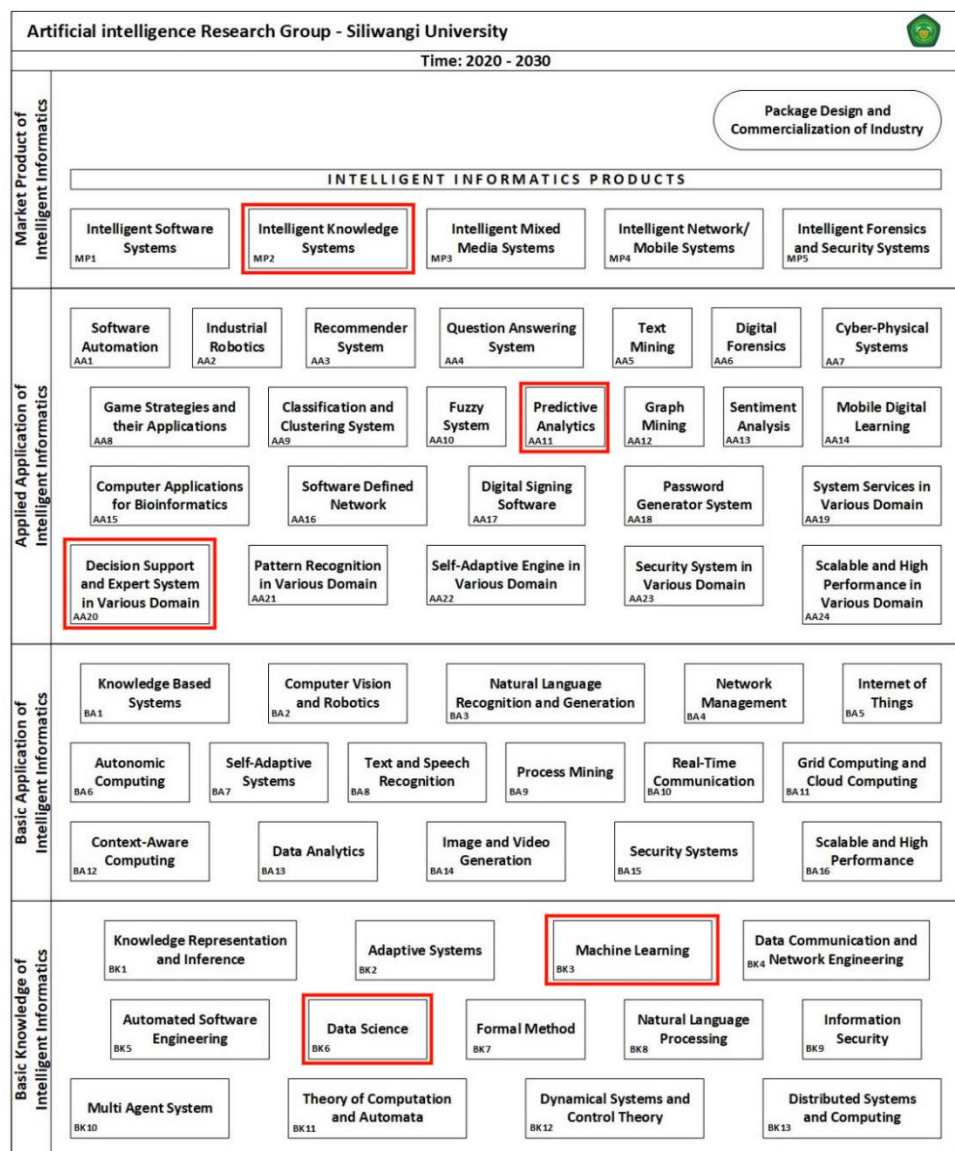


BAB III. METODOLOGI PENELITIAN

3.1 Roadmap Penelitian

Penelitian ini merupakan bagian dari peta jalan riset yang dikembangkan oleh Kelompok Keahlian Informatika dan Sistem Inteligen (KK ISI) di Universitas Siliwangi. Studi-studi sebelumnya dijadikan sebagai referensi utama untuk meningkatkan kualitas dan menyempurnakan hasil akhir dari keseluruhan roadmap penelitian, sebagaimana ditunjukkan pada Gambar 3.1.



