambahan fitur indikator *Exponential Moving Average (EMA)*. Proses tersebut menghasilkan representasi data yang informatif dan siap digunakan dalam pemodelan prediktif.

Penerapan model *informer* untuk melakukan peramalan harga pada berbagai horizon waktu menempatkan penelitian ini dalam bidang *Predictive Analytics*, karena model digunakan untuk mempelajari pola temporal dan menghasilkan prediksi nilai masa depan. Kemampuan model dalam menangkap dependensi jangka panjang dan hubungan antar waktu merepresentasikan pengetahuan implisit dari data, sehingga penelitian ini juga berada dalam lingkup *Knowledge Representation and Inference*. Integrasi indikator *EMA* dengan arsitektur *informer* menghasilkan sistem prediksi yang mampu mengolah data mentah menjadi pengetahuan yang bernilai guna mendukung pengambilan keputusan, yang secara keseluruhan mengukuhkan posisi penelitian ini dalam pengembangan *Intelligent Knowledge Systems* sesuai dengan roadmap penelitian.

Sementara itu, tujuan dari penelitian ini divisualisasikan dalam bentuk diagram *fishbone*, yaitu diagram yang menggambarkan berbagai faktor penyebab dari suatu fenomena spesifik (Farmasetika & Review, 2021). Diagram ini digunakan untuk mengidentifikasi dan mengelompokkan faktor-faktor utama yang memengaruhi permasalahan penelitian secara sistematis, sehingga hubungan antara

penyebab dan tujuan penelitian dapat dianalisis dengan lebih jelas dan terstruktur, sebagaimana ditunjukkan pada Gambar 3.2



Gambar 3. 2 Diagram *Fishbone*

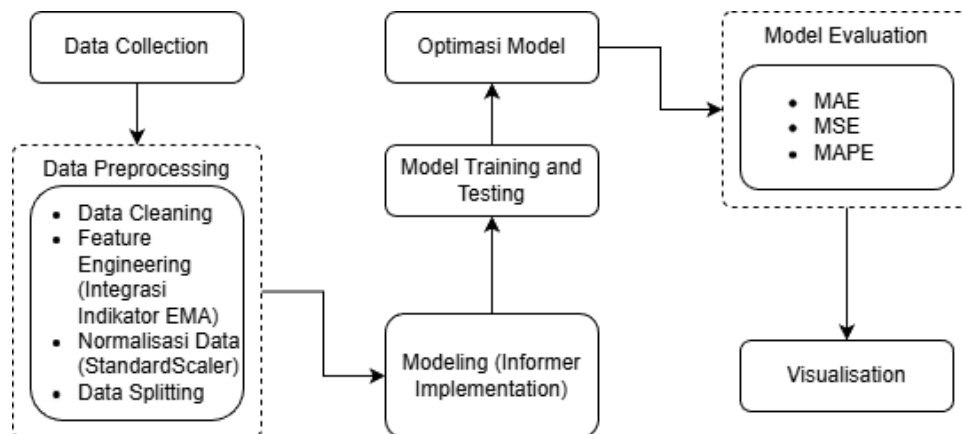
Gambar 3.2 menunjukkan *Fishbone Diagram* yang menggambarkan tahapan dan faktor-faktor utama dalam metodologi penelitian prediksi harga emas (*XAU/USD*). Diagram ini memvisualisasikan keterkaitan komponen penelitian yang terdiri atas data, *preprocessing*, model, parameter model, evaluasi, dan lingkungan komputasi yang secara kolektif memengaruhi hasil prediksi harga emas.

Tahap Data menggunakan data historis harga emas *XAU/USD* yang diperoleh dari situs Investing.com dengan karakteristik volatil dan non-linear. Tahap *preprocessing* meliputi pembersihan data, normalisasi menggunakan *StandardScaler*, serta pembentukan *sequence input-output* agar sesuai dengan kebutuhan model deret waktu. Penambahan indikator *Exponential Moving Average (EMA)* dengan periode *EMA-21* dan *EMA-90* dilakukan untuk membantu model dalam menangkap pola tren jangka pendek dan jangka panjang.

