

BAB I PENDAHULUAN

1.1 Latar Belakang

Dalam beberapa tahun belakangan ini, minat masyarakat Indonesia terhadap dunia investasi telah meningkat drastis. Berdasarkan laporan JakPat Insight Indonesia Investment Trends-2024, jumlah investor di pasar modal negara ini kini mencapai sekitar 12,16 juta orang. Survei yang melibatkan 2.088 responden di seluruh Indonesia menunjukkan bahwa tiga dari empat orang sudah terlibat aktif dalam investasi. Hal ini menandakan bahwa investasi tidak lagi sekadar sarana menyimpan uang, melainkan telah menjadi elemen kunci dalam pengelolaan keuangan pribadi. Akibatnya, kemampuan untuk memperkirakan fluktuasi *return* saham dengan tepat dan menyusun portofolio yang optimal semakin penting bagi para investor yang berusaha mencapai hasil maksimal di tengah dinamika pasar modal Indonesia.

Di bidang teori keuangan kontemporer, *Modern Portfolio Theory* (MPT) yang dikembangkan oleh Harry Markowitz pada 1952 menjadi fondasi utama untuk analisis pembentukan portofolio yang efisien. Teori ini menyoroti keterkaitan antara risiko dan imbal hasil melalui konsep *efficient frontier*, yang memungkinkan investor memilih kombinasi aset dengan *return* maksimal untuk tingkat risiko tertentu (Adnyana, 2020). Meski demikian, keberhasilan MPT sangat ditentukan oleh ketepatan estimasi *expected return*. Jika perkiraan *return* tersebut tidak sesuai dengan realitas pasar, maka hasil optimasi portofolio bisa melenceng dari kondisi ideal (Firmansyah dan Fathoni, 2025).

Perkembangan teknologi informasi dan data keuangan dalam beberapa tahun terakhir telah membuka peluang baru bagi penerapan *machine learning* (ML) dalam bidang keuangan. Model *machine learning* dapat mengenali pola *non-linear* dan hubungan kompleks antar variabel pasar yang sulit diidentifikasi oleh metode statistik tradisional. Berdasarkan tinjauan sistematis yang dilakukan oleh Nti, Adekoya, dan Weyori (2020), penerapan *machine learning* dalam analisis pasar saham terbukti mampu meningkatkan akurasi prediksi harga maupun *return* saham dibandingkan metode konvensional. Dua algoritma ML yang banyak digunakan dalam konteks prediksi *return* saham adalah *Extreme Gradient Boosting* (XGBoost) dan *Support Vector Regression* (SVR), karena keduanya mampu mempelajari pola *non-linear* dengan efisien serta menghasilkan prediksi yang lebih stabil terhadap dinamika pasar keuangan.

Model XGBoost dikenal karena cara kerjanya yang mengombinasikan sejumlah model pohon keputusan secara bertahap melalui *boosting*, sehingga meminimalkan error prediksi dan mencegah *overfitting* dengan mekanisme regularisasi yang efektif,



an oleh Wang et al. (2022). Adapun SVR, algoritma ini memiliki menangkap hubungan *non-linear* pada dataset keuangan yang menjaga stabilitas prediksi meski pasar sedang tidak stabil, kkan oleh Zhao et al. (2021).

ah membuktikan keefektifan XGBoost dan SVR dalam analisis), contohnya, menemukan bahwa XGBoost secara signifikan

lebih akurat dalam memprediksi harga saham daripada regresi linear biasa. Temuan serupa juga dilaporkan oleh Guennioui, Chiadmi, dan Amghar (2023), yang menunjukkan XGBoost menghasilkan RMSE lebih rendah dan akurasi lebih tinggi dibandingkan model ML lain seperti LightGBM atau regresi linier. Sementara itu, Thumu dan Nellore (2024) mengembangkan model *ensemble* SVR dengan berbagai kernel dan menegaskan bahwa SVR tetap stabil dalam kondisi pasar yang sangat bergejolak. Di Indonesia, Pratama dan Banowosari (2024) menerapkan XGBoost untuk memprediksi harga saham PT Bank Mandiri (BMRI) dan mendapatkan hasil yang lebih presisi dibandingkan LSTM. Begitu pula, Yuliana et al. (2023) menunjukkan bahwa SVR yang dioptimalkan dengan algoritma *Fruit Fly Optimization* (FOA) dapat meningkatkan akurasi prediksi harga saham BBRI berdasarkan sentimen berita keuangan. Semua ini menegaskan potensi kuat XGBoost dan SVR sebagai alat prediksi *return* saham, baik di pasar global maupun domestik.

Meskipun beberapa penelitian telah mencoba mengintegrasikan prediksi ML dengan optimasi portofolio *mean-variance*, sebagian besar masih terfokus pada aspek metodologi dan dilakukan di luar negeri. Misalnya, Lee et al. (2024) membangun model prediksi *return* berbasis *decision-focused learning* yang langsung dioptimalkan untuk tujuan portofolio *mean-variance*, dan hasilnya menunjukkan peningkatan efisiensi portofolio dibandingkan metode konvensional, meskipun terbatas pada data bulanan pasar global. Di Indonesia, studi seperti Pratama dan Banowosari (2024) dan Yuliana et al. (2023) lebih menekankan peningkatan akurasi prediksi harga saham dengan XGBoost dan SVR, tanpa menghubungkannya langsung ke pembentukan portofolio optimal. Oleh karena itu, masih ada ruang besar untuk mengembangkan integrasi model prediksi *return* berbasis ML ke dalam optimasi portofolio *mean-variance* di konteks pasar saham Indonesia.

Dengan demikian, penelitian ini tidak bermaksud menggantikan studi yang sudah ada, melainkan memperluas ruang lingkupnya dan mengadaptasinya ke konteks pasar saham Indonesia. Fokus utama penelitian ini adalah menguji efektivitas model XGBoost dan SVR dalam memprediksi *return* harian saham, kemudian menggunakan hasil prediksi tersebut sebagai input ke dalam proses pembentukan portofolio optimal dengan pendekatan *mean-variance* pada 15 saham yang terdaftar di Bursa Efek Indonesia (periode 2022–2024). Pendekatan ini juga akan menerapkan matriks kovarian bergulir (*rolling covariance matrix*) untuk memperbarui estimasi risiko portofolio secara dinamis, serta mencakup teknik *clipping* μ ($p10-p90$) untuk mengurangi pengaruh nilai ekstrem (*outlier*) dan menerapkan validasi *anti look-ahead* agar hasil evaluasi lebih realistis dan tidak terkontaminasi oleh data masa depan.



Optimized using
trial version
www.balesio.com

bungkan kemampuan prediksi model *machine learning* dan ortofolio klasik, penelitian ini diharapkan dapat memberikan literatur mengenai penerapan algoritma prediktif dalam lio saham di pasar Indonesia. Selain itu, hasil penelitian ini juga eferensi praktis bagi investor, analis keuangan, dan manajer ncang strategi investasi yang efisien, adaptif, dan berbasis data sar yang terus berubah.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana kinerja portofolio optimal berbasis prediksi *return* dari model XGBoost dan SVR dibandingkan dengan portofolio *equal-weight* dan *mean-return*?
2. Sejauh mana penerapan *rolling covariance matrix*, μ clipping (p10–p90), dan anti *look-ahead bias* dapat meningkatkan akurasi dan validitas hasil optimasi portofolio?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dikemukakan, maka tujuan penelitian ini adalah sebagai berikut:

1. Membandingkan kinerja portofolio optimal berbasis prediksi XGBoost dan SVR dengan portofolio *equal-weight* dan *mean-return*
2. Mengevaluasi pengaruh penerapan *rolling covariance matrix*, μ clipping (p10–p90), dan anti *look-ahead bias* terhadap akurasi serta validitas hasil optimasi portofolio.

1.3.2 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Bagi Penulis
Penelitian ini memberikan kesempatan untuk memperdalam pemahaman mengenai penerapan algoritma *machine learning* seperti XGBoost dan SVR dalam memprediksi *return* saham, serta penerapan pendekatan *mean-variance* dalam pembentukan portofolio optimal.
2. Bagi Pembaca dan Peneliti Selanjutnya
Hasil penelitian ini dapat menjadi referensi bagi pembaca dan peneliti yang tertarik mengembangkan studi mengenai integrasi *machine learning* dan teori portofolio modern, khususnya dalam konteks pasar saham Indonesia.
3. Bagi Universitas
Penelitian ini diharapkan dapat memperkaya literatur dan mendukung pengembangan kajian di bidang keuangan kuantitatif serta penerapan kecerdasan buatan dalam riset ekonomi dan investasi.

1.4 Landasan Teori



asar Modal

elaskan bahwa pasar modal adalah tempat terjadinya
stor dan emiten untuk memperjualbelikan surat berharga jangka
r modal, berbagai instrumen keuangan jangka panjang seperti
diperdagangkan, memberikan kesempatan kepada investor

untuk memperoleh imbal hasil sekaligus menyediakan dana bagi perusahaan yang memerlukan modal untuk ekspansi dan investasi.

Menurut Situmorang (2022), pasar modal di Indonesia berperan penting sebagai sumber pembiayaan utama bagi perusahaan. Melalui penawaran umum dan perdagangan efek di Bursa Efek Indonesia (BEI), perusahaan dapat memperoleh dana dengan biaya relatif lebih rendah dibandingkan pembiayaan dari sektor perbankan. Hal ini menunjukkan bahwa pasar modal mampu meningkatkan efisiensi alokasi dana serta memperkuat struktur keuangan korporasi di Indonesia.

Selain sebagai sarana pendanaan, pasar modal juga berperan dalam memperluas partisipasi masyarakat terhadap kegiatan investasi. Aktivitas transaksi efek di pasar sekunder memungkinkan investor memperoleh keuntungan melalui dividen dan *capital gain*, sehingga menumbuhkan minat investasi publik (Alam dkk., 2025). Kegiatan ini tidak hanya menggerakkan sektor keuangan, tetapi juga memiliki dampak nyata terhadap pertumbuhan ekonomi nasional melalui peningkatan investasi dan pembentukan modal. Oleh karena itu, perkembangan pasar modal dapat dijadikan indikator yang mencerminkan optimisme pelaku ekonomi terhadap prospek perekonomian suatu negara.

Dalam konteks Indonesia, BEI sebagai lembaga penyelenggara perdagangan efek memiliki fungsi strategis dalam menciptakan sistem perdagangan yang teratur, wajar, dan efisien. Otoritas Jasa Keuangan (OJK) turut berperan dalam menjaga integritas pasar dan melindungi kepentingan investor melalui kebijakan pengawasan yang ketat. Dengan demikian, keberadaan pasar modal tidak hanya berfungsi sebagai sarana investasi dan pembiayaan, tetapi juga sebagai barometer perekonomian nasional yang mencerminkan dinamika dan ekspektasi pertumbuhan ekonomi ke depan.

1.4.1.2 Investasi di Pasar Modal

Investasi merupakan kegiatan penanaman dana yang dilakukan oleh individu atau lembaga dengan tujuan memperoleh keuntungan di masa yang akan datang. Menurut Tandellin (2017), investasi adalah komitmen atas sejumlah dana atau sumber daya lainnya pada saat ini untuk memperoleh keuntungan di masa mendatang. Dalam konteks pasar modal, investasi dilakukan melalui pembelian instrumen keuangan seperti saham, obligasi, dan reksa dana yang diperdagangkan di bursa efek. Tujuan utama dari kegiatan investasi adalah memperoleh tingkat pengembalian (*return*) yang optimal dengan mempertimbangkan tingkat risiko yang bersedia ditanggung oleh investor.

Bodie, Kane, dan Marcus (2014) menjelaskan bahwa dalam pengambilan investor dihadapkan pada hubungan antara risiko dan imbalan yang memiliki karakteristik yang berbeda dalam hal potensi kerugian. Investor rasional akan memilih kombinasi aset yang pengembalian tertinggi untuk tingkat risiko tertentu atau tingkat pengembalian tertentu, yang dikenal dengan konsep portofolio di dasar dalam teori portofolio modern yang dikembangkan oleh



Dalam praktiknya, investasi di pasar modal dapat dibedakan menjadi dua jenis, yaitu investasi langsung dan investasi tidak langsung. Investasi langsung dilakukan dengan membeli instrumen pasar modal secara langsung seperti saham atau obligasi, sedangkan investasi tidak langsung dilakukan melalui lembaga perantara seperti reksa dana, di mana dana dari banyak investor dikumpulkan dan dikelola oleh manajer investasi. Kedua jenis investasi ini memberikan peluang bagi masyarakat untuk ikut berpartisipasi dalam kegiatan ekonomi produktif tanpa harus terlibat langsung dalam pengelolaan perusahaan (Husnan, 2015).

Situmorang (2022) menyatakan bahwa perkembangan pasar modal Indonesia telah mendorong peningkatan minat masyarakat terhadap investasi. Hal ini ditunjukkan oleh meningkatnya jumlah investor ritel dan transaksi saham di Bursa Efek Indonesia selama beberapa tahun terakhir. Faktor-faktor seperti kemudahan akses digital, edukasi keuangan, serta peran regulator dalam menciptakan iklim investasi yang kondusif turut berkontribusi terhadap pertumbuhan tersebut. Selain memberikan keuntungan bagi investor, peningkatan aktivitas investasi di pasar modal juga berdampak positif terhadap perekonomian nasional melalui penambahan sumber pembiayaan sektor riil.

Investasi di pasar modal juga memiliki peran penting dalam menciptakan stabilitas keuangan dan pertumbuhan ekonomi jangka panjang. Alam dkk. (2025) menegaskan bahwa peningkatan investasi di pasar modal berkorelasi positif dengan ekspansi sektor produktif dan pertumbuhan Produk Domestik Bruto (PDB). Semakin banyak investor yang menanamkan modalnya di pasar modal, semakin besar pula dana yang tersedia untuk mendukung kegiatan produksi dan penciptaan lapangan kerja. Oleh karena itu, pengembangan investasi di pasar modal bukan hanya berorientasi pada keuntungan individu, tetapi juga memiliki kontribusi terhadap pembangunan ekonomi nasional secara menyeluruh.

1.4.1.3 Saham sebagai Instrumen Investasi

Saham merupakan salah satu instrumen investasi utama di pasar modal yang menunjukkan kepemilikan atas suatu perusahaan. Menurut Husnan (2015), saham adalah surat berharga yang menunjukkan bagian kepemilikan suatu perusahaan, yang memberikan hak kepada pemilikinya untuk memperoleh bagian keuntungan berupa dividen dan hak suara dalam rapat umum pemegang saham. Dengan memiliki saham, seorang investor secara tidak langsung ikut memiliki bagian dari aset perusahaan dan menanggung risiko sesuai dengan jumlah saham yang dimilikinya.



Bodie, Kane, dan Marcus (2014) menjelaskan bahwa saham termasuk ke dalam *equities* yang memiliki karakteristik berbeda dibandingkan dengan *debt* tetap seperti obligasi. Keuntungan dari kepemilikan saham adalah *dividend yield* (imbalan periodik yang dibayarkan oleh perusahaan) dan *capital gain* (selisih positif antara harga jual dan harga beli saham). Namun, keuntungan yang tinggi tersebut juga disertai dengan risiko yang besar karena saham sangat sensitif terhadap berbagai faktor ekonomi, politik, dan sosial yang dapat mempengaruhi pasar modal.