Gambar 3. 1 Roadmap Penelitian (AIS Universitas Siliwangi, 2020)

Penelitian ini berfokus pada topik *Data Science* dan *Machine Learning* dengan penggunaan algoritma *Random Forest* dan *XGBoost* yang digabungkan menggunakan *Stacking Ensemble* serta *tuning* menggunakan *Bayesian Optimization* untuk diterapkan pada tujuan untuk memprediksi harga saham, dikembangkan dengan maksud untuk mendapatkan fungsi sebagai *Predictive Analytics* dan diharapkan menjadi *Decision Support and Expert System* sehingga dapat menjadi petunjuk dalam berbagai hal yang berkaitan pada domain saham, dengan begitu maka akan terbentuk *Intelligent Knowledge System* yang utuh dan mampu memberikan wawasan komprehensif, mendukung pengambilan keputusan yang lebih tepat, serta meningkatkan akurasi dalam menganalisis pola pergerakan harga saham.

3.2 Deskripsi Metodologi

Penelitian ini bersifat penelitian eksperimen dengan fokus pada penerapan pendekatan *Data Science* di sektor pergerakan saham. Data penelitian diambil dengan menggunakan *library yfinance* yang terhubung ke sumber data *Yahoo Finance*, sebuah platform penyedia data harga saham historis yang dapat diakses secara terbuka. Data yang diperoleh mencakup informasi numerik terkait harga saham pada periode tertentu, sehingga dapat dijadikan dasar dalam membangun model prediksi.

Dengan berbasis penelitian eksperimen (*experimental research*) yang menggunakan pendekatan kuantitatif (Penelitian Pendidikan IPA & Rahim, 2024),

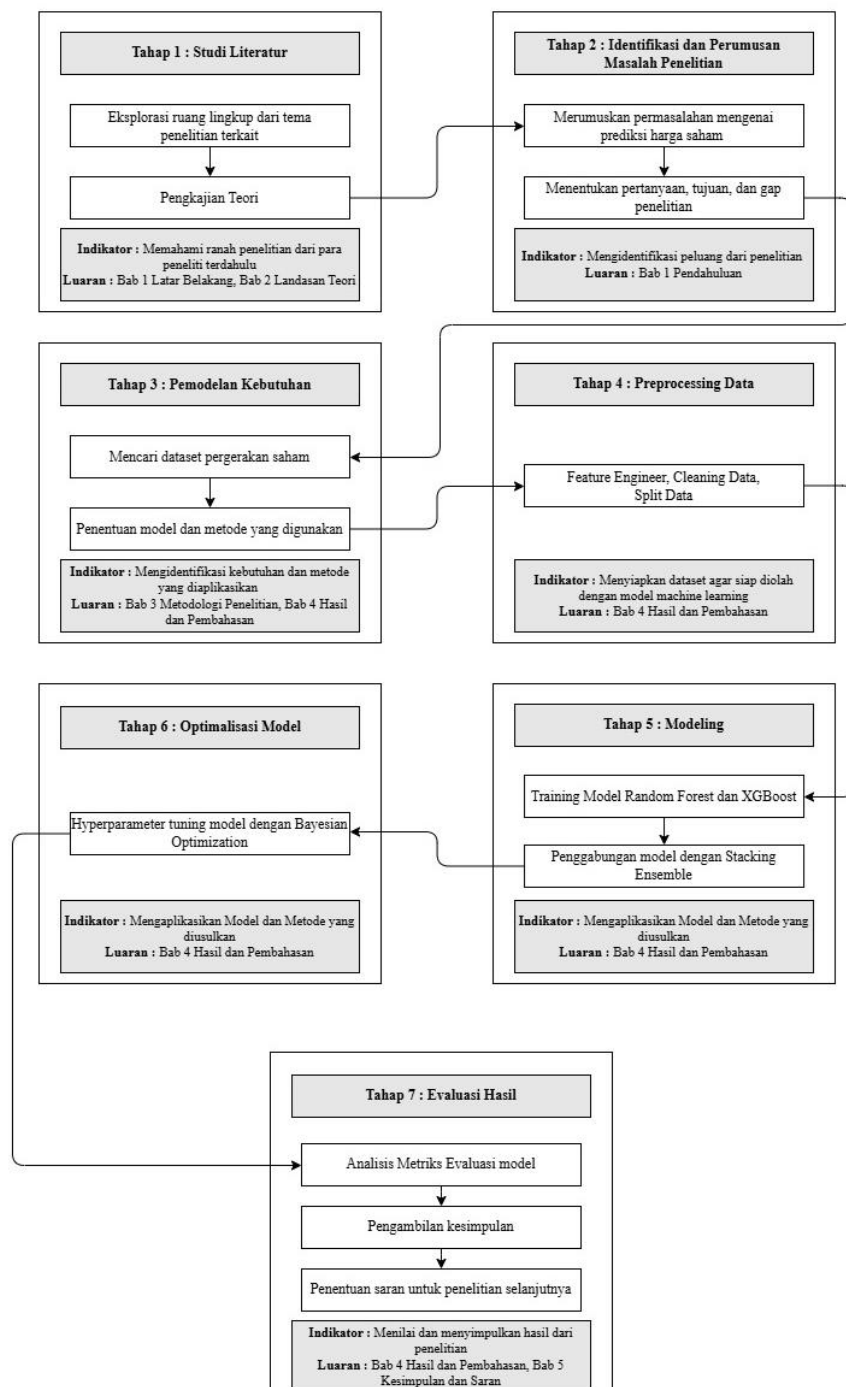
penelitian ini berupaya untuk menguji dan menganalisis kinerja beberapa algoritma *machine learning* yang diterapkan pada dataset pergerakan saham dengan menggunakan data historis saham. Algoritma-algoritma tersebut akan dibandingkan kinerjanya menggunakan metrik evaluasi *MSE*, *RMSE*, *MAE*, dan R^2 , sehingga dapat diketahui metode mana yang lebih efektif dalam menghasilkan prediksi yang lebih akurat.