Tahap Model menerapkan model *Informer* yang menggunakan mekanisme *ProbSparse Self-Attention* untuk meningkatkan efisiensi komputasi pada deret waktu berdimensi panjang. Tahap Parameter Model mencakup penentuan panjang *input sequence*, horizon prediksi, pengaturan *hyperparameter*, serta periode *EMA*. Tahap Evaluasi dilakukan menggunakan metrik *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, dan *Mean Absolute Percentage Error (MAPE)*. Lingkungan Komputasi menggunakan Python 3 dengan platform Google Colaboratory serta dukungan CPU dan RAM ± 12 GB untuk menjalankan seluruh proses penelitian.

3.2 Metode Penelitian

Capaian dalam penelitian ini diharapkan mampu menghasilkan model prediksi harga *forex* yang lebih akurat dan efisien dengan memanfaatkan data historis serta integrasi fitur teknikal *Exponential Moving Average (EMA)* kedalam arsitektur *Informer*. Pendekatan ini memungkinkan pemodelan yang lebih kuat dalam mengidentifikasi pola pergerakan harga berdasarkan informasi temporal yang terekam dalam histori data. Penelitian ini difokuskan pada dua tahap utama, yaitu pra-pemrosesan data historis, dan implementasi serta evaluasi model *Informer* dan *Informer-EMA* untuk peramalan deret waktu. Setelah data historis melalui proses normalisasi, pembagian data *train*, *validation* dan *test*, tahap berikutnya adalah pelatihan dan pengujian model prediksi, untuk kemudian dilakukan evaluasi performa menggunakan metrik *MAPE*, *MAE*, dan *MSE*. Tahapan lengkap dari penelitian ini dapat dilihat pada Gambar 3.3.



Gambar 3. 3 Tahapan Penelitian

3.2.1 Data Collection

Pengumpulan data merupakan tahap awal yang krusial dalam penelitian ini. Data yang digunakan berupa data historis pasangan komoditas emas terhadap mata uang dolar (XAU/USD) yang diambil dari situs *Investing.com*, yang menyediakan data harga *forex* secara publik. Data mencakup fitur-fitur utama, antara lain:

- 1) Harga hari ini (*price*)
- 2) Harga pembukaan (*open*)
- 3) Harga tertinggi (*high*)
- 4) Harga terendah (*low*)

Dalam penelitian ini, harga (*price*) dipilih sebagai target utama untuk proses peramalan, karena merepresentasikan harga akhir transaksi dalam satu periode waktu, serta umum digunakan sebagai referensi dalam analisis teknikal.

3.2.2 Data Preprocessing

Tahapan *preprocessing* bertujuan untuk mengolah data agar siap digunakan dalam proses pelatihan dan pemodelan. Proses ini sangat penting untuk memastikan bahwa input yang diterima model berada dalam format, skala, dan struktur yang

sesuai. *Pra-pemrosesan* Data Historis emas (*XAU/USD*) yang telah dikumpulkan diproses melalui beberapa tahapan berikut:

a. *Data Cleaning*

Proses *Data Cleaning* pada penelitian ini mencakup tahap penghapusan fitur (*feature removal*) yang dilakukan dengan menghilangkan kolom *Volume* (*Vol.*) dan *Change %* dari dataset. Langkah ini diterapkan karena kedua atribut tersebut tidak memberikan kontribusi yang relevan terhadap tujuan pemodelan deret waktu yang berfokus pada dinamika harga. Atribut *Change %* merupakan nilai turunan dari harga sehingga berpotensi menimbulkan redundansi, sedangkan atribut *Volume* tidak menunjukkan hubungan yang konsisten dengan pergerakan harga.

b. *Feature Engineering* (Integrasi Indikator *EMA*)

