BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan suatu analisis statistik yang digunakan untuk mengetahui hubungan antara variabel dependen (variabel respon) dengan satu atau lebih variabel independen (variabel prediktor). Pendekatan regresi dilakukan dengan tiga pendekatan, yaitu pendekatan parametrik, semiparametrik, dan nonparametrik. Regresi parametrik digunakan apabila bentuk kurva regresi diketahui. Regresi nonparametrik digunakan apabila kurva regresi tidak diketahui bentuk pola hubungannya. Adapun regresi semiparametrik merupakan gabungan antara regresi parametrik dan regresi nonparametrik (Sahidah dkk., 2022).

Apabila asumsi dalam analisis regresi parametrik terpenuhi maka analisis tersebut dapat digunakan, seperti asumsi bentuk kurva regresi yang harus diketahui. Selain dari asumsi bentuk kurva yang harus dipenuhi, regresi parametrik juga harus mempunyai sifat *Best Linier Unbiased Estimator* (BLUE) yaitu data berdistribusi normal, homoskedastisitas, non-autokorelasi, dan non-multikolinearitas. Namun, tidak semua data yang akan dijadikan penelitian memenuhi semua asumsi dari regresi parametrik (Rahayu & Wachidah, 2022).

Regresi nonparametrik digunakan sebagai alternatif dari regresi parametrik, karena regresi nonparametrik tidak memerlukan asumsi yang ketat. Regresi nonparametrik merupakan suatu analisis untuk mengetahui hubungan antara variabel terikat dengan variabel bebas tanpa harus diketahui bentuk kurva regresi serta bentuk fungsinya. Fungsi dari regresi nonparametrik diasumsikan mulus (*smooth*) dalam arti termuat dalam suatu ruang fungsi tertentu sehingga memberikan fleksibilitas yang tinggi dalam mengestimasi kurva regresi. Regresi nonparametrik merupakan analisis regresi yang sangat fleksibel dalam memodelkan pola data (Eubank, 1999).

Penggunaan regresi parametrik membutuhkan asumsi yang sangat ketat serta kurva regresi harus diketahui sehingga dianggap tidak mampu menangani data yang berdimensi tinggi. Kekurangan regresi parametrik dapat diatasi dengan penggunaan regresi nonparametrik. Regresi nonparametrik mempunyai kemampuan mengestimasi yang tinggi dan mempunyai sifat fleksibel dalam memodelkan data (Budiantara, 2005).

Pendekatan regresi nonparametrik secara *adaptive* banyak diminati. Beberapa contohnya adalah *Regression Tree*, *Recursive Partitioning Regression* (RPR) dan *Multivariate Adaptive Regression Spline* (MARS) (Breiman dkk., 1993). Model *Multivariate Adaptive Regression Spline* (MARS) menghasilkan prediksi variabel respon yang akurat, menghasilkan model yang kontinu pada knot (Friedman, 1991).

Metode Multivariate Adaptive Regression Spline (MARS) merupakan pengembangan dari metode *Recursive Partitioning Regression* (RPR) yang digabungkan dengan metode *Spline*. Namun, pendekatan dari RPR mempunyai kelemahan yaitu model yang didapatkan tidak kontinu pada knot. Pengembangan dari *Recursive Partitioning Regression* (RPR) yang dikombinasikan dengan metode *Spline* menghasilkan model yang kontinu pada knot yang berarti garis regresi selalu menyambung dan setiap knot menyambung dengan fungsi basisnya (Friedman, 1991). Kelebihan dari metode *Multivariate Adaptive Regression Splines* (MARS) lainnya yaitu penentuan knot dapat dilakukan secara otomatis menggunakan algoritma *stepwise forward* dan *backward* yang mana didasarkan pada nilai *Generalized Cross Validation* (GCV) minimum (Panggabean & Mansyur, 2023). Model *Multivariate Adaptive Regression Splines* (MARS) berguna dalam mengatasi permasalahan data yang berdimensi tinggi, yaitu data yang memiliki jumlah variabel prediktor $3 \leq x \leq 20$, dimana x adalah banyak variabel dan sampel data yang berukuran $50 \leq n \leq 1000$ dimana n untuk ukuran sampel (Friedman, 1991).

Pada penelitian sebelumnya yang telah melakukan analisis dengan metode *Multivariate Adaptive Regression Spline* (MARS), yaitu (Naser et al., 2022) dengan *Application of Multivariate Adaptive Regression Splines* (MARS) *approach in prediction of compressive strength of eco-friendly concrete*. Hasil dari penelitian ini, diperoleh nilai kombinasi BF = 13, dan MI = 2 sebagai model terbaik dengan nilai *Root Mean Square Error* (RMSE) sebesar 4,110. Selain itu, dalam penelitian tersebut kinerja MARS dibandingan dengan metode *Random Forest* dan SVM, diperoleh bahwa metode MARS menghasilkan prediksi yang lebih akurat dibandingkan dengan kedua metode tersebut.

