

BAB I

PENDAHULUAN

1.1 Latar Belakang

Stock market atau pasar saham telah lama menjadi salah satu instrumen investasi paling diminati oleh investor di seluruh dunia, termasuk di Indonesia. Saham memiliki potensi pengembalian yang tinggi, sehingga menarik minat dari individu hingga institusi besar (Hidayah & Alwi, 2022). Saham sebagai instrumen keuangan, memberikan kesempatan bagi pemegangnya untuk memiliki sebagian dari perusahaan. Di Indonesia, saham-saham yang tergabung dalam Indeks LQ45 yang terdiri dari 45 perusahaan dengan kapitalisasi pasar terbesar dan likuiditas tertinggi menjadi sorotan utama (Safitri *et al.*, 2021). Salah satu perusahaan yang tergabung dalam indeks ini adalah PT Astra International Tbk, merupakan konglomerat terbesar dan paling berpengaruh di Indonesia (Hidayah & Alwi, 2022).

Membeli saham pada dasarnya membeli sebagian kecil dari kepemilikan perusahaan tersebut, yang berarti berhak atas keuntungan yang dihasilkan perusahaan dalam bentuk dividen atau *capital gain*. Investasi saham memang memiliki potensi keuntungan yang besar, tetapi mengandung risiko. Salah satu tantangan terbesar yang dihadapi oleh para investor adalah volatilitas yang melekat pada pergerakan harga saham. Volatilitas ini ditandai oleh perubahan harga yang cepat dan tidak terduga, sering kali menciptakan risiko signifikan (Brealey *et al.*, 2020). PT Astra International Tbk, sebagai perusahaan multinasional yang bergerak di berbagai sektor bisnis, mulai dari otomotif, jasa keuangan, alat berat, agribisnis, hingga teknologi informasi, dapat mencerminkan dinamika ekonomi Indonesia secara umum.

Pergerakan harga saham Astra sering kali dipengaruhi oleh faktor eksternal seperti suku bunga, kebijakan pemerintah, kondisi perekonomian global, dan fluktuasi harga komoditas (Hidayah & Alwi, 2022). Peristiwa global tak terduga, seperti pandemi COVID-19, turut memberikan tekanan besar pada harga saham dan menciptakan ketidakpastian bagi investor. Ketidakpastian yang terjadi membuat kemampuan untuk memprediksi pergerakan harga saham dengan lebih akurat menjadi krusial, baik untuk mengurangi risiko maupun memaksimalkan keuntungan bagi investor (Dash & Maitra, 2022).

Analisis fundamental dan analisis teknikal adalah dua metode utama yang biasanya digunakan oleh investor untuk analisis prediksi harga saham (Hariazy, 2021). Analisis fundamental menilai kinerja perusahaan berdasarkan faktor seperti pendapatan, laba, dan struktur biaya, dengan tujuan untuk menentukan nilai intrinsik suatu saham serta mengidentifikasi apakah saham tersebut tergolong *undervalued* atau *overvalued* (Brealey *et al.*, 2020). Analisis ini sering kali tidak mampu menangkap perubahan cepat dalam sentimen pasar (Ashlihah *et al.*, 2022). Sebaliknya, analisis teknikal memanfaatkan pola pergerakan harga historis dan volume perdagangan untuk memproyeksikan masa depan (Hariazy, 2021). Penelitian sebelumnya menunjukkan bahwa analisis teknikal dapat menghasilkan portofolio investasi yang mengungguli strategi *buy and hold*, dengan hasil abnormal *return* yang tinggi (Han *et al.*, 2013). Penerapan analisis teknikal menggunakan berbagai indikator seperti, *Simple Moving Average* (SMA), *Relative Strength Index* (RSI), *Moving Average Convergence Divergence* (MACD), dan lain sebagainya. SMA merupakan metode yang memperhalus data harga saham dengan menghitung rata-rata harga selama periode tertentu, sehingga dapat menekan fluktuasi harga dan

mengidentifikasi tren dengan lebih jelas. RSI digunakan untuk mengidentifikasi tingkat kejenuhan pasar dengan menentukan batas atas dan bawah yang konstan, sehingga membantu investor menghindari pembelian saham pada harga yang terlalu tinggi atau penjualan pada harga yang terlalu rendah (Irahadi *et al.*, 2022).