Dalam praktiknya, saham dibedakan menjadi dua jenis utama, yaitu saham biasa (*common stock*) dan saham preferen (*preferred stock*). Saham biasa memberikan hak kepemilikan penuh termasuk hak suara dan pembagian dividen yang tidak tetap, sedangkan saham preferen memberikan prioritas pembayaran dividen tetapi biasanya tidak memiliki hak suara (Tandelilin, 2007). Kedua jenis saham tersebut memiliki peran penting dalam struktur modal perusahaan serta menjadi pilihan investasi yang berbeda tergantung pada profil risiko dan preferensi investor.

Harga saham di pasar modal ditentukan oleh kekuatan permintaan dan penawaran yang mencerminkan ekspektasi investor terhadap prospek kinerja perusahaan. Faktor-faktor yang memengaruhi harga saham antara lain kondisi fundamental perusahaan (seperti laba bersih, penjualan, dan dividen), faktor makroekonomi (tingkat suku bunga, inflasi, nilai tukar), serta faktor eksternal lainnya seperti stabilitas politik dan kebijakan pemerintah (Situmorang, 2022). Oleh karena itu, fluktuasi harga saham tidak hanya menggambarkan kondisi internal perusahaan, tetapi juga persepsi pasar terhadap situasi ekonomi secara keseluruhan.

Di Indonesia, perdagangan saham difasilitasi oleh Bursa Efek Indonesia (BEI) yang berfungsi sebagai pasar terorganisasi tempat investor dapat melakukan transaksi jual beli saham secara transparan dan efisien. Saham-saham yang tercatat di BEI dikategorikan dalam berbagai indeks seperti IHSG, LQ45, IDX30, dan lainnya. Indeks-indeks ini digunakan sebagai acuan untuk menilai kinerja pasar secara keseluruhan maupun membandingkan performa investasi antarperusahaan. Berdasarkan penelitian Alam dkk. (2025), kinerja indeks saham di Indonesia memiliki korelasi positif dengan pertumbuhan ekonomi, di mana peningkatan aktivitas perdagangan saham berkontribusi terhadap stabilitas dan perluasan sektor keuangan nasional.

Dengan demikian, saham berfungsi ganda dalam pasar modal, yaitu bagi perusahaan, sebagai sarana pengumpulan dana untuk ekspansi usaha, dan bagi investor, sebagai instrumen untuk memperoleh keuntungan melalui dividen dan *capital gain*. Pemahaman mengenai karakteristik, risiko, dan faktor yang memengaruhi harga saham sangat penting bagi investor untuk membentuk portofolio yang optimal sesuai dengan tujuan investasi dan toleransi risikonya.

1.4.1.4 Indeks Saham Indonesia

Indeks saham merupakan indikator yang menggambarkan pergerakan harga sekelompok saham yang dipilih berdasarkan kriteria tertentu. Indeks berfungsi sebagai ukuran kinerja pasar modal secara keseluruhan serta alat ukur bagi investor



hasil investasinya. Menurut Tandelilin (2007), indeks saham bahan nilai pasar dari kumpulan saham yang menjadi gga dapat digunakan untuk mengamati kecenderungan pasar menilai tingkat keuntungan (*return*) pasar modal dalam suatu

leks saham dikelola dan diterbitkan oleh Bursa Efek Indonesia ang digunakan sebagai tolok ukur adalah Indeks Harga Saham

Gabungan (IHSG), yang mencakup seluruh saham yang tercatat di BEI. IHSG memberikan gambaran umum mengenai arah pergerakan pasar saham nasional dan sering dijadikan acuan untuk menilai sentimen investor terhadap kondisi ekonomi makro. Menurut Husnan (2015), perubahan IHSG dapat dipengaruhi oleh berbagai faktor, seperti kondisi ekonomi domestik dan global, tingkat inflasi, suku bunga, nilai tukar, serta kebijakan pemerintah dalam bidang fiskal dan moneter.

Selain IHSG, terdapat pula indeks sektoral dan tematik yang dikembangkan untuk mengukur kinerja kelompok saham tertentu. Salah satu indeks yang populer adalah indeks LQ45, yang terdiri dari 45 saham dengan tingkat likuiditas dan kapitalisasi pasar terbesar di BEI. Indeks ini direvisi setiap enam bulan sekali untuk memastikan bahwa komposisinya tetap mencerminkan saham-saham paling aktif dan representatif di pasar. Indeks lainnya seperti IDX30, IDX80, dan Kompas100 juga dikembangkan untuk memberikan variasi tolok ukur sesuai kebutuhan analisis investasi (Bodie et al., 2014).

Menurut Alam dkk. (2025), keberadaan indeks saham seperti IHSG dan LQ45 memiliki peran penting dalam menjaga transparansi dan efisiensi pasar modal Indonesia. Indeks tidak hanya menjadi alat evaluasi performa portofolio, tetapi juga berfungsi sebagai dasar bagi pengembangan produk turunan seperti reksa dana indeks (*index fund*) dan kontrak berjangka indeks (*index futures*). Produk-produk ini memungkinkan investor untuk berinvestasi secara pasif mengikuti kinerja indeks tertentu tanpa harus membeli saham-saham penyusunnya satu per satu.

Dalam konteks penelitian ini, indeks saham menjadi acuan penting dalam pemilihan sampel dan evaluasi hasil portofolio. Penelitian ini menggunakan 15 saham IDX yang aktif diperdagangkan selama periode 2022-2024 dan memiliki kapitalisasi pasar yang besar. Pemilihan saham tersebut didasarkan pada relevansinya terhadap dinamika pasar dan likuiditas perdagangan di BEI, sehingga hasil analisis dapat mencerminkan kondisi pasar saham Indonesia secara representatif. Dengan menggunakan indeks sebagai tolok ukur, hasil perbandingan kinerja portofolio baik yang dibentuk dengan pendekatan *equal-weight*, *mean-return*, XGBoost, maupun SVR dapat dinilai secara objektif terhadap performa pasar yang sesungguhnya.

Dengan demikian, indeks saham tidak hanya berperan sebagai ukuran statistik, tetapi juga sebagai komponen penting dalam pengambilan keputusan investasi dan evaluasi strategi portofolio. Pemahaman mengenai struktur dan pergerakan indeks membantu investor maupun peneliti dalam menilai risiko sistematis serta mengembangkan model optimasi portofolio yang lebih akurat dan sesuai dengan kondisi pasar aktual.



Risiko

Risiko merupakan dasar utama dalam teori investasi dan portofolio. Setiap keputusan investasi selalu melibatkan pertimbangan tingkat keuntungan yang diharapkan dan risiko yang bersedia ditanggung. Menurut Bodie, Kane, dan Marcus (2014), *return* mencerminkan imbalan yang diterima investor atas risiko yang diambil, sedangkan risiko

menunjukkan kemungkinan hasil aktual investasi berbeda dari hasil yang diharapkan. Keduanya memiliki hubungan positif semakin tinggi risiko suatu aset, semakin tinggi pula *return* yang diharapkan (*expected return*) dari aset tersebut.

1.4.2.1 Return Saham Individual

Return merupakan hasil yang diperoleh dari suatu investasi selama periode tertentu. Menurut Tandililin (2007), *return* saham dapat berasal dari dua komponen utama, yaitu *dividend yield* dan *capital gain (loss)*. *Dividend yield* merupakan imbal hasil yang diterima investor dalam bentuk pembagian laba perusahaan, sedangkan *capital gain (loss)* merupakan keuntungan atau kerugian yang diperoleh dari selisih antara harga jual dan harga beli saham.

Secara matematis, *return* sederhana (*simple return*) dapat dituliskan sebagai:

$$R_{i(t)} = \frac{P_{i(t)} - P_{i(t-1)}}{P_{i(t-1)}} \quad (1)$$

Keterangan:

- $R_{i(t)}$ = *return* saham ke- i pada waktu ke- t
 $P_{i(t)}$ = harga penutupan saham ke- i pada waktu ke- t
 $P_{i(t-1)}$ = harga penutupan saham ke- i pada periode sebelumnya

Dalam penelitian ini, digunakan data *return* harian karena dapat menggambarkan fluktuasi pasar dengan lebih rinci dan menjadi dasar perhitungan parameter optimasi portofolio seperti *mean return* (μ) dan *covariance matrix* (Σ). *Return* harian juga digunakan sebagai input utama untuk model *machine learning* (XGBoost dan SVR) dalam memprediksi *return* esok hari (*next-day prediction*).

1.4.2.2 Expected Return dan Return Portofolio

Expected return atau tingkat pengembalian yang diharapkan adalah nilai ekspektasi dari kemungkinan hasil yang mungkin terjadi pada suatu investasi. Nilai *expected return* dihitung sebagai rata-rata tertimbang dari seluruh kemungkinan hasil:

$$E(R_i) = \sum_{j=1}^n P_j R_{ij} \quad (2)$$

Keterangan:

- $E(R_i)$ = *return* harapan saham ke- i
 P_j = Probabilitas terjadinya kondisi ke- j
 R_{ij} = *return* saham ke- i pada kondisi ke- j



yang terdiri dari beberapa aset, *expected return* dihitung dengan berdasarkan proporsi investasi masing-masing saham:

$$E(R_p) = \sum_{i=1}^n w_i E(R_i) \quad (3)$$

atau

$$E(R_p) = w^T \mu$$

Keterangan:

- $E(R_p)$ = *return* harapan portofolio dihitung dari μ historis
 w_i = bobot dana yang diinvestasikan pada saham ke- i
 $E(R_i)$ = *return* harapan saham ke- i

Dalam penelitian ini, nilai $E(R_i)$ tidak hanya dihitung berdasarkan rata-rata historis, tetap berdasarkan $\hat{\mu}$ hasil prediksi model *machine learning* (XGBoost dan SVR), yang kemudian digunakan dalam optimasi *mean-variance* untuk menentukan bobot portofolio optimal.

1.4.2.3 Risiko Saham dan Risiko Portofolio

Risiko merupakan ketidakpastian mengenai hasil investasi di masa mendatang. Salah satu ukuran risiko yang umum digunakan adalah variansi (σ^2) dan standar deviasi (σ). Menurut Bodie et al. (2014), risiko saham individu dapat dihitung sebagai:

$$\sigma_i^2 = \frac{1}{n-1} \sum_{t=1}^n (R_{i(t)} - \bar{R}_i)^2 \quad (4)$$

Keterangan:

- σ_i^2 = variansi *return* saham ke- i
 $R_{i(t)}$ = *return* saham ke- i pada waktu ke- t
 $\bar{R}_i$ = rata-rata *return* saham ke- i
 n = jumlah observasi (data harian)

Untuk portofolio yang terdiri atas beberapa saham, risiko total tidak hanya tergantung pada risiko individual masing-masing saham, tetapi juga pada hubungan antar saham yang diukur dengan kovariansi. Rumus risiko portofolio menurut Tandelilin (2007) adalah:

$$\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \sigma_{ij} \quad (5)$$

Keterangan:

- σ_p^2 = variansi *return* portofolio
 w_i, w_j = bobot investasi pada saham ke- i dan ke- j
 σ_{ij} = kovariansi antara *return* saham ke- i dan ke- j



Untuk matriks, dapat ditulis:

$$\begin{aligned} \sigma_p^2 &= W^T \Sigma W \\ \text{atau} \\ \sigma_p &= \sqrt{W^T \Sigma W} \end{aligned} \quad (6)$$

Keterangan:

W = vektor bobot portofolio
 Σ = matriks kovariansi antar *return* saham

Konsep ini menjelaskan pentingnya diversifikasi, karena risiko portofolio dapat dikurangi tanpa mengurangi *expected return*, selama korelasi antar saham tidak sempurna.

1.4.2.4 Hubungan antara *Return* dan Risiko

Hubungan antara *return* dan risiko bersifat positif, di mana semakin besar risiko yang bersedia diambil, semakin besar pula *return* yang diharapkan. Prinsip ini dikenal sebagai *risk-return trade-off* (Fabozzi, 1995). Investor dengan toleransi risiko tinggi cenderung memilih aset berfluktuasi besar seperti saham, sedangkan investor konservatif lebih memilih aset dengan risiko rendah seperti obligasi.

Dalam konteks pembentukan portofolio, keseimbangan antara *return* dan risiko menjadi dasar dalam menentukan portofolio optimal. Portofolio yang efisien adalah portofolio yang memberikan *return* maksimum untuk risiko tertentu atau risiko minimum untuk *return* tertentu.

1.4.3 Matriks Kovariansi dan *Rolling Covariance*

1.4.3.1 Definisi dan Fungsi Matriks Kovariansi

Matriks kovariansi (Σ) menggambarkan hubungan statistik antar *return* aset di dalam portofolio. Setiap elemen σ_{ij} dalam matriks ini menunjukkan kovariansi antara *return* saham ke- i dan ke- j , yang mengukur sejauh mana kedua aset tersebut bergerak bersama. Matriks ini digunakan untuk menghitung total risiko portofolio, yang tidak hanya bergantung pada risiko masing-masing aset, tetapi juga pada korelasi antar aset.

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{n2} & \dots & \sigma_{nn} \end{bmatrix}$$

Dengan setiap elemen didefinisikan sebagai:

$$\sigma_{ij} = E[(R_i - E[R_i])(R_j - E[R_j])] \quad (7)$$

Keterangan:

R_i dan R_j = *return* aktual saham ke- i dan ke- j
 $E[R_i], E[R_j]$ = *expected return* dari saham ke- i dan ke- j



Risiko Kovariansi (*Rolling-Ledoit Wolf*)

Is, hubungan antar aset tidak bersifat konstan. Korelasi dan at berubah seiring waktu akibat kondisi ekonomi, likuiditas, asar. Oleh karena itu, estimasi kovariansi dilakukan secara an pendekatan *rolling window*, bukan dengan seluruh data

historis sekaligus. Secara matematis, untuk jendela waktu sepanjang k hari, matriks kovariansi pada periode t dihitung sebagai:

$$\sum_t = \text{Cov}(R_{t-k+1}, R_{t-k+2}, \dots, R_t) \quad (8)$$

Pendekatan ini memungkinkan model menangkap perubahan volatilitas dan korelasi pasar dari waktu ke waktu. Dengan setiap pergeseran jendela satu periode, nilai $\sum_t$ diperbarui sehingga portofolio selalu dioptimasi menggunakan informasi terkini.

Estimasi matriks kovarians tradisional menggunakan *sample covariance* sering kali tidak stabil, terutama ketika jumlah aset yang dianalisis relatif besar. Kondisi ini menyebabkan matriks kovarians menjadi *ill-conditioned*, yang pada akhirnya mengganggu proses inversi yang diperlukan dalam optimasi portofolio Markowitz. Untuk mengatasi masalah tersebut, Ledoit dan Wolf mengusulkan metode *shrinkage covariance estimator*, yaitu pendekatan yang menggabungkan matriks kovarians sampel S dengan sebuah matriks target L yang lebih stabil melalui formula:

$$\sum_{\text{shrink}} = \delta L + (1 - \delta)S \quad (9)$$

Dengan $0 \leq \delta \leq 1$. Pendekatan ini terbukti mampu menghasilkan estimasi kovariansi yang lebih stabil dan *robust* terhadap kesalahan estimasi.

Dalam penelitian ini, estimasi matriks kovariansi dilakukan dengan menggabungkan kedua pendekatan tersebut. Pertama, kovariansi dihitung secara *rolling window* sepanjang 126 hari untuk menangkap dinamika volatilitas dan korelasi pasar dari waktu ke waktu. Selanjutnya, hasil estimasi tersebut distabilkan menggunakan metode Ledoit-Wolf *shrinkage* guna menghasilkan matriks kovariansi yang *well-conditioned* dan tidak menimbulkan bobot ekstrem pada proses optimasi *mean-variance*. Kombinasi kedua metode ini memungkinkan model portofolio yang dihasilkan bersifat adaptif terhadap perubahan pasar sekaligus stabil secara numerik.