Penelitian lainnya dilakukan oleh (Qayyum et al., 2023) dengan *Prediction of Inhibition Activity of Dihydrofolate Reductase Inhibitors with Multivariate Adaptive Regression Splines*. Hasil dari penelitian tersebut, diperoleh nilai kombinasi BF = 50, dan MI = 4 sebagai model terbaik dengan nilai GCV minimum sebesar 0,55 dan *Root Mean Square Error* (RMSE) sebesar 1,02. Penelitian ini juga membandingakan metode MARS dengan *metode Deep Neural Network* (DNN) dan *Partial Least Squares* (PLS). Berdasarkan hasil penelitian diperoleh bahwa metode MARS menghasilkan prediksi yang lebih akurat. Selain itu, terdapat juga penelitian (Otok et al., 2020) dengan *Bootstrap Aggregating Multivariate Adaptive Regression Spline for Observational Studies in Diabetes Cases*. Penelitian ini melakukan pengembangan dengan menggabungkan metode MARS dengan *Bootstrap Aggregating*. Berdasarkan penelitian tersebut diperoleh model terbaik dengan kombinasi BF = 28, MI = 3 and MO = 1 dengan nilai GCV sebesar 0,063. Penelitian tersebut juga membandingkan hasil akurasi klasifikasi antara model MARS dan *Bagging* MARS, diperoleh bahwa akurasi klasifikasi antara model MARS dan *Bagging* MARS menghasilkan akurasi klasifikasi yang sama yaitu sebesar 90,63%. Hal ini menunjukkan bahwa metode *bagging* tidak selalu bisa meningkatkan akurasi klasifikasi, tapi teknik ini dapat mengurangi peramalan *Mean Square Error* (MSE). Dalam analisisnya, *Bootstrap Aggregating* memerlukan banyak iterasi untuk membangun beberapa model MARS. Oleh karena itu, penelitian ini memiliki keterbatasan terkait komputasi dan waktu yang diperlukan untuk menjalankan analisis, terutama pada data yang besar dan kompleks.

Dalam pemodelan *Multivariate Adaptive Regression Spline* (MARS) dilakukan berdasarkan *trial and error* terhadap kombinasi antara jumlah basis fungsi, maksimum interaksi, dan minimum observasi. Dalam memudahkan pencarian kombinasi tersebut dilakukan penelitian untuk menggabungkan metode tersebut dengan teknik lain yang lebih canggih dan inovatif. Salah satunya adalah dengan mengintegrasikan algoritma *Differential Evolution* (DE). *Differential Evolution* (DE) merupakan suatu teknik optimasi yang dikembangkan oleh Kenneth Price yang dipublikasikan dalam majalah Dr. Dobb's Journal pada Oktober 1994 (Price dkk., 2005). Metode ini merupakan suatu pendekatan optimasi matematis untuk fungsi-fungsi multidimensi dan termasuk dalam kelompok algoritma evolusioner (Julianto, 2016). Algoritma *Differential Evolution* (DE) mengatasi masalah kompleks dengan mencari solusi terbaik melalui proses iterasi. Model *Multivariate Adaptive Regression Spline* (MARS) akan lebih baik digunakan apabila dikombinasikan dengan algoritma *Differential Evolution* (DE) karena teknik *Differential Evolution* (DE) dapat digunakan untuk mengoptimalkan hyperparameter dari *Multivariate Adaptive Regression Spline* (García–Nieto dkk., 2020). Oleh karena itu, kombinasi teknik ini diharapkan dapat mengoptimalkan model *Multivariate Adaptive Regression Spline*.

Penelitian sebelumnya yang telah melakukan analisis metode *Multivariate Adaptive Regression Spline* (MARS) dengan Algoritma *Differential Evolution* (DE), yaitu (Abdulelah Al-Sudani dkk., 2019) dengan Pengembangan *Multivariate Adaptive Regression Spline* terintegrasi dengan model *Differential Evolution* untuk simulasi aliran sungai, diperoleh kinerja terbaik model *Multivariate Adaptive Regression Spline* (MARS) dengan *Differential Evolution* yaitu dengan nilai MAPE sebesar 0,06. Hasil akurasi prediksi model menggunakan *Multivariate Adaptive Regression Spline* dengan *Differential Evolution* lebih baik. Penelitian yang dilakukan oleh (Abdulelah Al-Sudani dkk., 2019) menggunakan *Mean Absolute Percentage Error* (MAPE) dalam mengevaluasi performa model. Namun, penggunaan MAPE memiliki kelemahan yang signifikan yaitu menghasilkan nilai tak terhingga atau tidak terdefinisi untuk nilai aktual nol atau mendekati nol (Kim & Kim, 2016). Sehingga, untuk mengatasi masalah di MAPE tersebut, diusulkan metode evaluasi performa model dengan menggunakan *Mean Square Error* (MSE).