Pesatnya perkembangan teknologi dan meluasnya ketersediaan data, metode tersebut mulai digantikan atau setidaknya dilengkapi dengan teknologi yang lebih canggih, seperti *machine learning*. Selama beberapa tahun terakhir, *machine learning* telah menjadi salah satu pendekatan utama dalam analisis pasar keuangan, termasuk prediksi harga saham, karena kemampuannya yang luar biasa dalam mengolah data dalam jumlah besar dan kompleks (Triya *et al.*, 2024). Metode konvensional sering kali hanya berfokus pada pola linear atau menggunakan indikator tertentu, tetapi *machine learning* memiliki kemampuan untuk mengidentifikasi pola-pola tersembunyi dalam data. Hal ini sangat penting dalam pasar saham yang dinamis, di mana fluktuasi harga dipengaruhi oleh banyak faktor yang saling berinteraksi dengan cara yang tidak selalu dapat diprediksi secara sederhana (Maknickienė *et al.*, 2020).

Metode pada *machine learning* dibagi menjadi 4 tipe berdasarkan cara pembelajarannya (Sarker, 2021). Pertama, *supervised learning* merupakan metode yang memiliki fungsi dalam memetakan data masukan dan keluaran berdasarkan contoh dataset. Dataset *input* terdiri dari data *training* dan data *testing*, di mana data *testing* memiliki variabel *output* yang perlu diprediksi atau diklasifikasi. Kedua, *unsupervised learning* merupakan metode yang digunakan untuk melakukan pengelompokan data tanpa pengawasan, artinya dataset yang dikelola tidak mempunyai label tertentu. *Semi-supervised learning* merupakan kombinasi metode

supervised learning dan *unsupervised learning*, artinya data yang memiliki label dan tanpa label akan digunakan untuk melatih algoritma agar dapat melakukan proses pembelajaran sendiri. Terakhir, *reinforcement learning* merupakan metode yang berhubungan dengan pengambilan keputusan secara berurutan untuk memperoleh hasil terbaik. Pada prosesnya metode ini akan mempelajari struktur data yang dapat memodifikasi atau memperoleh perilaku dan keterampilan baru secara bertahap untuk mendapatkan *reward* berdasarkan *trial and error*.

Algoritma *machine learning* dapat memodelkan hubungan yang rumit antara berbagai variabel, baik dalam bentuk linear maupun non-linear, sehingga memungkinkan prediksi yang lebih akurat dan adaptif terhadap perubahan pasar yang cepat dan terkadang tidak menentu (Jiang *et al.*, 2020). Beberapa metode *machine learning* yang umum digunakan meliputi regresi (*Linear Regression*, *Support Vector Regression*), pohon keputusan (*Decision Tree*, *Random Forest*), *boosting* (XGBoost, LightGBM), jaringan saraf tiruan (*Artificial Neural Networks*, LSTM), serta metode *clustering* dan reduksi dimensi (K-Means, PCA) (Sarker, 2021). Algoritma *machine learning* yang sangat populer dan efektif untuk prediksi harga saham adalah *Support Vector Regression* (SVR) dan *eXtreme Gradient Boosting* (XGBoost) (Dash *et al.*, 2021; Jadhav *et al.*, 2021). SVR adalah algoritma *machine learning* yang berasal dari *Support Vector Machine* (SVM), yang biasanya digunakan untuk klasifikasi. Dalam hal prediksi, SVR digunakan untuk menangani masalah regresi. SVR efektif dalam menangani data dengan berbagai karakteristik, termasuk data linear maupun non-linear (Yulianto *et al.*, 2020). SVR mampu menangani regresi dengan kontrol terhadap error (ϵ – *insensitive loss function*), yang membuatnya lebih fleksibel dibandingkan regresi linear biasa. Tujuannya

adalah memaksimalkan margin antara data aktual dan data prediksi dalam ruang fitur. Teknik ini menggunakan prinsip kernel *trick*, yang memungkinkan SVR untuk memetakan data non-linear ke dalam ruang fitur berdimensi lebih tinggi agar pola kompleks dapat diidentifikasi dengan lebih baik (Isnaeni *et al.*, 2022). Namun, SVR juga dapat bekerja secara optimal pada data yang menunjukkan hubungan linear tanpa perlu transformasi ke ruang fitur yang lebih tinggi, seperti dalam penggunaan kernel linear (Bagaskara *et al.*, 2024).