1.4.3.3 Pemilihan Panjang *Window*

Pemilihan panjang *rolling window* merupakan aspek penting dalam estimasi kovariansi dinamis karena menentukan keseimbangan antara stabilitas dan responsivitas model terhadap perubahan pasar. Jendela yang terlalu pendek cenderung menghasilkan estimasi yang berisik dan tidak stabil, sedangkan jendela yang terlalu panjang dapat membuat model lambat menyesuaikan diri terhadap



Oleh karena itu, ukuran jendela harus ditetapkan secara mampu menangkap perubahan struktur risiko tanpa si estimasi.

ini digunakan *rolling window* sepanjang 126 hari perdagangan bursa). Pemilihan ini mengacu pada praktik empiris dan digunakan dalam literatur portofolio dinamis, karena ukuran memberikan keseimbangan antara stabilitas kovariansi dan

kemampuan model merespons perubahan pasar jangka menengah. Selain itu, panjang jendela ini sesuai dengan periode *rebalancing* bulanan yang digunakan dalam penelitian, sehingga estimasi risiko yang diperbarui tetap representatif terhadap kondisi pasar terkini.

1.4.3.4 Keunggulan dan Implementasi Pendekatan *Rolling Window*

Pendekatan *rolling window* memiliki peranan penting dalam pembentukan portofolio dinamis karena mampu menangkap perubahan struktur volatilitas dan korelasi antar saham dari waktu ke waktu. Tidak seperti estimasi kovariansi statis yang menggunakan seluruh data historis sekaligus, metode ini memperbarui perhitungan risiko pada setiap periode dengan menggunakan data sepanjang jendela waktu tertentu (*window size*). Dengan demikian, estimasi risiko portofolio lebih adaptif terhadap perubahan kondisi pasar yang fluktuatif dan tidak lagi bergantung pada asumsi bahwa hubungan antar aset bersifat konstan. Menurut Pfaff (2023), pendekatan *rolling statistics* seperti ini penting dalam keuangan karena karakteristik volatilitas dan korelasi antar aset bersifat *time-varying*, sehingga estimasi kovariansi perlu diperbarui secara dinamis untuk mencerminkan realitas pasar.

Selain itu, *rolling window* mencegah terjadinya *look-ahead bias* karena hanya menggunakan data masa lalu dalam perhitungan Σ_t . Hal ini memastikan bahwa keputusan investasi pada periode t tidak melibatkan informasi masa depan yang belum terjadi. Proses ini juga menjaga konsistensi dengan pendekatan prediksi ($\hat{\mu}$) yang dihasilkan dari model *machine learning* XGBoost dan SVR, karena baik ($\hat{\mu}$) maupun Σ_t sama-sama diperbarui setiap periode berdasarkan data terbaru.

Chiodini (2024) menjelaskan bahwa metode *rolling window* termasuk salah satu teknik klasik yang masih relevan untuk estimasi kovariansi portofolio karena kemampuannya mengikuti dinamika pasar *riil* dan mengurangi distorsi statistik akibat fluktuasi data historis yang tidak stasioner. Penelitian Chen et al. (2022) menunjukkan bahwa penggunaan *rolling window* dengan jendela sepanjang 70 hari dalam optimasi portofolio menghasilkan performa risiko-*return* yang lebih stabil dibanding model statis. Estimasi kovariansi yang diperbarui secara bergeser memungkinkan model untuk menyesuaikan diri dengan perubahan volatilitas pasar, menghasilkan bobot portofolio yang lebih realistis dan tidak ekstrem.

Pendekatan tersebut diadaptasi pula dalam penelitian ini, di mana matriks kovariansi Σ_t dihitung secara *rolling window* dari data *return* harian 15 saham IDX. Ukuran jendela digunakan untuk menghitung Σ pada periode ke- t , kemudian hasilnya digunakan untuk optimasi portofolio periode berikutnya ($t + 1$). Setiap kali jendela bergeser, model XGBoost dan SVR dilatih ulang untuk memprediksi



sa Σ_t diperbarui menggunakan data *return* terbaru. Hasilnya, sikan portofolio yang *time-varying*, di mana bobot optimal (w^*) erbarui secara berkala sesuai dengan kondisi pasar. Dengan n *rolling window* dalam penelitian ini tidak hanya berfungsi asi risiko, tetapi juga sebagai mekanisme pembaruan adaptif.

1.4.4 Teori Portofolio Modern (Markowitz)

Teori portofolio modern diperkenalkan oleh Harry Markowitz pada tahun 1952 melalui tulisan berjudul *Portfolio Selection*. Teori ini menjadi dasar utama dalam manajemen investasi modern karena memperkenalkan hubungan kuantitatif antara *return* dan risiko dalam proses pengambilan keputusan investasi.

Markowitz menyatakan bahwa investor yang rasional tidak hanya mempertimbangkan tingkat pengembalian (*return*) suatu aset secara individual, tetapi juga mempertimbangkan hubungan antar aset melalui korelasi dan kovariansi. Dengan melakukan diversifikasi, investor dapat mengurangi risiko total portofolio tanpa harus mengorbankan *return* secara signifikan.

Konsep utama teori ini adalah bahwa portofolio yang optimal adalah portofolio yang memberikan *return* ekspektasi tertinggi untuk risiko tertentu, atau sebaliknya, risiko terendah untuk *return* tertentu. Prinsip ini dikenal sebagai pendekatan *mean-variance*, karena mempertimbangkan rata-rata (*mean*) sebagai ukuran *return* dan varians (*variance*) sebagai ukuran risiko.

1.4.4.1 Mean-variance Optimization

Pendekatan *mean-variance optimization* (MVO) merupakan inti dari teori portofolio Markowitz. Portofolio diwakili oleh sekumpulan bobot $w = w_1, w_2, w_3 \dots w_n$ yang menunjukkan proporsi dana pada masing-masing aset. Vektor *return* ekspektasi tiap aset dinyatakan sebagai $\mu = \mu_1, \mu_2, \mu_3 \dots \mu_n$. Sedangkan matriks kovariansi antar aset ditulis sebagai,

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{n2} & \dots & \sigma_{nn} \end{bmatrix}$$

Return ekspektasi portofolio dirumuskan sebagaimana persamaan (4). Sementara risiko portofolio (varians) dinyatakan dengan persamaan (7). Optimasi dilakukan untuk menentukan kombinasi bobot yang efisien. Secara teori markowitz terdapat dua bentuk formulasi:

1. Meminimalkan risiko untuk target *return* tertentu

$$\begin{aligned} & \min_w \frac{1}{2} w^T \Sigma w \\ & \text{subject to } w^T \Sigma \mu = \mu_{\text{target}} \\ & \sum_{i=1}^n w_i = 1 \end{aligned} \quad (10)$$

2. Memaksimalkan fungsi utilitas dengan *trade-off* antara *return* dan risiko

$$\begin{aligned} & \max_w \left(w^T \mu - \frac{1}{2} \alpha w^T \Sigma w \right) \\ & \text{subject to } \sum_{i=1}^n w_i = 1 \end{aligned} \quad (11)$$



α = koefisien aversi risiko (*risk aversion coefficient*).

Nilai α yang besar menunjukkan investor yang sangat menghindari risiko (*risk-averse*), sedangkan α kecil menggambarkan investor yang lebih berani mengambil risiko (*risk-tolerant*).

Dalam perkembangannya, muncul formulasi ketiga yaitu maksimisasi rasio Sharpe yang merupakan turunan dari model utilitas Markowitz, digunakan untuk menentukan portofolio *tangency* dengan efisiensi risiko-*return* tertinggi. Fungsi objektifnya dinyatakan sebagai:

$$\begin{aligned} \max_w \quad & \frac{w^T (\mu - R_f \mathbf{1})}{\sqrt{w^T \Sigma w}} \\ \text{subject to} \quad & \sum_{i=1}^n w_i = 1 \\ & w_i \geq 0 \end{aligned} \quad (12)$$

Keterangan:

R_f = tingkat *return* bebas risiko

$\mathbf{1}$ = vektor satuan

Model ini mencari *tangency* portofolio, yaitu portofolio yang bersinggungan dengan *Capital Allocation Line* (CAL) dan memiliki rasio Sharpe maksimum. Semakin tinggi nilai *Sharpe ratio* semakin efisien kombinasi risiko-*return* portofolio. Dalam praktik empiris, portofolio ini sering menjadi basis bagi kombinasi dengan aset bebas risiko.

Hasil dari proses ini adalah bobot optimal w^* yang menggambarkan proporsi investasi pada tiap aset. Dalam penelitian ini, nilai μ digantikan oleh $\hat{\mu}$ yaitu hasil prediksi model *machine learning* XGBoost dan SVR, sedangkan Σ dihitung secara *rolling window* agar mencerminkan perubahan risiko dari waktu ke waktu. Dengan demikian, optimasi portofolio dilakukan secara adaptif sesuai kondisi pasar.

1.4.4.2 Efficient Frontier dan Capital Allocation Line

Setelah portofolio optimal diperoleh melalui proses optimasi *mean-variance*, hubungan antara risiko dan *return* dari sekumpulan portofolio efisien dapat digambarkan dalam sebuah kurva yang dikenal sebagai *efficient frontier*. Menurut Markowitz (1952), *frontier* ini menunjukkan himpunan semua portofolio yang memberikan *return* ekspektasi maksimum untuk setiap tingkat risiko, atau sebaliknya risiko minimum untuk setiap tingkat *return*. Bentuk kurva *frontier* secara teoretis bersifat cekung terhadap sumbu risiko, menandakan adanya *trade-off* positif antara



1. Portofolio modern, *frontier* tersebut dapat diperluas dengan aset bebas risiko (*risk-free asset*) sebagaimana dikembangkan oleh penambahan aset bebas risiko menghasilkan garis baru yang disebut garis CAL), yang menggantikan bentuk *frontier* menjadi garis lurus. Seluruh kombinasi antara aset bebas risiko dan portofolio modern, miringan garis CAL menunjukkan efisiensi portofolio dalam

menghasilkan *excess return* per satuan risiko, yang dikenal sebagai *Sharpe ratio*. Portofolio dengan rasio Sharpe maksimum disebut portofolio *tangency*, yaitu kombinasi aset berisiko yang paling efisien dalam menghasilkan *return* terhadap risiko.

Dalam penelitian ini, kumpulan titik portofolio efisien dibentuk secara empiris melalui hasil optimasi pada setiap periode *rolling window*. Setiap titik ($\sigma_p, E[R_p]$) merepresentasikan posisi efisien portofolio aktual pada waktu tertentu. Selanjutnya, prinsip CAL digunakan untuk menentukan portofolio *tangency*, yaitu portofolio dengan nilai rasio Sharpe tertinggi. Dengan demikian, proses optimasi *max Sharpe ratio* yang dilakukan dalam penelitian ini secara empiris merepresentasikan penentuan garis CAL dan titik *tangency* pada *efficient frontier* empiris yang terbentuk dari data historis (μ) dan hasil prediksi model XGBoost serta SVR ($\hat{\mu}$).

1.4.4.3 Asumsi dan Keterbatasan Teori Portofolio Modern

Teori Portofolio Modern (Markowitz, 1952) berasumsi bahwa investor bersifat rasional, menghindari risiko, dan membuat keputusan hanya berdasarkan dua parameter utama, yaitu *expected return* dan risiko (varians). Model ini juga menganggap *return* aset berdistribusi normal, informasi pasar sempurna, serta tidak ada biaya transaksi. Namun, dalam praktiknya asumsi tersebut jarang terpenuhi. *Return* saham nyata sering menunjukkan distribusi yang tidak normal, pasar bersifat dinamis, dan parameter statistik seperti *mean* dan kovariansi bersifat tidak stasioner. Kesalahan kecil dalam estimasi *expected return* dapat menyebabkan bobot portofolio yang ekstrem dan tidak stabil.

Oleh karena itu, penelitian ini melakukan penyesuaian terhadap model klasik Markowitz melalui penggunaan prediksi *return* ($\hat{\mu}$) dari model *machine learning* XGBoost dan SVR, estimasi matriks kovariansi dengan pendekatan *rolling window*, serta penerapan μ *clipping* untuk menstabilkan bobot portofolio. Penyesuaian ini bertujuan agar model portofolio lebih adaptif terhadap kondisi pasar saham Indonesia yang dinamis.

1.4.4.4 Sequential Least Squares Programming (SLSQP)

Optimasi portofolio Markowitz merupakan permasalahan optimasi kuadratik dengan kendala linear, seperti batasan jumlah bobot yang harus berjumlah satu dan larangan *short-selling*. Dalam praktiknya, penyelesaian masalah optimasi ini memerlukan metode numerik karena bentuk fungsi objektif dan kendala umumnya tidak memiliki solusi analitis, terutama ketika jumlah aset cukup besar. Salah satu metode yang umum digunakan untuk menyelesaikan optimasi terbatas adalah *Sequential Least Squares Programming* (SLSQP). Algoritma ini mengubah permasalahan optimasi

menjadi serangkaian submasalah kuadratik berbentuk *least-squares*. Pada setiap iterasi, SLSQP memperbarui solusi dengan memanfaatkan informasi *hessian* yang disederhanakan sehingga proses menjadi efisien namun tetap memenuhi kendala.

SLSQP antara lain:

1. Mengatasi kendala *linear* dan *non-linear*,



- mendukung *bound constraints* seperti bobot minimum 0 dan maksimum 1,
- stabil untuk masalah optimasi cembung seperti *mean–variance optimization*,
- tersedia secara luas dalam pustaka ilmiah seperti *SciPy*, yang membuatnya praktis untuk implementasi penelitian

1.4.5 Portofolio Pemandangan

1.4.5.1 *Equal Weight*

Portofolio *equal-weight* (EW) merupakan bentuk paling sederhana dari alokasi aset, di mana seluruh saham dalam portofolio diberi bobot yang sama besar tanpa mempertimbangkan nilai ekspektasi *return* maupun risiko masing-masing aset. Pendekatan ini sering digunakan sebagai *benchmark* karena tidak bergantung pada model statistik atau prediksi, sehingga mencerminkan kinerja pasar yang netral terhadap perbedaan karakteristik saham. Secara matematis, bobot portofolio *equal-weight* dengan jumlah aset n dapat ditulis sebagai:

$$w_i = \frac{1}{n}, \quad i = 1, 2, \dots, n$$

Portofolio jenis ini mengasumsikan bahwa semua aset memiliki kontribusi yang sama terhadap pembentukan portofolio. Dengan demikian, *expected return* dan risiko portofolio dapat dihitung menggunakan persamaan (3) dan (5). Kelebihan utama pendekatan *equal-weight* adalah kesederhanaan implementasinya dan ketahanannya terhadap kesalahan estimasi, karena tidak bergantung pada parameter seperti kovariansi atau prediksi ($\hat{\mu}$). Namun, kelemahannya adalah alokasi dana yang tidak efisien, karena tidak mempertimbangkan perbedaan risiko maupun potensi *return* antar saham. Dalam konteks penelitian ini, portofolio *equal-weight* berfungsi sebagai *benchmark* dasar untuk menilai sejauh mana model berbasis optimasi (*mean-return*, XGBoost, dan SVR) mampu memberikan peningkatan kinerja risiko-*return*.

1.4.5.2 *Mean-return Historis*

Portofolio *mean-return* atau portofolio *mean-variance* berbasis μ historis merupakan bentuk implementasi langsung dari teori Markowitz klasik. Dalam pendekatan ini, vektor ekspektasi *return* (μ) diperoleh dari rata-rata historis *return* saham pada periode tertentu tanpa melibatkan model prediktif. Nilai μ historis kemudian digunakan bersama matriks kovariansi Σ untuk menyelesaikan optimasi portofolio dengan kriteria maksimum rasio Sharpe.