Investasi di pasar modal saat ini sangat diminati bagi para investor. Hal tersebut disebabkan karena adanya perkembangan perekonomian dimana pertumbuhan ekonomi suatu negara dipengaruhi oleh pertumbuhan investasi di negara tersebut. Pasar modal merupakan salah satu instrumen investasi yang paling diminati karena memberikan imbal hasil yang sangat tinggi (Sutandi dkk., 2021). Sebelum melakukan investasi di pasar modal para investor mencari pertimbangan mengenai situasi pergerakan harga saham di pasar modal. Oleh karena itu para investor perlu mengetahui faktor-faktor apa yang mempengaruhi pergerakan harga saham. Salah satu indeks yang sering diperhatikan investor ketika berinvestasi di Bursa Efek Indonesia (BEI) adalah Indeks Harga Saham Gabungan (IHSG). Hal ini disebabkan indeks ini mencakup seluruh saham yang tercatat di Bursa Efek Indonesia. Indeks Harga Saham Gabungan (IHSG) merupakan salah satu indikator yang menunjukkan pergerakan harga saham (Astuti dkk., 2013).

Indeks Harga Saham Gabungan (IHSG) digunakan sebagai acuan bagi investor dalam melihat representasi dari harga saham keseluruhan. Oleh karena itu, untuk menganalisis kemungkinan kenaikan atau penurunan harga saham maka diperlukan suatu metode analisis. Variabel makro seperti tingkat suku bunga (SBI), nilai tukar (kurs) rupiah, dan inflasi sangat memengaruhi indeks harga saham. Selain itu, kondisi perekonomian negara maju mempunyai kekuatan untuk mempengaruhi perekonomian global, sehingga indeks saham di negara maju berdampak terhadap indeks saham di negara berkembang (Puteri & Rizal, 2024).

Pada tahun 2018, China yang terlibat perang dagang dengan Amerika Serikat memberikan dampak negatif terhadap perekonomian global. Sebagai salah satu negara tujuan ekspor terbesar Indonesia, China tentu memberikan dampak negatif terhadap perekonomian Indonesia sendiri. Indeks Shanghai Stock Exchange Composite (SSEC) digunakan untuk menggambarkan keadaan pasar modal dan ekonomi China. Indeks lain yang sering dipergunakan dalam pengambilan keputusan para investor yaitu Indeks Nikkei 225 yang merupakan indeks harga saham gabungan negara Jepang (Herlianto & Hafizh, 2020). Berdasarkan dari data dan informasi dari Kementerian Perdagangan Indonesia menyebutkan bahwa pada tahun 2019 – 2022 Jepang menjadi salah satu negara tujuan ekspor impor nonmigas Indonesia dengan nilai peran ekspor sebesar 7.78% dan peran impor sebesar 8.84%. Selain itu, ada juga Indeks Dow Jones yang merupakan cerminan dari kinerja perekonomian Amerika Serikat. Amerika Serikat adalah salah satu negara tujuan ekspor bagi negara Indonesia, jika ekonomi Amerika Serikat terus berkembang, pertumbuhannya dapat mendorong pertumbuhan ekonomi Indonesia baik melalui kegiatan ekspor, investasi langsung, dan investasi pada pasar modal Indonesia (Jayanti dkk., 2014).

Berdasarkan uraian latar belakang tersebut, maka penulis akan melakukan penelitian mengenai Pemodelan *Multivariate Adaptive Regression Spline* dengan Algoritma *Differential Evolution* pada data Indeks Harga Saham Gabungan (IHSG).

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, masalah yang dikaji dalam penelitian ini adalah:

1. Bagaimana hasil pemodelan dari *Multivariate Adaptive Regression Spline* dengan Algoritma *Differential Evolution*?
2. Bagaimana hasil dugaan Indeks Harga Saham Gabungan menggunakan MARS-DE?

1.3 Batasan Masalah

Pada penelitian ini, penulis memberikan batasan masalah sebagai berikut:

1. Data Indeks Harga Saham Gabungan (IHSG) periode tahun 2017-2022.
2. Pendekatan optimasi yang digunakan yaitu Algoritma *Differential Evolution.*

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut:

1. Mendapatkan hasil pemodelan dari *Multivariate Adaptive Regression Spline* dengan Algoritma *Differential Evolution.*
2. Mendapatkan hasil dugaan Indeks Harga Saham Gabungan menggunakan MARS-DE.