Pendekatan yang digunakan merupakan pendekatan kuantitatif dengan metode eksperimen berbasis *Machine Learning*. Penelitian ini membangun model prediksi yang memanfaatkan data numerik hasil pergerakan harga saham, yang nantinya hasil prediksi akan dievaluasi menggunakan beberapa metrik, seperti *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, *Mean Absolute Error (MAE)*, dan *R-squared* guna menilai tingkat ketepatan serta reliabilitas algoritma yang diaplikasikan.

3.3 Tahapan Penelitian

Penelitian ini menargetkan kebaruan dalam penerapan pendekatan *Stacking Ensemble* yang menggabungkan dua model *Machine Learning*, yaitu *Random Forest* dan *XGBoost*, dengan eksplorasi terhadap *Meta Learner* serta *Hyperparameter Tuning* menggunakan *Bayesian Optimization* untuk meningkatkan akurasi prediksi harga saham. Dengan kombinasi ini, diharapkan model yang dihasilkan memiliki performa lebih baik dibandingkan metode lainnya. Tahapan

penelitian secara keseluruhan disajikan pada Gambar 3.2 dan Gambar 3.3, yang menampilkan tahapan penelitian serta *Fishbone* diagram dari proses penelitian ini.



Gambar 3. 2 Tahapan Penelitian

3.3.1 Studi Literatur

Tahap ini bertujuan untuk mengeksplorasi ruang lingkup penelitian serta memahami teori-teori yang mendasari prediksi harga saham. Kajian literatur dilakukan dengan meninjau penelitian terdahulu yang relevan guna memperoleh wawasan yang lebih mendalam. Hasil dari tahap ini berupa landasan teoritis yang akan digunakan dalam penelitian dan didokumentasikan dalam Bab I (Latar Belakang) dan Bab II (Landasan Teori).