Pendekatan ini disebut juga sebagai *historical mean model*, karena mengandalkan data masa lalu untuk memproyeksikan ekspektasi *return* masa depan. Keunggulannya adalah sederhana, mudah diimplementasikan, dan tidak memerlukan rediktif yang kompleks. Namun, kelemahan utamanya terletak pada pola *return* masa lalu akan berulang di masa depan, padahal dinamis dan sering berubah.

Di sini, portofolio *mean-return* digunakan sebagai pembandingan dengan model berbasis prediksi ($\hat{\mu}$) dari XGBoost dan SVR. Dengan demikian, hasil optimasi berbasis μ historis dengan hasil berbasis ($\hat{\mu}$)



prediksi, dapat dievaluasi sejauh mana penggunaan model *machine learning* memberikan peningkatan efisiensi risiko-*return* dibanding pendekatan konvensional.

1.4.6 Data Preprocessing

Tahap data *preprocessing* merupakan langkah penting sebelum model *machine learning* dijalankan. Tujuannya adalah memastikan data harga saham yang digunakan memiliki kualitas yang baik, bebas dari kesalahan pencatatan, dan siap untuk dianalisis lebih lanjut. Mengingat data keuangan sering kali bersifat volatil, tidak stasioner, dan rentan mengandung *noise*, proses pembersihan serta transformasi data menjadi prasyarat agar hasil prediksi *return* dan optimasi portofolio bersifat valid dan bebas bias.

1.4.6.1 Pembersihan Data (*Missing* dan Duplikat)

Data harga saham kerap mengandung nilai kosong akibat hari libur perdagangan atau gangguan data. Dalam penelitian ini, nilai kosong diatasi dengan metode *forward fill*, yaitu mengisi data hilang menggunakan nilai terakhir yang tersedia. Selain itu, data duplikat dihapus, dan indeks waktu disortir agar urutan kronologisnya konsisten. Langkah ini memastikan bahwa perhitungan *return* harian tidak terganggu oleh kekosongan atau anomali data, sehingga hasilnya mencerminkan pergerakan harga aktual dari waktu ke waktu.

1.4.6.2 Standardisasi Fitur

Model *machine learning* seperti *Support Vector Regression* (SVR) sensitif terhadap perbedaan skala antar variabel input, sehingga dilakukan proses standardisasi menggunakan metode *z-score*, yang mengubah setiap x_i menjadi bentuk standar:

$$z_i = \frac{x_i - \bar{x}}{s} \quad (13)$$

Keterangan:

$\bar{x}$ = rata-rata nilai historis fitur
 s = simpangan baku

Melalui pendekatan ini, setiap fitur diubah agar memiliki rata-rata nol dan standar deviasi satu. Dalam implementasinya, standardisasi dilakukan secara otomatis di dalam *pipeline* model SVR dengan menggunakan *StandardScaler* dari pustaka *scikit-learn*. Sementara itu, model XGBoost tidak memerlukan normalisasi karena berbasis pohon keputusan yang tidak terpengaruh oleh skala data. Meskipun proses pembelajaran kedua model dilakukan dengan cara berbeda, hasil akhirnya yakni prediksi *return* harian ($\hat{\mu}$) tetap berada dalam satuan yang sama, yaitu tingkat μ dalam bentuk desimal. Proses ini membantu menjaga konsistensi antar fitur dan meningkatkan stabilitas hasil pembelajaran.



ering

merupakan tahapan penting dalam pembentukan variabel input untuk melatih model *machine learning*. Tujuannya adalah

mengekstraksi informasi relevan dari data harga dan volume saham agar model dapat mengenali pola yang berhubungan dengan pergerakan *return* di masa depan. Tahap ini menjadi krusial karena kualitas fitur secara langsung menentukan kemampuan model untuk menangkap dinamika pasar yang kompleks.

Dalam penelitian ini, *feature engineering* dilakukan terhadap data harga saham harian 15 emiten di Bursa Efek Indonesia periode 2022–2024. Seluruh fitur dibangun berdasarkan indikator teknikal yang umum digunakan dalam analisis pasar saham, seperti *Simple Moving Average* (SMA), *Exponential Moving Average* (EMA), *Moving Average Convergence Divergence* (MACD), *Relative Strength Index* (RSI), dan *Bollinger Bands*. Fitur-fitur ini dipilih karena mampu menggambarkan perilaku tren, momentum, dan volatilitas harga saham.

1.4.7.1 Indikator Teknikal

Penelitian ini menggunakan beberapa indikator teknikal populer yang dihitung dari data harga penutupan perdagangan. Indikator-indikator ini dipilih karena telah banyak digunakan dalam analisis teknikal untuk mengidentifikasi arah tren, momentum, dan tingkat volatilitas pasar. Menurut Investopedia (2024) dan Tradelink (2024), indikator seperti SMA, EMA, RSI, MACD, dan *Bollinger Bands* merupakan dasar utama dalam pembentukan strategi analisis teknikal modern.

1. *Simple Moving Average* (SMA5 dan SMA20)

SMA adalah rata-rata aritmetika dari harga penutupan dalam periode tertentu. SMA5 digunakan untuk mengamati tren jangka pendek, sedangkan SMA20 merepresentasikan tren jangka menengah. Persilangan antara kedua garis ini sering digunakan sebagai sinyal perubahan arah tren (*golden cross* atau *death cross*).

$$SMA_n(t) = \frac{1}{n} \sum_{i=0}^{n-1} P_{t-i}, \quad n \in \{5, 20\}. \quad (14)$$

Keterangan:

$SMA_n(t)$ = SMA periode n pada t

P_t = harga penutupan pada hari ke- t

2. *Exponential Moving Average* (EMA12 dan EMA26)

EMA merupakan bentuk rata-rata bergerak yang memberikan bobot lebih besar pada harga terbaru, sehingga lebih responsif terhadap perubahan harga dibanding SMA. Kombinasi antara EMA12 dan EMA26 digunakan untuk membentuk indikator MACD (*Moving Average Convergence Divergence*), di mana nilai MACD positif menunjukkan momentum naik, sementara nilai negatif menunjukkan tekanan jual.



$$EMA_n(t) = \alpha P_t + (1 - \alpha) EMA_n(t - 1), \quad n \in \{12, 26\}.$$

$$\alpha = \frac{2}{n + 1} \quad (15)$$

$EMA_n(t)$ = EMA periode n pada t

P_t = harga penutupan pada hari ke- t

3. MACD Signal Line

Garis sinyal MACD merupakan EMA9 dari nilai MACD. Ketika garis MACD memotong garis sinyal dari bawah ke atas, hal ini sering diinterpretasikan sebagai sinyal beli (*bullish crossover*), sedangkan pemotongan dari atas ke bawah dianggap sebagai sinyal jual (*bearish crossover*).

$$\begin{aligned} MACD(t) &= EMA_{12}(t) - EMA_{26}(t) \\ Signal(t) &= EMA_9(MACD(t)) \end{aligned} \quad (16)$$

4. Relative Strength Index (RSI14)

RSI merupakan indikator momentum yang mengukur kekuatan kenaikan dan penurunan harga dalam rentang 0-100. Nilai RSI di atas 70 menunjukkan kondisi *overbought* (harga terlalu tinggi), sedangkan di bawah 30 menunjukkan *oversold* (harga terlalu rendah). Indikator ini membantu mendeteksi kejenuhan pasar dan potensi pembalikan arah tren.

$$\begin{aligned} U_t &= \max(P_t - P_{t-1}, 0), \quad D_t = \max(P_{t-1} - P_t, 0) \\ \bar{U}_t &= \frac{13\bar{U}_{t-1} + U_t}{14}, \quad \bar{D}_t = \frac{13\bar{D}_{t-1} + D_t}{14} \\ RS_t &= \frac{\bar{U}_t}{\bar{D}_t} \\ RSI_{14}(t) &= 100 - \frac{100}{1 + RS_t} \end{aligned} \quad (17)$$

5. Bollinger Bands (BB_upper dan BB_lower)

Bollinger Bands terdiri atas tiga garis: garis tengah berupa SMA20, serta dua pita di atas dan bawah yang berjarak 2 deviasi standar dari SMA tersebut. *Band* atas (BB_upper) menunjukkan batas atas volatilitas, sedangkan *band* bawah (BB_lower) menunjukkan batas bawah. Harga yang mendekati pita atas menandakan tekanan beli tinggi, sedangkan pergerakan ke pita bawah mengindikasikan tekanan jual kuat.

$$\begin{aligned} Middle(t) &= SMA_{20}(t) \\ Upper(t) &= SMA_{20}(t) + 2\sigma_{20}(t) \\ Lower(t) &= SMA_{20}(t) - 2\sigma_{20}(t) \\ \sigma_{20}(t) &= \sqrt{\frac{1}{20} \sum_{i=0}^{19} (P_{t-i} - SMA_{20}(t))^2} \end{aligned} \quad (18)$$

6. Volume Perdagangan

Volume mencerminkan intensitas transaksi yang terjadi pada suatu saham dalam periode tertentu. Volume tinggi sering kali memperkuat sinyal dari harga, karena menandakan partisipasi pasar yang besar. Oleh karena itu, volume digunakan langsung sebagai fitur yang mewakili aktivitas pasar.



Indikator-indikator tersebut memberikan representasi komprehensif dari pergerakan harga. SMA dan EMA menangkap arah tren, MACD dan RSI mengukur momentum, *Bollinger Bands* memvisualisasikan volatilitas, dan volume menunjukkan tingkat aktivitas pasar yang mendasari pergerakan harga.

1.4.7.2 Seleksi Fitur dan Stasioneritas

Setelah fitur-fitur teknikal dibentuk, langkah berikutnya adalah memastikan bahwa variabel yang digunakan dalam proses pelatihan model memiliki relevansi terhadap target prediksi *return* ($\hat{\mu}$) dan bersifat stabil secara statistik. Proses seleksi fitur tidak dilakukan secara eksplisit menggunakan metode filter atau eliminasi manual, melainkan bergantung pada mekanisme internal dari masing-masing model.

Pada model XGBoost, proses seleksi fitur terjadi secara otomatis melalui pembentukan pohon keputusan, di mana setiap cabang pohon hanya memilih fitur dengan kontribusi tertinggi terhadap pengurangan error (*gain*). Dengan demikian, fitur yang kurang relevan secara alami akan jarang dipilih dan memiliki bobot kepentingan rendah. Sementara itu, pada model SVR, semua fitur teknikal digunakan sebagai input tanpa proses seleksi manual. Risiko *overfitting* diminimalkan karena SVR memiliki parameter regularisasi (C dan ε) yang mengontrol kompleksitas model, serta seluruh data input telah dinormalisasi menggunakan *StandardScaler()*. Selain itu, penelitian ini menerapkan skema *rolling training*, yang berfungsi sebagai validasi berulang untuk menjaga kestabilan model terhadap data baru.

Aspek stasioneritas juga diperhatikan dalam penelitian ini. Karena harga saham bersifat *non-stasioner*, model tidak menggunakan harga mentah sebagai target, melainkan *return* harian, yang cenderung memiliki *mean* dan varians lebih stabil. Fitur-fitur teknikal seperti SMA, EMA, RSI, MACD, dan *Bollinger Bands* juga merupakan transformasi relatif terhadap harga yang membantu menstabilkan skala data dan mengurangi pengaruh tren jangka panjang. Dengan demikian, baik target maupun fitur input dalam penelitian ini telah memenuhi prinsip stasioneritas secara implisit, sehingga model XGBoost dan SVR dapat belajar dari pola pasar yang konsisten dan tidak bias terhadap perubahan level harga.

1.4.8 Machine Learning dalam Keuangan

1.4.8.1 Perkembangan dan Tantangan Machine Learning di Pasar Modal

Perkembangan teknologi komputasi dan ketersediaan data pasar yang semakin besar telah mendorong adopsi *machine learning* (ML) secara luas dalam dunia keuangan. ML memungkinkan komputer untuk mempelajari pola dari data historis tanpa perlu pemrograman eksplisit, sehingga mampu menangani kompleksitas dan *non-linearitas* yang tinggi pada pergerakan harga saham. Menurut Nti, Adekoya, dan Weyori (2020), algoritma *machine learning* seperti *Support Vector Machine* (SVM) dan *Artificial Neural Network* (ANN) telah banyak digunakan untuk memprediksi harga dan *return* saham menggunakan pendekatan fundamental maupun teknikal. Hasil tinjauan sistematis mereka menunjukkan bahwa model ML secara umum



di prediksi yang lebih baik dibandingkan metode statistik klasik

pasar modal, ML telah banyak diterapkan untuk berbagai tujuan, n saham (*forecasting*), analisis tren pasar, manajemen risiko, portofolio optimal (*Machine Learning for Finance*, 2024). Model ata harga, volume, dan indikator teknikal untuk membangun adaptif terhadap perubahan pasar.

Namun, penerapan ML di pasar keuangan juga memiliki tantangan. Data pasar yang *non-stasioner* dan bersifat volatil membuat model rentan terhadap *overfitting* dan data *leakage*. Jika model dilatih tanpa memperhatikan urutan waktu, dapat terjadi *look-ahead bias*, yaitu ketika informasi masa depan secara tidak sengaja ikut digunakan dalam proses pelatihan. Untuk menghindari hal ini, penelitian ini menggunakan pendekatan *rolling window* dan validasi *out-of-sample*, sehingga setiap model hanya menggunakan data historis yang tersedia hingga tanggal prediksi dilakukan.

Selain itu, kondisi pasar yang terus berubah mengharuskan model diperbarui secara berkala. Dalam penelitian ini, proses retraining model XGBoost dan SVR dilakukan setiap periode *rebalancing* bulanan, sehingga hasil prediksi *expected return* ($\hat{\mu}$) selalu mencerminkan pola pasar terkini. Pendekatan ini memastikan bahwa model bersifat adaptif dan konsisten terhadap dinamika pasar saham Indonesia.

1.4.8.2 Kelebihan *Machine learning* dibanding Model Statistik Konvensional

Model statistik klasik seperti regresi linier memiliki keterbatasan dalam menangkap hubungan *non-linear* antar variabel dan sering mengasumsikan distribusi data yang stasioner. Sebaliknya, *machine learning* dapat mempelajari hubungan kompleks antar variabel tanpa asumsi distribusi tertentu serta menyesuaikan diri terhadap perubahan pasar.

Chen dan Guestrin (2016) menjelaskan bahwa algoritma XGBoost (*Extreme Gradient Boosting*) bekerja dengan menggabungkan sejumlah pohon keputusan (*decision trees*) secara bertahap untuk meminimalkan kesalahan prediksi. Selain itu, XGBoost menggunakan regularisasi untuk mengontrol kompleksitas model, sehingga mampu menghindari *overfitting* pada data keuangan yang bersifat bising dan tidak linear.

Sementara itu, *Support Vector Regression* (SVR) merupakan metode regresi yang menggunakan ϵ -insensitive loss function dan kernel function untuk menangkap hubungan *non-linear* antara variabel input dan target regresi. Dalam konteks prediksi saham, Dash et al. (2021) menggunakan SVR yang di tuning dengan *grid-search* untuk memilih kernel terbaik dan parameter (C, ϵ, γ) , sehingga model mampu menangani *non-linearitas* dalam data pasar saham.

