BAB I

PENDAHULUAN

1.1. Latar Belakang

Salah satu kegiatan yang dilakukan untuk memperoleh keuntungan dalam bidang perekonomian saat ini adalah investasi saham. Bursa Efek Indonesia (BEI) menurut Kementerian Keuangan Republik Indonesia merupakan lembaga penyedia sarana pertemuan aktivitas penawaran jual dan beli Efek berbagai pihak dengan tujuan memperdagangkan Efek diantara mereka. Salah satu contoh Efek adalah saham. Menurut Tandelilin (2017) investasi merupakan komitmen untuk mengalokasikan sejumlah modal saat ini dengan harapan akan memperoleh keuntungan di masa mendatang. Memperoleh *return* atau keuntungan yang maksimal merupakan tujuan utama investor melakukan investasi. Sektor bahan baku atau *basic materials* merupakan salah satu sektor dengan jumlah perusahaan tercatat paling banyak pada BEI. Sektor barang baku menurut Umam (2024) dan Ariantini (2023) merupakan sektor yang rentan terhadap fluktuasi harga komoditas, siklus bisnis global, kebijakan pemerintah, dan dinamika makroekonomi. Perubahan harga pada sektor ini akan mempengaruhi harga produk dari sektor lain yang menggunakan produk dari sektor bahan baku. Menunjukkan bahwa sektor bahan baku adalah salah satu tempat berinvestasi yang menjanjikan.

Laporan keuangan menurut Rustiana et al. (2022) bertujuan untuk melaporkan kinerja keuangan dan investasi pada waktu tertentu dan memberikan informasi terkait perubahan kondisi keuangan perusahaan. Analisis laporan keuangan disebut dengan analisis fundamental. Dahlquist & Knight (2022) menyebutkan bahwa dengan mengonversi informasi dalam laporan keuangan menjadi beberapa bentuk rasio keuangan yakni rasio aktivitas, likuiditas, profitabilitas, solvabilitas, dan penilaian pasar dapat membantu dalam pengambilan keputusan bisnis termasuk terkait investasi di masa depan karena hasil dari analisis rasio dapat digunakan untuk menganalisis tren kerja, menetapkan tolak ukur keberhasilan, menetapkan ekspektasi anggaran, dan membandingkan pesaing.

Disebutkan oleh Alfindy (2019) bahwa inflasi, nilai tukar, dan tingkat suku bunga merupakan faktor makroekonomi yang berpengaruh terhadap harga saham dan juga *return* yang diterima oleh investor, yakni bahwa semakin tinggi harga saham maka *return* yang dihasilkan juga akan tinggi. Karena adanya pengaruh dari faktor makroekonomi dan rasio keuangan yang sangat berdampak terhadap *return* saham pada perusahaan sektor bahan baku, kedua hal ini menjadi sesuatu yang harus dianalisis secara mendalam oleh para investor sebelum memutuskan untuk melakukan investasi.

Berbagai penelitian sebelumnya telah memanfaatkan faktor makroekonomi dan rasio keuangan sebagai prediktor dalam pemodelan *return* saham. Penelitian oleh Dewi et al. (2020) menggunakan rasio aktivitas yang diwakili oleh TATO (*Total Assets*

Turnover), likuiditas oleh CR *(Current Ratio)*, profitabilitas oleh ROE *(Return On Equity)*, solvabilitas oleh DER *(Debt to Equity Ratio)*, dan penilaian pasar oleh PER *(Price Earning Ratio)* untuk memodelkan *return* saham, dengan hanya rasio aktivitas TATO yang signifikan. Ramli (2021) melakukan analisis pemodelan *return* saham dengan prediktor rasio aktivitas yang diwakilkan oleh TATO, likuiditas oleh CR, profitabilitas oleh ROA, solvabilitas oleh DER, inflasi, suku bunga, dan nilai tukar, dengan hanya rasio likuiditas CR dan suku bunga yang signifikan. Sasmito (2023) mencari pengaruh rasio aktivitas yang diwakilkan oleh FATO *(Fixed On Assets)*, likuiditas oleh CR, profitabilitas oleh ROA *(Return On Assets)*, solvabilitas oleh DAR *(Debt to Total Assets),* dan penilaian pasar oleh PER *(Price Earning Ratio)*, inflasi, dan suku bunga terhadap *return* saham dengan hanya rasio profitabilitas ROA, penilaian pasar PER, dan suku bunga yang signifikan. Septiana et al. (2024) mencari implikasi faktor makroekonomi inflasi, nilai kurs, dan BI-Rate terhadap *return* saham dengan hanya nilai kurs yang signifikan. Penelitian-penelitian ini menunjukkan bahwa kedua kelompok prediktor ini secara umum relevan dalam pemodelan *return* saham.

Pemodelan dilakukan dengan metode analisis regresi. Model regresi yang melibatkan lebih dari satu variabel independen disebut regresi linear berganda. Syarat dalam menggunakan persamaan regresi linear berganda adalah model memenuhi uji asumsi klasik yang mencakup uji normalitas, multikolinearitas, autokorelasi dan heteroskedastisitas (Indartini & Mutmainah, 2024). Hasil pengujian menunjukkan syarat multikolinearitas tidak terpenuhi. Salah satu cara yang dapat digunakan untuk mengurangi dampak dari multikolinearitas adalah *model respecification* atau spesifikasi ulang model dengan pendekatan paling umum dilakukan adalah eliminasi variabel, dan pendekatan ini seringkali merupakan teknik yang efektif (Montgomery et al., 2012). Sejumlah jurnal penelitian seperti yang dihasilkan oleh Sulistianingsih et al. (2022), Hapsery & Lubis (2019), dan Subekti (2015) menyebutkan dan menggunakan regresi *stepwise* sebagai salah satu metode untuk menangani data dengan masalah multikolinearitas, data pengamatan adalah Indeks Pembangunan Manusia, *Track Quality Index*, dan tingkat pendidikan terhadap pengangguran. Selain itu jurnal penelitian oleh Kusuma & Wulansari (2019), Rahmawati & Suratman (2022), dan Nur et al. (2024) juga menyatakan dan menggunakan LASSO *(Least Absolute Shrinkage and Selection Operator)* sebagai salah satu metode yang dapat diterapkan pada data yang mengalami multikolinearitas sebagai tindakan penanganan, dengan data pengamatan adalah kemiskinan dan kerentanan kemiskinan, *infant mortality*, dan *college dataset* R Studio. Penggunaan data *return* saham dan keuangan dari sektor bahan baku yang meliputi rasio keuangan lengkap dan faktor makroekonomi juga masih sangat jarang digunakan untuk pemodelan *stepwise* dan LASSO. Serta kurangnya analisis *return* saham pada sektor bahan baku mendorong penulis untuk melakukan penelitian dengan judul **"Perbandingan Kinerja Regresi *Stepwise* Dan Lasso Dalam Pemodelan *Return* Saham Pada Sektor Bahan Baku BEI".**

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang, maka rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Faktor makroekonomi dan rasio keuangan apa yang terpilih sebagai prediktor utama dalam model *return* saham berdasarkan metode regresi *stepwise* dan LASSO?
2. Metode mana yang memberikan hasil terbaik berdasarkan nilai MSE *(Mean Squared Error)* dalam memodelkan variabel antara regresi *stepwise* dan LASSO?

1.3. Tujuan Penelitian

Berdasarkan uraian rumusan masalah, maka tujuan penelitian adalah sebagai berikut:

1. Untuk mengetahui faktor makroekonomi dan rasio keuangan apa yang terpilih sebagai prediktor utama dalam model *return* saham berdasarkan metode regresi *stepwise* dan LASSO.
2. Untuk mengetahui metode mana yang memberikan hasil terbaik berdasarkan nilai MSE *(Mean Squared Error)* dalam memodelkan variabel antara regresi *stepwise* dan LASSO.

1.4. Manfaat Penelitian

Penelitian dilakukan untuk mengembangkan literatur terkait pemodelan *return* saham dengan prediktor faktor makroekonomi dan rasio keuangan menggunakan metode seleksi variabel *stepwise* dan LASSO. Serta sebagai bahan rujukan bagi penelitian selanjutnya yang terkait dengan metode seleksi model. Selain itu juga untuk memberikan informasi terkait prediktor yang paling meningkatkan kesesuaian model *return* saham di antara faktor makroekonomi dan rasio keuangan pada sektor bahan baku yang dapat menjadi acuan dalam pengambilan keputusan investasi ataupun perencanaan operasional perusahaan di masa mendatang.

1.5. Batasan Masalah

Penelitian ini dibatasi hanya pada beberapa aspek tertentu, bertujuan agar hasil dan pembahasan yang diperoleh tetap sesuai dengan tujuan penelitian sebagai berikut:

1. Penelitian hanya mencakup perusahaan sektor bahan baku yang terdaftar pada Bursa Efek Indonesia.
2. Variabel prediktor terbatas pada rasio keuangan (aktivitas, likuiditas, profitabilitas, solvabilitas, dan penilaian pasar) serta faktor makroekonomi (inflasi, BI-Rate, dan nilai kurs).
3. Data laporan keuangan dan *return* saham diambil secara triwulan pada periode 2022 – 2024.
4. Proses seleksi metode *stepwise* menggunakan nilai ambang batas α *(alpha to enter and remove)* sebesar 0,15, sebagai kriteria seleksi untuk memasukkan atau mengeliminasi variabel prediktor.

