Griya Journal of Mathematics Education and Application
Volume 6 Nomor 1, Maret 2026
e-ISSN 2776-124X| |p-ISSN 2776-1258
https://mathjournal.unram.ac.id/index.php/Griya/index

Penanganan Multikolinieritas dalam Regresi Saham GOTO Menggunakan PCA, Ridge, LASSO, dan PLS

Nur El Hasanah[1], Revika Putri Asharia[1], Fachri Faisal[2*], Ratna Widayati[2], Zulfia Memi Mayasari[2], Siska Dwi Kumala[2], Aisyah Nooravieta Setiawan[2]

[1] *Mahasiswa Matematika, FMIPA, Universitas Bengkulu, Bengkulu*
[2] *Matematika, FMIPA, Universitas Bengkulu, Bengkulu*

fachrif@unib.ac.id

Diterima: 23-02-2026; Direvisi: 08-03-2026; Dipublikasi: 11-03-2026

Abstract
This study aims to address the problem of multicollinearity in a multiple regression model of the daily closing stock price of PT GoTo Gojek Tokopedia Tbk (GOTO) during the period from 2022 to early 2025. Multicollinearity occurs when independent variables are highly correlated, which can lead to inefficient and unreliable parameter estimates. GOTO's stock price has experienced high volatility since its Initial Public Offering (IPO) in April 2022. With an initial offering price of Rp338 per share, GOTO attracted significant attention from investors. However, by December 2023, GOTO's stock price had declined substantially to Rp95 per share, reflecting a decrease of Rp243 since the IPO. This highly fluctuating stock price movement requires an appropriate analytical approach to identify the factors influencing stock price movements. The study uses the closing price as the dependent variable, with opening price, high price, low price, and trading volume as independent variables. The methods employed include multiple regression and several approaches to handle multicollinearity, namely variable elimination, Principal Component Analysis (PCA), Ridge Regression, LASSO Regression, and Partial Least Squares (PLS) Regression. The initial multiple regression model achieved an R^2 of 0.9990 and an RMSE of 2.88, but Variance Inflation Factor (VIF) analysis indicated severe multicollinearity. After applying the alternative methods, PLS Regression demonstrated the best performance, with an R^2 of 0.9990 and an RMSE of 0.0318. Therefore, it can be concluded that PLS Regression is a more stable and accurate method for addressing multicollinearity and improving the prediction of GOTO's stock prices.

Keywords: stock prices; multicollinearity; LASSO regression; ridge regression; PLS regression.

Abstrak
Penelitian ini bertujuan menangani masalah multikolinearitas dalam model regresi berganda terhadap harga saham penutupan harian PT GoTo Gojek Tokopedia Tbk (GOTO) selama periode 2022 hingga awal 2025. Multikolinearitas terjadi ketika variabel bebas saling berkorelasi kuat sehingga menyebabkan estimasi parameter menjadi tidak efisien dan kurang akurat. Harga saham GOTO mengalami volatilitas tinggi sejak IPO April 2022, dengan harga saham perdana sebesar Rp338 per lembar, GOTO menarik perhatian besar dari investor. Namun, hingga Desember 2023, saham GOTO mengalami penurunan signifikan menjadi Rp95 per lembar, mencerminkan penurunan sebesar Rp243 sejak IPO. Pergerakan harga saham yang sangat fluktuatif ini memerlukan pendekatan analisis yang tepat untuk mengidentifikasi faktor-faktor yang memengaruhi pergerakan harga. Data penelitian menggunakan variabel Terakhir sebagai variabel dependen, serta Pembukaan, Tertinggi, Terendah, dan Volume sebagai variabel independen. Metode yang digunakan meliputi regresi berganda dan beberapa pendekatan penanganan multikolinearitas, yaitu penghapusan variabel, *Principal Component Analysis* (PCA), *Ridge Regression, LASSO Regression*, dan *Partial Least Squares* (PLS) *Regression*. Model awal menghasilkan R^2 sebesar 0,9990 dan RMSE 2,88, namun terindikasi multikolinearitas tinggi berdasarkan nilai VIF. Setelah

penerapan metode alternatif, PLS Regression memberikan performa terbaik dengan R^2 = 0,9990 dan RMSE = 0,0318. Dengan demikian, PLS Regression dinilai paling stabil dan akurat dalam mengatasi multikolinearitas serta meningkatkan ketepatan prediksi harga saham GOTO.

Kata Kunci: harga saham; multikolinearitas; LASSO *regression*; ridge *regression*; PLS *regression*.

1. PENDAHULUAN

Kondisi perekonomian global memiliki pengaruh yang signifikan terhadap dinamika pasar modal di berbagai negara. Pasar modal merupakan bagian dari sistem keuangan yang memiliki berbagai peran penting dalam perekonomian. Sebagai sarana penambah modal bagi perusahaan, pasar modal memungkinkan perusahaan memperoleh dana melalui penjualan saham yang dapat dibeli oleh masyarakat, lembaga, maupun pemerintah. Selain itu, pasar modal juga berfungsi sebagai sarana pemerataan pendapatan, karena dividen yang diterima pemegang saham dapat meningkatkan distribusi pendapatan (Ferdiansyah et al., 2016).

Melalui proses *Initial Public Offering* (IPO) pada April 2022 dengan harga saham perdana sebesar Rp338 per lembar, GOTO menarik perhatian besar dari investor. Namun, hingga Desember 2023, saham GOTO mengalami penurunan signifikan menjadi Rp95 per lembar, mencerminkan penurunan sebesar Rp243 sejak IPO. Pergerakan harga saham yang sangat fluktuatif ini menjadi objek yang relevan untuk dianalisis, terutama dengan menggunakan pendekatan statistik seperti analisis regresi.

Analisis regresi merupakan metode statistik yang digunakan untuk melihat hubungan antara satu variabel dependen dengan satu atau lebih variabel independen. Dalam konteks ini, regresi linier berganda digunakan untuk mengidentifikasi faktor-faktor yang memengaruhi harga saham secara simultan. Namun demikian, model regresi tidak terlepas dari berbagai tantangan teknis. Salah satu permasalahan yang sering muncul adalah multikolinearitas, yaitu kondisi ketika terdapat hubungan atau korelasi tinggi antar variabel bebas dalam model. Multikolinearitas dapat mengakibatkan estimasi parameter regresi menjadi tidak efisien karena bias dan variansnya menjadi besar, serta dapat memengaruhi akurasi hasil estimasi koefisien regresi (Sungkono & Nugrahaningsih, 2017; Sari, 2023). Masalah multikolinearitas telah menjadi perhatian dalam prediksi, karena mengganggu stabilitas model regresi (Al-Kassab & Ibrahim, 2022). Menurut Greene (2018), model regresi linear berganda memerlukan asumsi tidak adanya multikolinearitas yang kuat antar variabel independen agar estimasi parameter yang dihasilkan dapat diinterpretasikan secara tepat. Multikolinieritas melanggar asumsi penting dalam regresi linear klasik dan dapat merusak validitas inferensi statistik dari model yang digunakan (Montgomery, Peck, & Vining, 2012).

Melihat pentingnya membangun model regresi yang stabil dan bebas dari multikolinearitas, penelitian ini difokuskan pada analisis multikolinearitas dalam model regresi linear berganda terhadap harga saham PT GoTo Gojek Tokopedia Tbk.

Penelitian ini bertujuan untuk membangun model regresi berdasarkan beberapa variabel pasar dan fundamental perusahaan, mengidentifikasi keberadaan multikolinearitas, menganalisis dampaknya terhadap stabilitas model, serta menawarkan solusi apabila multikolinearitas terdeteksi. Solusi tersebut dapat berupa penghapusan variabel, transformasi data menggunakan *Principal Component Analysis* (PCA), atau penerapan metode regresi alternatif seperti *Ridge Regression*, LASSO *Regression*, dan *Partial Least Squares* (PLS). PCA digunakan untuk mengatasi masalah multikolinearitas dengan mengubah variabel asli yang berkorelasi menjadi sekumpulan variabel baru yang tidak berkorelasi, yang disebut komponen utama (Abdi & Williams, 2010). *Ridge Regression* adalah teknik regularisasi yang menambahkan penalti terhadap besar koefisien regresi dalam fungsi objektif. Menurut (Golam Kibria, 2003), estimator tipe *Ridge* dan tipe Liu merupakan metode *shrinkage* yang konsisten dan menarik untuk mengurangi efek multikolinearitas, baik dalam model regresi linier maupun non-linier. LASSO adalah metode shrinkage dan seleksi untuk regresi linear. Metode ini meminimalkan jumlah kuadrat galat seperti biasa, tetapi dengan batasan terhadap jumlah nilai absolut dari koefisien regresi (Tibshirani, 1996). Regresi *Partial Least Squares* (PLS) adalah metode regresi yang digunakan dalam berbagai ilmu terapan, terutama ketika terdapat banyak variabel namun jumlah sampel atau observasi relatif sedikit. PLS terbukti efektif dalam mengatasi masalah multikolinieritas seperti pada penelitian terdahulu (Abdi & Williams, 2010), (Swanson & Tayman, 2012). Penelitian ini menggabungkan lima pendekatan penanganan multikolinearitas dan membandingkan performanya dalam prediksi saham GOTO.

Secara keseluruhan, penelitian ini diharapkan memberikan kontribusi terhadap pengembangan ilmu statistik terapan, khususnya dalam bidang ekonometrika dan analisis regresi, dengan menunjukkan bagaimana gejala multikolinearitas memengaruhi model regresi serta bagaimana pendekatan alternatif dapat diterapkan untuk mengatasinya. Selain itu, penelitian ini juga memperkaya literatur empiris mengenai penerapan teknik statistik dalam menganalisis data keuangan di pasar modal Indonesia, khususnya terkait pergerakan harga saham emiten teknologi seperti GOTO.

2. METODE

2.1 Analisis Regresi

Analisis regresi adalah metode analisis yang bertujuan untuk mengetahui pengaruh variabel bebas terhadap variabel terikat. Menurut (Draper & Smith, 2014), analisis regresi merupakan metode yang dapat digunakan untuk menganalisis data dan mengambil kesimpulan tentang hubungan ketergantungan variabel terhadap variabel lainnya. Analisis regresi linier berganda merupakan suatu metode yang mengindentifikasi pengaruh antara dua atau lebih variabel bebas dengan satu

variabel terikat. Model regresi umum dengan k variabel bebasdapat dituliskan sebagai berikut:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \quad (2.1)$$

dengan:

Y	: Variabel dependen (respon)
β_0	: Intersep atau konstanta
$x_1, x_2, \ldots, x_k$	: Variabel independen (prediktor)
ε	: Galat (*error term*)

Estimasi parameter model regresi berganda pada Persamaan (1) dapat diperoleh dengan menggunakan metode kuadrat terkecil *Ordinary Least Square* (OLS) mendapatkan $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_k$yang *Best Linear Unbiased Estimator* (BLUE) sehingga menyebarnya persamaan regresi sedekat mungkin pada data aktualnya. Prinsip dasar metode OLS adalah meminimumkan jumlah kuadrat error, untuk mendapatkan nilai minimum dari fungsi maka syaratnya adalah differensiasi atau turunan pertama dari fungsi tersebut harus sama dengan nol. Nilai parameter β didapatkan dengan meminimumkan jumlah kuadrat error, yaitu sebagai berikut:

$$\boldsymbol{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \quad (2.2)$$

dengan:

$\boldsymbol{\beta}$	: Vektor koefisien regresi, yaitu $[\beta_0\ \beta_1 \ldots \beta_k]^T$
$\boldsymbol{X}$	: Matriks data variabel independen (termasuk kolom 1 untuk intersep)
$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$	: Invers dari hasil perkalian $\boldsymbol{X}^T\boldsymbol{X}$
$\boldsymbol{X}^T$	: Transpose dari matriks $\boldsymbol{X}$
$\boldsymbol{Y}$	: Vektor dari variabel dependen (nilai yang diamati)

2.2 Multikolinieritas

Istilah multikolinearitas awalnya diusulkan oleh Ragnar Frisch tahun 1934. Menurut (Greene, 2018), multikolinearitas atau kolinearitas ganda adalah adanya keterkaitan linier yang sangat erat antara beberapa atau semua variabel independen dalam model regresi. Multikolinearitas dapat menghasilkan koefisien regresi yang diperoleh dari analisis regresi berganda menjadi sangat lemah atau tidak memberikan hasil analisis yang mampu mewakili pengaruh variabel independen yang terlibat (Montgomery, Peck, & Vining, 2012). Selain itu, multikolinearitas dapat menyebabkan beberapa variabel yang sebenarnya signifikan menjadi tidak signifikan secara statistik (Shrestha, 2020). Dampak lain dari adanya multikolinearitas antara lain yaitu:

1. Nilai standar error cenderung akan semakin besar bersamaan dengan tingginya tingkat korelasi antarvariabel independen.