Pada tahap awal *preprocessing*, fitur teknikal *Exponential Moving Average* (*EMA*) ditambahkan ke dalam dataset. *EMA* dihitung dengan periode tertentu (misalnya *EMA-21* dan *EMA-90*) berdasarkan harga penutupan untuk menangkap tren jangka pendek dan menengah. Fitur ini diintegrasikan sebagai variabel input untuk menambah informasi sekuensial yang diterima oleh model (Rifai dkk., 2024). Selain nilai *EMA* utama, proses *preprocessing* juga menambahkan dua fitur turunan, yaitu *EMA Difference* dan *EMA Cross Signal*, yang berfungsi memperjelas dinamika tren pada data harga. Fitur *EMA Difference* diperoleh dari selisih antara *EMA-21* dan *EMA-90* sehingga menghasilkan indikator yang secara eksplisit menggambarkan kekuatan tren, nilai positif menunjukkan dominasi tren

naik jangka pendek terhadap tren menengah, sedangkan nilai negatif menandakan potensi pelemahan atau pergerakan turun. Sementara itu, *EMA Cross Signal* digunakan untuk merepresentasikan peristiwa persilangan (*crossover*) antara dua garis *EMA* dalam bentuk fitur numerik. Ketika *EMA-21* bergerak naik dan memotong *EMA-90* dari bawah, kondisi tersebut dikodekan sebagai indikasi perubahan arah tren, sedangkan persilangan sebaliknya dikodekan sebagai indikasi pelemahan tren. Apabila tidak terjadi persilangan, nilai fitur berada pada kondisi netral. Representasi ini tidak digunakan sebagai sinyal transaksi (beli atau jual), melainkan sebagai informasi tambahan bagi model untuk mempelajari pola dinamika tren harga dalam proses prediksi.

c. *Normalisasi Data (Standard Scaler)*

Skala data pada penelitian ini dinormalisasi menggunakan metode *StandardScaler*, yaitu teknik standarisasi yang mengubah setiap fitur agar memiliki nilai rata-rata (*mean*) sebesar 0 dan standar deviasi (*standard deviation*) sebesar 1. Proses ini dilakukan dengan mengurangi setiap nilai data terhadap nilai rata-ratanya, kemudian membaginya dengan standar deviasi (Rachmawati dkk., 2024).

d. *Data Splitting*

Dataset kemudian dibagi menjadi tiga bagian, yaitu data latih (*training set*), data validasi (*validation set*), dan data uji (*testing set*). Proporsi pembagian bertujuan agar model dapat dievaluasi secara adil pada data yang belum pernah dilihat sebelumnya.

3.2.3 *Modelling*

Setelah data diproses, dilakukan tahap pemodelan menggunakan arsitektur *Informer*, yang merupakan pengembangan dari model *Transformer* dan dirancang khusus untuk pemrosesan data deret waktu panjang (*long-sequence time series forecasting*). Prediksi Harga *forex* Menggunakan Model *Informer* Pemodelan dilakukan dengan langkah-langkah sebagai berikut:

3.2.3.1 Implementasi Model *Informer*

Model *Informer* diimplementasikan menggunakan framework *deep learning (PyTorch)*, dengan konfigurasi parameter yang mengacu pada hasil penelitian (Zhou dkk., 2021), termasuk panjang *sequence input/output*, *batch size*, dan *learning rate*.

3.2.3.2 Komponen Utama *Informer*

- a. *ProbSparse Self-Attention* digunakan untuk mengurangi kompleksitas komputasi matriks perhatian dari $O(L^2)$ menjadi $O(L \log L)$. Mekanisme ini menyaring dan hanya berfokus pada kueri (*Query*) yang paling relevan secara probabilistik, mempertahankan akurasi sekaligus menghemat memori dan waktu komputasi saat memproses urutan data yang panjang.
- b. *Self-Attention Distillation* sebuah mekanisme penyaringan yang diterapkan pada *Encoder* untuk mengurangi panjang urutan data secara progresif (misalnya, dari L menjadi $L/2$). Proses ini menggunakan Konvolusi 1D dan *Max Pooling* untuk mengekstrak fitur yang paling dominan dan ringkas, yang sangat penting untuk manajemen memori dalam *LSTF*.