Dibandingkan metode statistik konvensional, pendekatan ML memiliki beberapa keunggulan utama:

1. Mampu menangkap pola *non-linear* antar variabel keuangan.
2. Dapat menggabungkan banyak indikator teknikal secara simultan.

aptasi dengan data baru melalui pembaruan model (*rolling*



tung pada asumsi distribusi data normal.

g bukan hanya alat prediksi, tetapi juga fondasi penting dalam modern. Integrasi antara ML dan teori portofolio klasik seperti *modern portfolio theory* memberikan kerangka kerja yang adaptif, efisien, dan

1.4.9 Model *Extreme Gradient Boosting* (XGBoost)

1.4.9.1 Konsep *Boosting*

Extreme Gradient Boosting (XGBoost) merupakan algoritma *machine learning* berbasis *decision tree ensemble* yang dikembangkan oleh Chen dan Guestrin (2016). Algoritma ini termasuk dalam keluarga *ensemble learning*, khususnya metode *boosting*, yaitu proses menggabungkan beberapa *weak learners* secara bertahap menjadi satu *strong learner* yang mampu meningkatkan akurasi prediksi.

Konsep *boosting* bertujuan memperbaiki kesalahan (*error*) dari model sebelumnya melalui pembelajaran bertingkat (*sequential learning*). Setiap model baru berfokus pada observasi yang sebelumnya memiliki kesalahan prediksi besar, sehingga model akhir merupakan penjumlahan seluruh fungsi prediktor *tree* yang terbentuk:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathcal{F} \quad (19)$$

Keterangan:

- $\hat{y}_i$ = nilai prediksi akhir untuk observasi ke- i
- $f_k(x_i)$ = pohon keputusan (*decision tree*) ke- k
- K = jumlah pohon
- $\mathcal{F}$ = ruang fungsi yang berisi semua kandidat *regression trees*

Persamaan (19) menunjukkan bahwa prediksi akhir ($\hat{y}_i$) pada model XGBoost diperoleh dari penjumlahan output seluruh pohon keputusan (*decision trees*) yang dibangun secara bertahap. Setiap fungsi $f_k(x_i)$ merepresentasikan pohon keputusan ke- k yang memetakan himpunan fitur input x_i (seperti SMA, EMA, RSI, Volume, dll) menjadi nilai prediksi sementara. Nilai $f_1(x_i)$ berasal dari model pertama yang memberikan prediksi awal terhadap target y_i . Selanjutnya, model kedua $f_2(x_i)$ dilatih untuk mempelajari sisa kesalahan (*residual error*) dari model pertama, model ketiga $f_3(x_i)$ memperbaiki kesalahan kumulatif dari dua model sebelumnya, dan proses ini berlanjut hingga terbentuk K model.

1.4.9.2 Fungsi Objektif, Regulasi, dan Update Tree

Fungsi objektif XGBoost terdiri atas dua komponen, yaitu fungsi kerugian (*loss function*) dan fungsi regularisasi (*regularization term*), dituliskan sebagai:

$$\mathcal{L}(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (20)$$

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2$$



- si kerugian antara nilai aktual dan prediksi
- si regulasi untuk mengontrol kompleksitas model
- ih observasi
- h pohon keputusan (*tress*)

γ	= penalti untuk jumlah <i>leaf node</i>
λ	= koefisien regularisasi $L2$
T	= jumlah <i>leaf node</i> dalam pohon ke- k
w_j	= nilai prediksi pada <i>leaf node</i> ke- j

Persamaan (20) menggambarkan tujuan utama algoritma XGBoost, yaitu mencari parameter θ yang meminimalkan total kesalahan prediksi sekaligus menjaga agar model tidak terlalu kompleks. Pada persamaan (20) menunjukkan bahwa semakin banyak *leaf nodes* atau semakin besar bobot prediksi, maka nilai penalti meningkat. Hal ini membantu mencegah *overfitting* pada data keuangan yang fluktuatif. Agar pelatihan efisien, XGBoost menggunakan pendekatan *second-order Taylor expansion* terhadap fungsi kerugian (*loss function*), sebagaimana ditunjukkan oleh Chen, Zhang et al (2020):

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t)$$

$$g_i = \frac{\partial l}{\partial \hat{y}_i} = -2(y_i - \hat{y}_i^{(t-1)})$$

$$h_i = \frac{\partial^2 l}{\partial \hat{y}_i^2} = 2$$
(21)

Keterangan:

$\mathcal{L}^{(t)}$	= fungsi objektif pada iterasi ke- t
$f_t(x_i)$	= pohon baru yang sedang dibangun pada iterasi ke- t
g_i dan h_i	= turunan pertama dan kedua dari fungsi kerugian
$\Omega(f_t)$	= regulasi pohon baru pada iterasi tersebut
g_i	= gradien, menunjukkan arah perbaikan kesalahan
h_i	= hessian, menunjukkan kecepatan perubahan fungsi kerugian (<i>curvature</i>)

Persamaan ini digunakan untuk memperbarui model secara bertahap. Pohon baru $f_t(x_i)$ ditambahkan ke model sebelumnya sehingga menghasilkan model yang lebih akurat:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i)$$
(22)

Keterangan:

$\hat{y}_i^{(t)}$	= prediksi pada iterasi ke- t
$\hat{y}_i^{(t-1)}$	= prediksi dari model sebelumnya
$f_t(x_i)$	= pohon baru yang menyesuaikan kesalahan dari model sebelumnya



dan h_i dihitung untuk seluruh data, XGBoost menentukan titik ik menggunakan fungsi *gain*:

$$= \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma$$
(23)

Keterangan:

G_L dan H_L = jumlah gradien dan hessian pada cabang kiri
 G_R dan H_R = jumlah gradien dan hessian pada cabang kanan

Persamaan (23) digunakan untuk menghitung seberapa besar peningkatan akurasi (*gain*) jika suatu fitur digunakan sebagai titik pemisahan (*split point*). Fitur dengan nilai *gain* terbesar dipilih sebagai pemisahan terbaik pada pohon keputusan.

Secara keseluruhan cara kerja algoritma XGBoost dalam meminimalkan kesalahan prediksi dan menjaga kompleksitas model yaitu fungsi objektif menggabungkan kesalahan prediksi (*loss function*) dan penalti regularisasi untuk menghindari *overfitting*. Model diperbarui secara iteratif, di mana setiap pohon baru ($f_t(x_i)$) dibangun berdasarkan arah gradien (g_i) dan kelengkungan fungsi kerugian (h_i) dari iterasi sebelumnya. Proses ini menghasilkan penambahan model secara bertahap yang terus memperbaiki kesalahan hingga mencapai prediksi akhir. Selanjutnya, fungsi *gain* digunakan untuk memilih fitur terbaik seperti SMA, EMA, RSI, dan MACD yang memberikan penurunan kesalahan paling signifikan. Melalui mekanisme ini, XGBoost mampu membentuk model prediksi *expected return* yang stabil, efisien, dan adaptif terhadap pola *non-linear* pada data saham.

1.4.9.3 Parameter Utama

Dalam algoritma XGBoost, performa model sangat dipengaruhi oleh pengaturan parameter (*hyperparameters*) yang menentukan bagaimana proses pembelajaran dilakukan. Parameter-parameter ini mengontrol jumlah pohon keputusan yang dibentuk, kedalaman tiap pohon, laju pembelajaran, serta kekuatan penalti regularisasi. Secara umum, terdapat lima parameter utama yang paling menentukan kualitas model, yaitu sebagai berikut:

1. *n_estimators*

Menentukan jumlah iterasi *boosting* atau jumlah pohon keputusan (*trees*) yang akan dibangun. Semakin banyak iterasi, semakin tinggi kemampuan model mempelajari pola data yang kompleks. Namun, nilai yang terlalu besar dapat menyebabkan *overfitting* bila tidak disertai pengaturan regularisasi yang memadai.

2. *learning_rate* (η)

Mengontrol seberapa besar kontribusi setiap pohon baru terhadap model akhir. Nilai η yang kecil (biasanya antara 0.01-0.1) membuat proses pembelajaran lebih lambat tetapi meningkatkan kemampuan generalisasi model, sedangkan nilai yang terlalu besar dapat menyebabkan model cepat namun kurang stabil.



kedalaman maksimum setiap pohon keputusan (*maximum tree max_depth*) yang besar memungkinkan model menangkap *non-linear* yang lebih kompleks antar variabel, tetapi juga berisiko *overfitting*. Dalam data keuangan, nilai antara 3-8 memberikan hasil yang optimal.

4. γ (gamma)

Merupakan parameter regularisasi yang memberikan penalti terhadap jumlah *leaf nodes* pada pohon. Nilai γ yang lebih tinggi membuat model lebih konservatif dalam membentuk *split* baru, sehingga membantu menghindari pembentukan pola yang berasal dari *noise* pasar.

5. λ (lambda)

Parameter regularisasi *L2* yang mengontrol besar bobot (*weights*) pada setiap *leaf*. Nilai λ yang besar menurunkan kompleksitas model dan meningkatkan kestabilan prediksi, terutama ketika menghadapi data harga saham yang volatil.

Kelima parameter tersebut merupakan inti dalam pengendalian kinerja model XGBoost. Dalam penelitian ini, proses *hyperparameter tuning* menggunakan *GridSearchCV* dengan validasi berbasis waktu (*TimeSeriesSplit*) dilakukan terutama pada parameter *n_estimators*, *learning_rate*, dan *max_depth*, sedangkan parameter γ dan λ dipertahankan pada nilai default karena efeknya relatif kecil terhadap stabilitas model pada data keuangan harian. Selain parameter utama di atas, XGBoost juga memiliki beberapa parameter tambahan (*secondary parameters*) yang dapat digunakan untuk penyetelan lanjutan, antara lain:

- *Subsample*, yaitu proporsi data pelatihan yang digunakan untuk setiap iterasi, berguna untuk mengurangi *overfitting*.
- *Colsample_bytree*, yaitu proporsi fitur yang dipilih secara acak pada setiap pohon, membantu meningkatkan keragaman antar pohon (*diversity*).
- *Min_child_weight*, yaitu bobot minimum yang diperlukan agar suatu *split* terjadi, semakin besar nilainya, semakin kuat efek regularisasi yang dihasilkan.

Meskipun parameter tambahan ini tidak dikategorikan sebagai parameter utama, penyesuaiannya dapat meningkatkan efisiensi dan stabilitas model tanpa mengubah struktur utama pembelajaran *boosting*.

1.4.10 Model Support Vector Regression (SVR)

1.4.10.1 Hyperplane dan Kernel Function

Support Vector Regression (SVR) merupakan pengembangan dari Support Vector Machine (SVM) yang digunakan untuk regresi, bukan klasifikasi. Tujuan utama SVR adalah mencari *hyperplane* optimal yang dapat memprediksi nilai target (*return saham*) dengan kesalahan minimum dan margin kesalahan yang masih dapat ditoleransi (ϵ -insensitive margin) (Smola & Scholkopf, 2004). Secara matematis, fungsi regresi SVR didefinisikan sebagai:

$$f(x) = w^T x + b \quad (24)$$



tor bobot
tor fitur input

b = konstanta bias

Fungsi ini berperan sebagai *regression hyperplane* yang berusaha mendekati seluruh titik data dengan jarak seminimal mungkin di bawah batas toleransi kesalahan ε .

Untuk mengakomodasi hubungan *non-linear* antar variabel keuangan, SVR menggunakan *kernel function*, yaitu pemetaan data ke ruang berdimensi tinggi (*feature space*), sehingga hubungan *non-linear* di ruang input dapat ditangani secara linear di ruang fitur:

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) \quad (25)$$

Keterangan:

$\varphi(x)$ = fungsi pemetaan non-linear

$K(x_i, x_j)$ = fungsi *kernel* yang mengukur kesamaan antara dua titik data

Jenis *kernel* yang umum digunakan meliputi *linear*, *polynomial*, dan *radial basis function (RBF)*. Dalam penelitian ini digunakan *RBF kernel* karena kemampuannya menangkap pola *non-linear* dan volatilitas pada data harga saham (Dash et al., 2021).

1.4.10.2 ε – Insensitive Loss Function

Support Vector Regression (SVR) memperkenalkan batas toleransi ε yang disebut ε – *insensitive zone*. Kesalahan yang masih berada dalam batas ini diabaikan, sedangkan kesalahan yang melampaui margin akan dikenakan penalti. Fungsi kerugian SVR didefinisikan sebagai (Smola & Scholkopf, 2004):

$$L_{\varepsilon}(y_i, f(x_i)) = \begin{cases} 0, & \text{jika } |y_i - f(x_i)| \leq \varepsilon \\ |y_i - f(x_i)| - \varepsilon, & \text{jika } |y_i - f(x_i)| > \varepsilon \end{cases} \quad (26)$$

Keterangan:

y_i = nilai aktual *return* saham ke- i

$f(x_i)$ = nilai prediksi model

ε = margin kesalahan yang dapat ditoleransi (*insensitive zone*)

Fungsi ini mengabaikan fluktuasi kecil yang tidak signifikan dalam data pasar (*market noise*). Hal ini membuat SVR menghasilkan fungsi regresi yang lebih halus (*smooth regression function*) dan lebih stabil terhadap volatilitas harga saham jangka pendek.

1.4.10.3 Fungsi Objektif dan Kendala Optimasi

Setelah mendefinisikan fungsi kerugian, SVR membangun fungsi objektif untuk mencari parameter w dan b yang menghasilkan margin optimal dengan kompleksitas rendah. Tujuannya adalah menyeimbangkan antara *flatness* dan ε . Fungsi objektif SVR dirumuskan



$$\min_{w, b, \xi_i, \xi_i^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (27)$$

dengan kendala:

$$\begin{cases} y_i - (w^T x_i - b) \leq \varepsilon + \xi_i \\ (w^T x_i + b) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i \xi_i^* \geq 0 \end{cases}$$

Keterangan:

- ξ_i dan ξ_i^* = variabel *slack* yang mengizinkan pelanggaran batas margin
- C = parameter regularisasi yang mengontrol keseimbangan antara kompleksitas model dan jumlah pelanggaran margin
- n = jumlah data observasi

Bagian pertama dari fungsi objektif $\frac{1}{2} \|w\|^2$ menjaga agar model tetap sederhana (menghindari *overfitting*), sementara bagian kedua $C \sum_{i=1}^n (\xi_i + \xi_i^*)$ menambahkan penalti terhadap titik data yang melampaui margin ε . Dengan demikian, SVR menghasilkan *hyperplane* yang optimal, tidak terlalu kaku (*underfit*), namun tidak terlalu sensitif terhadap *noise* (*overfit*).

1.4.10.4 Parameter C , ε dan γ

Kinerja model *Support Vector Regression* (SVR) dalam memprediksi *expected return* saham sangat bergantung pada pemilihan dan pengaturan parameter yang tepat. Parameter-parameter ini berperan dalam mengendalikan kompleksitas model, sensitivitas terhadap data, serta kemampuan generalisasi terhadap pola yang tidak terlihat pada data pelatihan. Secara umum, terdapat tiga parameter utama yang paling berpengaruh, yaitu sebagai berikut:

1. C (*Regularization parameter*)
Menentukan kompromi antara kompleksitas model dan tingkat kesalahan. Nilai C besar membuat model berusaha menyesuaikan seluruh titik data (risiko *overfitting* meningkat), sedangkan nilai kecil membuat model lebih sederhana namun dapat mengurangi akurasi.
2. ε (*Epsilon-insensitive margin*)
Mengontrol lebar zona di sekitar *hyperplane* di mana kesalahan diabaikan. Nilai ε besar membuat model lebih toleran terhadap variasi kecil, sedangkan nilai kecil membuat model lebih sensitif terhadap fluktuasi data harga saham.
3. γ (*Gamma*)
Parameter pada *RBF kernel* yang menentukan jangkauan pengaruh setiap titik data. Nilai γ besar membuat model lebih fleksibel tetapi berpotensi *overfit*, sedangkan nilai kecil menghasilkan fungsi regresi yang lebih halus (*smooth*).