2. Nilai selang kepercayaan cenderung untuk lebih besar akibat besarnya standar *error* sehingga sangat sulit untuk menyangkal hipotesis nol dalam penelitian apabila terdapat multikolinearitas.
3. Dalam kasus multikolinearitas yang tinggi, mengakibatkan kemungkinan atau risiko gagal menolak hipotesis yang salah meningkat
4. Apabila multikolinearitas tinggi, memungkinkan diperoleh R^2 yang tinggi pula namun, tidak dapat menjelaskan sifat maupun pengaruh dari variabel independen yang bersangkutan.

Terdapat beberapa metode dalam mengidentifikasi adanya masalah multikolinearitas, antara lain sebagai berikut (Montgomery, Peck, & Vining, 2012):

1. Melihat korelasi antar variabel independen multikolinearitas dapat diduga dengan melihat tingginya korelasi antarvariabel independen. Apabila nilai dari koefisien korelasi antarvariabel independen diatas 0.5, diduga terjadi masalah multikolinearitas atau kolinearitas ganda antarvariabel independen.
2. Menggunakan *Variance Inflation Factor* (VIF), VIF merupakan alternatif cara untuk mengidetifikasi adanya masalah multikolinearitas. Peningkatan variansi bergantung dari σ^2 dan VIF itu sendiri. Untuk mencari nilai VIF, rumus yang digunakan adalah sebagai berikut:

$$VIF_{(j)} = \frac{1}{1-R_j^2} \tag{2.3}$$

merupakan koefisien determinasi ke-j (j = 1,2, … , k) yang diperoleh dari variabel independen Xj yang diestimasi dengan menggunakan variabel independen lainnya. Jika perolehan nilai VIF > 10, maka secara signifikan disimpulkan bahwa terjadi masalah multikolinearitas.

2.3 Metode Penanganan Multikolinieritas

2.3.1 *Principal Component Analysis* (PCA)

Principal Component Analysis (PCA) adalah teknik multivariat yang digunakan untuk menganalisis tabel data di mana pengamatan dijelaskan oleh beberapa variabel dependen kuantitatif yang saling berkorelasi. Tujuan PCA adalah untuk mengekstraksi informasi penting dari tabel tersebut, merepresentasikannya sebagai sekumpulan variabel ortogonal baru yang disebut komponen utama, serta menampilkan pola kesamaan pengamatan dan variabel dalam bentuk titik pada peta sebar (Verleysen & Verleysen, 2001). Kualitas model PCA dapat dievaluasi menggunakan teknik *cross-validation* seperti *bootstrap* dan *jackknife*. PCA dapat digeneralisasi sebagai *Correspondence Analysis* (CA) untuk menangani variabel kualitatif dan sebagai *Multiple Factor Analysis* (MFA) untuk menangani kumpulan variabel heterogen. Secara matematis, PCA bergantung pada dekomposisi eigen dari matriks positif semi-definit

dan pada dekomposisi nilai *singular* (*Singular Value Decomposition*/SVD) dari matriks persegi panjang (Abdi & Williams, 2010).

Komponen utama merupakan kombinasi linier dari variabel asli dengan bobot tertentu yang dihitung dari *eigen value*. PCA tidak mensyaratkan asumsi multivariat normal sehingga dapat diterapkan pada data dengan distribusi yang tidak normal. Banyaknya komponen utama ditentukan dengan beberapa metode, antara lain:

1. Berdasarkan proporsi kumulatif total keragaman yang mampu di jelaskan oleh k komponen utama minimal 80%, dan proporsi total variansi populasi bernilai cukup besar.
2. Berdasarkan nilai eigen dari komponen utama. Tapi hanya bisa diterapkan pada matriks korelasi. Jika nilai eigen lebih atau sama dengan satu.
3. Berdasarkan *scree plot*, dengan menggunakan metode ini banyaknya komponen utama yang di pilih yaitu k, adalah jika pada titik k tersebut *plot*-nya curam ke kiri , tetapi tidak curam ke kanan.

Dengan mereduksi dimensi melalui PCA, masalah multikolinearitas dapat diatasi karena komponen utama yang dihasilkan bebas dari korelasi.

2.3.2 *Ridge Regression*

Multikolinieritas yang terdapat dalam regresi linier berganda yang mengakibatkan matriks X^TX-nya hampir *singular* yang pada gilirannya menghasilkan nilai estimasi parameter yang tidak stabil. Regresi *Ridge* merupakan metode estimasi koefisien regresi yang diperoleh melalui penambahan konstanta bias c pada diagonal X^TX. Nilai c untuk koefisien regresi *Ridge* diantara 0 hingga 1.

Dalam analisis *Ridge Regression*, estimasi parameter Ridge (k) merupakan masalah penting. Terdapat banyak metode yang tersedia untuk memperkirakan parameter tersebut. Beberapa metode baru berbasis pendekatan regresi *Ridge* tergeneralisasi telah diusulkan dan dievaluasi melalui studi simulasi berdasarkan kriteria *Mean Squared Error* (MSE) minimum. Hasil simulasi menunjukkan bahwa dalam kondisi tertentu, estimator yang diusulkan memiliki performa lebih baik dibandingkan dengan *Least Squares Estimators* (LSE) dan estimator populer lainnya (Golam Kibria, 2003). Bentuk sederhana dari regresi *Ridge* adalah sebagai berikut:

$$\boldsymbol{\beta}^{(c)} = (\boldsymbol{X^T X} + c\boldsymbol{I})^{-1}\boldsymbol{X^T Y} \tag{2.4}$$

dengan:

$\boldsymbol{\beta}^{(c)}$: Estimasi parameter regresi hasil regularisasi dengan parameter c
$\boldsymbol{X}$: Matriks data variabel independen
$\boldsymbol{X^T}$: Transpose dari matriks $\boldsymbol{X}$
c : Parameter regularisasi (konstanta positif), yang mengontrol besarnya penalti

I : Matriks identitas berdimensi sama dengan $\boldsymbol{X^T X}$
$\boldsymbol{Y}$: Vektor dari variabel dependen (nilai yang diamati)

Umumnya sifat dari penafsiran *ridge* ini memiliki variansi yang minimum sehingga diperoleh nilai VIF-nya yang merupakan diagonal utama dari matriks

$$(\boldsymbol{X^T X} + c\boldsymbol{I})^{-1}\boldsymbol{X^T X}(\boldsymbol{X^T X} + c\boldsymbol{I})^{-1} \quad (2.5)$$

dengan:
$\boldsymbol{X^T X}$: Transpose dari matriks $\boldsymbol{X}$
$c\boldsymbol{I}$: Penambahan regularisasi (dari Ridge Regression)
$(\boldsymbol{X^T X} + cI)^{-1}$: Matriks identitas berdimensi sama dengan $\boldsymbol{X^T X}$
Pada dasarnya Regresi *Ridge* merupakan metode kuadrat terkecil. Perbedaannya adalah bahwa pada metode regresi *Ridge*, nilai variabel bebasnya ditransformasikan dahulu melalui prosedur *centering* dan *rescaling* (Wasilaine et al., 2014).

2.3.3 *Least Absolute Shrinkage and Selection Operator* (LASSO) *Regression*

Metode *Least Absolute Shrinkage and Selection Operator* (LASSO) diperkenalkan pertama kali oleh Tibshirani pada tahun 1996. LASSO menyusutkan koefisien regresi dari variabel prediktor yang memiliki korelasi tinggi dengan galat, menjadi tepat pada nol atau mendekati nol (Tibshirani, 1996).

Menurut (Zhao & Yu, 2006),persamaan secara umum LASSO dinyatakan sebagai berikut:

$$\boldsymbol{Y}^{**} = \boldsymbol{X}^{**}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^{**} \quad (2.6)$$

Keterangan
$\boldsymbol{Y}^{**}$ = Vektor variabel respon berukuran (n × 1)
$\boldsymbol{X}^{**}$ = Vektor variabel respon berukuran (n × p)
$\boldsymbol{\beta}$ = Vektor dari koefisien LASSO berukuran (k + 1) × 1
$\boldsymbol{\varepsilon}^{**}$ = Vektor galat berukuran $(n \times 1)$
Estimasi koefisien LASSO menggunakan pemrograman kuadratik dengan kendala pertidaksamaan. Estimasi lasso diperoleh dari persamaan berikut:

$$\hat{\beta}^{lasso} = argmin\left\{\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{k}\beta_j x_{ij})^2\right\} \quad (2.7)$$

dengan syarat $\sum_{j=1}^{k}|\beta_j| \leq t$. Nilai t merupakan parameter *tuning* yang mengontrol penyusutan koefisien LASSO dengan $t \geq 0$. Jika $t < t_0$ dengan $t_0 = \sum_{j=1}^{p}|\hat{\beta}_j|$ maka akan menyebabkan koefisien menjadi nol atau mendekati nol, sehingga LASSO akan berperan sebagai seleksi variabel. Akan tetapi jika $t > t_0$ maka penduga koefisien LASSO memberikan hasil yang sama dengan penduga kuadrat terkecil (Tibshirani, 1996).

Koefisien regresi LASSO ditentukan berdasarkan parameter *tuning* yang sudah dibakukan yaitu:

$$s = \frac{t}{\sum_{j=1}^{k}\left|\hat{\beta}_j^0\right|} \tag{2.8}$$

dengan $t = \sum_{j=1}^{p}\left|\hat{\beta}_j\right|$, $\hat{\beta}_j^0$ merupakan penduga terkecil untuk model penuh, nilai s optimal diperoleh melalui validasi silang (Dewi, 2010).

2.3.4 *Partial Least Squares (PLS) Regression*

Regresi *Partial Least Squares* (PLS) adalah metode regresi yang digunakan dalam berbagai ilmu terapan, terutama ketika terdapat banyak variabel namun jumlah sampel atau observasi relatif sedikit (Swanson & Tayman, 2012). PLS digunakan untuk membentuk komponen yang menangkap informasi dari variabel independen guna memprediksi variabel dependen. Komponen PLS ini diperoleh dari kombinasi linear variabel prediktor yang tidak saling berkorelasi sehingga dapat memaksimalkan kovariansi antara variabel independen dan dependen.