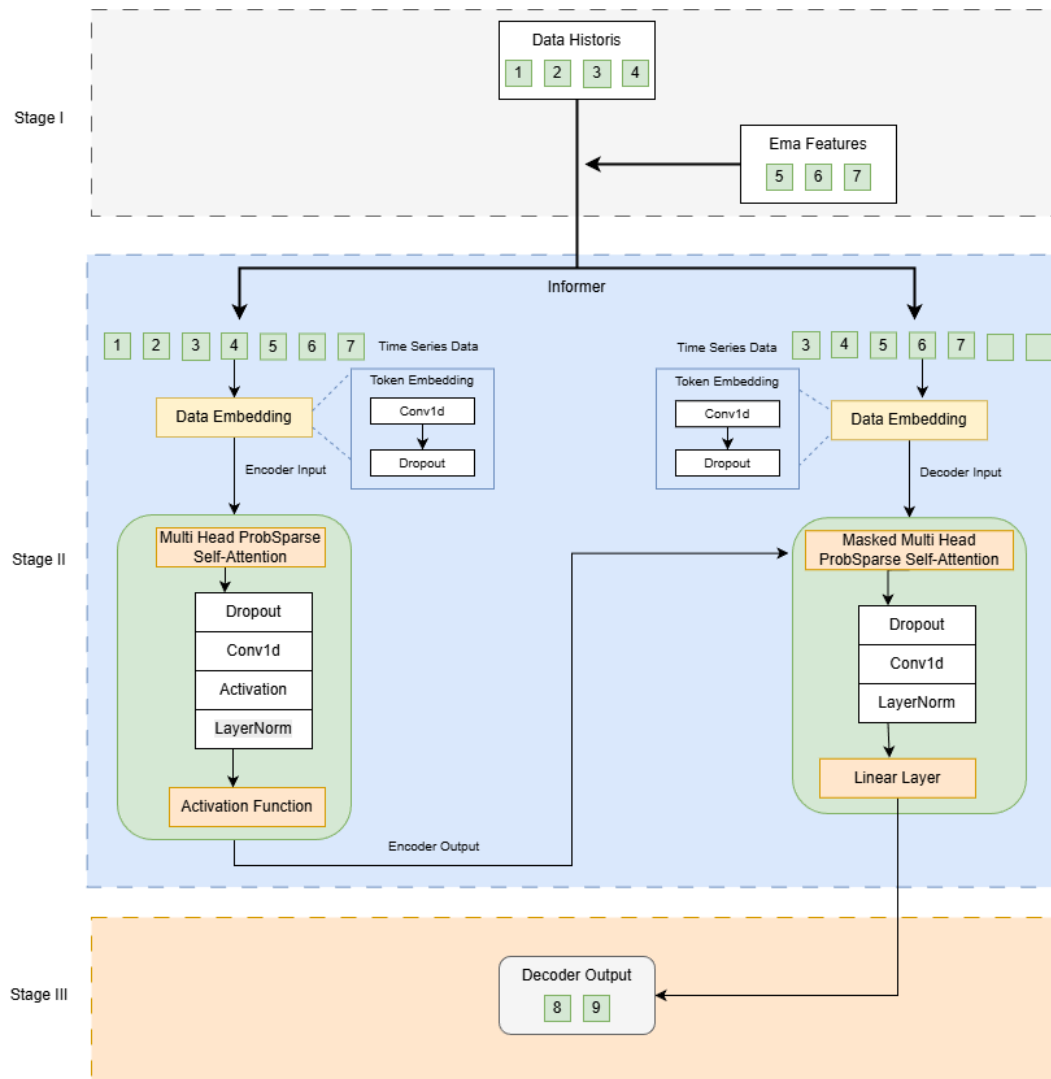
- c. *Generative Style Decoder* komponen ini memungkinkan prediksi seluruh urutan target (*Prediction Length*) dilakukan sekaligus (secara paralel) dalam satu langkah *forward*. Dengan mekanisme tersebut, proses inferensi menjadi lebih cepat dan peramalan jangka panjang dapat dilakukan secara lebih efisien, berbeda dengan metode tradisional yang memprediksi langkah waktu satu per satu (iteratif).

3.2.3.3 Model Training and Testing

Model dilatih menggunakan data yang telah diproses, baik versi dengan *EMA (Informer+EMA)* maupun tanpa *EMA (Informer baseline)* sebagai pembanding. Hasil pelatihan dari kedua model kemudian dibandingkan untuk mengukur pengaruh integrasi *EMA* terhadap akurasi prediksi.