Penelitian sebelumnya dilakukan oleh Huang pada tahun 2022 yang membahas penerapan SVR dalam memprediksi deret waktu keuangan yang bersifat tidak stasioner dan memiliki tingkat volatilitas tinggi. Penelitian ini bertujuan untuk meningkatkan akurasi prediksi harga saham dengan melakukan optimasi parameter SVR guna menangani kompleksitas data keuangan secara lebih efektif. SVR dipilih sebagai model utama karena kemampuannya dalam mengakomodasi hubungan non-linear serta menghasilkan prediksi yang lebih stabil dibandingkan dengan metode ARIMA. Performa SVR sangat bergantung pada pemilihan parameter yang optimal, termasuk C , gamma, dan epsilon (*margin error*). Penelitian tersebut mengadopsi *Randomized Search Cross Validation* sebagai metode optimasi untuk menyesuaikan parameter SVR secara otomatis.

XGBoost *Regressor* bekerja dengan membangun sekumpulan pohon keputusan regresi secara sekuensial untuk meminimalkan kesalahan dengan metode *gradient boosting*. XGBoost menggunakan fungsi kerugian seperti *Mean Square Error* (MSE) atau *Mean Absolute Error* (MAE) yang membantu dalam mengoptimalkan prediksi variabel kontinu seperti harga saham. Algoritma ini juga dilengkapi dengan regularisasi L1 (*Lasso*) dan L2 (*Ridge*) yang membantu

mengurangi risiko *overfitting*, sebuah masalah umum dalam prediksi harga saham akibat *overtraining* model pada data historis yang bervariasi (Chen & Guestrin, 2016). Keunggulan lain dari XGBoost adalah kemampuannya menangani dataset besar dan kompleks secara efisien, dengan teknik *shrinkage*, *sub-sampling*, dan *parallel processing* yang membuatnya sangat cepat dalam memproses data skala besar (Jiang *et al.*, 2020).

Penelitian sebelumnya dilakukan oleh Hafid *et al.* (2024) yang membahas penerapan XGBoost *Regressor* dalam memprediksi harga *cryptocurrency* dengan memanfaatkan berbagai indikator teknikal. Harga *cryptocurrency* memiliki volatilitas tinggi dan pola yang kompleks. Oleh karena itu, penelitian ini mengusulkan penggunaan XGBoost *Regressor* sebagai model prediksi harga Bitcoin, dengan memanfaatkan indikator teknikal seperti *Simple Moving Average* (SMA), *Relative Strength Index* (RSI), dan *Moving Average Convergence Divergence* (MACD) untuk menangkap perubahan harga secara lebih akurat. Hasil penelitian menunjukkan bahwa XGBoost *Regressor* mampu mengidentifikasi pola dalam data historis *cryptocurrency* dengan lebih baik dibandingkan metode lain. Teknik *boosting* dan regularisasi membuat model ini dapat meningkatkan akurasi prediksi serta mengurangi risiko *overfitting*.

Voting Regressor Ensemble digunakan sebagai metode yang efektif untuk meningkatkan akurasi prediksi dengan menggabungkan hasil dari beberapa model regresi. *Ensemble learning* merupakan teknik dalam *machine learning* yang menggabungkan beberapa model dengan tujuan untuk meningkatkan akurasi prediksi dibandingkan dengan menggunakan satu model saja (Arrohman & Arifudin, 2023). Konsep dasar dari *ensemble learning* adalah bahwa kombinasi dari

beberapa model, yang masing-masing memiliki perspektif atau pendekatan yang berbeda terhadap data, dapat menghasilkan prediksi yang lebih akurat dan lebih robust terhadap variabilitas data. Teknik ini berfungsi untuk mengatasi kelemahan yang ada pada setiap model individual, sehingga meningkatkan kinerja prediksi secara keseluruhan (Khan *et al.*, 2024).