1.6. Landasan Teori

1.6.1. *Return* Saham

Dalam kegiatan investasi, *return* merupakan salah satu faktor mempengaruhi keputusan berinvestasi para investor. Karena *return* merupakan tingkat keuntungan yang diperoleh investor dalam konteks investasi (Tandelilin, 2017). Tujuan dari seorang investor membeli saham menurut Acheampong et al. (2014) adalah karena mengharapkan adanya pengembalian yang akan diterima dalam bentuk pembayaran dividen dan keuntungan modal yang terjadi saat saham mengalami kenaikan harga atau *capital gain*. Livoreka et al. (2014) & Omerhodzic (2014) memberikan penjelasan terkait dengan dividen, kebijakan dividen, dan sumber keuntungan selain dari dividen, yakni bahwa dividen merupakan pembayaran yang dilakukan oleh perusahaan kepada investor sebagai imbalan atas investasi mereka di perusahaan, yang dapat diberikan dalam bentuk uang tunai atau bentuk lain, serta dibayarkan dari laba bersih yang dihasilkan oleh perusahaan pada tahun sebelumnya atau tahun berjalan. Disebutkan bahwa dividen menjadi salah satu yang menjadi pusat perhatian investor karena dividen dinilai mampu memberikan keuntungan yang lebih pasti dibandingkan dengan keuntungan modal (*capital gain*) yang belum pasti, juga menjadi indikasi kekuatan keuangan perusahaan, dan dianggap sebagai sumber pendapatan reguler untuk memenuhi biaya hidup mereka. Sehingga keberadaan dividen sebagai salah satu sumber keuntungan investor diatur melalui kebijakan dividen dan menjadi tanggung jawab dewan direksi perusahaan, karena kebijakan dividen yang tidak stabil akan mengakibatkan investor tidak tertarik untuk berinvestasi di perusahaan tersebut yang dapat mengakibatkan harga saham turun. Investor menunjukkan ketidakpuasannya dengan menjual saham mereka ketika tidak menerima dividen yang diharapkan. Salah satu faktor yang mempengaruhi kebijakan dividen adalah kondisi ekonomi yang tidak menentu, karena pada kondisi tersebut perusahaan dapat memutuskan untuk menahan sebagian besar laba dengan tujuan menciptakan cadangan untuk mengatasi guncangan atau ketidakstabilan di masa depan, namun sebaliknya jika situasi sedang stabil maka perusahaan dapat membayarkan dividen lebih besar, sehingga kebijakan terkait pembayaran dividen dapat mengalami perubahan atau mengalami penyesuaian tergantung pada kondisi keuangan yang dialami perusahaan. Maka dari itu, selain menjadikan dividen sebagai sumber keuntungan dari kegiatan investasi, keuntungan modal juga dijadikan sebagai salah satu sumber keuntungan oleh investor, yaitu ketika harga saham mengalami kenaikan (*capital gain*). Namun keuntungan yang diperoleh investor dari *capital gain* yaitu ketika investor menjual sahamnya pada saat harga sedang naik akan dibayarkan oleh investor lain bukan perusahaan tersebut, kecuali jika perusahaan membeli kembali sahamya sendiri.

Menurut Hartono (2022) *return* dibedakan menjadi dua berdasarkan waktu terjadinya, yakni *return* realisasi *(realized return)* dan *return* ekspektasi *(expected return)* dengan *return* realisasi adalah *return* yang diperoleh dari data historis sehingga menunjukkan *return* yang telah terjadi dan digunakan sebagai pengukur kinerja investasi, sedangkan *return* ekspektasi adalah *return* yang diharapkan oleh investor akan diperoleh di masa mendatang atau merupakan *return* yang belum terjadi saat ini.

Pada penelitian ini *return* yang digunakan adalah *return* realisasi atau yang sudah terjadi. *Return* realisasi dihitung berdasarkan *return* total *(total return)* atau yang umumnya hanya disebut sebagai *return*, dan merupakan *return* keseluruhan yang diperoleh pada suatu periode tertentu dari sebuah investasi yang terdiri dari *capital gain (loss)* dan *yield,* dituliskan sebagai berikut (Hartono, 2022):

$$Return = Capital\ gain\ (loss) + Yield$$

Dengan *capital gain (loss)* adalah kenaikan atau penurunan dari nilai yang diinvestasikan (perubahan harga) dan *yield* atau penghasilan merupakan pendapatan yang diperoleh dari investasi, sehingga keseluruhan *return* total dapat dituliskan sebagai berikut:

$$R_t = \frac{(P_t - P_{t-1})}{P_{t-1}} + \frac{I_t}{P_{t-1}} \quad (1)$$

$$= \frac{(P_t - P_{t-1}) + I_t}{P_{t-1}}$$

R_t = *Return* pada periode t
P_t = Harga pada periode t
P_{t-1} = Harga pada periode t-1
I_t = *Income* yang diperoleh dari aset
$\frac{(P_t - P_{t-1})}{P_{t-1}}$ = *Capital gain (loss)*
$\frac{I_t}{P_{t-1}}$ = *Yield*

Jika tidak ada dividen atau *yield* yang diberikan, maka *return* cukup dihitung dengan rumus *capital gain (loss).* Persamaan (1) oleh Ikatan Akuntan Indonesia (2019) disebutkan sebagai rumus untuk menghitung *return* saham yang membagikan dividen, sedangkan rumus untuk menghitung *return* saham yang tidak membagikan dividen ataupun untuk menyatakan *return* saham secara umum pada seluruh instrumen investasi selain saham dapat dinyatakan sebagai berikut:

$$R_t = \frac{(P_t - P_{t-1})}{P_{t-1}} = \frac{Investasi\ Awal\ - Investasi\ Akhir}{Investasi\ Awal} \quad (2)$$

Pada penelitian ini tidak terdapat pembayaran dividen yang dilakukan oleh perusahaan, karena perhitungan *return* yang dilakukan adalah untuk periode triwulan, dan dividen umumnya dibayarkan secara tahunan. Sehingga penghitungan *return* dilakukan dengan rumus *capital gain (loss)* atau persamaan (2).

1.6.2. Rasio Keuangan

Merupakan alat untuk mengukur kondisi keuangan perusahaan yang termuat dalam laporan keuangan dan terbagi atas 5 rasio, yakni rasio likuiditas, rasio aktivitas, rasio solvabilitas, rasio profitabilitas, dan rasio nilai pasar (Santoso et al., 2023). Setiap perusahaan yang menunjukkan hasil yang baik berpotensi menghasilkan *return* yang tinggi di masa depan bagi para investornya.

a. Rasio Likuiditas

Pengukuran untuk menilai seberapa baik suatu perusahaan dalam memenuhi kewajiban jangka pendeknya (liabilitas) pada saat jatuh tempo, dan salah satu rasio yang umum digunakan untuk mengukur likuiditas adalah *Current Ratio (CR)* yang mewakili aset lancar dibagi dengan liabilitas lancar, berikut merupakan rumus untuk menghitung *Current Ratio* (Dahlquist & Knight, 2022):

$$Current\ Ratio\ (CR) = \frac{Current\ Assets}{Current\ Liabilities}, \quad (3)$$

dengan Current Assets adalah jumlah aset lancar dan Current Liablities adalah jumlah liabilitas lancar. Rasio yang lebih dari satu (> 1) mengindikasikan bahwa perusahaan mampu membayar kewajiban jangka pendeknya, dan rasio yang kurang dari satu (< 1) menunjukkan bahwa perusahaan perlu mengevaluasi kembali total aset lancar yang dimiliki untuk membayarkan kewajiban jangka pendek. Aset lancar menurut Ikatan Akuntan Indonesia (2021) merupakan aset yang dapat berupa kas dan aset lain yang dalam waktu satu tahun atau satu siklus operasi normal (periode waktu yang dimulai dari mempersiapkan persediaan hingga menerima kas hasil penjualan) akan dikonversi menjadi kas, dijual, atau digunakan oleh perusahaan. Selain aset lancar, dijelaskan juga terkait dengan aset tidak lancar yaitu aset perusahaan yang dapat dimanfaatkan selama lebih dari satu tahun atau satu siklus operasi normal. Sedangkan liabilitas lancar menurut Binus University (2021) merupakan liabilitas atau utang yang diharapkan dapat dilunasi dalam kurun waktu tidak lebih dari satu tahun atau dalam satu siklus operasi normal perusahaan.

b. Rasio Aktivitas

Pengukuran untuk menilai seberapa baik suatu perusahaan dalam menggunakan asetnya terkait dengan pengelolaan penjualan, piutang, dan persediaan. Salah satu rasio yang umum digunakan untuk mengukur aktivitas adalah *Total Asset Turnover (TATO)* yang mengukur kemampuan perusahaan dalam menggunakan asetnya untuk menghasilkan pendapatan/penjualan, dan semakin tinggi nilai TATO mengindikasikan bahwa perusahaan mampu menggunakan asetnya dengan sangat efisien untuk menghasilkan penjualan (Dahlquist & Knight, 2022). Berikut merupakan rumus untuk menghitung *Total Asset Turnover (TATO)* (Santoso et al., 2023):

$$Total\ Asset\ Turnover\ (TATO) = \frac{Sales}{Total\ Assets}, \quad (4)$$

dengan *sales* adalah jumlah penjualan dan *total assets* adalah total aset yang dimiliki perusahaan. Penjualan menurut kamus hukum Kementerian Keuangan Republik Indonesia adalah ketika kepemilikan aset mengalami peralihan dari satu pihak ke pihak lain dengan menerima penggantian dalam bentuk uang. Sedangkan aset adalah seluruh sumber daya yang dimiliki oleh perusahaan untuk digunakan di masa mendatang, misalnya seperti uang tunai, bangunan, dan

tanah (Binus University, 2022). Total aset akan mencakup aset lancar dan aset tidak lancar.

c. Rasio Solvabilitas

Pengukuran untuk menilai seberapa baik perusahaan dalam memenuhi kewajiban (liabilitas) jangka panjangnya pada saat jatuh tempo agar mampu terus beroperasi di masa depan, dan salah satu rasio yang digunakan untuk mengukur solvabilitas adalah *Debt to Equity Ratio (DER)* yang menunjukkan hubungan dan kontribusi antara utang dan ekuitas dalam pembiayaan bisnis, berikut merupakan rumus untuk menghitung *Debt to Equity Ratio (DER)* (Dahlquist & Knight, 2022):

$$Debt\ to\ Equity\ Ratio\ (DER) = \frac{Total\ Liabilities}{Total\ Stockholder\ Equity}, \quad (5)$$

dengan *Total Liabilities* adalah total kewajiban/utang, dan *Total Stockholder Equity* adalah total ekuitas/modal. Sahetapy (2023) menyebutkan bahwa dalam total liabilitas terdiri atas liabilitas jangka pendek atau lancar dan liabilitas jangka panjang yang perbedaannya terletak pada durasi waktu pelunasan utang, dengan liabilitas jangka panjang memiliki tenggat waktu pelunasan yang cukup lama. Selanjutnya diberikan penjelasan terkait ekuitas yang adalah modal pemilik perusahaan atau investasi pihak lain yang ditanamkan untuk memperkuat struktur modal.

d. Rasio Profitabilitas

Pengukuran yang menilai seberapa efektif perusahaan dalam menghasilkan laba berdasarkan kinerjanya, dan salah satu rasio yang digunakan untuk mengukur profitabilitas adalah *Return On Assets (ROA)* yang menunjukkan kemampuan perusahaan untuk menghasilkan pengembalian *(return)* yang baik atas aset yang telah diinvestasikan, dan semakin besar nilai yang diperoleh menunjukkan semakin baik perusahaan dalam memanfaatkan asetnya menjadi laba (Dahlquist & Knight, 2022). Berikut merupakan rumus untuk menghitung *Return On Assets (ROA)* (Santoso et al., 2023):

$$Return\ On\ Assets\ (ROA) = \frac{Net\ Profit\ After\ Taxes}{Total\ Assets}, \quad (6)$$

dengan *Net Profit After Taxes* adalah laba bersih dan *Total Assets* adalah total aset yang dimiliki. Laba bersih adalah menurut Sahetapy (2023) adalah keuntungan akhir yang diperoleh perusahaan setelah dikurangkan dengan biaya-biaya lain dalam kegiatan operasional dan non-operasional.