PLS lebih unggul dibandingkan regresi linier berganda dan regresi *ridge* dalam kondisi data dengan banyak variabel tetapi sedikit sampel. Keunggulan PLS terletak pada stabilitas prediktor yang dihasilkan, karena metode ini memilih komponen dengan pengurangan maksimal pada kovariansi data (Swanson & Tayman, 2012). Selain itu, PLS cenderung menghasilkan jumlah variabel yang lebih sedikit dibandingkan metode lain seperti kriteria Akaike atau Mallows Cp. Model regresi *Partial Least Square* dengan m komponen dapat dituliskan sebagai berikut:

$$Y = \sum_{h=1}^{m} c_h t_h + \varepsilon \tag{2.9}$$

dengan Y adalah variabel dependent, c_h adalah koefisien regresi Y terhadap t_h, $t_h = \sum_{j=1}^{p} w_{h(j)} X_j$ adalah komponen utama ke-h yang tidak saling berkorelasi, $(h = 1, 2, \ldots, m)$ dengan syarat komponen PLS $t_h = \sum_{j=1}^{p} w_{h(j)} X_j$ *orthogonal*.

Menurut Menurut (Swanson & Tayman, 2012), PLS juga melibatkan analisis data yang cermat, sehingga dapat mendeteksi outlier dan kelompok data yang berbeda. Hal ini membantu dalam menyaring variabel yang relevan dan memastikan bahwa komponen yang dihasilkan memiliki interpretasi yang jelas. Partial least squares termasuk dalam metode statistik berbasis kovarians yang sering disebut sebagai *Structural Equation Modeling* (SEM). Teknik ini dirancang untuk menangani regresi

berganda ketika data memiliki sampel kecil, nilai hilang, atau masalah multikolinearitas (Pirouz, 2012).

2.4 Jenis Penelitian

Jenis penelitian yang akan digunakan adalah penelitian terapan. Penelitian terapan adalah salah satu jenis penelitian yang bertujuan untuk memberikan solusi atas permasalahan tertentu secara praktis. Dalam penelitian ini akan dilakukan pengecekan data sampel terhadap asumsi multikolinearitas. Jika asumsi multikolinearitas tidak terpenuhi maka akan dilakukan penanganan menggunakan metode *partial least squares* dalam mengatasi masalah multikolinearitas dalamanalisis regresi berganda.

2.5 Pengumpulan Data

Dataset yang digunakan dalam analisis ini merupakan data historis saham GOTO (Gojek Tokopedia) yang mencakup periode dari tahun 2022 hingga awal 2025. Data ini diambil dari sumber pasar modal dan merepresentasikan pergerakan harga saham harian selama periode tersebut. Setiap baris pada *dataset* menunjukkan informasi harga saham dan volume perdagangan untuk satu hari tertentu.

Dalam *dataset* ini terdapat beberapa variabel yang digunakan untuk analisis. Variabel terikat dalam model adalah Terakhir, yang merepresentasikan harga penutupan saham pada akhir hari perdagangan. Ini merupakan variabel yang ingin diprediksi atau dijelaskan dalam konteks analisis regresi. Sementara itu, variabel bebas yang digunakan sebagai prediktor terdiri dari:

1. Pembukaan : Harga saham pada saat pasar dibuka.
2. Tertinggi : Harga tertinggi yang dicapai saham dalam satu hari perdagangan.
3. Terendah : Harga terendah yang dicapai saham dalam satu hari perdagangan.
4. Vol. : Volume perdagangan saham pada hari tersebut, yakni jumlah saham yang diperjualbelikan, dinyatakan dalam satuan juta (M) atau miliar (B), yang kemudian dikonversi ke bentuk numerik murni.

Selain itu, pada data mentah awal juga terdapat kolom Perubahan% yang menunjukkan persentase perubahan harga saham dari hari sebelumnya, namun kolom ini tidak digunakan dalam analisis regresi karena tidak memenuhi format numerik yang bersih serta tidak relevan secara langsung terhadap model prediksi harga penutupan.

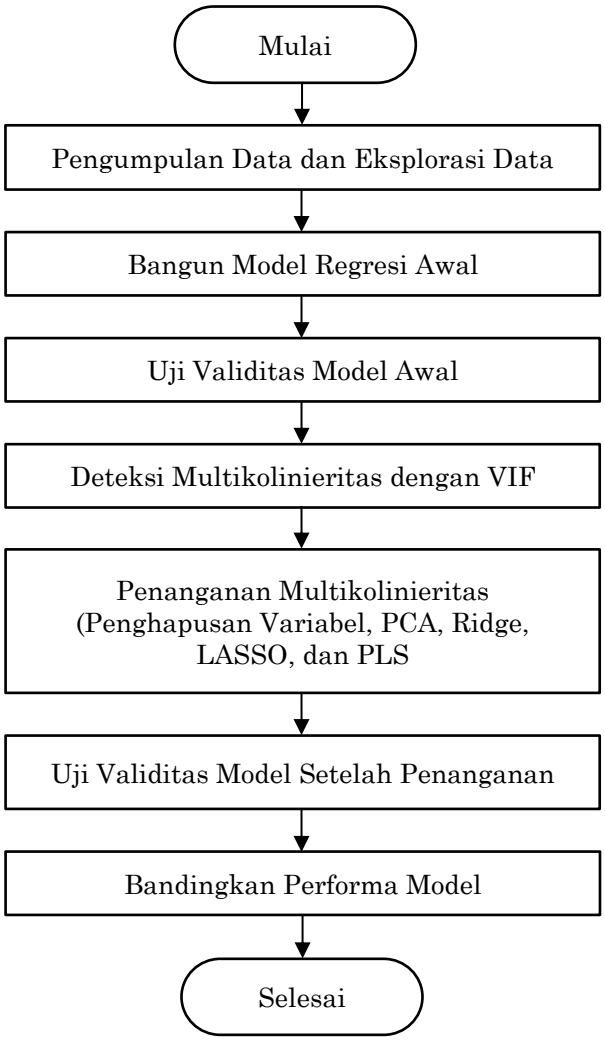
2.6 Pengolahan Data

Data yang digunakan dalam penelitian ini diolah menggunakan Python. Python dipilih karena kemampuannya yang fleksibel dan kuat dalam mengelola data, membangun model regresi, serta melakukan analisis statistik lanjutan. Berbagai pustaka, seperti pandas, numpy, dan matplotlib digunakan untuk manipulasi data dan visualisasi,

sedangkan *scikit-learn* dimanfaatkan untuk membangun model regresi berganda serta menerapkan teknik penanganan multikolinearitas seperti *Principal Component Analysis* (PCA), *Ridge Regression*, dan LASSO *Regression.*

2.7 *Research Step*

Berikut adalah langkah-langkah penelitian yang dilakukan:



3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Data

Data yang digunakan dalam penelitian ini adalah saham PT Gojek Tokopedia, Tbk (GOTO) dari tahun 2022 hingga 2025. Data ini berasal dari sumber investing.com, yaitu https://id.investing.com/equities/goto-gojek-tokopedia-pt-historical-data dan mencakup informasi pasar saham yang umum digunakan dalam analisis teknikal. Pada data ini mencakup informasi sebanyak 696 baris dan 7 kolom data harga saham harian.

Variabel-variabel penting *dataset* tersebut, yaitu Tanggal, Pembukaan, Tertinggi, Terendah, dan Vol. sebagai variabel independen, sedangkan variabel dependen, yaitu Terakhir.

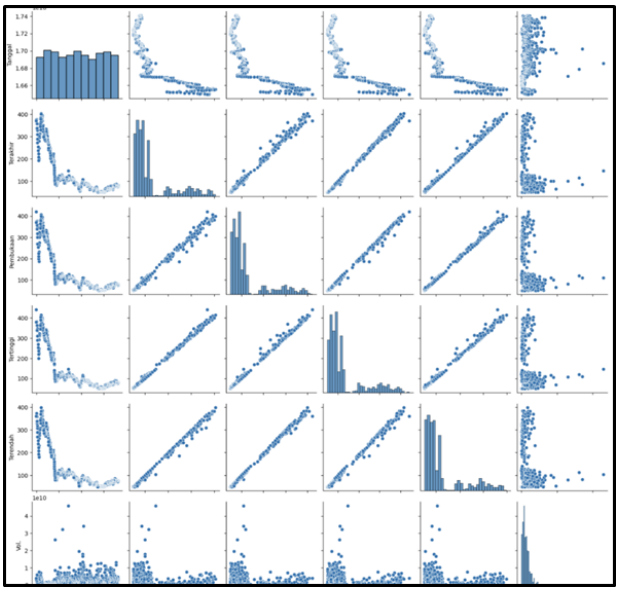
3.2 Eksplorasi Data dan Analisis Awal

3.2.1 Eksplorasi Data untuk Memahami Karakteristik *Dataset* (Statistik Deskriptif, Visualisasi Data)

Tahap awal analisis adalah eksplorasi data untuk memahami isi dan struktur *dataset*. Proses ini mencakup pemeriksaan jumlah observasi, jumlah variabel, dan tipe data (numerik, kategorik, atau waktu). Tujuannya adalah memperoleh gambaran awal sehingga dapat memilih pendekatan analisis yang tepat. Analisis statistik deskriptif dilakukan pada variabel numerik untuk menghitung nilai minimum, maksimum, rata-rata, *median*, dan standar deviasi. Hal ini membantu memahami sebaran data serta mendeteksi outlier atau anomali, seperti volume transaksi yang terlalu tinggi.

Tahap eksplorasi ini juga mencakup penilaian terhadap konsistensi dan kelengkapan data. Jika terdapat nilai kosong (*missing values*) atau duplikasi, maka perlu dipertimbangkan apakah akan dilakukan imputasi, penghapusan, atau perlakuan khusus terhadap data tersebut. Dengan memahami kondisi awal data, potensi masalah dalam analisis seperti multikolinearitas atau heteroskedastisitas dapat lebih dini terdeteksi.

Untuk memperjelas hubungan antar variabel numerik, disusun pula visualisasi berupa *scatter plot matrix* atau *pairplot* yang menunjukkan hubungan linier antar pasangan variabel. Visualisasi ini memungkinkan pengamatan pola keterkaitan secara lebih intuitif dan visual, sehingga mempermudah identifikasi awal terhadap adanya hubungan yang kuat atau sangat serupa antara dua atau lebih variabel. Pola yang sangat selaras atau tumpang tindih antar variabel dalam *plot* tersebut dapat menjadi indikasi awal adanya korelasi tinggi yang berpotensi menimbulkan multikolinearitas dalam model regresi yang akan dibangun.



Gambar 1. Plot Hubungan Linier antara Satu Variabel dengan Variabel Lainnya

Berdasarkan visualisasi *pairplot*, terlihat hubungan antar variabel dalam model, yaitu Tanggal, Terakhir, Pembukaan, Tertinggi, Terendah, dan Volume. *Pairplot* ini membantu mengidentifikasi pola hubungan linier atau *non*-linier antar variabel serta distribusi masing-masing variabel. Variabel numerik seperti Terakhir, Pembukaan, Tertinggi, dan Terendah memiliki distribusi mirip dengan pola miring ke kanan (*right-skewed*), menunjukkan sebagian besar nilai berada pada kisaran rendah dengan beberapa nilai jauh lebih tinggi. Hal ini umum pada data saham yang mengalami perubahan harga signifikan dalam waktu singkat.

Pada grafik sebar antar pasangan variabel seperti Terakhir vs Pembukaan, Tertinggi, dan Terendah, terlihat adanya pola hubungan linier yang sangat kuat. Titik-titik data membentuk garis lurus atau mendekati garis lurus, menunjukkan bahwa variabel-variabel ini memiliki hubungan yang erat satu sama lain. Hal ini mengindikasikan adanya potensi multikolinearitas, yaitu kondisi di mana dua atau lebih variabel independen dalam model regresi memiliki korelasi yang sangat tinggi. Ini bisa menjadi masalah dalam model karena dapat memengaruhi kestabilan estimasi koefisien regresi.