3.2.4 Model yang diusulkan

Tahapan pemodelan pada penelitian ini dirancang untuk menggambarkan alur kerja model *Informer* dan *Informer + EMA* dalam memprediksi harga berdasarkan data historis yang telah melalui proses *preprocessing* dan *feature engineering* berupa penambahan indikator *Exponential Moving Average (EMA)* hingga menghasilkan prediksi harga emas (*XAU/USD*), sebagaimana ditunjukkan pada Gambar 3.4. Gambar tersebut memperlihatkan arsitektur model *Informer* yang terintegrasi dengan indikator *EMA*, yang dibagi ke dalam tiga tahap utama, yaitu *Stage I (Integrasi EMA)* sebagai proses penambahan fitur indikator teknikal, *Stage II (Pemrosesan Model Informer)* untuk mempelajari pola temporal pada data deret waktu, serta *Stage III (Hasil Prediksi)* yang menghasilkan nilai prediksi harga emas sebagai keluaran model.

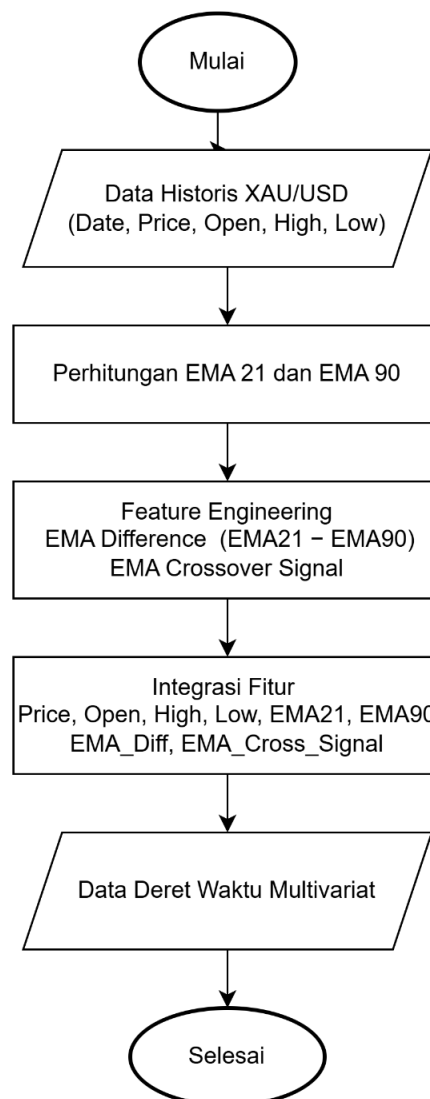


Gambar 3. 4 Model yang diusulkan

Pendekatan *Informer* + *EMA* didefinisikan sebagai integrasi indikator *Exponential Moving Average (EMA)* ke dalam model *Informer*. Integrasi dilakukan dengan memasukkan nilai *EMA 21*, *EMA 90*, serta fitur turunannya ke dalam data masukan bersama variabel harga historis, sehingga membentuk deret waktu multivariat. Indikator *EMA* menjadi bagian dari proses pembelajaran model dan berkontribusi langsung dalam pembentukan representasi data yang dipelajari oleh *Informer*. Metode yang digunakan tidak sekadar menerapkan indikator *EMA* sebagai alat analisis tambahan, melainkan mengintegrasikannya secara sistematis

pada tahap praproses dan pemodelan, sehingga istilah *Informer* terintegrasi *EMA* (*Informer* + *EMA*) digunakan dalam penelitian ini.

Stage I (Integrasi *EMA*) diawali dengan pengumpulan dan penyusunan data historis harga emas (*XAU/USD*) dalam bentuk deret waktu yang digunakan sebagai fitur utama. Selanjutnya, indikator *Exponential Moving Average (EMA)* dihitung menggunakan beberapa periode tertentu dan ditambahkan ke dalam dataset sebagai fitur tambahan.