3.3.2 Identifikasi dan Perumusan Masalah Penelitian

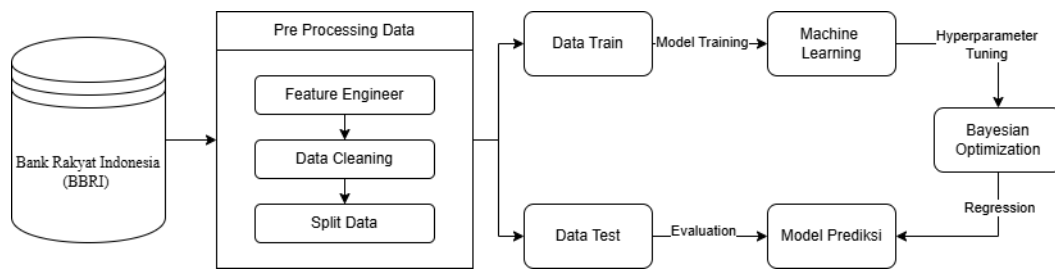
Pada tahap ini, dilakukan perumusan masalah penelitian terkait prediksi harga saham dengan mengidentifikasi gap penelitian dari literatur yang telah dikaji. Selain itu, ditetapkan pertanyaan penelitian (*RQ*), tujuan penelitian (*RO*), serta ruang lingkup penelitian. Hasil dari tahap ini dituangkan dalam Bab I (Pendahuluan).

3.3.3 Pemodelan Kebutuhan

Tahap ini mencakup pencarian *Dataset* terkait pergerakan harga saham serta penentuan model dan metode yang akan digunakan dalam penelitian. Pemodelan kebutuhan dilakukan untuk memastikan bahwa data yang diperoleh sesuai dengan tujuan penelitian. Data yang digunakan dalam penelitian diperoleh dari *Yahoo Finance* dengan bantuan *library yfinance* pada bahasa pemrograman *Python*. Saham yang dijadikan objek penelitian adalah Bank Rakyat Indonesia (BBRI) dengan rentang waktu yang disesuaikan pada penelitian terdahulu. Dataset ini berisi

variabel historis harga saham harian, antara lain *Date*, *Open*, *High*, *Low*, *Close*, *Adj Close*, dan *Volume*.

Arsitektur model yang diusulkan pada penelitian ini dirancang untuk menggambarkan alur keseluruhan proses dalam membangun model prediksi harga penutupan saham berbasis *Machine Learning*, mulai dari tahap pengumpulan data sampai evaluasi hasil *Tuning* pada gambar yang diadaptasi dan disesuaikan kembali dari penelitian (Manjunath dkk., 2024) yang terlihat pada gambar 3.3 berikut.



Gambar 3. 3 Arsitektur Model

Gambar tersebut menunjukkan alur sistematis dalam membangun model prediksi pada harga saham Bank Rakyat Indonesia (BBRI) menggunakan pendekatan *Machine Learning*. Proses dimulai dari pengumpulan data saham BBRI, yang kemudian melalui tahap *Pre Processing Data* yang mencakup *Feature Engineering*, *Data Cleaning*, dan *Split Data*. Setelah data dibersihkan dan diproses, dilakukan pemisahan antara *Data Train* dan *Data Test*. *Data Train* akan digunakan untuk proses *Model Training* dan *Hyperparameter Tuning* dalam tahap *Machine Learning*, sehingga menghasilkan model yang optimal. Model yang telah dilatih kemudian akan dievaluasi menggunakan *Data Test* untuk mengukur performa prediksi. Alur ini mencerminkan pendekatan terstruktur dalam pemodelan prediktif

berbasis data historis saham. Hasil dari tahap ini didokumentasikan dalam Bab III (Metodologi Penelitian) dan Bab IV (Hasil dan Pembahasan).

3.3.4 *Preprocessing Data*

Dataset yang telah dikumpulkan kemudian melalui tahap preprocessing untuk memastikan data dalam kondisi yang siap untuk diolah menggunakan model *Machine Learning*. Proses *Preprocessing* data ini meliputi:

- a. ***Feature Engineering***, penerapan *Feature Engineering* dilakukan dengan mentransformasi data harga mentah menjadi indikator teknikal (*SMA*, *EMA*, *MACD*, *RSI*, dan *Price Change*). Strategi ini didasarkan pada karakteristik data saham yang *noisy*, di mana penggunaan indikator teknikal terbukti lebih efektif dalam menonjolkan sinyal tren dan momentum dibandingkan sekadar menggunakan data harga historis mentah (Nti dkk., 2019). Selain itu, fitur *Date* diekstraksi menjadi *DayOfWeek* untuk menangkap anomali kalender (*calendar effect*) yang sering terjadi di pasar saham.
- b. ***Data cleaning***, yaitu menghapus data yang hilang (*missing values*) yang muncul dari hasil dilakukannya *Feature Engineering*, juga menghapus atribut data mentah seperti *Open*, *High*, *Low*, dan *Adj Close*. Keputusan eliminasi fitur ini diambil untuk menghindari masalah multikolinearitas yang tinggi, mengingat variabel variabel tersebut memiliki korelasi linear yang sangat kuat dengan variabel target (*Close Price*). Memasukkan variabel yang redundan dapat mendistorsi interpretasi *feature importance* dan meningkatkan

kompleksitas komputasi tanpa memberikan signifikansi informasi tambahan bagi model yang diuji (Bustos & Pomares-Quimbaya, 2020).

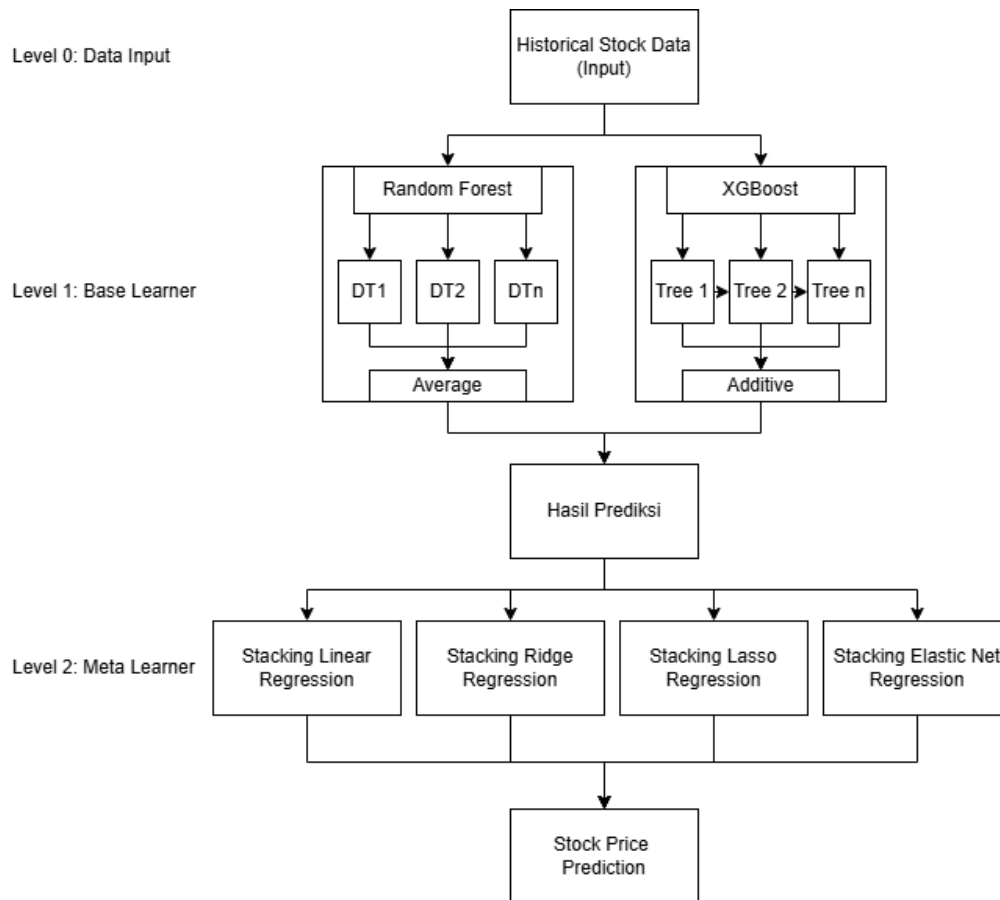
- c. *Split data*, yaitu membagi dataset menjadi data latih (training set) serta data uji (testing set) untuk kebutuhan eksperimen.

Hasil dari tahap ini dijelaskan dalam Bab IV (Hasil dan Pembahasan).