Dalam penelitian ini, parameter C , ε dan γ dioptimalkan menggunakan *TimeSeriesSplit* agar menghindari *look-ahead bias* dan prediksi stabil pada data *out-of-sample*.



han, proses pembentukan model SVR berfokus pada pencarian margin kesalahan minimum (ε - *insensitive margin*). Model ini akan menyeimbangkan antara kesederhanaan bentuk fungsi regresi dan menyesuaikan data. Fungsi kerugian memungkinkan model

mengabaikan kesalahan kecil yang dianggap sebagai *noise* pasar, sedangkan fungsi objektif menambahkan penalti untuk kesalahan yang melampaui batas ε . Melalui konsep *kernel trick* SVR juga dapat menangkap hubungan *non-linear* antar indikator teknikal seperti *SMA*, *EMA*, *RSI*, dan *MACD* terhadap *expected return* saham. Dengan mengoptimalkan parameter C , ε , dan γ , SVR mampu menghasilkan model yang akurat, stabil, dan adaptif terhadap dinamika harga saham yang volatil, tanpa kehilangan kemampuan generalisasi terhadap data baru.

1.4.11 Evaluasi Akurasi Model Prediksi

Evaluasi akurasi model merupakan tahap konseptual penting dalam proses peramalan (*forecasting*) berbasis *machine learning*. Menurut Hyndman dan Athanasopoulos (2021), tujuan utama evaluasi akurasi adalah menilai sejauh mana hasil prediksi (*predicted values*) mendekati nilai aktual (*actual values*), sehingga kualitas model dapat diukur secara objektif. Dalam konteks regresi deret waktu keuangan, pengukuran akurasi berperan untuk menentukan apakah model mampu menangkap dinamika pasar dengan baik sebelum hasil prediksinya digunakan pada tahap optimasi portofolio.

Terdapat beberapa ukuran kesalahan (*error metrics*) yang lazim digunakan untuk mengevaluasi performa model prediksi, antara lain *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Ketiga ukuran ini secara konseptual menggambarkan seberapa besar deviasi rata-rata antara nilai prediksi dan realisasi aktual. Secara matematis, masing-masing metrik dapat dirumuskan sebagai berikut:

$$MAE = \frac{1}{n} \sum_{t=1}^n |r_t - \hat{r}_t| \quad (28)$$

$$MSE = \frac{1}{n} \sum_{t=1}^n |r_t - \hat{r}_t|^2 \quad (29)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n |r_t - \hat{r}_t|^2} \quad (30)$$

Keterangan:

- r_t = *return* aktual saham pada periode ke- t
- $\hat{r}_t$ = *return* hasil prediksi model pada periode ke- t
- n = jumlah observasi



MAE mengukur rata-rata kesalahan absolut tanpa memperhatikan arah, sehingga lebih stabil terhadap data dengan fluktuasi tinggi. MSE mengukur kuadrat kesalahan dan memberikan bobot lebih besar terhadap *outlier*, menjadikannya sensitif terhadap *outlier*. Sementara itu, RMSE sama dengan MSE yang memiliki satuan yang sama dengan variabel target, sudah diinterpretasikan dalam konteks keuangan.

Dalam penelitian ini, konsep evaluasi akurasi model digunakan untuk menilai kinerja dua algoritma prediksi, yaitu XGBoost dan SVR, dalam memperkirakan *return* harian saham. Model dengan nilai RMSE, MSE, dan MAE yang lebih rendah dianggap memiliki kemampuan generalisasi yang lebih baik terhadap data baru (*out-of-sample*). Hasil prediksi yang akurat sangat penting karena nilai *expected return* yang dihasilkan ($\hat{\mu}$) akan menjadi dasar bagi pembentukan portofolio optimal pada tahap optimasi *mean-variance*. Dengan demikian, tahap evaluasi akurasi model berfungsi sebagai jembatan antara proses prediksi dan optimasi portofolio dalam kerangka penelitian ini.

1.4.12 Optimasi *Hyperparameter*

Dalam konteks pembelajaran mesin, *hyperparameter* merupakan parameter eksternal model yang mengatur perilaku proses pelatihan dan tidak dipelajari langsung dari data. Nilai-nilai ini harus ditentukan terlebih dahulu sebelum model dilatih (*training*). Optimasi *hyperparameter* berperan penting dalam meningkatkan akurasi, stabilitas, serta kemampuan generalisasi model, terutama pada data keuangan yang bersifat *non-linear* dan volatil. Pemilihan parameter yang tepat, seperti *learning rate*, *max depth*, dan *number of estimators* pada XGBoost, serta C , ε , dan γ pada SVR, menentukan sejauh mana model mampu menyeimbangkan antara bias dan *variance*. Kombinasi *hyperparameter* yang optimal akan menghasilkan prediksi *expected return* yang lebih andal dan stabil pada data *out-of-sample*.

Dalam penelitian ini, proses optimasi dilakukan menggunakan metode *grid search*, yaitu pendekatan pencarian menyeluruh (*exhaustive search*) yang menguji semua kombinasi nilai *hyperparameter* dari suatu himpunan nilai yang telah ditentukan sebelumnya. Setiap kombinasi parameter dievaluasi menggunakan metrik kinerja tertentu, seperti *Mean Squared Error* (MSE), untuk memilih konfigurasi terbaik yang menghasilkan kesalahan prediksi paling kecil. Secara umum, proses *grid search* dapat diformulasikan sebagai berikut:

$$\theta^* = \underset{\theta \in \Theta}{\operatorname{argmin}} \mathcal{L}(\theta) \quad (31)$$

Keterangan:

- θ = vektor *hyperparameter* yang akan diuji
 - Θ = himpunan semua kombinasi nilai *hyperparameter* yang akan dievaluasi
 - $\mathcal{L}(\theta)$ = fungsi kerugian (*loss function*) yang diukur dari hasil validasi model
 - θ^* = kombinasi *hyperparameter* yang menghasilkan nilai kerugian minimum
- Untuk meningkatkan akurasi hasil dan mencegah *overfitting*, setiap kombinasi parameter yang dihasilkan dari *grid search* dievaluasi dengan teknik validasi silang (*series cross-validation*). Berbeda dengan *cross-validation* acak, pendekatan ini mempertahankan urutan kronologis data agar tidak



terjadi *look-ahead bias*, yaitu kondisi ketika informasi masa depan secara tidak sengaja digunakan dalam proses pelatihan.

Salah satu skema validasi yang digunakan adalah *TimeSeriesSplit*, di mana data historis dibagi menjadi beberapa lipatan (*folds*) secara berurutan. Pada setiap iterasi, model dilatih menggunakan data masa lalu dan diuji pada data masa depan. Secara konsep, data dengan panjang T dapat ditulis sebagai:

$$D = \{(X_1, y_1), (X_2, y_2), \dots, (X_P, y_P)\}$$

Setiap lipatan ke- k terdiri atas:

$$Train\ set_k = \{(X_1, y_1), \dots, (X_t, y_t)\}$$

$$Test\ set_k = \{(X_{t+1}, y_{t+1}), \dots, (X_{t+h}, y_{t+h})\}$$

(32)

Keterangan:

- t = batas akhir data pelatihan pada lipatan ke- k
- h = panjang horizon data pengujian
- P = total observasi dalam dataset

Pendekatan ini memastikan model dilatih dengan data masa lalu dan dievaluasi pada data masa depan sesuai prinsip kausalitas. Cerqueira et al (2019) menunjukkan bahwa metode *cross-validation* berbasis waktu seperti ini menghasilkan estimasi performa model yang lebih realistis dan bebas *look-ahead bias* dibandingkan pembagian data secara acak pada deret waktu.

Secara keseluruhan, kombinasi antara *grid search* dan *cross-validation* berbasis waktu merupakan pendekatan yang efektif untuk menemukan parameter optimal pada model prediksi keuangan. *Grid search* menjamin eksplorasi menyeluruh terhadap ruang parameter, sedangkan *TimeSeriesSplit* menjaga urutan waktu agar hasil validasi mencerminkan kondisi pasar yang sebenarnya.

Dalam penelitian ini digunakan skema *TimeSeriesSplit* dengan 5 lipatan (5-fold). Jumlah lipatan ini dipilih sebagai kompromi antara akurasi evaluasi dan efisiensi komputasi, di mana jumlah lipatan yang lebih banyak memang dapat memberikan estimasi performa yang lebih stabil, namun juga meningkatkan waktu pelatihan secara signifikan. Dengan lima lipatan, model memperoleh cukup banyak variasi data pelatihan dan pengujian untuk mengukur konsistensi performa tanpa mengorbankan efisiensi. Pendekatan ini memastikan bahwa konfigurasi *hyperparameter* XGBoost dan SVR yang dihasilkan bersifat stabil, tidak bias, serta adaptif terhadap dinamika harga saham pada periode pengujian (*out-of-sample*).

1.4.13 Integrasi $\hat{\mu}$ Prediksi ke Optimasi Portofolio

Dalam teori optimasi portofolio modern, *expected return* yang digunakan pada model Markowitz dapat diperoleh dari berbagai pendekatan estimasi. Salah satu



ru dalam literatur keuangan kuantitatif adalah pemanfaatan *ring* untuk menghasilkan prediksi *return* masa depan, yang kan ke dalam kerangka *mean-variance*. Nilai prediksi *return* dari SVR dilambangkan dengan $\hat{\mu}$, yang berfungsi menggantikan μ sebagai estimasi *expected return* dalam proses optimasi

Pendekatan ini didasarkan pada asumsi bahwa model *machine learning* mampu menangkap hubungan *non-linear* antara variabel teknikal (seperti SMA, EMA, RSI, dan MACD) dengan pergerakan harga saham, sehingga $\hat{\mu}$ dapat merepresentasikan ekspektasi *return* yang lebih dinamis dan adaptif terhadap kondisi pasar. Dengan memasukkan $\hat{\mu}$ ke dalam fungsi tujuan *mean-variance*, proses optimasi portofolio tidak lagi bergantung pada data historis semata, melainkan pada estimasi prediktif yang diperoleh dari model pembelajaran mesin. Secara matematis, fungsi tujuan optimasi portofolio dengan *return* prediksi $\hat{\mu}$ dituliskan sebagai:

$$\max_w \frac{w^T(\hat{\mu} - R_f)}{\sqrt{w^T \Sigma w}}$$

dengan kendala:

$$\sum_{i=1}^n w_i = 1$$

$$0 \leq w_i \leq 1$$
(33)

dengan:

- $\hat{\mu}$ = vektor *return* prediksi dari model XGBoost atau SVR
- R_f = tingkat *return* bebas risiko harian
- Σ = matriks kovariansi *return* saham
- N = jumlah saham dalam portofolio
- w_i = proporsi bobot saham ke- i

Proses ini memungkinkan portofolio menyesuaikan diri terhadap dinamika pasar secara lebih responsif dibandingkan penggunaan rata-rata historis μ . Kendala pertama memastikan total bobot portofolio bernilai satu, sedangkan kendala kedua mencegah posisi *short selling*. Bobot optimal w^* yang diperoleh dari optimasi ini kemudian menjadi dasar untuk menghitung *return* dan risiko aktual pada tahap *backtesting*.

Integrasi $\hat{\mu}$ ke dalam kerangka optimasi portofolio memberikan dasar teoritis bagi penggunaan pendekatan data driven dalam pembentukan portofolio. Dengan menggantikan μ historis menggunakan estimasi prediksi $\hat{\mu}$, investor dapat mengantisipasi perubahan ekspektasi pasar secara lebih cepat dan membentuk portofolio dengan efisiensi yang lebih tinggi.

1.4.14 μ Clipping

1.4.14.1 Definisi

Dalam konteks optimasi portofolio berbasis *mean-variance*, kesalahan kecil dalam estimasi nilai *expected return* (μ) dapat menyebabkan bobot portofolio yang ekstrem



ini terutama terjadi ketika nilai μ terlalu besar atau terlalu kecil di normal historisnya. Untuk mengatasi permasalahan tersebut, *clipping*, yaitu membatasi (men-clip) nilai ($\hat{\mu}$) prediksi dalam dasarkan persentil statistik. Pendekatan ini bertujuan untuk outlier dan menjaga stabilitas numerik pada proses optimasi. μ akan sebagai:

$$\hat{\mu}_i^c = \min(\max(\hat{\mu}_i, P_{10}), P_{90}) \quad (34)$$

dengan:

- $\hat{\mu}_i^c$ = nilai $\hat{\mu}$ yang telah di-*clipping* untuk saham ke- i
- $\hat{\mu}_i$ = nilai prediksi *return* awal dari model
- P_{10}, P_{90} = batas bawah dan atas (persentil ke-10 dan ke-90)

Teknik ini mencegah *expected return* ekstrem dari mempengaruhi fungsi tujuan optimasi secara berlebihan, sehingga menghasilkan bobot portofolio yang lebih realistis.

1.4.14.2 Batas Persentil (p10-p90)

Pemilihan batas persentil pada proses μ *clipping* didasarkan pada pertimbangan statistik bahwa sebagian besar distribusi *return* saham berada di antara rentang persentil ke-10 hingga ke-90. Nilai di luar rentang tersebut umumnya merepresentasikan pergerakan ekstrem yang jarang terjadi dan berpotensi menimbulkan distorsi dalam perhitungan portofolio.

Penetapan batas *clipping* pada persentil ke-10 (P10) dan ke-90 (P90) dianggap sebagai kompromi yang seimbang antara penghilangan *outlier* dan pelestarian informasi relevan dalam data. Pendekatan ini serupa dengan konsep *winsorization* yang digunakan dalam statistik untuk mengontrol pengaruh data ekstrem tanpa menghapus observasi tersebut. Jiang et al. (2021) juga menyarankan penggunaan batas kuantil yang serupa untuk menjaga kestabilan estimasi *return* prediktif dalam portofolio berbasis *machine learning*.

Dalam penelitian ini, pembatasan dilakukan terhadap seluruh vektor $\hat{\mu}$ hasil prediksi XGBoost dan SVR sebelum digunakan dalam proses optimasi *mean-variance*. Dengan demikian, hanya nilai ($\hat{\mu}$) yang berada dalam rentang $[P_{10}, P_{90}]$ yang digunakan untuk menghitung fungsi objektif rasio Sharpe.

1.4.14.3 Dampak terhadap Stabilitas Bobot

Penerapan μ *clipping* memiliki dampak langsung terhadap hasil bobot portofolio optimal (w^*). Dengan membatasi nilai $\hat{\mu}$ ekstrem, hasil optimasi menjadi lebih stabil, portofolio cenderung memiliki bobot yang lebih seimbang antar aset, serta tingkat *turnover* yang lebih rendah pada proses *rolling optimization*. DeMiguel et al. (2009) menyimpulkan bahwa kesalahan dalam estimasi *expected return* dan *covariance matrix* dapat menghasilkan bobot yang sangat ekstrem dan performa *out-of-sample* yang buruk, sehingga pengendalian terhadap kesalahan estimasi menjadi aspek penting dalam optimasi portofolio.

Selain itu, pendekatan ini juga berkontribusi terhadap peningkatan nilai rasio Sharpe yang lebih konsisten antar periode. Chen, Zhang, dan Mehlawat (2021)



dan penerapan pembatasan prediksi ekstrem dalam model *mean-variance Optimization* menghasilkan portofolio dengan kinerja risiko yang lebih baik. μ *clipping* dapat dianggap sebagai regularisasi sederhana pada *expected return* dalam optimasi Markowitz. Dengan mengurangi pengaruh nilai ekstrem, fungsi objektif optimasi menjadi lebih stabil secara numerik.

dan menghasilkan bobot w^* yang realistis serta portofolio yang lebih reliabel pada data *out-of-sample*.

