

BAB I. PENDAHULUAN

1.1 Latar Belakang

Saham didefinisikan sebagai objek spesifik dari hak milik dan hak kontraktual dalam modal dasar suatu entitas bisnis, yang memungkinkan kepemilikan, perdagangan, dan perlindungan hukum (Ponomarenko dkk., 2023). Saham ini menjadi unit kepemilikan dalam sebuah perusahaan, yang diwakili oleh saham yang diterbitkan oleh perusahaan swasta untuk mengumpulkan modal. Saham memungkinkan investor untuk berpartisipasi dalam pertumbuhan dan keuntungan perusahaan, sementara juga mengekspos mereka terhadap volatilitas pasar dan perubahan valuasi (Patterson & Kuthy, 2024).

Dalam dunia saham, memprediksi harga saham menjadi aspek yang krusial sebagai faktor keberhasilan bagi para investor yang akan melakukan investasi saham, yang mana prediksi harga saham ini harus dilakukan dengan tepat agar tidak terjadi kerugian. Penelitian oleh Putratama, dkk. dalam penelitiannya pada sektor saham menjelaskan bahwa pergerakan harga saham merupakan fenomena yang kompleks dan dipengaruhi oleh banyak faktor, termasuk kondisi pasar, maupun sentimen investor (Putratama dkk., 2023). Dengan adanya pergerakan saham yang fluktuatif meskipun dalam rentan waktu yang cukup singkat, ini membuat prediksi pergerakan saham menjadi tantangan yang menarik dalam dunia investasi.

Maka dari itu prediksi pada metode *Machine Learning* menjadi alat yang efektif dalam mengatasi masalah ini. Saat ini teknik prediksi dalam analisis data sudah banyak dipakai dalam berbagai data pada sektor saham sehingga pemanfaatan dapat digunakan untuk mendukung pengambilan keputusan yang baik

berdasarkan data yang ada. *Machine Learning* sendiri merupakan metode yang dapat mendeteksi pola dari data secara otomatis, yang mana nantinya pola tersebut dapat dipakai untuk *forecasting* atau memprediksi data di masa depan (Ramadani, 2023). Algoritma *Machine Learning* ini sudah digunakan untuk meramalkan dan analisis harga saham dari 1980-an (Vachhani dkk., 2020).

Terdapat beberapa metode yang dapat dilakukan untuk mendeteksi pola dalam data dengan menggunakan algoritma *Machine Learning*, yaitu seperti *Random Forest*, *XGBoost*, dan banyak metode lainnya yang dapat diaplikasikan. *Random Forest* ini unggul dalam menangani data berdimensi tinggi dan efektif dalam mengurangi *Overfitting*, menjadikannya pilihan yang andal untuk metode *Ensemble* (Zhang dkk., 2024), tetapi dapat mengalami *Overfitting* terutama dengan data yang *noise* atau ketika terlalu banyak *tree* yang digunakan, sehingga akan menangkap *noise* daripada pola data (Fatima dkk., 2023). Sedangkan *XGBoost* meskipun kuat dalam mengoptimalkan kinerja model dengan teknik *Boosting*, model algoritma ini mempunyai tantangan dalam interpretabilitasnya karena mempunyai model yang kompleks. *XGBoost* adalah metode *Ensemble* yang canggih yang menggabungkan banyak pohon keputusan, membuatnya kurang dapat diinterpretasikan dibandingkan dengan pohon tunggal (Wieland dkk., 2020).

Untuk mengatasi kelemahan masing-masing model tunggal tersebut, diperlukan pendekatan *Stacking Ensemble*. Berbeda dengan metode *ensemble* sederhana seperti *voting* atau rata-rata (*averaging*) yang memperlakukan setiap model dengan bobot setara, *Stacking* bekerja dengan melatih *Meta Learner* untuk mempelajari cara terbaik menggabungkan prediksi berdasarkan pola kesalahan dari

Base Learner (Ganaie dkk., 2022). Mekanisme pembelajaran bertingkat ini memungkinkan sistem untuk mengambil keunggulan spesifik dari masing-masing algoritma, sehingga menghasilkan prediksi yang lebih presisi dan memiliki kemampuan generalisasi yang lebih tangguh (*robust*) terhadap perubahan perilaku pasar (Huang dkk., 2022).

Berbeda dengan pendekatan *Ensemble* pada *Bagging* dan *Boosting* cenderung hanya menggabungkan beberapa model pada satu lapisan saja, metode pada *Ensemble Stacking* ini menambahkan lapisan pembelajaran lainnya dari hasil penggabungan model. Dengan demikian, hasil dan temuan yang diberi bobot ditampilkan. Pendekatan ini sangat berguna dalam mengalokasikan bobot secara cerdas pada model-model dasar (N. Sharma dkk., 2021).