3.3.5 *Modeling*

Setelah data siap, tahap selanjutnya adalah pembangunan model dengan menggunakan algoritma *Random Forest* dan *XGBoost*. Selain itu, dilakukan eksplorasi penggabungan model dengan pendekatan *Stacking Ensemble* untuk meningkatkan performa prediksi, yang mana akan mengaplikasikan *Linear Regression*, *Lasso*, *Ridge*, dan *Elastic Net*. Arsitektur ini terdiri atas tiga lapisan, yakni lapisan input, lapisan *Base Learner*, dan lapisan *Meta Learner*, sebagaimana ditunjukkan pada Gambar 3.4.

Lapisan input akan menerima data historis saham Bank Rakyat Indonesia (BBRI) yang sudah melalui tahap *Feature Engineering* dan *Preprocessing*. Data tersebut digunakan secara bersamaan oleh *Random Forest* dan *XGBoost*, yang berperan sebagai *Base Learner* untuk menghasilkan prediksi awal. Hasil prediksi dari kedua model ini kemudian dikombinasikan oleh lapisan *Meta Learner*, yaitu algoritma *Ridge*, untuk menghasilkan prediksi yang lebih akurat dan stabil.



Gambar 3. 4 Tahapan Pemodelan yang Diusulkan

Tahapan pemodelan yang diusulkan tersebut menggambarkan struktur internal dari pendekatan *Stacking Ensemble* yang digunakan pada penelitian ini. Model terdiri atas tiga lapisan utama, yaitu lapisan input pada Level 0 yang menerima data hasil preprocessing, lapisan *Base Learner* pada Level 1 yang terdiri dari algoritma *Random Forest* dan *XGBoost*, serta lapisan *Meta Learner* pada Level 2 yang menggabungkan hasil prediksi dari kedua *Base Learner* dengan berbagai *Meta Learner* yang diuji untuk menghasilkan output akhir berupa prediksi harga saham. Hasil dari tahap ini berupa implementasi model yang terdokumentasi dalam Bab IV (Hasil dan Pembahasan).

3.3.6 Optimasi Model

Tahapan optimasi model ini dilakukan dengan menggunakan *Hyperparameter Tuning* dengan *Bayesian Optimization*. Optimasi ini bertujuan untuk mendapatkan parameter terbaik yang dapat meningkatkan akurasi model. *Parameter* dari *Random Forest*, *XGBoost*, serta *Meta Learner* terbaik akan dioptimasi, hasil optimasi ini kemudian dibandingkan dengan konfigurasi default sebagai pembanding. Hasil dari tahap ini akan disajikan dalam Bab IV (Hasil dan Pembahasan).

3.3.7 Evaluasi Hasil

Tahap terakhir dalam penelitian ini adalah evaluasi kinerja model yang dilakukan secara komprehensif menggunakan empat metrik standar: *RMSE* (*Root Mean Square Error*) dan *MSE* (*Mean Squared Error*) digunakan untuk mengukur sensitivitas model terhadap *outlier*, di mana kesalahan prediksi yang besar diberikan penalti lebih berat guna mengantisipasi risiko volatilitas harga ekstrem (Hodson, 2022). Penggunaan kedua metrik ini dilengkapi dengan *MAE* (*Mean Absolute Error*) yang memberikan gambaran rata-rata penyimpangan prediksi secara intuitif dalam satuan mata uang asli, yang lebih tahan (*robust*) terhadap data pencilan.