e. Rasio Nilai Pasar

Pengukuran untuk menilai harga pasar suatu perusahaan secara keseluruhan di pasar modal (Dahlquist & Knight, 2022). Karena pada rasio ini diamati sejumlah informasi yang berkaitan dengan jumlah saham yang beredar, total keuntungan, dividen, dan modal dialokasikan pada tiap lembar saham, dan *Price to Book Value (PBV)* merupakan salah satu rasio yang digunakan untuk mengukur nilai saham dengan rumus sebagai berikut (Santoso et al., 2023):

$$Price\ to\ Book\ Value\ (PBV) = \frac{Market\ Price\ of\ Common\ Stock}{Book\ Value\ Per\ Share}, \quad (7)$$

Book Value Per Share atau Nilai Buku Per Lembar Saham dapat dihitung dengan rumus berikut (Santoso et al., 2023):

$$\text{Nilai Buku Per Lembar Saham} = \frac{Total\ Ekuitas}{Jumlah\ Saham\ Beredar}, \quad (8)$$

dengan *Market Price of Common Stock* adalah harga pasar persaham, dan *Book Value Per Share* adalah nilai buku per lembar saham. Menurut Sitompul et al. (2022) nilai buku per lembar saham menunjukkan besaran nilai ekuitas investor pada setiap lembar saham perusahaan yang mereka miliki, dan merupakan nilai jaminan yang akan diperoleh investor jika perusahaan dilikuidasi. Hasil dari *PBV* digunakan untuk menilai apakah harga suatu saham *overvalued* atau *undervalued* (Rustiana et al., 2022). Disebutkan oleh Dahlquist & Knight (2022) bahwa jika harga pasar per saham berada di bawah nilai buku maka saham dianggap terlalu rendah (*undervalued*) dan jika harga pasar per saham berada di atas nilai buku akan dianggap terlalu tinggi (*overvalued*). Selain itu, nilai PBV yang diperoleh juga dapat diartikan sebagai berapa kali lipat harga saham yang beredar dari nilai buku perusahaan dan untuk mengukur kelayakan saham tersebut untuk dibeli (Binus University, 2022).

1.6.3. Faktor Makroekonomi

Merupakan cabang dari teori ekonomi yang mempelajari ekonomi secara keseluruhan atau yang mempertimbangkan seluruh lingkungan yang berkaitan aktivitas bisnis (Dahlquist & Knight, 2022). Membahas mengenai seluruh keadaaan dan aktivitas perekonomian sejumlah pihak mulai dari konsumen, pengusaha, hingga pemerintah serta bagaimana aktivitas-aktivitas tersebut mempengaruhi perekonomian dalam skala besar, dalam kaitannya terhadap tingkat pertumbuhan pasar modal, beberapa indikator makroekonomi seperti inflasi, suku bunga, dan nilai kurs turut berperan (Santoso et al., 2023). Iswanto (2017) menyebutkan bahwa pendapatan dan penjualan suatu perusahaan yang berdampak pada harga saham dan laba dipengaruhi oleh kondisi makroekonomi yang sedang terjadi, dan mempengaruhi pilihan investor dalam membuat pilihan untuk membeli atau menjual saham perusahaan tersebut.

a. Inflasi

Merupakan fenomena kenaikan harga banyak barang atau kenaikan harga secara umum yang mengakibatkan daya beli mata uang menurun, pengukuran inflasi umumnya dilakukan dengan *Consumer Price Index (CPI)* atau Indeks Harga Konsumen (IHK) (Dahlquist & Knight, 2022). Meningkatnya biaya produksi dan operasional perusahaan merupakan dampak dari terjadinya inflasi, yang apabila tidak disertai dengan adanya peningkatan pendapatan dan laba, akan mengakibatkan *return* saham menurun (Alfindy, 2019). Penyebab inflasi menurut Bank Indonesia adalah karena adanya peningkatan biaya produksi, meningkatnya permintaan barang dan jasa relatif terhadap kesediaannya, dan

ekspektasi inflasi. Atmadja (1999) menyebutkan bahwa terdapat empat kelompok inflasi, yaitu inflasi ringan yang berada di bawah 10%, inflasi sedang di kisaran 10% - 30%, inflasi tinggi di kisaran 30% - 100%, dan hyperinflasion yang berada di atas 100%.

b. Suku Bunga

Bank Indonesia (2024) menyebutkan bahwa Indonesia menetapkan suku bunga yang menjadi acuan atau patokan bagi seluruh lembaga keuangan yang berlaku di seluruh wilayah Indonesia, yang kemudian disebut sebagai suku bunga acuan atau BI-Rate, dan diumumkan secara rutin setiap bulannya, tujuannya adalah untuk menentukan suku bunga yang ditawarkan kepada nasabah. Dijelaskan bahwa ketika BI-Rate mengalami kenaikan, maka suku bunga perbankan akan cenderung naik, begitupun sebaliknya, dengan demikian salah satu sektor yang ikut dipengaruhi oleh perubahan BI-Rate adalah sektor investasi karena akan berdampak pada tingkat keuntungan investasi yang diperoleh, saat BI-Rate rendah investor perlu mencari instrumen investasi lain yang menawarkan *return* lebih tinggi.

c. Nilai Kurs

Menurut KBBI kurs merujuk kepada nilai mata uang suatu negara yang dinyatakan dalam bentuk mata uang negara lain. Fluktuasi yang terjadi pada nilai tukar atau kurs terhadap mata uang asing akan berdampak pada kondisi investasi dalam negeri (Iswanto, 2017). Pada penelitian ini kurs yang digunakan adalah kurs Rupiah Indonesia (IDR) terhadap Dolar Amerika (USD).

1.6.4. Analisis Regresi

Menurut Montgomery et al. (2012) analisis regresi adalah teknik statistik untuk memeriksa dan memodelkan hubungan antara variabel, serta merupakan prosedur iteratif di mana data akan menghasilkan suatu model yang kemudian dievaluasi untuk dilakukan penyesuaian dengan tujuan memodifikasi atau menerima model yang terbentuk, dan tujuan pentingnya adalah untuk mengestimasi parameter dalam model yang tidak diketahui. Terbagi atas regresi linear sederhana dan regresi linear berganda. James et al. (2021) menjelaskan bahwa regresi linear sederhana adalah pendekatan yang sangat sederhana untuk memprediksi respon kuantitatif Y berdasarkan variabel prediktor tunggal X, yang secara matematis dituliskan sebagai berikut:

$$Y \approx \beta_0 + \beta_1 X + \varepsilon,$$

$Y = variabel\ respon$
$X = variabel\ prediktor$
$\beta_0 = intercept$
$\beta_1 = koefisien\ parameter\ variabel\ prediktor\ 1$
$\varepsilon = error\ acak\ dengan\ rata - rata\ 0$

namun dalam praktiknya seringkali variabel prediktor X yang digunakan ada lebih dari satu termasuk pada penelitian ini, dan pendekatan itulah yang disebut sebagai regresi linear berganda, dengan persamaannya adalah sebagai berikut:

$$Y \approx \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_j X_j + \varepsilon,$$

$X_j = variabel\ prediktor\ ke - j$
$\beta_j = koefisien\ parameter\ variabel\ prediktor\ ke - j$

dengan simbol ≈ dapat dibaca "dapat didekati dengan", β_0 dan β_j masing-masing belum diketahui nilainya dan mewakili *intercept* dan *slope* yang dikenal sebagai koefisien model atau parameter, keduanya untuk mengukur hubungan antara variabel prediktor dan respon, dengan interpretasi β_0 adalah nilai rata-rata pada Y ketika semua nilai $X = 0$, dan β_j adalah efek rata-rata pada Y akibat peningkatan pada satu unit X_j dengan semua variabel prediktor tetap konstan. Untuk mengestimasi nilai dari β_0 dan β_j digunakan data dengan persamaan:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_j x_j, \tag{9}$$

$\hat{y} = estimasi\ dari\ Y$
$\hat{\beta}_0 = estimasi\ dari\ \beta_0$
$\hat{\beta}_j = estimasi\ dari\ \beta_j$
$x_j = variabel\ prediktor\ ke - j$

dengan $\hat{y}$ menunjukkan estimasi Y berdasarkan $X = x$, dan simbol ˆ atau dibaca hat/topi untuk menunjukkan nilai estimasi. Dijelaskan lebih lanjut oleh Montgomery et al. (2012) bahwa parameter $\hat{\beta}_1$ menunjukkan perubahan yang diharapkan pada respon ($\hat{y}$) untuk setiap kenaikan satu satuan pada x_1, dengan anggapan bahwa x_2 dan variabel prediktor lainnya adalah konstan, begitu pula dengan $\hat{\beta}_2$ hingga $\hat{\beta}_j$. Juga menyebutkan bahwa model regresi tidak menyiratkan hubungan sebab akibat antara variabel, melainkan hanya dapat membantu mengonfirmasi hubungan tersebut, karena butuh dasar di luar sampel seperti pertimbangan teoritis untuk menetapkan kausalitas. Berikut merupakan persamaan yang digunakan untuk mengestimasi nilai koefisien parameter:
persamaan (9) dapat dituliskan menjadi:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i, i = 1,2,\ldots,p$$

pada analis regresi linear berganda, penggunaan notasi matriks akan mempermudah untuk menyajikan model, data, dan hasil estimasi secara lebih ringkas, dengan notasi matriks sebagai berikut:

$$y = X\beta + \varepsilon,$$

yang diharapkan dapat meminimumkan jumlah kuadrat residual (jika dinyatakan dalam bentuk matriks):

$$S(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta) = y'y - 2\beta'X'y + \beta'X'X\beta,$$

bentuk jumlah kuadrat residual secara umum adalah

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i-1}^{n} (y_i - \hat{y}_i)^2 \tag{10}$$