Sementara itu, Volume (Vol.) tidak menunjukkan pola hubungan yang jelas dengan variabel lainnya. Sebaran titiknya terlihat acak dan menyebar, baik saat dibandingkan dengan Terakhir, Tertinggi, maupun variabel harga lainnya. Ini mengisyaratkan bahwa volume perdagangan tidak memiliki hubungan linier yang kuat dengan harga saham pada data ini, atau hubungan tersebut bisa saja bersifat non-linier.

Variabel Tanggal tampak tidak memberikan informasi hubungan yang berarti jika disertakan dalam analisis korelasi. Sebaran titik pada kolom dan baris yang melibatkan Tanggal terlihat tidak menunjukkan pola jelas, sehingga variabel ini

kemungkinan besar hanya digunakan untuk urutan kronologis, bukan sebagai bagian dari prediktor dalam model.

3.2.2 Membangun Model Regresi Berganda Awal Menggunakan Semua Variabel Independen

Setelah variabel dibersihkan dan dipersiapkan, model regresi linier berganda dibangun untuk menganalisis pengaruh variabel bebas terhadap harga penutupan saham. Model ini digunakan untuk menguji apakah harga pembukaan, harga tertinggi, harga terendah, dan Volume perdagangan memiliki keterkaitan yang signifikan terhadap harga akhir harian. Hasil pemodelan menghasilkan koefisien regresi yang menggambarkan arah dan kekuatan pengaruh masing-masing variabel bebas. Selain itu, nilai koefisien determinasi menunjukkan seberapa baik model mampu menjelaskan variasi pada harga penutupan. Informasi mengenai tingkat signifikansi dan ketidakpastian estimasi juga disertakan untuk mengidentifikasi variabel yang memiliki pengaruh nyata terhadap harga saham GOTO selama periode pengamatan. Berikut adalah hasil regresi berganda:

Hasil Regresi Berganda:

OLS Regression Results

Dep. Variable:	Terakhir	R-squared:	0.999
Model:	OLS	Adj. R-squared:	0.999
Method:	Least Squares	F-statistic:	1.705e+05
Date:	Mon, 12 May 2025	Prob (F-statistic):	0.00
Time:	07:04:32	Log-Likelihood:	-1722.8
No. Observations:	696	AIC:	3456.
Df Residuals:	691	BIC:	3478.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.4446	0.249	-1.786	0.075	-0.933	0.044
Pembukaan	-0.5975	0.028	-21.500	0.000	-0.652	-0.543
Tertinggi	0.7753	0.022	35.685	0.000	0.733	0.818
Terendah	0.8213	0.029	28.253	0.000	0.764	0.878
Vol.	1.407e-10	3.51e-11	4.011	0.000	7.18e-11	2.1e-10

Omnibus:	336.190	Durbin-Watson:	1.715
Prob(Omnibus):	0.000	Jarque-Bera (JB):	17337.106
Skew:	1.381	Prob(JB):	0.00
Kurtosis:	27.294	Cond. No.	1.12e+10

Gambar 2. Hasil Regresi Linier Berganda

Berdasarkan Gambar 2, hasil regresi linier berganda menunjukkan bahwa model yang digunakan memiliki kemampuan yang sangat baik dalam menjelaskan variasi pada variabel dependen, yaitu harga penutupan saham. Hal ini dapat dilihat dari nilai R-*squared* sebesar 0,999 dan *Adjusted* R-*squared* sebesar 0,999, yang menunjukkan bahwa sekitar 99,9% variasi harga penutupan saham dapat dijelaskan oleh variabel harga pembukaan, harga tertinggi, harga terendah, dan volume perdagangan.

Selanjutnya, hasil uji F menunjukkan nilai Prob(F-statistic) sebesar 0,000, yang berarti model regresi secara simultan signifikan pada tingkat signifikansi 5%. Dengan demikian, dapat disimpulkan bahwa secara bersama-sama variabel harga pembukaan, harga tertinggi, harga terendah, dan volume perdagangan berpengaruh signifikan terhadap harga penutupan saham.

Berdasarkan hasil uji parsial, variabel harga pembukaan memiliki koefisien sebesar -0,5975 dengan nilai p-*value* 0,000, yang menunjukkan bahwa variabel tersebut berpengaruh signifikan terhadap harga penutupan saham. Variabel harga tertinggi memiliki koefisien sebesar 0,7753 dengan nilai p-*value* 0,000, sehingga juga berpengaruh signifikan dan memiliki hubungan positif dengan harga penutupan saham. Selanjutnya, variabel harga terendah memiliki koefisien sebesar 0,8213 dengan nilai p-value 0,000, yang menunjukkan adanya pengaruh positif dan signifikan terhadap harga penutupan saham. Variabel volume perdagangan juga menunjukkan pengaruh yang signifikan dengan koefisien sebesar $1{,}407 \times 10^{-10}$ dan p-*value* 0,000.

Selain itu, nilai Durbin-Watson sebesar 1,715 menunjukkan bahwa tidak terdapat indikasi autokorelasi yang kuat pada model regresi. Secara keseluruhan, hasil analisis regresi linier berganda ini menunjukkan bahwa variabel harga pembukaan, harga tertinggi, harga terendah, dan volume perdagangan memiliki kontribusi yang signifikan dalam menjelaskan pergerakan harga penutupan saham GOTO selama periode pengamatan.

Berikut adalah hasil tabel ANOVA:

Tabel ANOVA:

	SS	df	MS	F
Regression	5.680172e+06	4.0	1420042.995937	170479.160889
Residual	5.755834e+03	691.0	8.329716	
Total	5.685928e+06	695.0		

Gambar 3. Tabel ANOVA

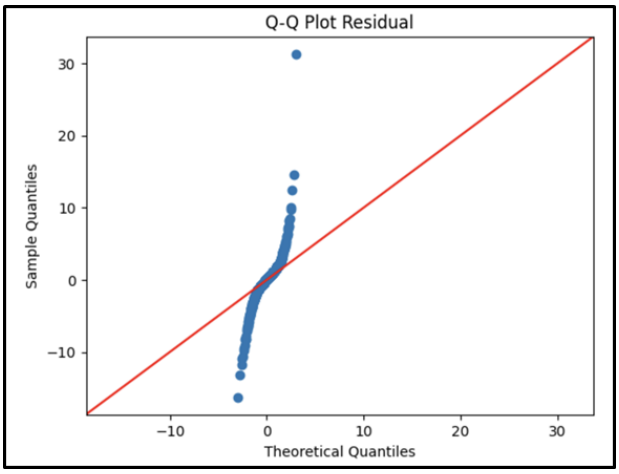
Jadi, model regresi yang dihasilkan adalah

$$Terakhir = -0.4446 - 0.5975Pembukaan + 0.7753Tertinggi + 0.8213Terendah + 1.407 \times 10^{-10}Vol$$

Selanjutnya, dilakukan pengujian validitas model regresi untuk memastikan bahwa model yang dihasilkan memenuhi asumsi-asumsi klasik regresi. Uji validitas bertujuan untuk mengevaluasi apakah residual dari model memenuhi syarat kenormalan, tidak terdapat autokorelasi, dan tidak mengalami heteroskedastisitas. Berikut adalah hasil dari uji validitas model awal:

Tabel 1. Hasil Uji Validitas Model Awal

Uji Validitas	Statistik	Nilai	$p - value$	Kesimpulan
Uji Normalitas Residual	JB *Statistic*	17337.1057	0.0000	Residual tidak berdistribusi normal ($p < 0.05$)
Uji Autokorelasi	Durbin-Watson	1.7154	-	Tidak ada autokorelasi (nilai mendekati 2)
Uji Heterokedastisitas	LM *Statistic*	174.7988	0.0000	Terdapat heteroskedastisitas ($p < 0.05$)
	F *Statistic*	57.9363	0.0000	



Gambar 4. Q-Q *Plot* Residual Model Awal

Berdasarkan Gambar 4, Q-Q *Plot* residual digunakan untuk melihat apakah residual model mengikuti distribusi normal. Pada *plot* terlihat bahwa sebagian titik residual berada di sekitar garis diagonal, namun terdapat beberapa titik yang menyimpang terutama pada bagian ekor distribusi. Hal ini menunjukkan bahwa residual model awal belum sepenuhnya mengikuti distribusi normal.

3.2.2 Interpretasi Hasil Regresi (Koefisien, R-*squared*, *p-value*)

Nilai R-*squared* yang mencapai 0.999 menunjukkan bahwa hampir seluruh perubahan pada harga penutupan dapat dijelaskan oleh kombinasi dari keempat variabel bebas: harga pembukaan, harga tertinggi, harga terendah, dan volume perdagangan. Nilai ini juga menunjukkan bahwa model memiliki tingkat akurasi yang sangat tinggi terhadap data yang digunakan. Dari sisi interpretasi masing-masing variabel, terlihat bahwa harga tertinggi dan harga terendah memiliki koefisien positif yang cukup besar, yaitu masing-masing 0.7753 dan 0.8213, serta nilai p-value yang sangat kecil (0.000), menunjukkan bahwa keduanya memberikan pengaruh signifikan secara statistik terhadap harga penutupan. Artinya, semakin tinggi nilai tertinggi atau terendah dalam suatu hari perdagangan, maka cenderung akan diikuti oleh harga penutupan yang juga tinggi. Sebaliknya, variabel harga pembukaan menunjukkan koefisien negatif sebesar -0.5975 dan juga signifikan secara statistik. Hal ini menunjukkan bahwa dalam data yang diamati, ketika harga pembukaan meningkat, justru harga penutupan cenderung mengalami penurunan, yang bisa mengindikasikan adanya tren koreksi atau penurunan setelah pembukaan yang tinggi. Untuk variabel volume perdagangan, meskipun koefisiennya sangat kecil secara absolut (1.407e-10), nilainya tetap signifikan secara statistik dengan p-value yang juga 0.000. Ini menunjukkan bahwa volume juga mempengaruhi harga penutupan, meskipun pengaruhnya secara numerik sangat kecil karena satuannya besar (dalam satuan miliar). Namun, ada catatan penting dari hasil ini. Nilai *condition number* yang sangat besar (1.12e+10) mengindikasikan adanya kemungkinan multikolinearitas yang kuat di antara variabel bebas. Ini bisa

menyebabkan estimasi koefisien menjadi tidak stabil dan perlu penanganan lebih lanjut.

3.3 Deteksi Multikolininearitas

3.3.1 Nilai VIF (*Variance Inflation Factor*)

Setelah model regresi terbentuk, langkah selanjutnya adalah memeriksa adanya korelasi yang terlalu kuat antar variabel bebas. Hal ini penting karena jika terdapat hubungan yang sangat erat antara dua atau lebih variabel bebas, model dapat mengalami kesulitan dalam membedakan pengaruh masing-masing variabel secara akurat. Untuk mendeteksi multikolinearitas, digunakan *Variance Inflation Factor* (VIF), yang mengukur seberapa besar variabilitas koefisien regresi meningkat akibat adanya korelasi antar variabel bebas. Jika nilai VIF melebihi 10, hal tersebut mengindikasikan adanya multikolinearitas yang serius. Apabila ditemukan variabel dengan nilai VIF yang tinggi, maka perlu dipertimbangkan apakah variabel tersebut akan tetap digunakan dalam model, dimodifikasi, atau bahkan dihilangkan. Hal ini penting agar interpretasi hasil regresi lebih akurat dan stabil.