Voting Regressor Ensemble dapat meningkatkan robustness dan akurasi hasil prediksi melalui dua pendekatan utama, yaitu *Average Voting* (AV) dan *Weighted Voting* (WV) (Erdebilli & İçtenbaş, 2022). Pendekatan AV, prediksi akhir dihitung sebagai rata-rata sederhana dari prediksi semua model yang tergabung dalam *Voting Regressor*. Sebaliknya, pada pendekatan WV setiap model diberikan bobot tertentu berdasarkan kinerja atau keandalannya. Prediksi akhir adalah rata-rata tertimbang dari prediksi masing-masing model. Model yang memiliki bobot lebih tinggi akan memberikan pengaruh lebih besar pada prediksi akhir. Oleh karena itu, metode *ensemble* yang digunakan dalam penelitian ini adalah WV. WV, memberikan keunggulan tambahan karena dapat memaksimalkan kontribusi model-model yang memiliki kinerja lebih baik, sekaligus mengurangi pengaruh model yang kurang akurat. Pendekatan WV, setiap model diberi bobot yang proporsional dengan tingkat keandalannya. Prediksi akhir dihitung sebagai rata-rata tertimbang dari prediksi yang dihasilkan oleh setiap model, dengan model yang memiliki bobot lebih tinggi memberikan kontribusi lebih besar terhadap hasil akhir. Pendekatan WV menghasilkan prediksi yang lebih akurat dan stabil, karena metode ini mengoptimalkan kontribusi model-model yang lebih andal, sehingga dapat mengurangi fluktuasi yang mungkin timbul dari model-model yang kurang akurat (Pham & Vu, 2021).

Voting Regressor Ensemble memiliki kelebihan seperti peningkatan akurasi, robustness, dan kemampuan generalisasi. Metode ini mampu mengurangi kesalahan individu yang mungkin terjadi dalam model tunggal, sehingga menghasilkan model akhir yang lebih akurat dan stabil (Erdebilli & İçtenbaş, 2022). Penerapan *Voting Regressor Ensemble* dengan menggabungkan SVR dan XGBoost diharapkan dapat menghasilkan prediksi harga saham PT Astra International Tbk, yang lebih akurat dan memberikan informasi yang lebih berharga bagi para investor dalam pengambilan keputusan investasi yang tepat.

Salah satu penelitian yang relevan dilakukan oleh Arrohman dan Arifudin (2023), yang mengkaji prediksi pengembalian saham menggunakan metode pembelajaran *ensemble*, khususnya algoritma *Voting Regressor* yang dikombinasikan dengan berbagai algoritma *machine learning*. Penelitian ini menggabungkan beberapa algoritma, termasuk ABR, GBR, RFR, dan SVR, dengan *Voting Regressor* sebagai metode *ensemble*. Penelitian ini menggunakan data saham *Microsoft* tahun 2012 sampai 2022, dengan melakukan *pre-processing* data dan transformasi. Pengembalian saham mingguan dihitung berdasarkan periode perdagangan 5 hari. Evaluasi model dilakukan menggunakan metrik MSE, RMSE, dan MAE. Hasil penelitian menunjukkan bahwa RFR memperoleh kinerja terbaik untuk MSE sebesar 0,00109 dan MAE sebesar 0,02378, sementara *Voting Regressor* mencapai skor RMSE terbaik 0,0325. Pendekatan *ensemble* yang diterapkan berhasil mengurangi risiko *overfitting* dan meningkatkan akurasi prediksi. Penelitian ini menegaskan bahwa kombinasi beberapa algoritma melalui pembelajaran *ensemble* dapat menghasilkan prediksi saham yang lebih baik

dibandingkan dengan penggunaan algoritma tunggal, meskipun terdapat variasi performa antar metode berdasarkan metrik evaluasi yang digunakan.