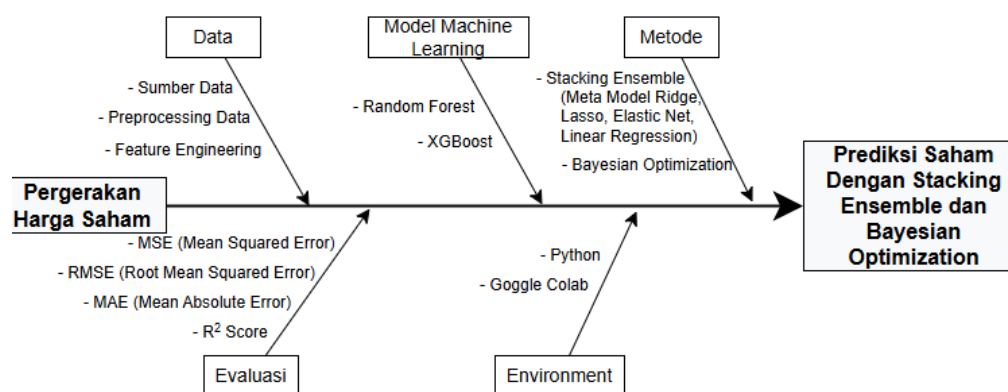
Selanjutnya, untuk mengukur seberapa baik variabilitas data harga saham dapat dijelaskan oleh model, digunakan metrik *R² Score* (*Coefficient of Determination*). Berdasarkan studi komparatif menyatakan bahwa pendekatan *multi-metric* yang menggabungkan dimensi kesalahan (*RMSE/MAE*) dan dimensi

kecocokan (R^2) sangat direkomendasikan dalam pemodelan regresi keuangan untuk memastikan bahwa model tidak hanya meminimalkan bias, tetapi juga mampu menangkap pola tren pasar secara akurat (Chicco dkk., 2021; Yao, 2025).

Selain itu, dilakukan analisis terhadap hasil prediksi serta perumusan kesimpulan dan saran untuk penelitian selanjutnya. Hasil dari tahap ini dituangkan dalam Bab IV (Hasil dan Pembahasan) dan Bab V (Kesimpulan dan Saran).

3.3.8 *Fishbone* Diagram

Diagram *Fishbone* pada Gambar 3.5 dibawah ini mengilustrasikan tujuan utama penelitian serta langkah-langkah yang diperlukan untuk mencapainya, diagram *Fishbone* ini di adaptasi dari penelitian sebelumnya yang dilakukan (Nuno Rodrigues Ferreira dkk., 2021), dengan beberapa perubahan yang menyesuaikan penelitian yang dilakukan.



Gambar 3. 5 *Fishbone* Diagram

Pada bagian ekor dari diagram *fishbone* menampilkan fokus utama penelitian yang menjadi titik awal dari keseluruhan proses, yakni analisis

pergerakan harga saham berdasarkan data historis yang diperoleh dari sumber resmi. Sementara itu, pada bagian kepala diagram direpresentasikan target sekaligus capaian hasil yang ingin dicapai, yaitu untuk membangun model prediksi harga saham dengan memanfaatkan pendekatan *Stacking Ensemble* serta peningkatan performa melalui metode *Bayesian Optimization*.

Struktur utama dari diagram ini menggambarkan secara menyeluruh faktor-faktor penting yang memiliki kontribusi langsung terhadap jalannya penelitian. Faktor-faktor tersebut mencakup aspek data, yang terdiri atas pemilihan dan pengambilan data dari sumber terpercaya, proses *preprocessing data* untuk memastikan kualitas serta konsistensi data, hingga *engineering data* yang dilakukan guna menghasilkan fitur-fitur relevan yang dapat meningkatkan kinerja model.

Selanjutnya, pada tahapan pemodelan digunakan algoritma *Machine Learning* utama, yaitu *Random Forest* dan *XGBoost*, yang kemudian dikombinasikan melalui metode *Stacking Ensemble* dengan eksplorasi terhadap beberapa *Meta Learner*, antara lain *Ridge*, *Lasso*, *Elastic Net*, serta *Linear Regression*, guna mencari *Meta Learner* terbaik dalam menghasilkan prediksi. Setelah itu, dilakukan proses *Hyperparameter Tuning* dengan memanfaatkan pendekatan *Bayesian Optimization* untuk memperoleh parameter terbaik pada model utama.

Kinerja model yang dibangun kemudian dievaluasi menggunakan sejumlah metrik pengukuran, yaitu *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, *Mean Absolute Error (MAE)*, dan *R-squared (R^2)*, sehingga performa

model dapat diukur secara kuantitatif. Seluruh rangkaian eksperimental penelitian ini diimplementasikan dalam lingkungan komputasi *Google Colab* menggunakan bahasa pemrograman *Python*, yang menyediakan fleksibilitas sekaligus mendukung kebutuhan komputasi yang intensif dalam pengolahan data dan pemodelan.