Tabel 2. Hasil Analisis Multikolinieritas

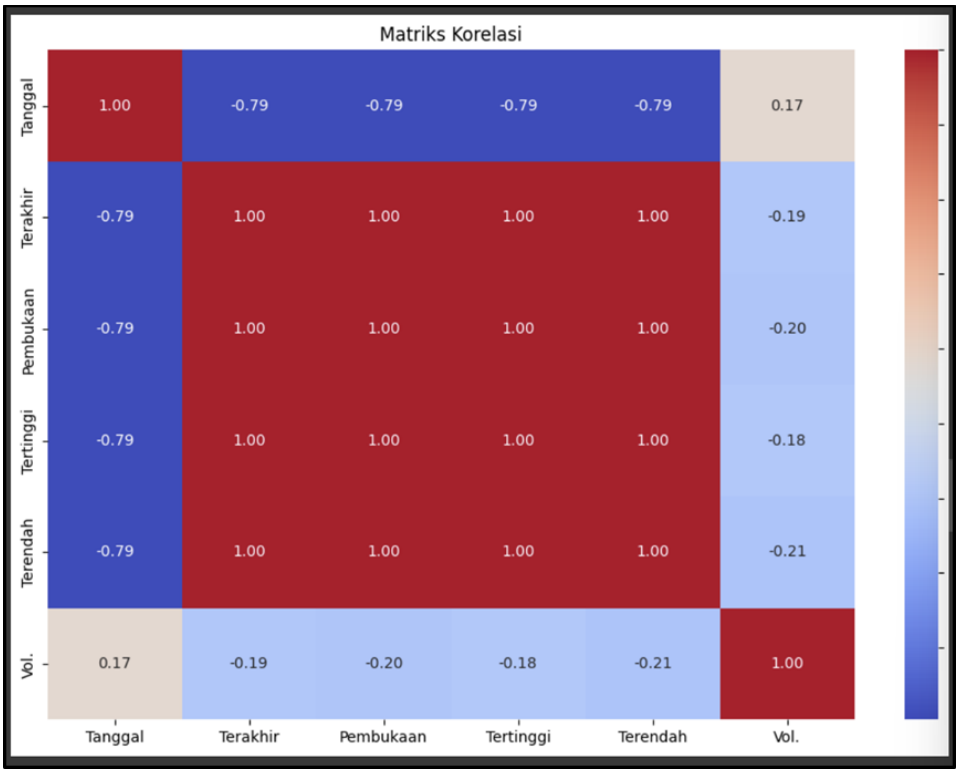
Variable	VIF
Pembukaan	1602.096486
Tertinggi	951.095247
Terendah	1488.741362
Vol.	1.451374

Berdasarkan hasil perhitungan *Variance Inflation Factor* (VIF), ditemukan adanya masalah serius pada model regresi, terutama pada tiga variabel bebas yaitu Pembukaan, Tertinggi, dan Terendah. Ketiga variabel ini memiliki nilai VIF yang sangat tinggi, bahkan mencapai ribuan. Hal ini menunjukkan adanya korelasi yang sangat kuat antar variabel, sehingga sulit membedakan pengaruh masing-masing secara independen. Kondisi ini mengindikasikan bahwa variabel tersebut membawa informasi yang hampir sama, sehingga dapat menyebabkan estimasi koefisien menjadi tidak stabil dan sulit diinterpretasikan. Di sisi lain, variabel Volume memiliki nilai VIF yang rendah, sekitar 1.45, yang menandakan tidak adanya korelasi kuat dengan variabel bebas lainnya. Oleh karena itu, variabel Volume aman dari indikasi multikolinearitas dan dapat digunakan dalam model tanpa risiko distorsi estimasi.

3.3.2 Korelasi Antar Variabel

Untuk memahami hubungan antar variabel dalam dataset saham PT GoTo Gojek Tokopedia Tbk (GOTO), dilakukan analisis korelasi antar variabel numerik. Analisis ini penting untuk mengidentifikasi sejauh mana hubungan linear terjadi antara variabel-variabel tersebut dan mendeteksi potensi multikolinearitas yang dapat memengaruhi kestabilan model regresi. Visualisasi korelasi menggunakan heatmap dipilih karena

mampu menampilkan kekuatan dan arah hubungan secara jelas. Pada heatmap, warna merah tua menunjukkan korelasi positif yang sangat kuat, sedangkan warna biru tua menunjukkan korelasi negatif yang kuat. Nilai korelasi berkisar dari -1 hingga 1, di mana nilai mendekati 1 berarti hubungan positif sempurna, nilai mendekati -1 berarti hubungan negatif sempurna, dan nilai sekitar 0 menunjukkan tidak adanya hubungan linear yang signifikan. Dengan demikian, analisis ini tidak hanya memberikan gambaran awal mengenai struktur data, tetapi juga berfungsi sebagai alat diagnosis untuk mengidentifikasi kemungkinan adanya multikolinearitas dalam model regresi linear berganda. Hasil dari analisis korelasi melalui visualisasi *heatmap* adalah sebagai berikut:



Gambar 5. Matriks Korelasi

Output di atas adalah *heatmap* korelasi yang menunjukkan kekuatan dan arah hubungan antar variabel numerik dalam dataset saham GOTO. Warna merah menandakan korelasi positif kuat, sedangkan warna biru menunjukkan korelasi negatif. Nilai korelasi berkisar antara -1 hingga 1, dengan nilai mendekati 1 atau -1 menunjukkan hubungan sangat erat.

Terlihat bahwa variabel Pembukaan, Tertinggi, dan Terendah memiliki korelasi sempurna yaitu 1.00 dengan variabel Terakhir, serta antar ketiga variabel tersebut juga sangat berkorelasi. Hal ini mengindikasikan multikolinearitas tinggi di antara variabel bebas tersebut. Sebaliknya, variabel Volume menunjukkan korelasi lemah dengan variabel harga, berkisar antara -0.18 hingga -0.21, yang berarti volume

perdagangan tidak berkaitan kuat secara linier dengan harga saham. Variabel Tanggal memiliki korelasi negatif sekitar -0.79 terhadap variabel harga, mengindikasikan tren penurunan harga seiring waktu, meski ini bisa dipengaruhi oleh representasi tanggal sebagai data numerik yang meningkat *linear*.

Secara keseluruhan, *heatmap* ini menguatkan adanya multikolinearitas tinggi antara beberapa variabel bebas, yang dapat menyebabkan ketidakstabilan dalam estimasi model regresi dan perlu penanganan lebih lanjut. Multikolinearitas seperti ini dapat membuat koefisien regresi menjadi tidak konsisten dan sulit diinterpretasikan. Oleh karena itu, langkah pengurangan variabel atau teknik transformasi data sangat dianjurkan agar model menjadi lebih stabil dan hasil analisis lebih dapat dipercaya.

3.4 Penanganan Multikolinieritas

3.4.1 Penghapusan Variabel dengan VIF Tinggi atau Korelasi Tinggi

Salah satu cara paling sederhana dan efektif untuk mengatasi multikolinearitas adalah dengan menghapus variabel yang memiliki korelasi sangat tinggi atau nilai *Variance Inflation Factor* (VIF) melebihi ambang batas. Dalam analisis regresi, VIF di atas 10 biasanya menandakan multikolinearitas serius yang bisa mengganggu kestabilan estimasi parameter model. Oleh karena itu, variabel dengan VIF > 10 sebaiknya dipertimbangkan untuk dihapus, terutama jika informasi yang dibawanya sudah tumpang tindih dengan variabel lain. Berikut adalah nilai VIF sebelum penghapusan:

Tabel 3. Nilai VIF Sebelum Penghapusan

Variable	VIF
Pembukaan	1602.096486
Tertinggi	951.095247
Terendah	1488.741362
Vol.	1.451374

Berdasarkan tabel di atas, variabel Pembukaan, Tertinggi, dan Terendah memiliki nilai VIF yang sangat tinggi, masing-masing di atas 900, bahkan Pembukaan mencapai lebih dari 1600. Hal ini menunjukkan adanya multikolinearitas yang sangat kuat di antara variabel-variabel tersebut, dengan hubungan hampir *linear* sempurna. Sebaliknya, variabel Volume memiliki VIF rendah sekitar 1.45, menandakan tidak adanya korelasi linier kuat dengan variabel lain dan relatif bebas dari multikolinearitas. Kondisi ini mengindikasikan perlunya penghapusan variabel yang terlalu berkorelasi agar model regresi lebih stabil dan koefisiennya lebih dapat dipercaya. Oleh karena itu, variabel Pembukaan yang memiliki VIF tertinggi dihapus dari model, sebagaimana berikut:

Tabel 4. Nilai VIF Setelah Menghapus Variabel Pembukaan

Variable	VIF
Const	5.163415
Tertinggi	292.354362
Terendah	296.005523

Vol.	1.337031

Setelah Penghapusan Pembukaan, nilai VIF Tertinggi dan Terendah masih sangat tinggi, masing-masing sekitar 292 dan 296. Artinya, keduanya juga masih memiliki korelasi yang sangat kuat satu sama lain, sehingga belum cukup untuk mengatasi multikolinearitas hanya dengan menghapus satu variabel. Langkah berikutnya adalah menghapus variabel Terendah.

Tabel 5. Nilai VIF Setelah Menghapus Variabel Terendah

Variable	VIF
Const	4.423684
Tertinggi	1.033383
Vol.	1.033383

Setelah penghapusan variabel Pembukaan, nilai VIF untuk variabel bebas yang tersisa turun drastis menjadi sekitar 1.03. Ini menandakan bahwa multikolinearitas sudah tidak lagi menjadi masalah, karena variabel-variabel tersebut kini cukup independen satu sama lain. Proses ini menunjukkan bahwa penghapusan satu variabel saja belum cukup untuk mengatasi multikolinearitas; perlu dilakukan eliminasi berkelanjutan hingga semua variabel memenuhi batas VIF yang wajar. Berikut adalah hasil analisis regresi linier berganda setelah penghapusan variabel dengan nilai VIF tinggi:

Hasil Regresi Setelah Penghapusan Variabel dengan VIF Tinggi:
OLS Regression Results

Dep. Variable:	Terakhir	R-squared:	0.998
Model:	OLS	Adj. R-squared:	0.998
Method:	Least Squares	F-statistic:	1.574e+05
Date:	Mon, 12 May 2025	Prob (F-statistic):	0.00
Time:	10:07:14	Log-Likelihood:	-1992.4
No. Observations:	696	AIC:	3991.
Df Residuals:	693	BIC:	4004.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.6035	0.338	4.737	0.000	0.939	2.268
Tertinggi	0.9631	0.002	551.145	0.000	0.960	0.967
Vol.	-1.858e-10	4.54e-11	-4.097	0.000	-2.75e-10	-9.68e-11

Omnibus:	690.567	Durbin-Watson:	1.488
Prob(Omnibus):	0.000	Jarque-Bera (JB):	71545.484
Skew:	-4.157	Prob(JB):	0.00
Kurtosis:	51.969	Cond. No.	1.03e+10

Gambar 6. Hasil Regresi Setelah Menghapus Variabel dengan VIF Tinggi

Berdasarkan hasil regresi setelah penghapusan variabel dengan nilai VIF tinggi, model menunjukkan performa yang sangat baik dengan nilai R-*squared* sebesar 0.998, yang berarti 99,8% variasi harga penutupan dapat dijelaskan oleh variabel Tertinggi dan Volume. Nilai *Adjusted* R-*squared* yang sama memperkuat kualitas model meskipun menggunakan lebih sedikit variabel. Koefisien intersep sebesar 1.6035 signifikan secara statistik ($p < 0.001$), begitu pula koefisien variabel Tertinggi (0.9631) dan Volume (-1.858e-10) yang keduanya juga signifikan dengan $p < 0.001$. Artinya, kenaikan satu satuan pada Tertinggi akan menaikkan harga penutupan sekitar 0.96, sementara

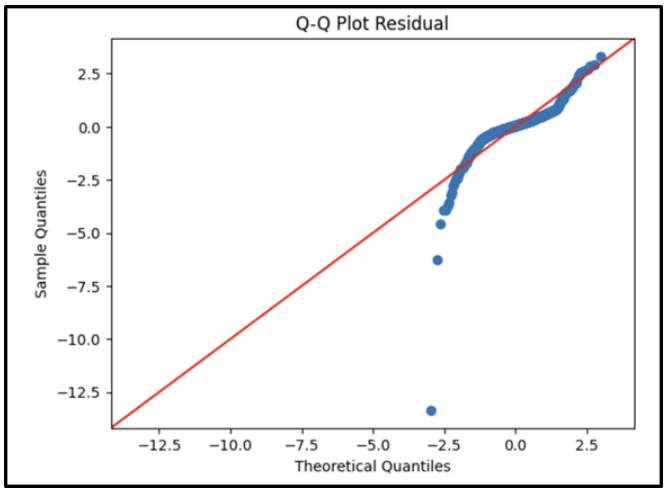
peningkatan Volume sedikit menurunkan harga penutupan, dengan asumsi variabel lain konstan.