Untuk meminimumkan $S(\beta)$ maka turunan fungsi S terhadap β harus sama dengan nol:

$$\frac{\partial S}{\partial \beta}\Big|_{\hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0,$$

sehingga diperoleh:

$$X'X\hat{\beta} = X'y$$

dan dengan mengalikan kedua sisi dengan invers dari $X'X$, maka estimator $\hat{\beta}$ dapat dituliskan demikian:

$$\hat{\beta} = (X'X)^{-1}X'Y,$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}, \tag{11}$$

sehingga untuk mengestimasi nilai koefisien $\hat{\beta}_0$ hingga $\hat{\beta}_p$ adalah sebagai berikut:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_j \end{bmatrix} = (X'X)^{-1}X'Y \tag{12}$$

dengan X adalah matriks berukuran $n \times p$ yang berisi variabel prediktor, dan X' adalah transpose dari matriks X, sedangkan Y merupakan vektor berukuran $n \times 1$ yang memuat nilai observasi dari variabel respon (Montgomery et al., 2012). Syarat dalam melakukan analisis regresi adalah memenuhi uji asumsi klasik yang mencakup uji normalitas, multikolinearitas, autokorelasi dan heteroskedastisitas (Indartini & Mutmainah, 2024).

a. Uji Normalitas

Pengujian ini menurut Iba & Wardhana (2024) dimaksudkan untuk melihat apakah residu dari model statistik mengikuti distribusi normal, karena data diharapkan untuk mengikuti distribusi normal yang ditunjukkan dengan sebagian besar nilai residu berkumpul di sekitar nilai tengah atau *mean,* beberapa metode untuk melakukan pengukuran normalitas adalah uji Shapiro Wilk dan uji Kolmogorov-Smirnov. Dijelaskan bahwa uji Shapiro Wilk digunakan untuk sampel kecil (≤ 50) dan uji Kolmogorov Smirnov untuk sampel besar (> 50), jika hasil pengujian yang menunjukkan:

- Nilai P (signifikansi) $> 5\%$, maka data berdistribusi normal.
- Nilai P (signifikansi) $< 5\%$, maka data tidak berdistribusi normal.

Vikaliana et al. (2022) menyebutkan bahwa uji normalitas hanya diberlakukan untuk nilai-nilai residual dan bukan terhadap seluruh variabel prediktor yang digunakan, dengan hipotesis yang diuji adalah sebagai berikut:

H_0: data berdistribusi normal.

H_1: data tidak berdistribusi normal.

b. Uji Multikolinearitas

Multikolinearitas dijelaskan oleh Iba & Wardhana (2024) adalah situasi saat dua atau lebih variabel prediktor dalam model memiliki korelasi tinggi yang berdampak pada sulitnya untuk memisahkan pengaruh dari masing-masing prediktor tersebut terhadap variabel respon, sehingga berakibat pada hasil estimasi yang tidak dapat diandalkan, maka dari itu harus dilakukan pengujian untuk melihat keberadaan korelasi antar variabel dengan salah satu metode yang disebut uji VIF *(Variance Inflation Error)*, nilai VIF > 10 menunjukkan keberadaaan multikolinearitas dalam model.

- Nilai VIF > 10, terdapat multikolinearitas.
- Nilai VIF < 10, tidak terdapat multikolinearitas.

Dengan hipotesis yang diuji adalah sebagai berikut:

H_0: tidak terjadi multikolinearitas.

H_1: terjadi multikolinearitas.

Selanjutnya dijelaskan juga beberapa cara untuk mengatasi multikolinearitas, yakni sebagai berikut:

- Menghapus variabel prediktor yang teridentifikasi mengalami multikolinearitas.
- Menggabungkan variabel-variabel yang teridentifikasi mengalami multikolinearitas menjadi satu variabel prediktor baru.
- Menggunakan metode regresi lain yang dapat membantu mengendalikan multikolinearitas seperti regresi ridge dan LASSO.
- Menambah jumlah data yang digunakan sebagai sampel untuk menaikkan keragaman nilai pada variabel prediktor.

c. Uji Autokorelasi

Pengujian ini dilakukan untuk mengetahui hubungan antara residu pada periode saat ini (t) dan periode sebelumnya $(t-1)$ (Iba & Wardhana, 2024). Salah satu metode yang dapat digunakan untuk mengindentifikasi keberadaaan autokorelasi adalah uji *Breusch-Godfrey* dengan dengan kriteria pengambilan keputusan berdasarkan *p-value* adalah sebagai berikut (Binus University, 2021):

- Jika nilai signifikansi $> 0{,}05$, maka tidak terjadi autokorelasi.
- Jika nilai signifikansi $< 0{,}05$, maka terjadi autokorelasi.

Dengan hipotesis yang diuji adalah sebagai berikut:

H_0: tidak terjadi autokorelasi.

H_1: terjadi autokorelasi.

d. Uji Heteroskedatisitas

Heteroskedatisitas adalah situasi dimana varians (sebaran) residu tidak sama untuk semua tingkat nilai prediktor, sedangkan asumsi yang diharapkan adalah situasi dimana varians residu tetap konstan untuk semua tingkat nilai prediktor atau yang disebut dengan homoskedatisitas ((Iba & Wardhana, 2024). Pengujian heteroskedatisitas menurut (Binus University, 2021) salah satunya dapat dilakukan dengan menggunakan uji *Breusch-Pagan,* dengan kriteria pengambilan keputusan adalah sebagai berikut

- Jika nilai signifikansi > 0.05, maka tidak terjadi heteroskedatisitas.
- Jika nilai signifikansi < 0.05, maka terjadi heteroskedatisitas.

Dengan hipotesis yang diuji adalah sebagai berikut:

H_0: tidak terjadi heteroskedatisitas.

H_1: terjadi heteroskedatisitas.

1.6.5. Regresi *Stepwise*

Hastie et al. (2008) menyebutkan bahwa metode *least square* (estimasi kuadrat terkecil) yang umumnya digunakan dalam analisis regresi linear memiliki beberapa kelemahan yang berdampak pada ketidakpuasan terhadap hasil yang diperoleh, alasan yang pertama adalah terkait akurasi prediksi yaitu bahwa estimasi *least square* seringkali memiliki bias yang rendah namun varians tinggi, sehingga akurasinya menjadi kurang stabil, dan untuk meningkatkannya dapat dengan mengurangi nilai koefisien atau membuatnya menjadi nol, dengan cara ini bias akan bertambah namun varians akan berkurang sehingga akurasi model meningkat. Namun James et al. (2021) menyebutkan jika nilai $n > p$ (kondisi berdimensi rendah), dengan n adalah jumlah observasi dan p adalah jumlah variabel prediktor yang digunakan, maka estimasi dengan *least square* akan menghasilkan varians yang rendah, dan dengan begitu mampu menghasilkan hasil yang baik pada pengamatan uji atau *test observation.* Sebaliknya, jika $p > n$ (kondisi berdimensi tinggi) maka estimasi dengan *least square* akan menghasilkan varians yang sangat tinggi karena tidak lagi dapat menghasilkan satu solusi estimasi koefisien, dan berakibat pada buruknya hasil prediksi untuk pengamatan uji yang menggunakan data uji, yakni data yang tidak digunakan dalam pelatihan model.

Alasan yang kedua disebutkan oleh Hastie et al. (2008) adalah interpretasi, yaitu bahwa dengan jumlah variabel prediktor yang banyak, ingin ditemukan subset atau sebagian dari variabel prediktor yang paling berpengaruh terhadap model, dan untuk memperoleh itu perlu mengurangi jumlah variabel prediktor. Penjelasan selanjutnya menyebutkan bahwa dalam memilih subset akan dipertahankan beberapa prediktor dan mengeliminasi lainnya, salah satu pendekatan yang dapat dilakukan untuk melakukan seleksi variabel dalam regresi linear adalah dengan metode *stepwise,* dan kemudian *least square* akan digunakan untuk mengestimasi koefisien dari variabel prediktor yang bertahan. Metode *least square* mengestimasi nilai koefisien $\beta_0, \beta_1, \dots, \beta_p$ menggunakan nilai-nilai yang meminimumkan jumlah kuadrat residual (James et al., 2021):

$$RSS = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \sum_{j=1}^{p} \hat{\beta}_j x_{ij})^2, \tag{13}$$

dengan RSS adalah *Residual Sum Of Square* atau Jumlah Kuadrat Residual, nilai $\beta_0, \beta_1, \dots, \beta_p$ yang meminimumkan persamaan (13) adalah estimasi koefisien regresi berganda.

Selain untuk mengatasi permasalahan ketidakpuasan terhadap hasil dari estimasi *least square*, Montgomery et al. (2012) juga menyebutkan bahwa metode eliminasi

variabel juga dapat digunakan untuk mengurangi dampak dari multikolinearitas, yaitu dengan *model respecification* atau spesifikasi ulang model, dan pendekatan ini seringkali merupakan teknik yang efektif.

Prosedur *stepwise* merupakan salah satu metode telah dikembangkan untuk mengevaluasi hanya sejumlah kecil model regresi subset dengan menetapkan dua nilai ambang batas untuk memasukkan atau mengeluarkan variabel melalui statistik uji F (atau t) parsial, hal ini karena mengevaluasi semua regresi dapat menjadi beban komputasi yang berat (Montgomery et al., 2012). Berikut merupakan langkah-langkah dalam menjalankan regresi *stepwise* dangan menggunakan uji t parsial menurut Montgomery et al. (2012), James et al. (2021), dan Zhou et al. (2024):

1. Memulai dari model kosong (seperti pada *forward selection*). Diasumsikan bahwa belum terdapat satu pun prediktor di dalam model, selain dari *intercept*. Sehingga prediktor akan ditambahkan satu per satu ke dalam model.
2. Pada setiap langkah akan diperoleh *p-value* untuk setiap prediktor kandidat berdasarkan nilai t statistik, dengan persamaan berikut:

$$t = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}}, \tag{14}$$

dengan,

t = nilai t statistik.

$SE(\hat{\beta}_j)$ = *standard error* dari $\hat{\beta}_j$ yang menunjukkan jumlah rata-rata perbedaan antara hasil estimasi $\hat{\beta}_j$ dengan nilai β_j yang sebenarnya. Semakin banyak observasi, maka nilai *standard error* akan semakin kecil.

$\hat{\sigma}^2$ = adalah estimasi varians error, dengan n adalah jumlah observasi dan p adalah jumlah prediktor, yang dinyatakan demikian:

$$\hat{\sigma}^2 = \frac{RSS}{n-p} \tag{15}$$

C_{jj} = elemen diagonal ke *j* dari matriks $(X'X)^{-1}$.