Şimsek A. pada penelitiannya pada bidang saham pada *Dataset BIST100*, digunakan metode *Stacking* dengan memakai algoritma *Random Forest* dan *XGBoost*, yang menghasilkan capaian R^2 sebesar 98% dalam memprediksi harga saham (Şimşek, 2025). Meskipun berdasarkan penelitiannya mengenai metode *Stacking* dengan *Random Forest* dan *XGBoost* mendapatkan hasil yang baik, terdapat potensi pengembangan pada metode *Hyperparameter Tuning* dan *Meta Learner* yang dipakai, penelitian tersebut masih menggunakan *Grid Search* dan *Random Search* untuk optimasi *hyperparameter*. Penggunaan metode konvensional ini memiliki kelemahan signifikan dari sisi efisiensi komputasi.

Hal ini dikuatkan oleh studi empiris terdahulu yang membandingkan performa *Grid Search* dan *Bayesian Optimization* pada model *Random Forest*.

Penelitian tersebut menemukan bahwa meskipun kedua metode menghasilkan akurasi metrik evaluasi yang setara, Bayesian Optimization mampu memangkas waktu eksekusi hingga 42.19% lebih cepat (selisih sekitar 3.798 detik) dibandingkan *Grid Search* pada parameter yang sama (Darmawan dkk., 2025).

Pentingnya penggunaan teknik optimasi dalam prediksi saham juga telah dibuktikan pada penelitian sebelumnya yang berhasil menerapkan *Particle Swarm Optimization* (PSO) dan *Bagging* pada metode SVR untuk memprediksi saham sektor teknologi di era pandemi COVID-19 (Sulastri dkk., 2023). Mengacu pada keberhasilan penelitian tersebut serta penelitian pada Şimsek A, penelitian ini mengembangkan model prediksi menggunakan *Bayesian Optimization* pada arsitektur *Stacking Ensemble* dengan eksplorasi terhadap *Meta Learner* untuk menangani dinamika pasar pasca-pandemi hingga tahun 2025.

Penelitian ini akan mengarah pada bagaimana perbandingan dan eksplorasi algoritma dalam *Meta Learner* untuk *Stacking* dengan pendekatan *Machine Learning Random Forest* dan *XGBoost* serta peningkatan hasil prediksi model memakai *Hyperparameter Tuning Bayesian Optimization* dalam memprediksi pergerakan saham. Dengan *output* hasil yang direncanakan berupa model yang dapat memprediksi angka harga awal saham, sehingga akan memberikan wawasan tentang penggunaan algoritma *Random Forest* dan *XGBoost* dalam metode *Stacking Ensemble Learning* dengan memakai *Bayesian Optimization* untuk prediksi harga saham.

1.2 Rumusan Masalah

Masalah utama yang diangkat dalam penelitian ini adalah bagaimana mengembangkan model yang dapat memprediksi harga saham dengan memanfaatkan *Machine Learning* dan mencapai tingkat akurasi serta kinerja yang tinggi, mengingat bahwa fluktuasi harga saham merupakan tantangan yang signifikan dalam sektor keuangan..

Berdasarkan dengan hal tersebut, penelitian ini akan menjawab pertanyaan berikut:

1. Apakah *Stacking Ensemble* dapat menghasilkan peningkatan performa dibandingkan pada model tunggal *Random Forest* dan *XGBoost*?
2. Sejauh mana *Bayesian Optimization* dapat meningkatkan akurasi prediksi harga saham pada model *Random Forest*, *XGBoost*, dan *Stacking Ensemble* dibandingkan dengan tanpa dilakukan optimasi *Hyperparameter Tuning*?
3. Bagaimana perbandingan *Meta Learner Linear Regression*, *Lasso*, *Ridge*, dan *Elastic Net* dalam pengaplikasiannya pada *Stacking*?

1.3 Tujuan Penelitian

Penelitian ini bertujuan mengembangkan model prediksi harga saham berbasis *Stacking Ensemble* dengan *Random Forest* dan *XGBoost*, menerapkan *Bayesian Optimization* untuk mencari konfigurasi *parameter* terbaik guna meningkatkan akurasi prediksi, dan membandingkan performa pada setiap

eksperimen yang dilakukan pada ketiga algoritma yang akan diuji untuk *Meta Learner*.