Statistik F yang sangat besar (1.574e+05) dan $p-value$ 0.000 menunjukkan model secara keseluruhan signifikan. Dari sisi diagnostik, nilai Durbin-Watson sebesar 1.488 mengindikasikan tidak ada autokorelasi serius pada residual, tetapi nilai statistik Omnibus dan Jarque-Bera yang tinggi dengan p = 0.000 menunjukkan residual tidak memenuhi asumsi normalitas. Hal ini perlu diperhatikan dalam evaluasi model selanjutnya. Secara keseluruhan, model regresi ini berhasil menjelaskan variasi harga penutupan dengan baik dan mengurangi masalah multikolinearitas pada variabel bebas.

Selanjutnya, dilakukan pengujian validitas model regresi setelah penghapusan variabel dengan VIF tinggi untuk memastikan bahwa model yang dihasilkan memenuhi asumsi-asumsi klasik regresi. Uji validitas bertujuan untuk mengevaluasi apakah residual dari model memenuhi syarat kenormalan, tidak terdapat autokorelasi, dan tidak mengalami heteroskedastisitas. Berikut adalah hasil dari uji validitas model setelah penghapusan nilai VIF tinggi:

Tabel 6. Hasil Uji Validitas Model Setelah Penghapusan Variabel

Uji Validitas	Statistik	Nilai	$p-value$	Kesimpulan
Uji Normalitas Residual	JB *Statistic*	71545.4840	0.0000	Residual tidak berdistribusi normal ($p < 0.05$)
	Shapiro-Wilk *Statistic*	0.7239	0.0000	
Uji Autokorelasi	Durbin-Watson	1.4875	-	Tidak ada autokorelasi (nilai mendekati 2)
Uji Heterokedastisitas	LM *Statistic*	55.6129	0.0000	Terdapat heteroskedastisitas ($p < 0.05$)
	F *Statistic*	30.0909	0.0000	



Gambar 7. Q-Q Plot residual Model Setelah Penghapusan Variabel

3.4.2. Transformasi Data menggunakan Teknik PCA (*Principal Component Analysis*)

Selain metode penghapusan variabel berdasarkan nilai VIF yang tinggi, teknik lain yang dapat digunakan untuk mengatasi multikolinearitas adalah *Principal Component Analysis* (PCA). PCA bekerja dengan mentransformasikan kumpulan variabel yang saling berkorelasi menjadi sekumpulan komponen utama yang bebas dari korelasi satu sama lain. Komponen utama ini merupakan kombinasi linear dari variabel-variabel asli dan disusun sedemikian rupa sehingga masing-masingnya menjelaskan variansi data secara maksimal. Dengan menggunakan PCA, dimensi data dapat direduksi tanpa kehilangan banyak informasi penting, sekaligus menghilangkan multikolinearitas dalam model regresi.

```
Mengatasi Multikolinearitas dengan PCA:

Proporsi Varian yang Dijelaskan oleh Komponen PCA:
Komponen 1: 0.7632
Komponen 2: 0.2360
Komponen 3: 0.0005
Komponen 4: 0.0003

Jumlah komponen utama yang dipilih untuk menjelaskan ≥95% varian: 2
```

Gambar 8. Hasil Transformasi Data Menggunakan PCA

Berdasarkan hasil transformasi PCA pada dataset saham GOTO, dua komponen utama menjelaskan lebih dari 95% variansi total, dengan komponen pertama sebesar 76,32% dan komponen kedua 23,60%. Ini berarti sebagian besar informasi variabel asli dapat direpresentasikan oleh kedua komponen tersebut. Selanjutnya, kedua komponen utama ini digunakan sebagai variabel prediktor dalam model regresi *linear* berganda untuk memprediksi harga penutupan (Terakhir).

Selanjutnya, model regresi dibangun dengan kedua komponen utama sebagai variabel independen dan harga penutupan sebagai variabel dependen. Analisis dilakukan untuk mengevaluasi kemampuan kedua komponen tersebut dalam menjelaskan variasi harga penutupan, termasuk pemeriksaan koefisien regresi, nilai $p - value$, dan nilai R-squared sebagai ukuran kecocokan model. Berikut disajikan hasil regresi linear berganda menggunakan kedua *komponen* utama, yang menunjukkan kontribusi masing-masing komponen dalam memprediksi harga penutupan serta keakuratan dan reliabilitas model.

Hasil Regresi dengan PCA:

OLS Regression Results

Dep. Variable:	Terakhir	R-squared:	0.997
Model:	OLS	Adj. R-squared:	0.997
Method:	Least Squares	F-statistic:	1.257e+05
Date:	Mon, 12 May 2025	Prob (F-statistic):	0.00
Time:	10:15:22	Log-Likelihood:	-2070.6
No. Observations:	696	AIC:	4147.
Df Residuals:	693	BIC:	4161.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	128.4440	0.180	713.380	0.000	128.090	128.797
x1	51.3904	0.103	498.710	0.000	51.188	51.593
x2	9.4466	0.185	50.977	0.000	9.083	9.810

Omnibus:	306.759	Durbin-Watson:	1.638
Prob(Omnibus):	0.000	Jarque-Bera (JB):	14214.606
Skew:	1.213	Prob(JB):	0.00
Kurtosis:	25.006	Cond. No.	1.80

Gambar 9. Hasil Regresi Menggunakan PCA

Berdasarkan hasil tersebut, regresi linear berganda menggunakan dua komponen utama PCA menunjukkan performa yang sangat baik. Nilai R-*squared* sebesar 0,997 mengindikasikan bahwa 99,7% variasi harga penutupan dapat dijelaskan oleh kombinasi linear kedua komponen utama, dan *Adjusted* R-*squared* yang sama memperkuat bahwa model tidak *overfitting* dan efisien dalam menjelaskan pola data.

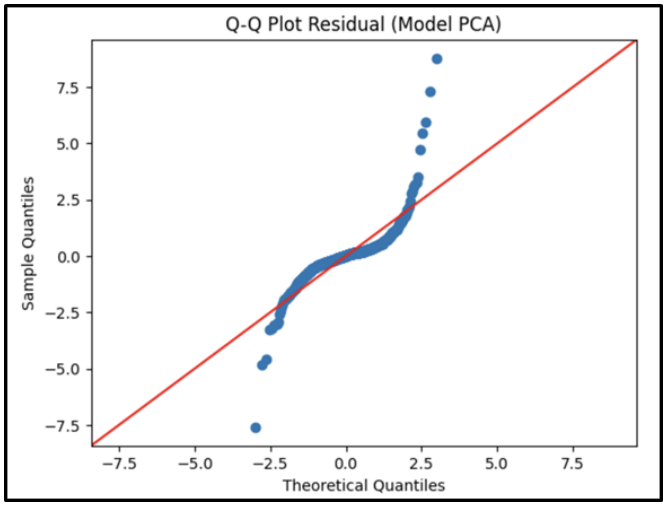
Koefisien regresi menunjukkan intersep sebesar 128,4440 yang signifikan secara statistik, dengan komponen pertama (x_1) berkontribusi positif sebesar 51,3904 dan komponen kedua (x_2) sebesar 9,4466, keduanya signifikan dengan p-value < 0,001. Ini berarti setiap kenaikan satu unit pada x1 atau x2, dengan asumsi variabel lain tetap, akan meningkatkan harga penutupan sesuai nilai koefisien masing-masing. Statistik F yang sangat besar dengan p-value 0,000 menunjukkan model secara keseluruhan signifikan. Nilai Durbin-Watson 1,638 menandakan tidak ada autokorelasi serius pada residual, meskipun tes Omnibus dan Jarque-Bera menunjukkan adanya pelanggaran asumsi normalitas residual.

Secara keseluruhan, penerapan PCA berhasil mereduksi dimensi dan mengatasi multikolinearitas, sekaligus menghasilkan model regresi yang sangat baik dalam menjelaskan variasi harga penutupan saham. Pendekatan ini juga menyederhanakan interpretasi model dengan mengurangi jumlah variabel prediktor tanpa kehilangan informasi penting. Dengan demikian, PCA menjadi solusi efektif untuk meningkatkan stabilitas dan keandalan model regresi dalam konteks data yang kompleks.

Selanjutnya, dilakukan pengujian validitas model regresi setelah penanganan multikolinieritas menggunakan transformasi PCA untuk memastikan bahwa model yang dihasilkan memenuhi asumsi-asumsi klasik regresi. Berikut adalah hasil dari uji validitas model setelah transformasi PCA:

Tabel 7. Hasil Uji Validitas Model Setelah Transformasi PCA

Uji Validitas	Statistik	Nilai	$p-value$	Kesimpulan
Uji Normalitas Residual	JB *Statistic*	14214.6064	0.0000	Residual tidak berdistribusi normal ($p < 0.05$)
	Shapiro-Wilk *Statistic*	0.7239	0.0000	
Uji Autokorelasi	Durbin-Watson	1.6379	-	Tidak ada autokorelasi (nilai mendekati 2)
Uji Heterokedastisitas	LM *Statistic*	89.6343	0.0000	Terdapat heteroskedastisitas ($p < 0.05$)
	F *Statistic*	51.2204	0.0000	



Gambar 10. Q-Q Plot residual Model PCA

3.4.3 Mengatasi Multikolinieritas menggunakan *Ridge* Regression

Metode lain untuk mengatasi multikolinearitas adalah *Ridge Regression*, yaitu regresi linear dengan penalti pada besar koefisien dalam fungsi loss. *Ridge Regression* tidak menghapus variabel, tetapi mengurangi pengaruh variabel yang berkorelasi tinggi, sehingga mencegah *overfitting* dan meningkatkan generalisasi model. Pada analisis ini, *Ridge Regression* diterapkan pada dataset saham GOTO dengan dua variabel prediktor yang sudah ditentukan. Metode ini menekan nilai koefisien secara otomatis, membuat model lebih stabil saat multikolinearitas tinggi terjadi antar variabel bebas.

```
Mean Squared Error (MSE) untuk Ridge Regression: 12.9835

Koefisien dari model Ridge Regression:
[-4.84980308 50.6040211  45.08051517  0.912643  ]
```