Semakin besar nilai t statistik yang diperoleh, akan menghasilkan *p-value* yang kecil. Nilai tersebut menyatakan seberapa besar nilai $\hat{\beta}_j$ menyimpang dari nol dalam satuan simpangan baku *(standard deviation)*. Berikut merupakan hipotesis yang berlaku:

- H_0: $\beta_j = 0$ (tidak terdapat hubungan variabel prediktor dengan respon).
- H_1: $\beta_j \neq 0$ (terdapat hubungan variabel prediktor dengan respon),

hipotesis H_0 yang terpenuhi akan menunjukkan bahwa tidak ada pengaruh dari variabel prediktor *j* terhadap respon, karena koefisien sama dengan nol. Maka dari itu ketika nilai t statistik besar atau menyimpang jauh dari nol, akan menunjukkan bahwa $\hat{\beta}_j$ juga akan menyimpang jauh dari nol sehingga dapat dispastikan bahwa $\beta_j \neq 0$, yang artinya adalah terdapat pengaruh dari variabel prediktor *j* tersebut terhadap respon, atau koefisien parameter tidak sama dengan nol (menolak H_0). Sehingga t statistik akan membantu untuk melihat kontribusi dari setiap prediktor dengan mengendalikan prediktor lain yang sudah berada di dalam model. Hal tersebut juga akan mempengaruhi *p-value* yang dihasilkan,

karena *p-value* menyatakan peluang akan diperoleh nilai sama dengan $|t|$ atau lebih besar dalam nilai absolut jika hipotesis H_0 terpenuhi, dan nilai yang kecil menandakan bahwa peluang tidak terdapat hubungan antara variabel prediktor dan respon (H_0) adalah kecil terjadi, yang sekaligus menunjukkan bahwa terdapat hubungan antara variabel prediktor dan respon. *P-value* dapat dihitung dengan menggunakan fungsi pt pada perangkat lunak R Studio. Bobbit (2020) menyebutkan bahwa fungsi ini merupakan fungsi *cumulatif distribution function (CDF)* dari distribusi *t-Student* yang menghitung luas daerah untuk suatu nilai *t* dan derajat kebebasan *df.* Distribusi ini memiliki bentuk lonceng *(bell shape)*, sehingga untuk menghitung luas daerah di sisi kiri suatu nilai $-t$ dan sisi kanan nilai *t* dalam distribusi ini adalah sebagai berikut:

$$pt(-t, df = n - p - 1) + pt(t, df = n - p - 1, lower.tail = FALSE) \quad (16)$$

dengan ruas kiri adalah untuk menghitung daerah di sisi kiri nilai t statistik sehingga digunakan nilai *-t*, dan ruas kanan untuk menghitung daerah di sisi kanan nilai t statistik. Disebutkan juga oleh Sulistianingsih et al. (2022) bahwa kriteria penolakan H_0 adalah jika *p-value* < α (nilai ambang batas).

3. Variabel prediktor dengan *p-value* terkecil akan terpilih untuk masuk ke dalam model (jika lebih kecil dari nilai ambang batas yakni α yang ditentukan).
4. Prediktor kedua dan seterusnya yang dimasukkan ke dalam model adalah yang saat ini memiliki *p-value* terkecil ketika diregresikan dengan respon, dan telah disesuaikan dengan pengaruh dari prediktor yang sudah berada terlebih dahulu di dalam model.
5. Pada saat proses memasukkan model, akan bersamaan dilakukan evaluasi ulang terhadap semua prediktor yang sudah masuk ke dalam model berdasarkan *p-value*, karena variabel prediktor yang telah ditambahkan sebelumnya mungkin tidak lagi diperlukan setelah variabel lainnya masuk ke dalam model.
6. Prediktor dengan *p-value* yang lebih besar dari ambang batas α akan dihapus dari model.
7. Proses ini akan berhenti ketika tidak ada lagi prediktor yang memenuhi nilai ambang batas α yang ditentukan.

Selain dengan menggunakan kriteria berdasarkan *p-value,* Li et al. (2026) menyebutkan bahwa algoritma *stepwise* dapat juga menjalankan proses penambahan dan pengurangan prediktor berdasarkan kriteria lain yang telah ditentukan sebelumnya seperti kriteria informasi misalnya AIC, ataupun metrik kinerja lainnya. Oleh Montgomery et al. (2012) juga menyebutkan bahwa dapat juga dilakukan prosedur *stepwise* berdasarkan nilai korelasi parsial antara prediktor dan respon sebagai urutan memasukkan prediktor ke dalam model, dan keputusan memasukkan atau mengeluarkan prediktor diambil berdasarkan perbandingan nilai absolut t statistik dengan nilai t yang ditentukan atau yang disebut sebagai t_{in}. Nilai t_{in} atau *cutoff value* yang ditetapkan untuk menambahkan atau mengeluarkan prediktor dinyatakan sebagai $t_{\frac{\alpha}{2},n-p}$, dengan $n - p$ adalah derajat kebebasan, dan dalam jumlah prediktor juga diperhitungkan *intercept*. Nilainya dapat ditemukan pada tabel t atau kalkulator nilai t jika tidak tersedia pada tabel t. Dengan tahapan adalah sebagai berikut:

1. Variabel pertama yang masuk ke dalam model adalah yang memiliki korelasi terbesar dengan variabel respon. Prediktor dengan korelasi terbesar akan menghasilkan nilai absolut t statistik terbesar juga.
2. Jika nilai t statistik yang dihasilkan melebihi nilai t yang ditentukan atau yang disebut sebagai t_{in}, maka prediktor tersebut akan masuk ke dalam model, jika lebih kecil maka akan dikeluarkan dari model.
3. Prediktor kedua yang dimasukkan ke dalam model adalah yang saat ini memiliki korelasi terbesar dengan respon setelah menyesuaikan dengan efek dari prediktor pertama dalam model, atau yang disebut dengan korelasi parsial.
4. Lalu kemudian nilai absolut t statistik setiap predikor akan dievaluasi kembali apakah masih memenuhi kriteria nilai t_{in} yang ditentukan atau tidak, dan akan terus berlanjut hingga tidak ada lagi prediktor yang dapat dimasukkan atau dikeluarkan dari model.

Dijelaskan oleh Montgomery et al. (2012) dan Li et al. (2026) bahwa urutan masuk prediktor ke dalam model tidak menunjukkan urutan pentingnya sebuah prediktor, karena pada regresi *stepwise* prediktor yang awalnya masuk ke dalam model bisa saja kemudian dikeluarkan ketika terdapat variabel lain yang masuk, karena proses penambahan dan pengurangan prediktor akan terus dilakukan secara berulang-ulang hingga tidak terdapat lagi perbaikan lebih lanjut yang dapat dilakukan untuk prediktor berdasarkan kriteria yang ditentukan. Sehingga hanya akan terbentuk satu model akhir dan hasil yang diperoleh oleh *stepwise* tidak menjamin bahwa model tersebutlah yang paling optimal karena bisa saja terdapat model lain yang sama baiknya. Karena pada regresi *stepwise* tidak dilakukan evaluasi untuk untuk semua kandidat prediktor seperti pada metode *all possible regression* atau *best subset selection* yang jika terdapat *p* prediktor, maka akan terdapat 2^p persamaan atau kombinasi prediktor yang harus dievaluasi untuk menemukan kombinasi prediktor terbaik. Sehingga disebutkan bahwa model *stepwise* merupakan model lebih sederhana yang hanya mengevaluasi sebagian kecil model regresi untuk mengurangi beban komputasi, dan mencakup variabel independen yang memberikan kinerja prediksi untuk variabel respon. James et al. (2021) menyebutkan bahwa prediktor yang ditambahkan ke dalam model *stepwise* adalah prediktor yang mampu memberikan peningkatan tambahan terbesar pada kesesuaian model.

Li et al. (2026) menjelaskan bahwa hasil dari regresi *stepwise* tidak boleh digunakan untuk inferensi statistik atau menarik kesimpulan kausal, karena proses seleksi variabel telah menghilangkan validitas inferensi statistik seperti *p-value* dan interval kepercayaan akibat dari adanya pengujian berulang-ulang, sehingga model yang diperoleh hanya digunakan untuk prediksi, karena tujuan utama pemodelan prediktif adalah untuk memaksimalkan akurasi bukan menarik kesimpulan kausal. Kelemahan *stepwise* menurut Kuhn & Johnson (2019) adalah adanya peningkatan terjadinya *false positive* atau positif palsu, yaitu memilih prediktor yang sebenarnya salah, karena dalam regresi *stepwise* digunakan banyak pengujian berulang untuk memutuskan memasukkan atau mengeluarkan prediktor berdasarkan nilai t statistik dan *p-value* yang terus berubah-ubah, sehingga akan meningkatkan peluang salah memilih prediktor. Untuk mengatasi hal ini dapat dilakukan dengan pengujian setiap

model berdasarkan nilai AIC. Kelemahan selanjutnya adalah *overfitting* model, yaitu statistik model yang dihasilkan sangat optimis atau sangat sesuai pada data yang digunakan, karena tidak memperhitungkan bahwa proses seleksi dilakukan secara berulang-ulang pada data yang sama, maka dari itu untuk mengatasi hal ini dapat dilakukan estimasi yang akan mencerminkan kinerja prediktif pada data baru. Dengan metode *stepwise*, prediktor yang telah dikeluarkan dapat dimasukkan kembali, karena semua keputusan memasukkan dan mengeluarkan prediktor dilakukan berdasarkan kriteria yang ditentukan sebelumnya.

Regresi *stepwise* memerlukan dua nilai ambang batas atau *cutoff value*, satu untuk menambahkan prediktor dan satu lagi untuk mengeliminasi prediktor, besaran nilai ini menentukan kebebasan seleksi, dan tidak ada aturan baku terkait nilai ambang batas ini, beberapa analisis memilih untuk memakai ambang batas yang sama untuk memasukkan dan mengeliminasi prediktor (namun hal ini tidak wajib). Montgomery et al. (2012) dan Zhou et al. (2024) menggunakan nilai ambang batas (α) sebesar 0,15 untuk memasukkan dan mengeluarkan variabel, selain itu Chowdhury & Turin (2020) menjelaskan bahwa pemilihan nilai ambang batas yang kecil seperti 0,05 dapat menyebabkan prediktor yang sebenarnya penting jadi terlewat, sehingga diberikan rekomendasi untuk menggunakan nilai dalam kisaran 0,15 – 0,20, karena nilai yang lebih tinggi dapat membantu agar tidak ada prediktor penting yang tereleminasi, dan menghindari adanya penghapusan prediktor yang kurang signifikan namun memiliki alasan praktis misalnya secara teori atau pengalaman ternyata berpengaruh, namun karena penerapan ambang batas yang terlalu ketat sehingga tereliminasi. Dengan adanya *cutoff value*, akan diberi batasan toleransi untuk *p-value*. Karena *p-value* yang besar mengindikasikan bahwa peluang diterimanya H_0 juga akan besar, dan semakin besar *p-value* artinya koefisien mendekati nol atau tidak berpengaruh terhadap variabel respon. Maka dari itu dijelaskan oleh Montgomery et al. (2012) dan Zhou et al. (2024) bahwa selain menggunakan nilai ambang batas dapat juga menggunakan indeks kecocokan berbasis informasi, salah satunya yaitu *Akaike Information Criterion* (AIC) yang memberikan pinalti terhadap jumlah prediktor yang dimasukkan ke dalam model sebagai alternatif dalam pemilihan prediktor yang bertujuan untuk memaksimalkan kecocokan model dengan penggunaan prediktor yang sesedikit mungkin, pada setiap langkah ke-*t* nilai AIC dihitung sebagai berikut:

$$AIC = n \times ln\left(\frac{RSS}{n}\right) + 2p \tag{17}$$

dengan n adalah jumlah observasi, p menyatakan jumlah prediktor dalam model, dan AIC menerapkan pinalti tetap sebesar dua terhadap prediktor.