1.5 Manfaat Penelitian

Penelitian ini dapat bermanfaat untuk menambah khasanah pengetahuan terkait metode *Multivariate Adaptive Regression Spline* dan algoritma *Differential Evolution* yang dapat menjadi acuan dan literatur tambahan bagi peneliti selanjutnya. Selain itu, penelitian ini dapat memberikan pemahaman mendalam tentang Indeks Harga Saham Gabungan (IHSG) serta dapat membantu identifikasi pola dan tren dalam pergerakan harga saham yang dapat membantu investor dan analis dalam membuat suatu keputusan investasi yang lebih baik.

1.6 Kajian Teori

1.6.1 Analisis Regresi

Analisis regresi menjadi salah satu metode statistika yang sangat berperan pada perkembangan ilmu statistika (Islamiyati, 2018). Analisis regresi merupakan suatu metode analisis yang digunakan untuk mencari pengaruh antara variabel bebas dan variabel terikat (Tiro, 2011). Dalam mengestimasi suatu kurva regresi ada tiga pendekatan, yaitu regresi parametrik, regresi nonparametrik, dan regresi semiparametrik. Regresi parametrik digunakan jika pola kurva regresi diketahui, regresi nonparametrik digunakan jika pola kurva regresi tidak diketahui, adapun regresi semiparametrik digunakan jika sebagian bentuk pola kurva regresi diketahui dan sebagian lagi tidak diketahui (Ente, Islamiyati dan Raupong, 2021). Model umum regresi yang menjelaskan hubungan antara variabel terikat (y_i) dengan variabel bebas (x_i) sebagai berikut (Eubank, 1999) :

$$y_i = f(x_i) + \varepsilon_i \,;\, i = 1,2, \dots, n \tag{1}$$

dimana:

y_i : variabel respon ke-i

$f(x_i)$: fungsi regresi yang tidak diketahui bentuk kurvanya

ε_i : error atau galat yang saling menyebar $\sim N(0, \sigma^2)$

Regresi parametrik digunakan apabila penyebaran pola data membentuk suatu pola tertentu atau diketahui. Analisis regresi parametrik akan lebih menguntungkan apabila diketahui bentuk kurva atau fungsi regresi, khususnya metode inferensi dan interpretasi parameternya akan lebih sederhana. Sehingga, analisis regresi parametrik lebih sering digunakan jika terdapat informasi tentang bentuk kurva regresinya. Pendekatan regresi parametrik memiliki sifat yang sangat baik, yaitu sederhana, interpretasinya yang mudah, tak bias, estimator linier, parsimoni, konsisten, efisien, dan *Best Linier Unbiased Estimator* (BLUE) yang mana hal tersebut tidak dimiliki dalam pendekatan regresi nonparametrik

(Budiantara, 2009). Namun, tidak semua pola hubungan dapat didekati dengan pendekatan parametrik. Jika tidak ada informasi apapun tentang bentuk kurva, maka pendekatan yang dapat digunakan adalah pendekatan nonparametrik.

Regresi nonparametrik merupakan salah satu metode yang digunakan untuk mengestimasi pola hubungan antara variabel respon dan variabel prediktor dimana bentuk kurva regresinya tidak diketahui. Dalam regresi nonparametrik kurva regresi hanya diasumsikan mulus (*smooth*) dalam arti termuat dalam suatu ruang fungsi tertentu sehingga mempunyai sifat fleksibilitas yang tinggi (Eubank, 1999). Beberapa model regresi nonparametrik yang banyak digunakan yaitu Spline, *Multivariate Adaptive Regression Splines* (MARS), Kernel, Deret *Fourier*, Deret Orthogonal, *Neural Network* (NN), Polinomial Lokal, Histogram, *Wavelets*, k-NN, dan yang lainnya.

1.6.2 Regresi Spline

Regresi spline merupakan jenis analisis regresi yang memiliki kemampuan untuk mengestimasi data yang tidak memiliki pola tertentu dan cenderung mencari sendiri estimasi dari pola yang terbentuk (Budiantara, 2009). Pendekatan spline memiliki kemampuan dalam mengatasi pola data yang menunjukkan naik atau turun yang tajam dengan bantuan titik-titik knot, serta menghasilkan kurva yang relatif mulus (Hardle, 1990). Titik perpaduan bersama dari potongan-potongan tersebut atau titik yang menunjukkan terjadinya perubahan perilaku kurva pada interval-interval yang berbeda dikenal sebagai titik knot (Jao, Islamiyati dan Sunusi, 2022). Spline merupakan potongan (*piecewise*) polinomial orde q dan memiliki turunan yang kontinu dengan knot sampai orde (q-1) (Friedman, 1991). Dalam spline univariat dengan K knot memiliki fungsi basis sebagai berikut (Wicaksono dkk., 2014):