Gambar 11. Hasil *Ridge Regression*

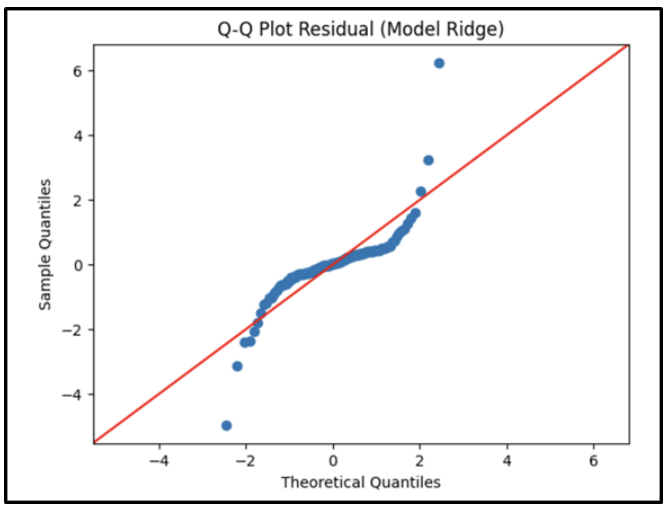
Hasil regresi *Ridge* menunjukkan bahwa nilai *Mean Squared Error* (MSE) sebesar 12.9835, yang menggambarkan tingkat kesalahan rata-rata kuadrat antara prediksi

dan nilai aktual variabel target. Nilai MSE ini menunjukkan performa model yang cukup baik. Koefisien regresi yang diperoleh adalah -4.8498 untuk konstanta, serta 50.6040, 45.0805, dan 0.9126 untuk tiga variabel independen. Koefisien positif menunjukkan hubungan searah dengan variabel target, sedangkan koefisien negatif menunjukkan hubungan berlawanan. *Ridge Regression* diterapkan untuk mengatasi multikolinearitas dan *overfitting* dengan menambahkan penalti pada besarnya koefisien, sehingga model menjadi lebih stabil dan mampu melakukan generalisasi lebih baik.

Selanjutnya, dilakukan pengujian validitas model regresi setelah penanganan multikolinieritas menggunakan *Ridge Regression* untuk memastikan bahwa model yang dihasilkan memenuhi asumsi-asumsi klasik regresi. Berikut adalah hasil dari uji validitas model setelah menggunakan *Ridge Regression*:

Tabel 8. Hasil Uji Validitas Model Setelah Penanganan Multikolinieritas Menggunakan *Ridge Regression*

Uji Validitas	Statistik	Nilai	$p-value$	Kesimpulan
Uji Normalitas Residual	JB *Statistic*	1278.6311	0.0000	Residual tidak berdistribusi normal ($p < 0.05$)
	Shapiro-Wilk *Statistic*	0.7504	0.0000	
Uji Autokorelasi	Durbin-Watson	2.0056	-	Tidak ada autokorelasi (nilai mendekati 2)
Uji Heterokedastisitas	BP *Statistic*	92.7084	0.0000	Terdapat heteroskedastisitas ($p < 0.05$)



Gambar 12. Q-Q *Plot* residual Model *Ridge*

3.4.4 Mengatasi Multikolinieritas menggunakan LASSO *Regression*

Selanjutnya, metode yang dapat digunakan untuk mengatasi permasalahan multikolinearitas adalah LASSO *Regression*. LASSO (*Least Absolute Shrinkage and Selection Operator*) *Regression* merupakan teknik regularisasi yang tidak hanya menambahkan penalti terhadap besar koefisien regresi, tetapi juga mampu mengecilkan beberapa koefisien menjadi nol. Dengan demikian, LASSO tidak hanya mengurangi kompleksitas model, tetapi juga melakukan seleksi variabel secara otomatis, menjadikannya sangat efektif dalam mengidentifikasi prediktor yang paling relevan. Berikut adalah hasil dari LASSO *Regression*.

```
Mean Squared Error (MSE) untuk Lasso Regression: 14.9595

Koefisien dari model Lasso Regression:
[ 0.      0.     59.5909 29.7642  0.    ]

Intercept:
128.444
```

Gambar 13. Hasil LASSO Regression

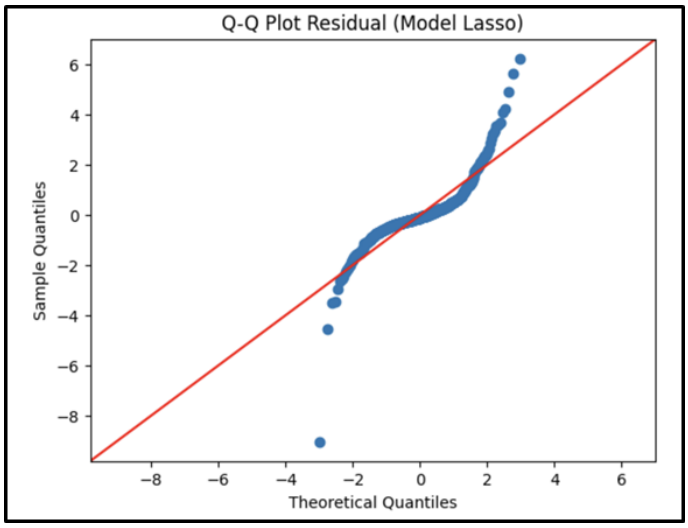
Berdasarkan hasil LASSO *Regression* pada data distandarisasi, nilai *Mean Squared Error* (MSE) sebesar 14.9595 menunjukkan tingkat kesalahan prediksi model. Intercept model adalah 128.444, yang berarti nilai prediksi saat semua prediktor nol. Dari empat variabel prediktor, hanya dua yang memiliki koefisien signifikan, yaitu prediktor ketiga dan keempat dengan nilai masing-masing 59.5909 dan 29.7642, sedangkan dua lainnya koefisiennya nol. Hal ini menunjukkan LASSO melakukan seleksi variabel otomatis dengan mengeliminasi variabel yang kurang berkontribusi, efektif mengatasi multikolinearitas. Model ini menghasilkan prediksi yang cukup akurat sekaligus menyederhanakan struktur model tanpa perlu transformasi data seperti PCA, sehingga cocok untuk mempertahankan interpretabilitas variabel asli.

Selanjutnya, dilakukan pengujian validitas model regresi setelah penanganan multikolinieritas menggunakan LASSO *Regression* untuk memastikan bahwa model yang dihasilkan memenuhi asumsi-asumsi klasik regresi. Berikut adalah hasil dari uji validitas model setelah menggunakan *Ridge Regression*:

Tabel 9. Hasil Uji Validitas Model Setelah Penanganan Multikolinieritas Menggunakan LASSO *Regression*

Uji Validitas	Statistik	Nilai	$p-value$	Kesimpulan
Uji Normalitas Residual	JB *Statistic*	7373.1105	0.0000	Residual tidak berdistribusi normal ($p < 0.05$)
	Shapiro-Wilk *Statistic*	0.7938	0.0000	
Uji Autokorelasi	Durbin-Watson	1.5448	-	Tidak ada autokorelasi (nilai mendekati 2)
Uji	LM *Statistic*	420.3968	0.0000	Terdapat

Heterokedastisitas	F *Statistic*	263.5076	0.0000	heteroskedastisitas ($p < 0.05$)



Gambar 14. Q-Q *Plot* residual Model LASSO

3.4.5 Mengatasi Multikolinieritas menggunakan *Partial Least Squares* (PLS) *Regression*

Metode lain untuk mengatasi multikolinearitas adalah *Partial Least Squares* (PLS) *Regression*. PLS bekerja dengan membentuk komponen baru sebagai kombinasi *linear* dari variabel asli yang memaksimalkan kovariansi antara prediktor dan variabel target. Pada analisis ini, PLS *Regression* diterapkan pada dataset saham GOTO menggunakan empat variabel prediktor yang sudah distandarisasi untuk menjaga kestabilan model dan menghilangkan pengaruh perbedaan skala antar variabel. Berikut adalah hasil dari PLS *Regression*.

```
Mean Squared Error (MSE) untuk PLS Regression: 0.0010
Koefisien dari model PLS Regression:
[ 0.         -0.60202853  0.80295555  0.79953891  0.00561513]
```

Gambar 15. Hasil PLS Regression

Berdasarkan hasil PLS *Regression*, nilai *Mean Squared Error* (MSE) sebesar 0.0010 menunjukkan tingkat kesalahan prediksi yang sangat rendah dan akurasi model yang tinggi. Koefisien regresi menunjukkan konstanta sebesar 0, yang berarti prediksi juga nol saat semua variabel prediktor nol. Variabel pertama berkontribusi negatif dengan koefisien -0.6020, sementara variabel kedua dan ketiga memberikan pengaruh positif dengan koefisien masing-masing 0.8029 dan 0.7995. Variabel keempat memiliki pengaruh yang sangat kecil, dengan koefisien 0.0056.

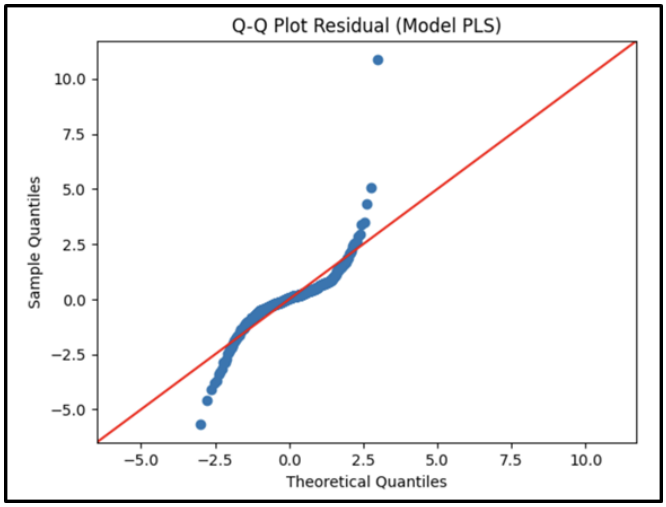
Berbeda dengan LASSO yang dapat mengecilkan koefisien menjadi nol, PLS tidak mengeliminasi variabel, melainkan mengurangi pengaruh variabel kurang relevan

melalui komponen baru. Dengan demikian, PLS mempertahankan semua variabel namun memberikan bobot berbeda sesuai kontribusinya. Metode ini efektif untuk mengatasi multikolinearitas sekaligus menjaga interpretabilitas variabel asli.