1.6.6. Regresi LASSO

Pada metode sebelumnya yaitu regresi *stepwise* (bagian dari *subset selection*) disebutkan oleh James et al. (2021) dan Hastie et al. (2008) merupakan proses diskrit yang pemodelannya dilakukan dengan menggunakan metode *least square* (kuadrat terkecil) yang seringkali menunjukkan varians yang tinggi, dan oleh karena itu tidak mengurangi kesalahan prediksi *(prediction error)* model lengkap. Selanjutnya dijelaskan bahwa sebagai alternatifnya, pemodelan dapat dilakukan dengan

melibatkan semua prediktor yang dimiliki dengan menggunakan teknik yang membatasi estimasi koefisien, atau yang menyusutkan estimasi koefisien menuju nol, karena metode penyusutan *(shrinkage)* yang menyusutkan estimasi koefisien dapat secara signifikan mengurangi varians, lebih kontinu dan lebih tidak terpengaruh oleh variabilitias yang tinggi, dengan demikian juga mampu bekerja dengan baik untuk meningkatkan akurasi prediksi terhadap data baru yang tidak digunakan pada proses pelatihan. Dua teknik yang umum digunakan untuk menyusutkan koefisien regresi menuju nol adalah regresi *ridge* dan LASSO. Dengan regresi *ridge* yang akan selalu menghasilkan model yang penuh, sedangkan LASSO hanya akan membentuk model yang mencakup prediktor-prediktor yang paling penting saja, dan karena itu tantangan dalam interpretasi model merupakan kelemahan regresi ridge dan regresi LASSO adalah alternatif untuk mengatasi kelemahan ini. Pada penelitian ini akan berfokus kepada regresi LASSO.

Berkaitan dengan jumlah prediktor dan sampel yang digunakan dalam penelitian, sebelumnya pada regresi *stepwise* disebutkan bahwa jika $n > p$ maka estimasi dengan *least square* akan menghasilkan varians yang rendah, sehingga mampu menghasilkan hasil yang baik pada pengamatan uji, dan *stepwise* masih menggunakan metode least square untuk mengestimasi koefisiennya. Sedangkan jika $p > n$ maka metode *least square* tidak dapat memberikan hasil yang baik, karena estimasi koefisien yang dihasilkan tidak mampu menghasilkan satu solusi yang pasti. Maka dari itu disebutkan oleh James et al. (2021) untuk kasus $p > n$ dapat digunakan metode yang membatasi atau menyusutkan koefisien yang di estimasi, karena dengan begitu varians akan berkurang, dan dengan demikian meningkatkan akurasi prediksi terhadap data uji.

Zhou et al. (2024) menyebutkan bahwa regresi LASSO merupakan pendekatan umum untuk mengurangi kesalahan prediksi dengan menemukan keseimbangan antara bias dan varians, bias dalam konteks regresi merujuk kepada jarak antara koefisien yang diestimasi dan koefisien yang sebenarnya, sedangkan varians merujuk kepada ketidakpastian estimasi koefisien akibat variabilitas dalam pengambilan sampel, kedua elemen ini sering saling bertentangan satu sama lain, dan dikenal sebagai *bias-variance tradeoff.*

Regresi LASSO juga melakukan seleksi variabel sama seperti pada regresi *stepwise* yang akan menghasilkan model yang lebih mudah untuk diinterpretasikan. James et al. (2021) menyebutkan bahwa dalam kasus LASSO diterapkan penalti ℓ_1 yang berdampak pada menyusutnya beberapa estimasi koefisien menjadi tepat nol ketika parameter penyesuaian atau juga disebut *tuning parameter* (λ) cukup besar, dan karena itu regresi LASSO dapat menghasilkan model dengan jumlah variabel yang berbeda-beda tergantung pada nilai λ yang diperoleh, saat nilai λ meningkat maka varians akan berkurang namun bias meningkat. Dijelaskan juga bahwa pinalti penyusutan pada koefisien hanya diterapkan pada $\beta_1, \beta_2, \ldots, \beta_p$ sedangkan pada koefisien *intercept* (β_0) tidak, karena penyusutan hanya berlaku pada estimasi hubungan setiap variabel terhadap respon, sedangkan *intercept* tidak mewakili hubungan antar variabel melainkan hanya merupakan ukuran nilai rata-rata respon

ketika semua $X = 0$, dan sebelum menjalankan proses regresi LASSO perlu dilakukan standarisasi pada semua prediktor, agar semuanya berada dalam skala yang sama, karena antar variabel terdapat perbedaan skala satu sama yang lain, dengan standarisasi semua variabel memiliki *mean* 0 dan *standard deviation* 1. Zhou et al. (2024) juga menambahkan bahwa saat λ meningkat maka koefisien prediktor dengan pengaruh yang kecil terhadap respon akan cepat menyusut tepat ke nol, sedangkan koefisien prediktor dengan pengaruh yang besar terhadap respon tidak akan menyusut tepat ke nol, sehingga semakin besar λ yang diperoleh maka akan semakin sulit prediktor masuk ke dalam model, dan semakin kecil nilai λ akan semakin mudah prediktor untuk memasuki model. Maka dari itu diperlukan tingkat λ yang sesuai (optimal) untuk mencapai model yang diinginkan, dan salah satu metode yang paling sering digunakan untuk menentukan λ yang optimal adalah metode validasi silang *k-fold.*

Estimasi koefisien LASSO yakni $\hat{\beta}_{\lambda}^{L}$ diperoleh dari persamaan yang meminimumkan:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j\, x_{ij})^2 + \lambda\sum_{j=1}^{p}|\beta_j| = RSS + \lambda\sum_{j=1}^{p}|\beta_j|, \tag{18}$$

dengan $\lambda\sum_{j=1}^{p}|\beta_j|$ disebut sebagai pinalti penyusutan *(shrinkage pinalty)* yang akan mengecil jika $\beta_1, \ldots, \beta_p$ mendekati nol, karena itu pinalti ini mendorong estimasi β_j untuk semakin mendekati nol, dan meskipun menggunakan cara yang berbeda untuk melakukan estimasi koefisien, LASSO masih merupakan bagian dari model linear (James et al., 2021). Sehingga penulisan model akhir akan mengikuti penulisan pada hasil model linear, sama seperti *stepwise*.

Jadi pada regresi LASSO selain meminimalkan RSS atau *Residual Sum Of Square* (Jumlah Kuadrat Residual), ditambahkan juga pinalti pada nilai mutlak koefisien beta (Zhou et al., 2024). Langkah-langkah dalam melakukan analisis regresi LASSO adalah sebagai berikut menurut Nur et al. (2024), Zhou et al. (2024) dan Cardall et al. (2023):

1. Mencari nilai λ yang optimal menggunakan metode *k-fold cross validation.* Akan dihasilkan dua nilai λ, yakni λ_{min} dan λ_{1se}. Untuk seleksi variabel lebih disarankan untuk memilih λ_{min} yang mampu meminimumkan MSE *(Mean Squared Error).* Sedangkan λ_{1se} merupakan nilai lambda yang paling besar, dapat digunakan sebagai alternatif untuk memperoleh model yang lebih sederhana lagi. Dengan syarat adalah $\lambda > 0$. Karena ketika $\lambda = 0$ maka artinya pinalti tidak akan memberikan dampak apapun kepada koefisien atau nilai koefisien tidak akan mengalami penyusutan, dan hasilnya akan sama dengan yang diperoleh oleh metode *least square* (James et al., 2021).
2. Melakukan analisis regresi LASSO berdasarkan hasil estimasi koefisien LASSO yang diperoleh oleh persamaan (18) dengan menggunakan nilai λ optimal yang telah diperoleh melalui metode *k-fold cross validation.* Sehingga hasil estimasi koefisien prediktor yang diperoleh akan meminimalkan RSS *(Residual Sum Square)* atau Jumlah Kuadrat Residual, dan menambahkan pinalti pada nilai absolut dari koefisien beta yang adalah koefisien dari variabel prediktor, sehingga

koefisien dari semua prediktor akan mengecil sesuai dengan nilai yang ditentukan oleh parameter penyesuaian (λ).