Berdasarkan dengan penjelasan tersebut, tujuan penelitian ini yaitu:

1. Melakukan evaluasi peningkatan performa *Stacking Ensemble* yang telah dioptimasi dibandingkan dengan model tunggal yang juga telah dioptimasi.
2. Penerapan *Bayesian Optimization* dalam *Hyperparameter Tuning* untuk meningkatkan performa model *Random Forest*, *XGBoost*, dan *Stacking Ensemble* dalam prediksi harga saham.
3. Menganalisis perbandingan antara *Meta Learner Linear Regression*, *Lasso*, *Ridge*, dan *Elastic Net* dalam pengaplikasiannya dalam *Stacking*.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat secara teoritis dan praktis. Secara teoritis, penelitian ini berkontribusi dalam memperkaya kajian di bidang *Machine Learning*, khususnya dalam analisis performa model prediksi harga saham menggunakan pendekatan *Stacking Ensemble* yang dikombinasikan dengan *Bayesian Optimization*. Hasil penelitian ini dapat menjadi referensi metodologis dalam memahami bagaimana pemilihan algoritma dan *Hyperparameter Tuning* memengaruhi performa model pada data saham.

Secara praktis, penelitian ini tidak menghasilkan model siap implementasi, tetapi dapat memberikan wawasan awal bagi analis pasar, praktisi keuangan, atau

peneliti yang ingin memilih metode prediksi harga saham dengan performa lebih baik berdasarkan perbandingan empiris. Hasil penelitian ini dapat dipakai sebagai dasar atau rujukan untuk pengembangan sistem prediksi pada penelitian atau implementasi selanjutnya.

1.5 Batasan Masalah

Penelitian ini memiliki beberapa batasan agar memastikan fokus dan ruang lingkup yang jelas, yaitu sebagai berikut:

1. Penelitian hanya menggunakan data historis pergerakan harga saham yang bersumber dari *Yahoo Finance* melalui *library yfinance*, dengan rentang waktu tertentu. Data dari sumber lain atau dengan rentang waktu berbeda tidak termasuk dalam analisis.
2. Model prediksi yang digunakan terbatas pada algoritma *Random Forest*, *XGBoost*, serta *Stacking Ensemble Learning* yang menggabungkan keduanya. Algoritma lain di luar itu tidak digunakan.
3. *Hyperparameter Tuning* hanya dilakukan menggunakan *Bayesian Optimization*. Teknik optimasi lain seperti Random Search tidak digunakan.
4. Eksplorasi *Meta Learner* hanya dibatasi pada *Linear Regression*, *Lasso*, *Ridge*, dan *Elastic Net*.
5. Evaluasi performa model dilakukan menggunakan metrik evaluasi *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, *Mean Absolute Error (MAE)*, dan *R-squared*. Metrik evaluasi lain tidak menjadi fokus penelitian.

6. Faktor eksternal yang bersifat non-kuantitatif, seperti sentimen pasar, isu geopolitik, maupun faktor makroekonomi, tidak menjadi bagian dari penelitian.
7. Pendekatan *Stacking Ensemble* yang digunakan untuk *Base Learner* hanya mencakup kombinasi dua algoritma, yaitu *Random Forest* dan *XGBoost*, serta eksplorasi terhadap *Linear Regression*, *Lasso*, *Ridge*, dan *Elastic Net* untuk *Meta Learner* tanpa melibatkan variasi lain.
8. Penelitian ini hanya berfokus pada perbandingan algoritma tanpa mengembangkan model menjadi sistem prediksi yang siap digunakan secara praktis.

1.6 Sistematika Penulisan

Sistematika penulisan pada penelitian ini disusun agar dapat memberikan gambaran yang jelas bagaimana alur penelitian. Adapun sistematika penulisan yang digunakan yaitu :

BAB I Pendahuluan : yang berisi latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan skripsi.

BAB II Tinjauan Pustaka : yang menyajikan teori-teori yang relevan dengan penelitian, landasan konseptual yang menjadi dasar dalam pengembangan penelitian ini, dan hasil-hasil penelitian terdahulu.

BAB III Metodologi Penelitian : menjelaskan metode penelitian yang digunakan, meliputi roadmap penelitian, deskripsi metodologi, tahapan penelitian

yang dilakukan, serta diagram alir dan fishbone diagram yang mendukung pemahaman alur penelitian.

BAB IV Hasil dan Pembahasan : berisi penyajian hasil dan eksperimen penelitian yang telah dilakukan, dimulai dari pengolahan *dataset*, hingga evaluasi kinerja model, serta pembahasan terhadap hasil yang diperoleh dengan mengaitkannya pada teori maupun penelitian terdahulu.

BAB V Kesimpulan dan Saran : berisi rangkuman kontribusi penelitian, menjawab rumusan masalah dengan sinkronisasi pada tujuan penelitian, juga rangkuman mengenai kelebihan dan kekurangan dari penelitian yang telah dilakukan serta saran untuk penelitian selanjutnya.