Selanjutnya, dilakukan pengujian validitas model regresi setelah penanganan multikolinieritas menggunakan PLS *Regression* untuk memastikan bahwa model yang dihasilkan memenuhi asumsi-asumsi klasik regresi. Berikut adalah hasil dari uji validitas model setelah menggunakan PLS *Regression*:

Tabel 10. Hasil Uji Validitas Model Setelah Penanganan Multikolinieritas Menggunakan PLS *Regression*

Uji Validitas	Statistik	Nilai	$p-value$	Kesimpulan
Uji Normalitas Residual	JB *Statistic*	17337.1065	0.0000	Residual tidak berdistribusi normal ($p < 0.05$)
	Shapiro-Wilk *Statistic*	0.8003	0.0000	
Uji Autokorelasi	Durbin-Watson	1.7154	-	Tidak ada autokorelasi (nilai mendekati 2)
Uji Heterokedastisitas	LM *Statistic*	174.7988	0.0000	Terdapat heteroskedastisitas ($p < 0.05$)
	F *Statistic*	57.9363	0.0000	



Gambar 16. Q-Q *Plot* residual Model PLS

3.5 Membandingkan Performa Model Regresi Sebelum dan Setelah Penanganan Multikolinearitas

Untuk mengevaluasi efektivitas penanganan multikolinearitas, dilakukan perbandingan performa model regresi sebelum dan setelah penanganan multikolinieritas. Perbandingan ini menggunakan metrik evaluasi umum yaitu R-Squared, Adjusted R-Squared, dan Root Mean Squared Error (RMSE). Dengan

membandingkan ketiga metrik ini, dapat dilihat apakah model yang telah dikoreksi dari multikolinearitas mampu memberikan hasil prediksi yang lebih baik, lebih stabil, dan lebih akurat dibandingkan model awal. Berikut adalah tabel perbandingan setiap metode sebelum dan setelah penanganan multikolinearitas:

Tabel 11. Perbandingan Setiap Metode Sebelum dan Setelah Penanganan Multikolinieritas

Metode	R-*Squared*	*Adjusted* R-*Squared*	RMSE	Komponen	Catatan
Model Awal	0.9990	0.9990	2.8861	-	Multikolinieritas tinggi
Penghapusan Variabel	0.9978	0.9978	4.2364	-	Penghapusan variabel dengan VIF tinggi
PCA	0.9978	0.9978	4.2364	2	Reduksi variabel
Ridge	0.9983	0.9983	3.6033	-	Penalti moderat
LASSO	0.9982	0.9982	3.8677	2 dipilih	Variabel diseleksi
PLS	0.9990	0.9990	0.0318	4	Terbaik

Berdasarkan tabel diatas, dapat dilihat bahwa sebelum penanganan multikolinearitas, model menunjukkan R-Squared dan Adjusted R-Squared sangat tinggi yaitu sebesar 0.9990, menandakan model mampu menjelaskan variabilitas data, namun potensi multikolinearitas masih ada. RMSE relatif rendah yaitu sebesar 2.8861, namun belum mencerminkan stabilitas model. Setelah penanganan dengan penghapusan variabel VIF tinggi dan PCA, R-Squared turun sedikit menjadi 0.9978 dan RMSE naik menjadi 4.2364, ini menunjukkan penurunan akurasi prediksi.

Ridge dan LASSO *Regression* menurunkan RMSE menjadi 3.6033 dan 3.8677, masing-masing, dengan sedikit penurunan R-*Squared*, menunjukkan peningkatan stabilitas dan akurasi dibandingkan penghapusan variabel atau PCA. PLS *Regression* mempertahankan R-*Squared* sebesar 0.9990 namun berhasil menurunkan RMSE secara signifikan menjadi 0.0318, menunjukkan efektivitas terbaik dalam mengatasi multikolinearitas sekaligus menghasilkan prediksi yang sangat akurat dan stabil. Hasil ini konsisten dengan temuan dalam (Chand & Kamal, 2011) dan (Zhang et al., 2021) yang menunjukkan efektivitas LASSO dan PLS dalam prediksi saham. Model LASSO menunjukkan seleksi variabel serupa seperti dalam (Chand & Kamal, 2011), sedangkan PLS menunjukkan keakuratan tinggi seperti dilaporkan dalam (Swanson & Tayman, 2012), (Chin et al., 2010).

Temuan dalam penelitian ini sejalan dengan studi oleh (Rajput & Kaulwar, 2018) yang menggunakan jaringan saraf NARX dan PCA untuk prediksi harga saham, di mana PCA digunakan untuk mereduksi dimensi dan mengatasi multikolinearitas sebelum pelatihan model. Meskipun model NARX menunjukkan kinerja yang baik dalam prediksi, pendekatan PLS *Regression* dalam penelitian ini terbukti lebih unggul dalam menjaga kestabilan model tanpa kehilangan akurasi, sebagaimana ditunjukkan oleh

RMSE yang sangat rendah. Selain itu, hasil ini juga sejalan dengan tinjauan literatur oleh (Gupta et al., 2023), yang menyimpulkan bahwa metode *machine learning* seperti LSTM, CNN, dan SVR memiliki performa prediksi yang menjanjikan. Namun, pendekatan regresi terkontrol seperti *Ridge*, LASSO, dan khususnya PLS dalam penelitian ini menunjukkan bahwa solusi yang lebih sederhana secara matematis tetap mampu bersaing dari segi akurasi dan stabilitas, terutama dalam konteks multikolinearitas. Dengan demikian, pendekatan regresi berbasis regularisasi dan transformasi variabel tetap relevan dan dapat menjadi alternatif efektif di samping metode machine learning kompleks.

Secara keseluruhan, PLS *Regression* menjadi metode paling efektif, sementara Ridge dan LASSO memberikan perbaikan signifikan dibanding penghapusan variabel dan PCA. Hal ini menunjukkan pentingnya pemilihan metode yang tepat dalam mengatasi multikolinearitas agar model tetap akurat dan stabil. Dengan pendekatan yang tepat, prediksi yang dihasilkan dapat lebih dapat diandalkan untuk pengambilan keputusan.

4. SIMPULAN

Berdasarkan analisis yang dilakukan, variabel-variabel independen yang berpengaruh signifikan terhadap harga saham PT GoTo Gojek Tokopedia Tbk (GOTO) meliputi harga pembukaan, harga tertinggi, harga terendah, dan volume perdagangan. Keempat variabel ini mencerminkan faktor teknikal yang secara langsung memengaruhi pergerakan harga saham. Namun, terdapat gejala multikolinearitas yang signifikan antar variabel independen tersebut, yang terdeteksi dari nilai *Variance Inflation Factor* (VIF) yang tinggi. Multikolinearitas ini menyebabkan ketidakstabilan estimasi koefisien regresi dan menurunkan keandalan model meskipun nilai R-*Squared* tinggi. Untuk mengatasi masalah tersebut, beberapa metode diterapkan, yaitu penghapusan variabel dengan VIF tinggi, *Principal Component Analysis* (PCA), *Ridge Regression*, LASSO *Regression*, dan PLS *Regression*. Hasil menunjukkan bahwa PLS *Regression* adalah metode paling efektif dalam mengatasi multikolinearitas dengan memberikan model yang lebih stabil dan akurat dalam memprediksi harga saham GOTO. Secara keseluruhan, penanganan multikolinearitas secara tepat dapat meningkatkan keandalan dan akurasi model regresi dalam memprediksi harga saham. Namun, Penelitian ini belum mempertimbangkan variabel makroekonomi seperti inflasi, suku bunga, atau nilai tukar, yang menurut (Rouf et al., 2021), (Zhao & Yu, 2006), memiliki pengaruh signifikan terhadap pergerakan harga saham. Ketidakhadiran variabel tersebut dapat membatasi kemampuan model dalam menangkap dinamika pasar yang lebih luas. yang perlu diperhatikan untuk pengembangan penelitian selanjutnya. Studi lanjutan disarankan untuk memasukkan variabel makroekonomi seperti tingkat suku bunga, inflasi, dan sentimen pasar agar model dapat mencerminkan faktor fundamental yang mempengaruhi harga saham. Selain itu, penggunaan metode yang lebih canggih dan robust terhadap multikolinearitas seperti model *Long Short-Term Memory* (LSTM)

atau *Hybrid* PCA-LSTM, sebagaimana disarankan oleh (Zhao & Yu, 2006), dapat diuji untuk meningkatkan akurasi dan stabilitas prediksi harga saham.

5. REFERENSI

Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: *Computational Statistics*, 2(4), 433–459. https://doi.org/10.1002/wics.101

Chand, S., & Kamal, S. (2011). Variable selection by lasso-type methods. *Pakistan Journal of Statistics and Operation Research*, 7(2 SPECIAL ISSUE), 451–464. https://doi.org/10.18187/pjsor.v7i2-sp.389

Chin, W. W., Henseler, J., & Ringle, P. (2010). *Handbook of Partial Least Squares*. In Handbook of Partial Least Squares (Issue January 2010). https://doi.org/10.1007/978-3-540-32827-8

Dewi, Y. S. (2010). OLS, LASSO dan PLS Pada data Mengandung Multikolinearitas. *Jurnal ILMU DASAR*, 11(1), 83–91.

Draper, N. R., & Smith, H. (2014). *Applied regression analysis*. In Applied Regression Analysis (pp. 1–716). https://doi.org/10.1002/9781118625590

Ferdiansyah, Tin, S., & Anthonius. (2016). Globalisasi Ekonomi, Integrasi Ekonomi Global, Dinamika Pasar Modal & Kebutuhan Standar Akuntansi Internasional Ferdiansyah Se Tin Anthonius. *Jurnal Akuntansi*, 8(1), 119–130.

Golam Kibria, B. M. (2003). Performance of some New Ridge regression estimators. Communications in Statistics Part B: *Simulation and Computation*, 32(2), 419–435. https://doi.org/10.1081/SAC-120017499

Greene, W. H. (2018). *Econometric analysis* (8th ed.). Pearson.

Gupta, A., Akansha, Joshi, K., Patel, M., & Pratap, V. (2023). Stock Market Prediction using Machine Learning Techniques: A Systematic Review. *International Conference on Power, Instrumentation, Control and Computing*, PICC 2023, 1–6. https://doi.org/10.1109/PICC57976.2023.10142862

Montgomery, D.C., Peck, E.A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. (Fifth Edit). John Wiley & Sons, Hoboken.

Pirouz, D. M. (2012). An Overview of Partial Least Squares. *SSRN Electronic Journal*, March. https://doi.org/10.2139/ssrn.1631359

Rajput, G. G., & Kaulwar, B. H. (2018). Predicting Stock Prices in National Stock Exchange of India using Principal Component Analysis and Neural Networks. *International Journal of Computer Sciences and Engineering*, 6(6), 746–752. https://doi.org/10.26438/ijcse/v6i6.746752

Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H. C. (2021). Stock market prediction using machine learning techniques: A decade survey on methodologies, recent developments, and future directions. *Electronics (Switzerland)*, 10(21). https://doi.org/10.3390/electronics10212717

Sari, D. R. P. (2023). Metode Principal Component Analysis (PCA) sebagai penanganan asumsi multikolinearitas (studi kasus: data produksi tapioka). *Parameter: Jurnal Matematika, Statistika dan Terapannya*, 2(2), 115–124. https://doi.org/10.30598/parameterv2i02pp115-124

Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*, 8(2), 39–42. https://doi.org/10.12691/ajams-8-2-1

Al-Kassab, M. M., & Ibrahim, S. (2022). Using ridge regression to estimate factors affecting the number of births. A comparative study. In *International Conference on Mathematics and Computations* (pp. 183-194). Singapore: Springer Nature Singapore.

Sungkono, J., & Nugrahaningsih, T. K. (2017). Simulasi Dampak Multikolinearitas Pada Kondisi Penyimpangan Asumsi Normalitas. *Magistra*, XXIX (101), 45–50.

Swanson, D.A., & Tayman, J. (2012). *Regression Methods*. In: Subnational Population Estimates. The Springer Series on Demographic Methods and Population Analysis, vol 31. Springer, Dordrecht. https://doi.org/10.1007/978-90-481-8954-0_8

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society*. Series B: Methodological, 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Verleysen, M., & Verleysen, M. (2001). Principal component analysis (PCA). *Université catholique de*.

Wasilaine, T. L., Talakua, M. W., & Lesnussa, Y. A. (2014). Model Regresi Ridge untuk Mengatasi Model Regresi Linier Berganda yang Mengandung Multikolinieritas (Studi Kasus: Data Pertumbuhan Bayi di Kelurahan Namaelo RT 001, Kota Masohi). *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 8(1), 31–37.

Zhang, Y., Shen, D., & Huang, L. (2021). Predicting stock market returns using deep learning and technical indicators. *Neurocomputing*, 432, 347–364.

Zhao, P., & Yu, B. (2006). On model selection consistency of Lasso. *Journal of Machine Learning Research*, 7, 2541–2563.