$$\{x^j\}_1^q,\ \left\{(x-t_k)_+^q\right\}_1^K \tag{2}$$

Dimana $\{t_k\}_1^K$ adalah titik knot, diharapkan adanya kontinuitas dari fungsi basis antara satu garis regresi (*region*) dengan *region* lainnya. Secara umum spline dengan orde q dapat dituliskan sebagai berikut:

$$f(x) = \beta_0 + \beta_1 x + \cdots + \beta_q x^q + \sum_{k=1}^{K} \gamma_k (x-t_k)_+^q \tag{3}$$

Dimana $q \geq 1$ dan $(t_1, t_2, \ldots . t_k)$ merupakan titik-titik knot. Adapun fungsi dari t*runcated power* $(x-t_k)_+^q$ adalah:

$$(x-t_k)_+^q = \begin{cases} (x-t_k)^q & ; x-t_k > 0 \\ 0 & ; x-t_k \leq 0 \end{cases} \tag{4}$$

dimana q menunjukkan orde polinomial dari fungsi spline. Di setiap titik knot, diharapkan adanya dari fungsi basis antar satu region dengan region lainnya. Oleh karena itu pada umumnya fungsi basis yang dipilih adalah berbentuk polinomial dengan turunan $(q-1)$ yang kontinu disetiap titik knot.

1.6.3 Multivariate Adaptive Regression Spline (MARS)

Multivariate Adaptive Regression Spline (MARS) merupakan salah satu pendekatan regresi nonparametrik, yakni suatu model yang mengasumsikan fungsi bentuk hubungan antara variabel terikat dan variabel bebas tidak diketahui. Metode *Multivariate Adaptive Regression Spline* (MARS) yang diusulkan oleh Jerome H. Friedman pada tahun 1991 mempunyai bentuk fungsi yang fleksibel. Metode *Multivariate Adaptive Regression Spline* (MARS) digunakan untuk memprediksi variabel terikat bernilai kontinu yang didasarkan pada beberapa variabel bebas. Metode *Multivariate Adaptive Regression Spline* (MARS) menghasilkan suatu prediksi respon yang akurat dan dapat mengatasi kelemahan dari regresi partisi rekursif (RPR) yakni menghasilkan suatu metode yang kontinu pada knot (Panggabean & Mansyur, 2023).

Metode *Multivariate Adaptive Regression Spline* (MARS) difokuskan untuk mengatasi permasalahan data berdimensi tinggi dengan jumlah variabel dan observasi yang cukup banyak dan menghasilkan model yang kontinu pada knot. Prinsip dasar MARS adalah memberikan fleksibilitas tinggi untuk mengeksplorasi hubungan non linier yang terjadi diantara variabel respon dan variabel prediktor melalui fungsi yang berbeda untuk setiap interval yang berbeda. Metode ini juga memungkinkan untuk mengetahui interaksi yang terjadi diantara variabel prediktor (Oktora, 2016).

Hasil modifikasi model *Recursive Partitioning Regression* (RPR) dengan kombinasi spline merupakan estimasi model *Multivariate Adaptive Regression Splines* (MARS) (Friedman, 1991) sebagai berikut:

$$\hat{f}(x) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} \left[s_{km}.\left(x_{v(k,m)} - t_{km}\right)\right]_+ \tag{5}$$

dimana:

a_0 : konstanta regresi dari fungsi basis
a_m : koefisien fungsi basis ke-m
M : maksimum fungsi basis
K_m : maksimum derajat interaksi
s_{km} : tanda fungsi basis, yang bernilai ± 1 jika $x_{v(k,m)} \geq t_{km}$ dan bernilai -1 jika $x_{v(k,m)} < t_{km}$
$x_{v(k,m)}$: variabel bebas ke-v
t_{km} : nilai knot dari variabel bebas $x_{v(k,m)}$
v : banyaknya variabel bebas
m : banyaknya basis fungsi
k : banyaknya interaksi

Penjabaran dari persamaan (5) disajikan sebagai berikut:

$$\hat{f}(x) = \alpha_0 \sum_{m=1}^{M} \alpha_m \left[s_{1m} \cdot \left(x_{v(1,m)} - t_{(1,m)}\right)\right]_+ + \sum_{m=1}^{M} a_m \left[s_{1m} \cdot \left(x_{v(1,m)} - t_{(1,m)}\right)\right]_+ \left[s_{2m} \cdot \left(x_{v(2,m)} - t_{(2,m)}\right)\right]_+ + \sum_{m=1}^{M} a_m \left[s_{1m} \cdot \left(x_{v(1,m)} - t_{(1,m)}\right)\right]_+ \left[s_{2m} \cdot \left(x_{v(2,m)} - t_{(2,m)}\right)\right]_+ \left[s_{3m} \cdot \left(x_{v(3,m)} - t_{(3,m)}\right)\right]_+ + \cdots \tag{6}$$