Gambar 3. 5 *Flowchart* Integrasi *EMA*

Alur proses integrasi data historis dan indikator EMA ditunjukkan pada Gambar 3.5 yang menggambarkan tahapan perhitungan Exponential Moving Average (EMA), penggabungan fitur, hingga pembentukan data deret waktu multivariat. Perhitungan *EMA 21* dan *EMA 90* pada penelitian ini merujuk pada Persamaan (1). Selanjutnya, dilakukan proses *feature engineering* dengan membentuk fitur selisih *EMA (EMA Difference)*, yang diperoleh dari perbedaan antara nilai *EMA 21* dan *EMA 90* pada setiap waktu pengamatan. Selain itu, dibentuk fitur sinyal persilangan *EMA (EMA Crossover Signal)* untuk merepresentasikan perubahan arah tren, di mana nilai sinyal ditentukan berdasarkan kondisi persilangan antara *EMA 21* dan *EMA 90* pada waktu ke- t dan waktu sebelumnya. Seluruh fitur yang terdiri atas harga (*Price*), *Open*, *High*, *Low*, *EMA21*, *EMA90*, *EMA Difference*, dan *EMA Crossover Signal* kemudian diintegrasikan menjadi data deret waktu multivariat. Data multivariat yang dihasilkan selanjutnya digunakan sebagai masukan (input) bagi model *Informer*.

Stage II (Pemrosesan Model *Informer*) merupakan tahap inti pemodelan yang terdiri atas komponen *encoder* dan *decoder*. Pada sisi *encoder*, data deret waktu yang telah dibentuk diproses melalui tahap *data embedding* yang mencakup *token embedding*, proyeksi menggunakan *Conv1d*, serta penerapan *dropout* untuk meningkatkan stabilitas representasi data. Representasi hasil *embedding* kemudian diproses menggunakan mekanisme *Multi-Head ProbSparse Self-Attention* untuk menangkap ketergantungan temporal jangka panjang secara efisien. Hasil pemrosesan *encoder* selanjutnya menghasilkan *encoder output* yang merepresentasikan konteks historis global.

Pada sisi decoder, *decoder input* dibentuk dari sebagian data historis terakhir yang digeser sesuai dengan horizon prediksi. Data ini juga melalui tahap *data embedding* yang serupa dengan *encoder*. Selanjutnya, *decoder* memanfaatkan *Masked Multi-Head ProbSparse Self-Attention* untuk memastikan bahwa proses prediksi hanya menggunakan informasi dari waktu sebelumnya, sehingga mencegah kebocoran informasi. Output dari mekanisme tersebut dikombinasikan dengan *encoder output* melalui proses *cross-attention* untuk mengaitkan informasi historis dengan prediksi nilai masa depan. Hasil pemrosesan *decoder* kemudian dilewatkan ke lapisan linear untuk menghasilkan keluaran model.

Stage III (Hasil Prediksi) merupakan tahap akhir yang menghasilkan nilai prediksi harga emas (*XAU/USD*) sesuai dengan horizon waktu yang telah ditentukan. Nilai prediksi ini selanjutnya digunakan sebagai keluaran utama model dan menjadi dasar dalam proses evaluasi performa model pada tahap berikutnya.

3.2.5 Evaluation

Tahap evaluasi bertujuan untuk mengukur performa model dalam memprediksi harga emas (*XAU/USD*). Evaluasi dilakukan menggunakan beberapa metrik evaluasi, yaitu *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, dan *Mean Absolute Percentage Error (MAPE)*. *MAE* digunakan untuk mengukur rata-rata kesalahan absolut antara nilai prediksi dan nilai aktual, sehingga memberikan gambaran besarnya kesalahan prediksi dalam satuan harga yang mudah diinterpretasikan. *MSE* berfungsi untuk mengukur kesalahan prediksi dengan memberikan penalti yang lebih besar terhadap kesalahan bernilai ekstrem melalui proses pengkuadratan, sehingga metrik ini sensitif terhadap fluktuasi harga

yang tinggi dan mampu mencerminkan stabilitas model. Sementara itu, *MAPE* digunakan untuk mengukur tingkat kesalahan dalam bentuk persentase, sehingga memudahkan interpretasi akurasi prediksi secara relatif dan memungkinkan perbandingan kinerja model pada berbagai horizon prediksi. Dengan menggunakan ketiga metrik tersebut secara bersamaan, evaluasi performa model menjadi lebih komprehensif dan mampu menggambarkan akurasi serta stabilitas prediksi secara menyeluruh.