1.4.15 Backtesting dan Rebalancing Portofolio

1.4.15.1 Konsep dan Tujuan

Backtesting merupakan proses evaluasi performa suatu strategi investasi dengan menerapkan strategi tersebut pada data historis. Tujuannya adalah untuk menilai konsistensi, profitabilitas, dan tingkat risiko strategi sebelum diterapkan pada kondisi pasar aktual. Melalui *backtesting*, dapat dinilai sejauh mana pendekatan yang digunakan mampu menghasilkan kinerja yang stabil dan efisien pada berbagai kondisi pasar. Dengan demikian, *backtesting* menjadi alat penting untuk menguji validitas dan ketahanan strategi portofolio sebelum digunakan pada data *out-of-sample*.

Dalam implementasinya, *backtesting* harus menjaga urutan waktu (*temporal order*) agar hasil evaluasi tidak bias. Keputusan investasi pada periode t hanya boleh didasarkan pada informasi yang tersedia hingga periode tersebut, tanpa melibatkan data masa depan. Hal ini bertujuan untuk menghindari *look-ahead bias*, yaitu kondisi ketika strategi tampak menghasilkan kinerja lebih baik karena secara tidak sengaja menggunakan informasi yang belum tersedia pada saat pengambilan keputusan. Dengan menjaga prinsip kausalitas waktu, hasil *backtesting* dapat mencerminkan kinerja strategi yang realistis terhadap dinamika pasar yang sebenarnya.

Salah satu komponen penting dalam *backtesting* adalah proses *rebalancing portofolio*. Dalam teori manajemen portofolio *rebalancing* adalah proses penyesuaian kembali proporsi aset portofolio agar sesuai dengan bobot target yang telah ditetapkan sebelumnya. Tujuan utamanya adalah menjaga keseimbangan antara tingkat risiko dan *expected return* yang diharapkan, terutama setelah terjadi perubahan harga saham yang menyebabkan bobot aktual portofolio menyimpang dari bobot optimal (w^*). Melalui *rebalancing*, investor dapat memastikan bahwa portofolio tetap berada pada garis efisien sesuai dengan prinsip *mean-variance optimization*.

Terdapat beberapa pendekatan dalam menentukan frekuensi *rebalancing*, dua di antaranya adalah harian dan bulanan. *Rebalancing* harian dilakukan untuk merespons perubahan pasar dengan cepat, tetapi memiliki kelemahan berupa biaya transaksi yang tinggi akibat frekuensi penyesuaian yang lebih sering. Sebaliknya, *rebalancing* bulanan cenderung lebih efisien secara biaya, walaupun memiliki risiko keterlambatan dalam merespons perubahan kondisi pasar.

Penelitian Alfath dan Dewi (2022) pada portofolio saham indeks Kompas100 di masa COVID-19 menunjukkan bahwa *rebalancing* bulanan pada Kompas100 menghasilkan kinerja yang lebih optimal selama periode tersebut. Penelitian ini menyeimbangkan antara efisiensi biaya dan adaptivitas terhadap perubahan pasar. Berdasarkan temuan tersebut, penelitian ini menggunakan *rebalancing* bulanan untuk menjaga keseimbangan antara efisiensi biaya dan kemampuan portofolio merespons perubahan pasar.



Dengan demikian, *backtesting* dan *rebalancing* dalam penelitian ini dipandang sebagai satu kesatuan proses yang saling melengkapi. *Backtesting* berfungsi mengevaluasi kinerja strategi secara historis, sedangkan *rebalancing* memastikan strategi tetap adaptif terhadap perubahan risiko dan *return* dari waktu ke waktu. Kombinasi keduanya memberikan kerangka yang menyeluruh untuk menilai konsistensi, stabilitas, dan efisiensi strategi portofolio berbasis *machine learning* pada pasar saham Indonesia.

1.4.15.2 Metode *Walk-Forward* dan *Out-of-Fold Prediction*

Walk-forward analysis merupakan pendekatan konseptual dalam teori evaluasi strategi investasi yang menekankan pentingnya urutan waktu (*temporal dependency*) dalam data keuangan. Konsep ini berangkat dari prinsip bahwa setiap keputusan investasi harus didasarkan hanya pada informasi yang tersedia hingga periode pengambilan keputusan, tanpa melibatkan data masa depan. Dengan demikian, *walk-forward analysis* menjadi kerangka evaluasi yang bebas dari *look-ahead bias* dan lebih representatif terhadap kondisi pasar yang sesungguhnya.

Pendekatan ini menggunakan kerangka *rolling window*, di mana proses pembelajaran, estimasi parameter, dan evaluasi kinerja dilakukan secara berurutan dalam dimensi waktu. Pada setiap jendela observasi, model membentuk estimasi ekspektasi *return* dan risiko berdasarkan data historis hingga periode t , kemudian hasil estimasi tersebut digunakan untuk menilai performa portofolio pada periode berikutnya ($t + 1$). Setelah itu, jendela waktu diperbarui dengan menambahkan data terbaru dan menghapus data terlama, menciptakan mekanisme evaluasi yang berkelanjutan dan adaptif terhadap perubahan kondisi pasar.

Pendekatan ini didasari oleh pemahaman bahwa pasar keuangan bersifat *non-stasioner*, di mana distribusi *return* dan volatilitas dapat berubah dari waktu ke waktu. Oleh karena itu, strategi portofolio dan model prediksi perlu diuji dengan mempertahankan struktur temporal agar hasilnya merefleksikan kinerja yang realistis dan kontekstual terhadap dinamika pasar. Chen et al (2022) menjelaskan bahwa pendekatan berbasis *rolling window* memberikan kerangka evaluasi yang realistis terhadap strategi portofolio, karena mempertahankan urutan temporal data dan memungkinkan penilaian kinerja secara berurutan sebagaimana praktik investasi aktual. Sementara itu, Guo dan Ryan (2023) menekankan pentingnya pembaruan portofolio secara dinamis untuk mencapai efisiensi terhadap perubahan risiko dan peluang pasar. Pandangan ini sejalan dengan prinsip dasar *walk-forward analysis*, di mana proses penilaian dan penyesuaian strategi dilakukan secara berkelanjutan mengikuti dinamika pasar.



look-ahead bias pada pemodelan berbasis *Machine Learning*, dan *Out-of-Fold (OOF) prediction*. OOF merupakan teknik yang menggunakan data pelatihan hanya pada data hingga periode t untuk menghasilkan prediksi untuk periode t . Proses ini dilakukan secara iteratif menggunakan skema *walk-forward*, sehingga setiap prediksi bersifat *sample*. Deret prediksi OOF ini memastikan bahwa estimasi yang digunakan dalam optimasi portofolio bebas dari kebocoran

data. Dengan demikian, hasil *backtesting* dapat mencerminkan kondisi operasi strategi yang lebih realistis.

Dengan demikian, *walk-forward framework* dan pendekatan OOF berfungsi sebagai fondasi metodologis yang menjaga integritas temporal, memastikan bahwa seluruh proses prediksi, estimasi, dan evaluasi kinerja dilakukan tanpa melanggar kausalitas waktu.

1.4.15.3 Rolling Retrain dan Pembaruan Bobot

Pendekatan *rolling retrain* digunakan untuk memastikan bahwa model prediksi dan estimasi risiko diperbarui secara dinamis sepanjang periode penelitian. Konsep ini sejalan dengan prinsip *rolling rebalancing* yang dijelaskan oleh Guo dan Ryan (2023), di mana estimasi *return* dan risiko diperbarui secara bergulir agar portofolio tetap adaptif terhadap dinamika pasar. Pendekatan ini memungkinkan model *machine learning* beradaptasi dengan kondisi pasar terkini tanpa melanggar prinsip kausalitas, karena hanya data hingga periode t yang digunakan untuk memprediksi *return* periode $t + 1$. Dengan demikian, risiko *look-ahead bias* dapat dihindari.

Prosedur *rolling retrain* dan *rebalancing* dapat dijelaskan melalui langkah-langkah berikut:

1. Mengumpulkan data harga penutupan harian hingga akhir bulan t untuk seluruh saham dalam portofolio.
2. Melatih ulang model XGBoost dan SVR menggunakan data hingga periode t untuk menghasilkan prediksi $\hat{\mu}$ pada periode $t + 1$.
3. Mengestimasi matriks kovariansi Σ dengan *rolling window*.
4. Melakukan optimasi portofolio menggunakan $\hat{\mu}$ dan Σ ke dalam fungsi objektif *mean-variance* untuk memperoleh bobot optimal (w^*).
5. Menahan bobot tersebut hingga periode *rebalancing* berikutnya, lalu mengulangi proses yang sama.

Pendekatan *rolling retrain* ini sejalan dengan prinsip *walk-forward validation*. Cerqueira et al (2020) menjelaskan bahwa metode evaluasi berbasis urutan temporal seperti ini penting untuk mengatasi karakteristik data yang bersifat *non-stasioner*, sehingga hasil prediksi lebih representatif terhadap kondisi pasar yang dinamis.

Dalam konteks pasar Indonesia, hasil penelitian Alfath dan Dewi (2022) menunjukkan bahwa strategi *rebalancing* berkala dengan frekuensi bulanan memberikan kinerja portofolio yang lebih stabil dan efisien dibandingkan frekuensi penyesuaian yang lebih tinggi. Hal itu memperkuat gagasan bahwa pembaruan model dan penyesuaian bobot portofolio secara periodik dapat menjaga keseimbangan antara akurasi prediksi, stabilitas risiko, dan efisiensi biaya transaksi.



ombinasi antara *rolling retrain* dan *rebalancing* bulanan dalam i dasar teoritis yang kuat serta relevan untuk diterapkan pada sia.

Biaya Transaksi dalam Rebalancing

Transaction cost) merupakan komponen penting yang muncul setiap ukan penyesuaian bobot atau *rebalancing*. Biaya transaksi

mencakup seluruh biaya yang timbul dari aktivitas perdagangan, seperti *bid-ask spread*, komisi broker, *slippage*, serta biaya implisit akibat pergerakan harga saat eksekusi. Dalam konteks manajemen portofolio, biaya transaksi menjadi faktor pembatas frekuensi *rebalancing* karena semakin sering portofolio disesuaikan, semakin besar biaya yang harus ditanggung investor. Strategi dengan turnover tinggi cenderung menghasilkan kinerja *out-of-sample* yang lebih rendah setelah memperhitungkan biaya transaksi.

Banyak penelitian menggunakan pendekatan biaya transaksi proporsional dalam rentang 5 hingga 20 basis poin (bps) untuk mensimulasikan kondisi pasar yang realistis, terutama pada aset dengan tingkat likuiditas tinggi. Dalam penelitian ini digunakan biaya transaksi sebesar 10 bps, karena nilai tersebut merepresentasikan kisaran rata-rata *bid-ask spread* dan komisi broker di Bursa Efek Indonesia untuk saham-saham berkapitalisasi besar seperti anggota indeks LQ45. Nilai ini juga konsisten dengan praktik empiris dalam literatur yang menilai 10 bps sebagai asumsi moderat antara efisiensi biaya dan realisme pasar.

Penambahan biaya transaksi ini penting agar evaluasi strategi portofolio mencerminkan kinerja *riil* dan tidak bias terhadap strategi yang terlalu agresif dalam melakukan *rebalancing*. Dengan demikian, pemilihan frekuensi *rebalancing* tidak hanya dipengaruhi oleh kebutuhan adaptasi terhadap dinamika pasar, tetapi juga oleh pertimbangan biaya transaksi yang dapat mengurangi nilai akhir portofolio. Strategi yang mampu menyeimbangkan antara stabilitas bobot, akurasi sinyal, dan efisiensi biaya transaksi akan memberikan hasil yang lebih optimal dalam jangka panjang.

1.4.16 Evaluasi Kinerja Portofolio

Evaluasi kinerja portofolio merupakan tahapan penting dalam teori manajemen investasi yang bertujuan untuk menilai efektivitas strategi portofolio dalam menghasilkan *return* dibandingkan dengan risiko yang ditanggung. Menurut Bodie, Kane, dan Marcus (2021), pengukuran kinerja portofolio tidak hanya berfokus pada tingkat keuntungan, tetapi juga pada efisiensi risiko-*return* yang dicapai selama periode investasi. Dengan demikian, evaluasi dilakukan untuk mengetahui sejauh mana suatu strategi portofolio mampu memberikan kombinasi optimal antara tingkat pengembalian dan risiko.

Dalam penelitian ini, evaluasi kinerja digunakan untuk membandingkan empat jenis portofolio, yaitu *equal-weight*, *mean-return*, serta dua portofolio prediktif berbasis XGBoost dan SVR. Perbandingan ini bertujuan untuk menilai apakah integrasi hasil prediksi model *machine learning* ($\hat{\mu}$) ke dalam optimasi *mean-variance* kinerja yang lebih baik dibandingkan pendekatan konvensional.



Portofolio

representasikan tingkat keuntungan yang diperoleh dari suatu periode tertentu. Nilai ini diperoleh dari penjumlahan tertimbang aset dengan bobot portofolio yang bersangkutan (Markowitz,

$$R_{p,t} = \sum_{i=1}^N w_i R_{i(t)} \quad (35)$$

dengan:

$R_{p,t}$	= <i>return</i> portofolio pada hari ke- t
$R_{i(t)}$	= <i>return</i> saham ke- i pada hari ke- t
w_i	= bobot saham ke- i
N	= jumlah saham dalam portofolio

Nilai *return* berfungsi sebagai indikator utama hasil investasi dan mencerminkan kemampuan portofolio dalam menghasilkan keuntungan dari aset-aset yang dimiliki.

1.4.16.2 Risiko Portofolio

Risiko menunjukkan tingkat ketidakpastian dari hasil investasi dan merupakan aspek penting dalam teori portofolio modern. Risiko portofolio diukur menggunakan simpangan baku dari *return* portofolio, yang dihitung berdasarkan matriks kovariansi antar aset seperti persamaan (5). Semakin kecil nilai standar deviasi portofolio (ukuran risiko total), semakin stabil kinerja portofolio terhadap perubahan pasar. Oleh karena itu, *return* dan risiko aktual menjadi indikator utama dalam menilai konsistensi serta stabilitas strategi portofolio.

1.4.16.3 Rasio Sharpe

Rasio Sharpe digunakan untuk mengukur efisiensi antara risiko dan hasil yang diperoleh, yang dikembangkan oleh William F. Sharpe (1966). Rasio ini menilai kelebihan *return* portofolio dibandingkan tingkat bebas risiko dengan memperhitungkan volatilitas portofolio sebagai ukuran risiko. Nilai rasio Sharpe yang tinggi menunjukkan bahwa portofolio mampu menghasilkan kelebihan *return* yang lebih besar untuk setiap satuan risiko yang diambil. Fabozzi (2020) menegaskan bahwa rasio Sharpe merupakan alat evaluasi paling komprehensif untuk menilai efisiensi portofolio, karena memperhitungkan secara simultan hubungan antara keuntungan dan volatilitas. Secara matematis, *Sharpe ratio* dirumuskan sebagai:

$$SR = \frac{E(R_p) - R_f}{\sigma_p} \quad (37)$$

Keterangan:

SR	= <i>Sharpe ratio</i>
$E(R_p)$	= <i>return</i> harapan portofolio
σ_p	= standar deviasi <i>return</i> portofolio



o yang lebih tinggi menunjukkan bahwa portofolio memberikan i baik per unit risiko yang diambil. Rasio Sharpe menggunakan gai pembanding dalam perhitungan *excess return*.

te Harian

merupakan tingkat pengembalian bebas risiko yang digunakan banding dalam perhitungan *excess return* pada rasio Sharpe.