Disebutkan oleh Pleiss (2025) bahwa regresi LASSO tidak memiliki solusi tertutup atau solusi eksplisit untuk mengestimasi koefisien regresi LASSO, sehingga harus dilakukan optimasi secara numerik dengan menggunakan salah satunya adalah algoritma *glmnet* yang bekerja secara iteratif hingga tercipta konvergensi. Penyebabnya disebutkan oleh Emmert-Streib & Dehmer (2019) adalah karena fungsi objektifnya tidak *differentiable* atau tidak dapat diturunkan. Breheny (2021) menyebutkan bahwa algoritma yang diterapkan oleh *glmnet* untuk mengestimasi koefisien regresi LASSO adalah *coordinate descent algorithm,* yang mengoptimalkan fungsi terhadap satu parameter saja pada satu waktu dengan menganggap koefisien regresi lainnya tetap konstan, dan seluruh parameter secara akan diiterasi hingga tercapai konvergensi. Persamaan (18) akan diminimumkan terhadap β_j dengan menganggap koefisien lainnya yaitu β_{-j} konstan, namun dalam *coordinate descent* fungsi objektif yang digunakan dalam bentuk berikut ini:

$$Q(\beta_j|\beta_{-j}) = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \sum_{k\neq j} x_{ik}\beta_k - x_{ij}\beta_j)^2 + \lambda|\beta_j| + Constant$$

tidak terdapat *intercept* (β_0) karena *coordinate descent* tidak mengestimasi *intercept*, karena *intercept* tidak terdampak pinalti. Selanjutnya dijelaskan terkait prosesnya sebagai berikut yang dilakukan untuk setiap $j = 1,2,\ldots,p$ adalah:

1. Mendefinisikan residual parsial terhadap prediktor ke-*j*.

   $$\tilde{r}_{ij} = y_i - \sum_{k\neq j} x_{ik}\tilde{\beta}_k\,, i = 1,\ldots,n$$

   tanda "~" untuk menunjukkan kondisi sementara. Jadi $\tilde{r}_{ij}$ menyatakan residual parsial sementara dari prediktor ke-*j* pada observasi ke-i. Hasil dari $\tilde{r}_{ij}$ merupakan bagian dari y yang belum dijelaskan oleh prediktor lain selain prediktor ke-*j*, atau merupakan nilai kontribusi prediktor ke-*j*.
2. Menghitung statistik korelasi parsial berdasarkan residual parsial prediktor ke-j, dinyatakan sebagai $\tilde{z}_j$ yang berasal dari konsep yang sama dengan penduga *least square* pada regresi linear sederhana (satu prediktor), namun berdasarkan korelasi antara x_j dan residual, karena pada algoritma ini juga dilakukan perhitungan untuk satu persatu koefisien prediktor. Pada *coordinate descent* penduga estimator dinyatakan dalam $\tilde{z}_j$:

   $$\tilde{z}_j = n^{-1}\sum_{i=1}^{n} x_{ij}\tilde{r}_{ij},$$

   tambahan $\tilde{\beta}_j^{(s)}$ untuk mengembalikan kontribusi dari prediktor ke-*j*.

   $$\tilde{z}_j = n^{-1}\sum_{i=1}^{n} x_{ij}r_i + \tilde{\beta}_j^{(s)},$$

   $\tilde{\beta}_j^{(s)}$ adalah koefisien ke-*j* yang meminimumkan persamaan (18) pada iterasi ke-*s*.

   $$\tilde{\beta}_j^{(s)} = S(\tilde{z}_j|\lambda)$$

3. Memperbaharui koefisien dengan operator ***soft-thresholding***

$$\tilde{\beta}_j^{(s+1)} \leftarrow S(\tilde{z}_j|\lambda)$$

dengan fungsi $S(\cdot\,|\lambda)$ adalah operator ***soft-thresholding*** yang dinyatakan sebagai estimasi LASSO (Breheny, 2019):

$$S(\tilde{z}_j|\lambda) = \begin{cases} \tilde{z}_j - \lambda, & jika\ \tilde{z}_j > \lambda \\ 0, & jika\ |\tilde{z}_j| \leq \lambda \\ \tilde{z}_j + \lambda, & jika\ \tilde{z}_j < -\lambda \end{cases}$$

4. Memperbaharui residual karena akan terdapat perbedaan antara residual yang diperoleh pada iterasi ke-***s*** dan ***s+1***.

$$r_i \leftarrow r_i - (\tilde{\beta}_j^{(s+1)} - \tilde{\beta}_j^{(s)})x_{ij}, i = 1, \ldots, n$$

dengan r_i adalah residual total, dan tanda "←" pada algoritma menunjukkan *update* atau berarti nilai pada sisi kiri diganti dan diperbarui dengan nilai yang diperoleh pada sisi kanan.

5. Dilakukan pengulangan untuk semua koefisien hingga tercipta konvergensi.

Fatih (2024) dan Ranstam & Cook (2018) menyebutkan bahwa hasil yang diperoleh oleh regresi LASSO merupakan model yang terdiri dari kombinasi prediktor-prediktor dan estimasi koefisien relevan yang mampu meminimimalkan kesalahan prediksi, sehingga memastikan kinerja yang baik pada data yang belum pernah digunakan (data uji), dan memberikan estimasi parameter individu dengan prediksi keseluruhan yang lebih baik. Dijelaskan juga bahwa sebagai akibat dari penyusutan nilai koefisien menjadi nol, kekurangan LASSO adalah terdapat kemungkinan akan mengeliminasi prediktor dengan nilai koefisien kecil namun signifikan, sehingga pada regresi LASSO tidak dapat dilakukan interpretasi terkait kontribusi prediktor secara individu, karena fokusnya adalah pada prediksi gabungan atau model terbaik. Terkait dengan prediktor yang tidak masuk ke dalam model disebutkan juga oleh Ranstam & Cook (2018) adalah prediktor dengan koefisien regresi nol setelah mengalami penyusutan.

1.6.7. *Adjusted* R^2

Chung (2010) menyebutkan bahwa *adjusted* R^2 merupakan bentuk penyesuaian dari R^2. Nilai dari R^2 atau koefisien determinasi menunjukkan proporsi variabilitas dalam variabel respon (Y) yang mampu dijelaskan oleh variabel prediktor (X) atau model. Seiring dengan pertambahan prediktor ke dalam model, maka nilai dari R^2 juga akan meningkat sehingga tidak dapat digunakan sebagai ukuran kesesuaian model. Untuk mengatasi permasalah tersebut digunakan *adjusted* R^2 yang akan melakukan penyesuaian terhadap jumlah prediktor yang digunakan dalam model. Berikut merupakan persamaan untuk menghitung nilai adjusted R^2:

$$Adjusted\ R^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1} \tag{19}$$

dengan,

$R^2 = \frac{SST - SSE}{SST}$

$SST = total\ sum\ squares\ yang\ dinyatakan\ dengan \sum_{i=1}^{n}(Y_i - \bar{Y})^2$, dan SSE terdapat pada persamaan (10)

$\bar{Y} = rata - rata\ dari\ Y\ yang\ dinyatakan\ dengan\ \frac{1}{n}\sum_{i=1}^{n} Y_i$

Total sum squares mengukur variabilitas total dalam variabel Y, dan SSE (*Sum of Squared Errors*) mengukur jumlah variabilitas dalam variabel Y yang tidak dijelaskan oleh model.

1.6.8. *K-Fold Cross Validation*

Dalam menjalankan proses regresi digunakan sejumlah data untuk memprediksi respon pada suatu pengamatan. James et al. (2021) menjelaskan penggunaan metode pemodelan akan dianggap tepat jika metode tersebut mampu menghasilkan nilai rata-rata kesalahan yang terjadi dalam proses pemodelan pada suatu pengamatan baru (tidak digunakan selama proses latih/*training)* atau *test error* yang rendah, namun seringkali data *test* tidak tersedia sehingga sulit untuk menghitung *test error*. Sedangkan *training error* dapat dihitung dengan lebih mudah karena hanya melibatkan data-data yang digunakan dalam pemodelan (data latih/*training*). Untuk mengestimasi *test error* dibutuhkan himpunan data *test* yang sangat besar, namun apabila tidak tersedia, maka terdapat beberapa teknik yang dapat digunakan untuk mengestimasi *test error* dengan menggunakan data latih, *validation set* dan *k-fold cross validation* adalah dua contoh teknik yang dapat dimanfaatkan.

James et al. (2021) menjelaskan *validation set* adalah melibatkan pembagian secara acak data menjadi dua bagian, yakni *training set* (kelompok data latih) dan *validation set* (kelompok data uji/validasi). Model akan dilatih dengan data latih, lalu kemudian model yang telah terbentuk tersebut digunakan untuk memprediksi pengamatan yang terdapat di dalam data uji/validasi. Hasil dari validation set ini dapat sangat bervariasi, tergantung secara tepat pada pengamatan mana yang termasuk ke dalam kelompok data latih, dan pengamatan mana yang dimasukkan ke dalam data uji. Selain itu, metode statistik cenderung bekerja dengan buruk ketika hanya menggunakan sedikit data latih, akibatnya *test error* dengan data uji juga akan cenderung tinggi, karena tidak menggunakan seluruh himpunan data (karena adanya pembagian antara data latih dan data uji di awal). *K-fold cross validation* adalah teknik yang dapat menyempurnakan *validation set.*

K-fold cross validation bekerja dengan membagi secara acak data ke dalam *k* kelompok dengan ukuran yang kurang lebih sama. Kelompok pertama digunakan sebagai data uji, dan sisanya yakni *k-1* kelompok lainnya akan digunakan sebagai data latih. *Mean Squared Error* (MSE) kemudian akan dihitung pada setiap kelompok tersebut, dan berulang sebanyak *k* kali dengan setiap kelompok akan bergantian menjadi data uji. Proses ini akan menghasilkan estimasi *test error* yakni MSE sebanyak *k*, yaitu MSE_1, MSE_2,..., MSE_k, dengan *k* adalah jumlah *fold* atau lipatan yang ditentukan, umumnya digunakan *k*=5 atau *k*=10 karena lebih menguntungkan dalam hal komputasi, akan diperoleh estimasi *k-fold cross validation* dengan merata-ratakan nilai MSE tersebut, berikut adalah persamaan yang digunakan (James et al., 2021):

$$CV_k = \frac{1}{k}\sum_{i=1}^{k} MSE_i, \tag{20}$$

Dalam menentukan nilai lambda optimal pada regresi LASSO juga digunakan *k-fold cross validation.* Setiap proses iterasi *training* dan validasi, berbagai nilai λ

digunakan untuk menghasilkan nilai MSE dari *cross validation* dan dibandingkan satu sama lain, nilai λ yang mampu meminimumkan MSE *cross validation* yaitu λ_{min} adalah yang direkomendasikan untuk seleksi variabel (Zhou et al., 2024). Berikut adalah persamaan *k-fold cross validation* untuk memperoleh nilai λ yang mampu meminimumkan MSE:

$$CV(\lambda) = \frac{1}{k}\sum_{i=1}^{k} MSE_i(\lambda) \tag{21}$$

berlaku konsep yang sama dengan persamaan (20) namun dengan penambahan komponen λ, untuk membedakan antara hasil yang diperoleh oleh *cross validation* untuk MSE *test* dan λ optimal. Nilai λ optimal diperoleh dari persamaan berikut (Emmert-Streib & Dehmer, 2019):

$$\lambda_{min} = arg\ min\ CV(\lambda) \tag{22}$$

1.6.9. MSE

Mean Squared Error (MSE) disebutkan oleh James et al. (2021) adalah ukuran yang paling umum untuk digunakan dalam mengevaluasi kinerja suatu metode statistik, yang menunjukkan sejauh mana hasil prediksi sesuai dengan data yang diamati atau sejauh mana nilai respon yang diprediksi suatu pengamatan mendekati nilai respon pengamatan yang sebenarnya. Model yang memiliki MSE terkecil adalah model yang terbaik (Kusuma & Wulansari, 2019). Dengan persamaan adalah sebagai berikut (Zhou et al., 2024):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2, \tag{23}$$

dengan *N* adalah jumlah sampel, $\hat{y}_i$ adalah nilai hasil estimasi dan y_i merupakan nilai yang sebenarnya.