Penelitian yang dilakukan oleh Erdebilli dan İçtenbaş (2022), membahas penerapan *Voting Regression Ensemble* berbasis *machine learning* untuk memprediksi jumlah limbah medis di Istanbul, Turki. Penelitian ini bertujuan untuk meningkatkan akurasi prediksi dalam manajemen limbah medis, yang menjadi tantangan penting bagi kota besar. Menggunakan pendekatan *ensemble learning*, penelitian ini menggabungkan beberapa model regresi. Metode yang digunakan dalam penelitian ini melibatkan tiga model dasar yaitu, *Random Forest* (RF), *Gradient Boosting Machine* (GBM), dan *Adaptive Boosting* (AdaBoost). Model-model ini kemudian dikombinasikan menggunakan *Voting Regression* dengan pendekatan *weighted voting*, di mana bobot diberikan berdasarkan performa masing-masing model dalam memprediksi data. Evaluasi dilakukan menggunakan beberapa metrik, termasuk MAE, RMSE, dan R^2 untuk mengukur tingkat akurasi prediksi yang dihasilkan oleh setiap model.

Berdasarkan fenomena dan pendekatan yang telah dijelaskan, maka penelitian ini dilakukan untuk mengkaji efektivitas metode *Support Vector Regression*, *eXtreme Gradient Boosting*, dan kombinasi keduanya melalui pendekatan *Voting Regressor Ensemble* dalam memprediksi harga saham PT Astra International Tbk. Fokus penelitian diarahkan pada peningkatan akurasi prediksi dengan memanfaatkan indikator teknikal RSI dan SMA sebagai variabel *input*. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pengambilan keputusan investasi yang lebih tepat serta menjadi alternatif metode prediksi harga saham yang andal dan aplikatif.

1.2 Rumusan Masalah

Rumusan masalah yang muncul pada penelitian ini berdasarkan latar belakang yang telah dipaparkan adalah sebagai berikut:

1. Bagaimana akurasi dan hasil prediksi harga saham menggunakan *Support Vector Regression* (SVR)?
2. Bagaimana akurasi dan hasil prediksi harga saham menggunakan *eXtreme Gradient Boosting Regressor* (XGBoost)?
3. Bagaimana akurasi dan hasil prediksi harga saham dengan menggunakan kombinasi model SVR dan XGBoost dalam metode *Voting Regressor Ensemble*?
4. Bagaimana hasil prediksi harga saham untuk periode mendatang berdasarkan model terbaik?

1.3 Batasan Masalah

Batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan data sekunder berupa harga penutupan harian (*daily close*) saham PT Astra International Tbk selama periode Agustus 2020 hingga Agustus 2024, yang diperoleh melalui situs www.finance.yahoo.com.
2. Pemilihan model dasar (*base-learners*) dibatasi pada 2 model *machine learning*, yaitu *Support Vector Regression* (SVR) dan *eXtreme Gradient Boosting Regressor* (XGBoost).

3. Penelitian ini berfokus pada penerapan teknik *Voting Regressor Ensemble* dengan pendekatan *Weighted Voting* untuk meningkatkan akurasi prediksi harga saham.
4. Variabel *input* dalam penelitian ini dibatasi pada 2 indikator teknikal, yaitu *Relative Strength Index* (RSI) dan *Simple Moving Average* (SMA), guna menjaga kesederhanaan model serta mengurangi risiko *overfitting*.
5. Evaluasi performa model dalam penelitian ini akan menggunakan metrik *Root Mean Square Error* (RMSE), *Mean Absolute Percentage Error* (MAPE), dan R^2 untuk mengukur tingkat akurasi prediksi.

1.4 Tujuan Penelitian

Adapun tujuan penelitian ini adalah sebagai berikut:

1. Mengetahui hasil prediksi harga saham dengan menggunakan metode *Support Vector Regression* (SVR) serta mengevaluasi tingkat akurasinya.
2. Mengetahui hasil prediksi harga saham dengan menggunakan metode *eXtreme Gradient Boosting Regressor* (XGBoost) serta mengevaluasi tingkat akurasinya.
3. Mengevaluasi hasil prediksi dan tingkat akurasi harga saham ketika model SVR dan XGBoost dikombinasikan dalam metode *Voting Regressor Ensemble*.
4. Melakukan prediksi harga saham untuk periode mendatang berdasarkan model terbaik yang diperoleh dari evaluasi SVR, XGBoost, dan *Voting Regressor Ensemble*.