Dalam konteks pasar Indonesia, *risk-free rate* umumnya diwakili oleh BI Rate sebagai suku bunga acuan yang merepresentasikan pengembalian instrumen berisiko sangat rendah seperti Sertifikat Bank Indonesia (SBI). Penggunaan BI Rate memastikan bahwa evaluasi kinerja portofolio sesuai dengan kondisi pasar domestik, karena analisis *return* portofolio dalam penelitian ini dilakukan pada frekuensi harian, BI Rate perlu dikonversi menjadi *risk-free rate* harian. Konversi dilakukan menggunakan persamaan:

$$r_{f,harian} = (1 + r_{f,tahunan})^{\frac{1}{252}} - 1 \quad (38)$$

dengan asumsi terdapat 252 hari perdagangan dalam satu tahun. Konversi ini penting agar perhitungan *excess return* dan rasio Sharpe konsisten dengan frekuensi data yang digunakan. Dengan demikian, penggunaan *risk-free rate* harian memberikan dasar evaluasi yang lebih akurat dalam mengukur kinerja portofolio yang disesuaikan dengan risiko.

1.4.16.4 Compound Annual Growth Rate (CAGR)

CAGR mengukur tingkat pertumbuhan tahunan gabungan portofolio selama periode pengujian, dihitung berdasarkan perubahan *Net Asset Value* (NAV) dari awal hingga akhir periode investasi. Metrik ini menunjukkan tingkat pertumbuhan rata-rata per tahun jika investasi diasumsikan mengalami pertumbuhan majemuk secara konstan.

$$CAGR = \left(\frac{NAV_n}{NAV_0} \right)^{1/n} - 1 \quad (39)$$

dengan:

- NAV_n = nilai aktiva bersih portofolio pada akhir periode investasi
- NAV_0 = nilai aktiva bersih portofolio pada awal periode investasi
- n = jumlah tahun periode observasi

1.4.16.5 Volatilitas

Volatilitas tahunan menggambarkan tingkat penyimpangan atau fluktuasi *return* portofolio yang ditahunkan, sehingga berfungsi sebagai ukuran risiko total yang sejalan dengan simpangan baku harian. Nilai volatilitas yang rendah menunjukkan stabilitas kinerja portofolio terhadap perubahan pasar.

$$\sigma_{p,annual} = \sigma_p \times \sqrt{252} \quad (40)$$

dengan:

- $\sigma_{p,annual}$ = volatilitas tahunan portofolio
- σ_p = simpangan baku (standar deviasi) harian portofolio
- 252 = hari perdagangan dalam satu tahun bursa



Drawdown (MaxDD)

Penurunan maksimum nilai portofolio dari titik puncak ke titik nadir selama periode observasi. Metrik ini memberikan informasi mengenai risiko

kerugian ekstrem yang mungkin dialami portofolio, yang sering kali tidak tercermin hanya dari nilai simpangan baku.

1.4.16.7 Turnover Portofolio

Turnover portofolio merupakan ukuran seberapa besar perubahan komposisi bobot aset yang terjadi pada setiap periode *rebalancing*. Nilai turnover yang tinggi menunjukkan bahwa portofolio melakukan penyesuaian bobot secara agresif, sehingga aktivitas perdagangan menjadi lebih sering. Sebaliknya, turnover rendah menandakan bahwa portofolio relatif stabil dari waktu ke waktu, seperti yang terjadi pada strategi *equal-weight* atau *mean-return* konvensional. Turnover juga mencerminkan stabilitas bobot portofolio. Strategi dengan bobot yang terlalu sering berubah dapat mengalami ketidakstabilan performa karena sensitif terhadap fluktuasi jangka pendek. Sebaliknya, strategi dengan turnover moderat atau rendah umumnya lebih konsisten dalam jangka menengah dan panjang.

Evaluasi kinerja portofolio juga perlu mempertimbangkan dampak biaya transaksi. Biaya transaksi dapat menurunkan *return* bersih portofolio secara signifikan, terutama pada strategi dengan tingkat turnover tinggi. Oleh karena itu, pengukuran kinerja seperti *return* tahunan, *Sharpe ratio*, dan CAGR umumnya dihitung setelah biaya transaksi dikurangkan untuk mencerminkan kondisi pasar yang sesungguhnya. Dengan memasukkan biaya transaksi dalam evaluasi, penilaian performa strategi portofolio menjadi lebih objektif dan *reliable*, serta mampu memberikan gambaran yang lebih akurat mengenai efisiensi keputusan investasi.

Hubungan antar ukuran evaluasi bersifat hierarkis. *Return* mencerminkan hasil investasi, risiko portofolio menunjukkan tingkat ketidakpastian, sedangkan *rasio Sharpe* menggabungkan keduanya menjadi indikator efisiensi portofolio secara menyeluruh. CAGR dan *Maximum Drawdown* memperluas perspektif evaluasi dengan menambahkan aspek pertumbuhan jangka panjang dan ketahanan portofolio terhadap penurunan nilai ekstrem. Dengan demikian, kombinasi metrik *return*, risiko, *Sharpe ratio*, CAGR, *volatilitas tahunan*, dan *Maximum Drawdown* memberikan penilaian menyeluruh terhadap kinerja portofolio baik dari sisi hasil, risiko, maupun kestabilan pertumbuhan.



BAB II METODE PENELITIAN

2.1 Jenis dan Pendekatan Penelitian

Penelitian ini merupakan penelitian kuantitatif yang bertujuan untuk menganalisis dan menguji kinerja model *machine learning* dalam memprediksi *return* saham serta menerapkan hasil prediksi tersebut pada proses optimasi portofolio berbasis teori *mean-variance* Markowitz. Pendekatan yang digunakan bersifat deskriptif kuantitatif dengan analisis empiris, di mana hasil perhitungan dan simulasi dilakukan secara komputasional menggunakan bahasa pemrograman Python. Penelitian ini berfokus pada pembentukan portofolio optimal dengan mempertimbangkan *return*, risiko, dan efisiensi risiko-*return* (rasio Sharpe).

2.2 Lokasi dan Waktu Penelitian

Penelitian ini dilaksanakan di Ruang Belajar Ilmu Aktuaria, Program Studi Ilmu Aktuaria, Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin. Pelaksanaan penelitian berlangsung selama Agustus 2025 hingga Oktober 2025, mencakup seluruh tahapan pengumpulan data, pemrosesan, pelatihan model, optimasi portofolio, serta evaluasi hasil. Data penelitian diambil melalui *Yahoo Finance* pada tautan <https://finance.yahoo.com>.

2.3 Jenis dan Sumber Data

Jenis data yang digunakan adalah data sekunder berupa harga penutupan harian (*closing price*) dan volume perdagangan saham, yang diperoleh dari platform *Yahoo Finance* dengan periode observasi 1 Januari 2022 hingga 31 Desember 2024. Selain data harga saham, digunakan pula data suku bunga bebas risiko (*risk-free rate*) yang bersumber dari Bank Indonesia melalui suku bunga acuan *BI 7-Day Reverse Repo Rate* (BI7DRR), yang kemudian dikonversi ke tingkat harian untuk keperluan perhitungan rasio Sharpe pada tahap optimasi portofolio.

2.4 Populasi dan Sampel Penelitian

Populasi penelitian mencakup seluruh saham yang terdaftar di Bursa Efek Indonesia selama periode 2022–2024. Sampel penelitian ditentukan menggunakan *purposive sampling*, yaitu pemilihan saham berdasarkan kriteria tertentu, meliputi:

1. Saham aktif diperdagangkan selama periode penelitian.
2. Memiliki data harga penutupan dan volume perdagangan harian yang



dalam kategori saham dengan tingkat likuiditas dan kapitalisasi

2.5 Tahapan Penelitian

Berikut ini adalah langkah-langkah analisis yang dilakukan dalam penelitian ini:

1. Mengunduh data harga penutupan (*close price*) dan volume perdagangan 15 saham yang aktif dan konsisten terdaftar di Bursa Efek Indonesia periode 1 Januari 2022 hingga 31 Desember 2024 melalui *Yahoo Finance* menggunakan pustaka *yfinance*.
2. Membersihkan data dari nilai kosong (*missing values*) dan data duplikat, serta memastikan kesesuaian tanggal antar saham.
3. Menghitung *daily return* untuk setiap saham.
4. Menghitung indikator teknikal (*feature engineering*) berupa SMA5, SMA20, EMA12, EMA26, RSI4, MACD, Signal MACD, *Bollinger Upper*, dan *Bollinger Lower* untuk menggambarkan pola harga jangka pendek dan menengah.
5. Membentuk dataset untuk setiap saham, di mana variabel X berisi fitur teknikal dan variabel Y berisi *return* saham $t + 1$ (*next-day return*).
6. Melakukan *scaling* terhadap seluruh fitur menggunakan *StandardScaler* khusus untuk model SVR agar setiap fitur berada dalam skala yang seragam dan meningkatkan stabilitas perhitungan *kernel*.
7. Melatih dua model *machine learning*, yaitu *Extreme Gradient Boosting* (XGBoost) dan *Support Vector Regression* (SVR), dengan skema *TimeSeriesSplit* untuk menjaga urutan waktu dan mencegah *look-ahead bias*.
8. Menentukan parameter terbaik dari masing-masing model melalui *Grid Search* dengan metrik evaluasi *Mean Squared Error* (MSE) menggunakan pendekatan *time-series cross-validation*.
9. Melakukan prediksi *next-day return* ($\hat{\mu}$) untuk setiap saham berdasarkan model terbaik.
10. Menerapkan μ *clipping* dengan batas persentil ke-10 dan ke-90 untuk menghindari nilai ekstrem pada hasil prediksi yang dapat menyebabkan ketidakstabilan bobot portofolio.
11. Menghitung matriks kovariansi antar saham (Σ) menggunakan metode *rolling covariance* dengan jendela waktu 126 hari untuk memperbarui estimasi risiko secara dinamis.
12. Melakukan optimasi portofolio berdasarkan model *mean-variance* Markowitz dengan tujuan memaksimalkan rasio Sharpe.
13. Menggunakan tingkat bebas risiko (*risk-free rate*) yang diperoleh dari Bank Indonesia 7-Day Reverse Repo Rate (BI7DRR) dan dikonversi ke bentuk harian.



monthly rebalancing terhadap bobot portofolio serta melatih (*rolling retrain*) setiap akhir bulan menggunakan data terbaru tersebut.

biaya transaksi sebesar 10 basis poin pada setiap proses untuk mencerminkan kondisi pasar yang realistis.

16. Melakukan *backtesting* secara *out-of-sample* untuk mengukur kinerja portofolio berdasarkan *return aktual rata-rata*, risiko (volatilitas), dan rasio Sharpe.
17. Membandingkan hasil empat strategi portofolio, yaitu *equal-weight*, *mean-return historis*, *XGBoost-based*, dan *SVR-based*, untuk menilai efektivitas pendekatan *machine learning* dalam membentuk portofolio optimal.

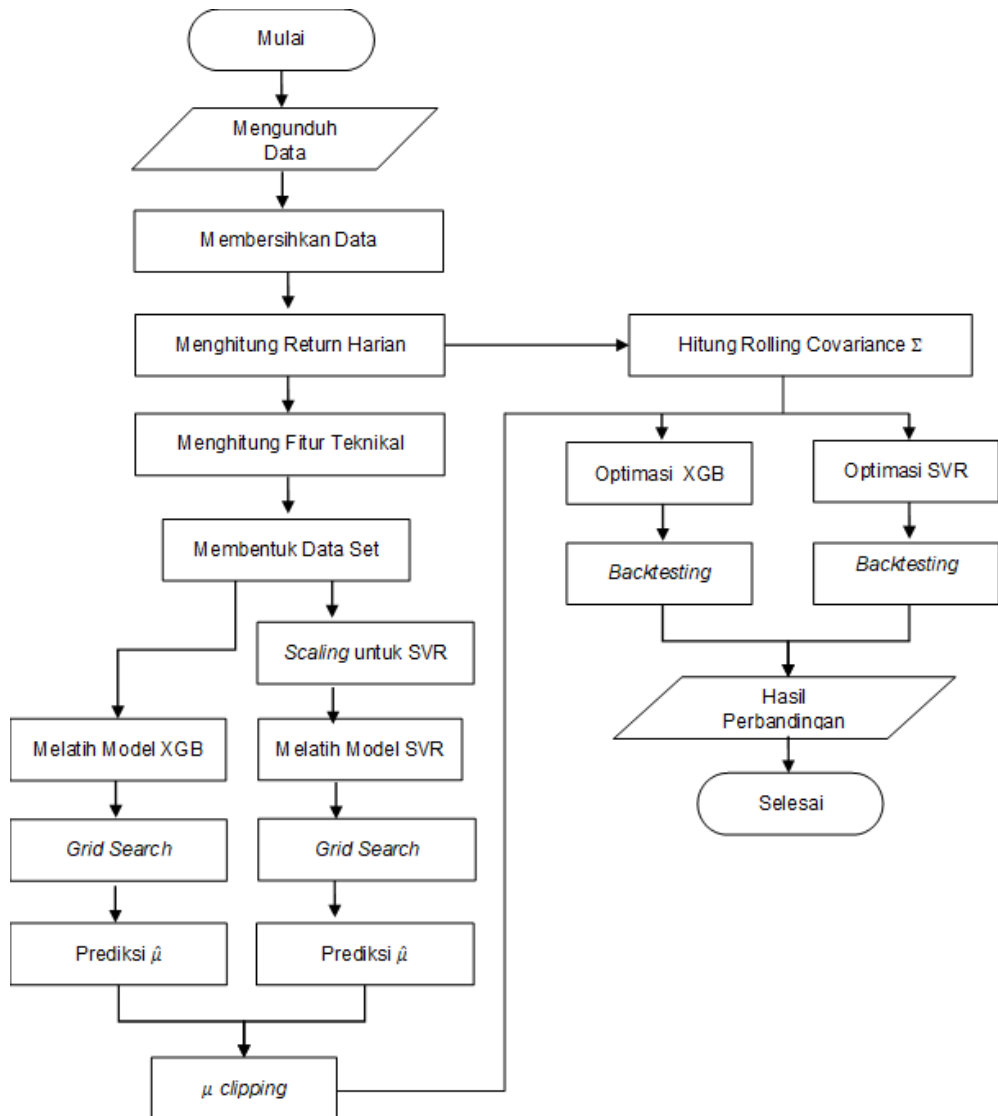
2.6 Metode Analisis Data

Metode analisis yang digunakan terbagi menjadi dua tahap utama, yaitu:

1. Analisis Prediksi *Return* Saham,
Tahap ini bertujuan menghasilkan estimasi *expected return* ($\hat{\mu}$) per saham menggunakan algoritma XGBoost dan SVR.
 - Data diubah menjadi format deret waktu (*time series*) dan dinormalisasi.
 - Model dilatih menggunakan data historis hingga periode ke $-t$, lalu diuji untuk memprediksi periode ke $-t + 1$.
 - Pengujian dilakukan secara bergulir (*rolling window*).
2. Analisis Optimasi Portofolio
Tahap ini menggunakan hasil prediksi ($\hat{\mu}$) dan matriks kovariansi Σ untuk menentukan bobot optimal (w^*) melalui fungsi maksimisasi rasio Sharpe. Hasil dari tahap ini adalah bobot portofolio optimal untuk masing-masing strategi, yang kemudian digunakan dalam perhitungan *return*, risiko, dan rasio Sharpe secara *out-of-sample*.



2.7 Alur Kerja



Gambar 1. Alur Kerja