Ditambahkan oleh James et al. (2021) bahwa secara umum peneliti lebih tertarik pada akurasi prediksi yang dapat diperoleh saat menerapkan metode ini pada data uji *(test data)* yang belum pernah dilihat sebelumnya, atau kepada kemampuan prediksi yang dihasilkan oleh suatu model, namun yang menjadi permasalahan adalah ketidaktersediaan data uji *(test data),* sehingga yang digunakan adalah cukup data pelatihan *(training data)* karena MSE *training* dan MSE *test* tampaknya terkait erat, namun karena tidak ada jaminan bahwa metode dengan nilai MSE *training* rendah juga akan memiliki nilai MSE *test* yang rendah, maka dapat digunakan salah satu metode untuk mengatasi permasalahan ini yaitu validasi silang *(cross validation)* untuk memperkirakan nilai MSE *test* menggunakan data pelatihan/*training*. James et al. (2021) juga menyebutkan bahwa MSE *test* berkaitan erat dengan varians ditambah dengan bias kuadrat, sehingga metode yang mampu menghasilkan bias sekaligus varians yang rendah akan menghasilkan kinerja yang baik pada pengujian dengan data uji.

1.6.10. Sektor Bahan Baku

Bursa Efek Indonesia (BEI) membagi perusahaan-perusahaan yang terdaftar di dalamnya ke dalam 11 sektor bergantung pada jenis usaha yang dijalankan oleh

perusahaan tersebut, salah satunya adalah sektor bahan baku atau *basic materials.* Sektor bahan baku merupakan salah satu sektor dengan jumlah perusahaan tercatat paling banyak diantara sektor-sektor lainnya yang terdapat di BEI, sehingga opsi saham yang ditawarkan juga beragam. Sektor bahan baku menurut Umam (2024) dan Ariantiani (2023) merupakan sektor yang menyediakan bahan baku atau bahan mentah yang diperlukan oleh perusahaan dari sektor lain untuk memproduksi barang jadi yang siap digunakan oleh konsumen, menjadikannya sebagai sektor yang sangat strategis, penting dalam perekonomian, dan mengalami perkembangan yang pesat, juga sektor ini rentan terhadap fluktuasi harga komoditas, siklus bisnis global, kebijakan pemerintah, dan dinamika makroekonomi, sehingga kenaikan harga pada sektor ini akan mempengaruhi harga produk dari sektor lain yang menggunakan produk dari sektor bahan baku, menjadikannya sebagai salah satu tempat berinvestasi yang menjanjikan. Beberapa contoh produk yang dihasilkan sektor bahan baku adalah emas, baja, besi, plastik, semen, serta berbagai barang kimia.

BAB II

METODE PENELITIAN

2.1. Jenis Penelitian

Penelitian ini dilakukan dengan pendekatan kuantitatif, karena proses penelitian yang sejalan dengan konsep penelitian kuantitatif, yakni menggunakan data berbentuk angka, sistematis dan terukur, mengacu pada teori yang telah tersedia, bersifat objektif, menggunakan sampel yang hasilnya dapat diterapkan ke dalam populasi (generalisasi), dan baik dalam menginterpretasi analisis statistik (Soesana et al., 2023), dan analisis statistik merupakan pokok utama dalam penelitian yang dilakukan.

2.2. Waktu dan Tempat Penelitian

Penelitian dilakukan sejak September 2025 yang diawali dengan meninjau penelitian-penelitian terdahulu dan teori terkait, serta mengindetifikasi jenis data yang diperlukan. Pengolahan data dengan menggunakan Microsoft Excel dan perangkat lunak R Studio.

2.3. Sumber Data

Penelitian dilakukan dengan mengolah data sekunder, dan mengambil lima perusahaan yang tergabung di dalam sektor bahan baku BEI sebagai sampel, yakni Archi Indonesia Tbk. (ARCI), Merdeka Copper Gold Tbk. (MDKA), Aneka Tambang Tbk (ANTM), Saraswanti Anugerah Makmur Tbk. (SAMF), dan Semen Baturaja Tbk. (SMBR). Kelima sampel dipilih berdasarkan jumlah lembar saham yang relatif banyak beredar, dan berdasarkan ketersediaan informasi yang lengkap terkait laporan keuangan triwulan dan laporan tahunan, serta mampu memenuhi uji asumsi model. Penelitian menggunakan data triwulan dari tahun 2022 – 2024. Sejumlah data dikumpulkan dari berbagai situs seperti:

- Laman resmi BEI untuk informasi terkait perusahaan yang bergabung di dalam sektor bahan baku, laporan keuangan dan tahunan perusahaan.
- Laman resmi perusahaan masing-masing untuk melengkapi laporan keuangan yang tidak tersedia pada laman resmi BEI.
- Laman *Investing.com* untuk informasi terkait harga penutupan saham setiap perusahaan terkait.
- Laman resmi Bank Indonesia (BI) untuk data BI-Rate dan inflasi .
- Lama resmi Satu Data Perdagangan dari Kementrian Perdagangan untuk data nilai tukar/kurs.

2.4. Variabel Penelitian

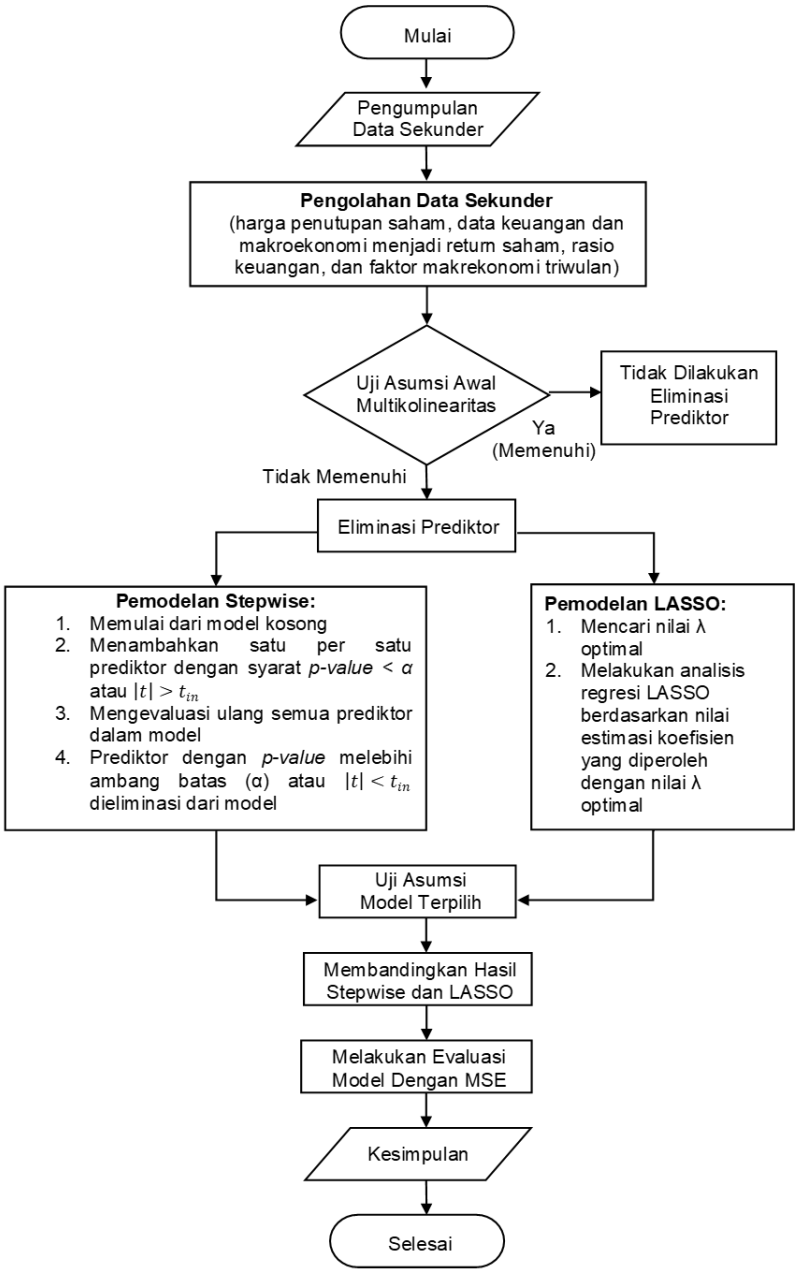
Variabel yang digunakan dalam penelitian ini mencakup *return* saham sebagai variabel respon, dan faktor fundamental perusahaan serta faktor makroekonomi sebagai variabel prediktor. Faktor fundamental yakni rasio keuangan, terdiri dari rasio likuiditas yang diwakili oleh CR, rasio aktivitas oleh TATO, rasio solvabilitas oleh DER, rasio profitabilitas oleh ROA, dan rasio nilai pasar oleh PBV. Faktor makroekonomi diwakili oleh inflasi, BI-Rate, dan nilai tukar/kurs.

2.5. Langkah Analisis Data

Berikut merupakan langkah-langkah analisis data yang dilakukan dalam penelitian ini:

1. Mengumpulkan data sekunder berupa harga penutupan saham perusahaan, laporan keuangan perusahaan, data inflasi, BI-Rate, dan nilai tukar selama 12 triwulan sejak Januari 2022 hingga Desember 2024.
2. Melakukan pengolahan data awal seperti menghitung *return* saham triwulan dari harga penutupan saham, menghitung rasio-rasio keuangan perusahaan yakni CR, TATO, DER, ROA, dan PBV secara triwulan, menghitung inflasi, BI-Rate, dan nilai kurs triwulan untuk digabungkan secara utuh sebagai informasi data triwulan tahun 2022 hingga 2024 ke dalam kelompok variabel masing-masing.
3. Melakukan pengujian awal asumsi multikolinearitas pada model penuh (yakni model yang belum mengalami proses seleksi prediktor).
4. Melakukan pemodelan dengan regresi *stepwise*.
5. Melakukan pemodelan dengan regresi LASSO.
6. Melakukan uji asumsi pada model yang berisi prediktor terpilih (setelah proses seleksi).
7. Mengevaluasi model yang diperoleh dengan MSE.
8. Membandingkan hasil antara regresi *stepwise* dan LASSO, serta menarik kesimpulan

2.6. Diagram Alur Penelitian



Gambar 1. Diagram Alur Penelitian