Dari persamaan (6) dapat dituliskan kembali menjadi (Mattalunru dkk., 2022) dan (Raupong, 2010):

$$\hat{f}(x) = a_0 + \sum_{K_m=1} f_i(x_i) + \sum_{K_m=2} f_{ij}(x_i, x_j) + \sum_{K_m=3} f_{ijk}(x_i, x_j, x_k) + \cdots . \quad (7)$$

Penjumlahan pertama pada persamaan (7) meliputi semua basis fungsi untuk satu variabel, penjumlahan kedua meliputi semua basis fungsi untuk interaksi yang melibatkan dua variabel dengan cara yang sama penjumlahan ketiga meliputi semua basis fungsi untuk interaksi antara tiga variabel dan seterusnya.

Misalkan jika $V(m) = \{v(k,m)\}_1^{K_m}$ merupakan himpunan dari variabel yang dihubungkan dengan basis fungsi B_m ke-m, maka setiap penjumlahan pertama pada persamaan (7) dapat dituliskan sebagai berikut (Friedman, 1991):

$$f_i(x_i) = \sum_{K_m=1} a_m B_m(x_i) \quad (8)$$

Fungsi tersebut merupakan penjumlahan semua basis fungsi untuk satu variabel x_i dan merupakan Spline dengan derajat q=1 yang mewakili fungsi univariat. Setiap fungsi bivariat pada penjumlahan kedua dan fungsi trivariat pada penjumlahan ketiga dengan menjumlahkan semua basis fungsi untuk tiga variabel dalam persamaan (7), secara berturut dapat dituliskan sebagai berikut (Friedman, 1991):

$$f_{ij}(x_i, x_j) = \sum_{K_m=2} a_m B_m(x_i, x_j), \quad (9)$$

$$f_{ijk}(x_i, x_j, x_k) = \sum_{K_m=3} a_m B_m(x_i, x_j, x_k) \quad (10)$$

Fungsi (9) yang mempresentasikan penjumlahan semua basis fungsi dua variabel x_i dan x_j. Penambahan tersebut untuk menghubungkan kontribusi univariat, yang dituliskan sebagai berikut (Friedman, 1991):

$$f_{ij}^*(x_i, x_j) = f_i(x_i) + f_j(x_j) + f_{ij}(x_i, x_j) \quad (11)$$

Pada fungsi (10), penambahan fungsi univariat dan bivariat mempunyai kontribusi dalam bentuk yang dituliskan sebagai berikut (Friedman, 1991):

$$f_{ijk}^*(x_i, x_j, x_k) = f_i(x_i) + f_j(x_j) + f_k(x_k) + f_{ij}(x_i, x_j) + f_{ik}(x_i, x_k),$$
$$+ f_{jk}(x_j, x_k) + f_{ijk}(x_i, x_j, x_k) \quad (12)$$

Diberikan data berpasangan $(x_{1i}, x_{2i}, \dots, x_{pi}, y_i)$ $i = 1,2,\dots,n$ dimana $x_1, x_2, \dots, x_p$ adalah prediktor dan y adalah respon sedangkan n merepresentasikan banyaknya pengamatan. Jika hubungan antara variabel prediktor dan respon dinyatakan dalam fungsi regresi f yang bentuknya tidak diketahui dan dapat didekati dengan model regresi nonparametrik, maka hubungan ini dapat ditulis dalam bentuk persamaan regresi sebagai berikut:

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{pi}) + \varepsilon_i \quad (13)$$

Fungsi regresi f dalam persamaan (13) merupakan fungsi regresi nonparametrik yang diasumsikan tidak diketahui bentuknya. Karena fungsi ini didekati dengan fungsi regresi MARS sebagaimana dinyatakan pada persamaan (5), maka fungsi regresi MARS dapat dinyatakan untuk fungsi regresi f sebagai berikut:

$$f(x_{1i}, x_{2i}, \dots, x_{pi}) = \alpha_0 + \sum_{m=1}^{M} \alpha_m \prod_{k=1}^{K_m} \left[s_{km}(x_{v(k,m)i} - t_{km}) \right]_+ \quad (14)$$

dengan:

Jika $s_{km} = +1$, maka $\left[+(x_{v(k,m)i} - t_{km})\right]_+ = \begin{cases} x_{v(k,m)i} - t_{km}, jika\ x_{v(k,m)i} > t_{km} \\ \quad 0 \qquad , sebaliknya \end{cases}$

Jika $s_{km} = -1$, maka $\left[-(x_{v(k,m)i} - t_{km})\right]_+ = \begin{cases} t_{km} - x_{v(k,m)i}, jika\ t_{km} > x_{v(k,m)i} \\ \quad 0 \qquad , sebaliknya \end{cases}$

Fungsi regresi MARS yang terdapat pada persamaan (14) dapat disederhanakan untuk fungsi regresi f sebagai berikut:

$$f\left(x_{1i}, x_{2i}, \dots, x_{pi}\right) = \alpha_0 + \sum_{m=1}^{M} \alpha_m B_{mi}(\boldsymbol{x}, \boldsymbol{t}) \qquad (15)$$

Dengan $B_{mi}(\boldsymbol{x}, \boldsymbol{t})$ pada persamaan (15) adalah fungsi basis untuk $f\left(x_{1i}, x_{2i}, \dots, x_{pi}\right)$ dalam model MARS, yang dapat dinyatakan dalam bentuk persamaan sebagai berikut:

$$B_{mi}(\boldsymbol{x}, \boldsymbol{t}) = \prod_{k=1}^{K_m} \left[s_{km}(x_{v(k,m)i} - t_{km})\right]_+ \qquad (16)$$

Jika diuraikan untuk setiap observasi $(i = 1,2,..,n)$, maka fungsi regresi MARS pada persamaan (14) dapat diuraikan dalam bentuk berikut:

$$\begin{aligned} f\left(x_{11}, x_{21}, \dots, x_{p1}\right) &= \alpha_0 + \sum_{m=1}^{M} \alpha_m \prod_{k=1}^{K_m} \left[s_{km}(x_{v(k,m)1} - t_{km})\right]_+, \\ f\left(x_{12}, x_{21}, \dots, x_{p2}\right) &= \alpha_0 + \sum_{m=1}^{M} \alpha_m \prod_{k=1}^{K_m} \left[s_{km}(x_{v(k,m)2} - t_{km})\right]_+, \\ &\vdots \\ f\left(x_{1n}, x_{2n}, \dots, x_{pn}\right) &= \alpha_0 + \sum_{m=1}^{M} \alpha_m \prod_{k=1}^{K_m} \left[s_{km}(x_{v(k,m)n} - t_{km})\right]_+, \end{aligned} \qquad (17)$$

Berdasarkan banyaknya fungsi basis $(m = 1,2,..,M)$, maka fungsi regresi MARS pada persamaan (17) dapat diuraikan sebagai berikut:

$$\begin{aligned} f\left(x_{11}, \dots, x_{p1}\right) &= \alpha_0 + \alpha_1 \prod_{k=1}^{K} \left[s_{k1}(x_{v(k,1)1} - t_{k1})\right]_+ + \cdots + \alpha_M \prod_{k=1}^{K_M} \left[s_{kM}(x_{v(k,M)1} - t_{kM})\right]_+, \\ f\left(x_{12}, \dots, x_{p2}\right) &= \alpha_0 + \alpha_1 \prod_{k=1}^{K} \left[s_{k1}(x_{v(k,1)2} - t_{k1})\right]_+ + \cdots + \alpha_M \prod_{k=1}^{K_M} \left[s_{kM}(x_{v(k,M)2} - t_{kM})\right]_+, \\ &\vdots \\ f\left(x_{1n}, \dots, x_{pn}\right) &= \alpha_0 + \alpha_1 \prod_{k=1}^{K} \left[s_{k1}(x_{v(k,1)n} - t_{k1})\right]_+ + \cdots + \alpha_M \prod_{k=1}^{K_M} \left[s_{kM}(x_{v(k,M)n} - t_{kM})\right]_+, \end{aligned}$$

$$(18)$$

Berdasarkan fungsi regresi nonparametrik, model MARS dinyatakan dalam persamaan berikut:

$$y_i = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} \left[s_{km}(x_{v(k,m)} - t_{km})\right] + \varepsilon_i \qquad (19)$$

Dari model MARS persamaan (19) dalam bentuk matriks dapat ditulis sebagai berikut (Prihastuti Yasmirullah dkk., 2021):

$$Y = Ba + \varepsilon \qquad (20)$$

dimana:

Y : variabel respon yang berukuran $(n \times 1)$

B : $\left[1, \left(x_{v(k,m)} - t_{km}\right)_1^K\right]$ = basis fungsi yang berukuran $\left(n \times (M+1)\right)$

a : koefisien regresi yang berukuran $\left((M+1) \times 1\right)$

ε : error dari variabel acak yang diasumsikan independen serta menyebar normal dengan rataan nol dan ragam tertentu σ^2 yang berukuran $(n \times 1)$

dengan, $Y = (y_1, \dots, y_n)^T$, $a = (a_0, \dots, a_M)^T$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$, dan

$$B = \begin{pmatrix} 1 & \prod_{k=1}^{K_1}\left[s_{k1}(x_{v1(k,1)} - t_{k1})\right] & \cdots & \prod_{k=1}^{K_M}\left[s_{kM}(x_{v1(k,M)} - t_{kM})\right] \\ 1 & \prod_{k=1}^{K_1}\left[s_{k1}(x_{v2(k,1)} - t_{k1})\right] & \cdots & \prod_{k=1}^{K_M}\left[s_{kM}(x_{v2(k,M)} - t_{kM})\right] \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1}\left[s_{k1}(x_{vn(k,1)} - t_{k1})\right] & \cdots & \prod_{k=1}^{K_M}\left[s_{kM}(x_{vn(k,M)} - t_{kM})\right] \end{pmatrix}$$

(Prihastuti Yasmirullah dkk., 2021)

Untuk memperoleh estimator $\hat{a}$ dengan menggunakan metode estimasi *Ordinary Least Square* (OLS), yaitu dengan meminimumkan jumlah kuadrat errornya atau Sum Square Error (SSE). Estimasi modelnya diperoleh dari persamaan:

$$\varepsilon = Y - B\alpha \tag{21}$$

$$\begin{aligned} JKE = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon^T\varepsilon &= (Y - B\alpha)^T(Y - B\alpha) \\ &= (Y^T - \alpha^T B^T)(Y - B\alpha) \\ &= Y^T Y - Y^T B\alpha - \alpha^T B^T Y + \alpha^T B^T B\alpha \\ &= Y^T Y - 2\alpha^T B^T Y + \alpha^T B^T B\alpha \end{aligned} \tag{22}$$

Penyelesaian optimasi pada persamaan (22) dilakukan dengan menurunkan atau melakukan diferensial JKE secara parsial terhadap α dengan hasil sebagai berikut:

$$\begin{aligned} \left.\frac{\partial(\varepsilon^T\varepsilon)}{\partial\alpha}\right|_{\alpha=\hat{\alpha}} &= \frac{\partial(Y^T Y - 2\hat{\alpha}^T B^T Y + \hat{\alpha}^T B^T B\hat{\alpha})}{\partial\hat{\alpha}} = 0 \\ -2B^T Y + 2B^T B\hat{\alpha} &= 0 \\ -B^T Y + B^T B\hat{\alpha} &= 0 \\ B^T B\hat{\alpha} &= B^T Y \\ \hat{\alpha} &= (B^T B)^{-1} B^T Y \end{aligned} \tag{23}$$

Sehingga diperoleh taksiran α dapat dirumuskan sebagai berikut:

$$\hat{\alpha} = (B^T B)^{-1} B^T Y \tag{24}$$

Beberapa hal yang perlu diperhatikan dalam membuat model MARS yaitu (Febriyanti dkk., 2013):

a. Knot, adalah nilai variabel prediktor (variabel bebas) ketika *slope* suatu garis regresi mengalami perubahan yang dapat didefinisikan sebagai akhir dari satu segmen sekaligus merupakan awal dari segmen yang lain. Pada setiap titik knot, diharapkan adanya kontinuitas dari fungsi basis satu region dengan region lainnya. Minimum observasi (MO) antara knot adalah 0, 1, 2, dan 3 observasi.

b. *Basis Function* (BF) atau fungsi basis, merupakan suatu fungsi yang digunakan untuk menunjukkan hubungan antara variabel terikat dan variabel bebas (Pintowati & Otok, 2012). Fungsi basis yang dipilih biasanya berbentuk polinomial dengan turunan yang kontinu pada setiap titik knot. Maksimum fungsi basis yang disarankan adalah 2-4 kali dari jumlah variabel bebasnya.

c. *Interaction* (interaksi), merupakan hasil perkalian silang antar variabel yang saling berkorelasi (saling berhubungan). Adapun jumlah maksimum interaksi (MI) yang diperbolehkan adalah 1, 2, atau 3. Jika MI lebih dari 3, model yang dihasilkan akan semakin kompleks dan lebih sulit untuk diinterpretasikan.

1.6.4 Pemilihan Model MARS

Pemilihan model terbaik *Multivariate Adaptive Regression Splines* (MARS) dilakukan dengan dua pendekatan. Pendekatan pertama yaitu menggunakan *forward stepwise* untuk memperoleh jumlah basis fungsi, pemilihan basis fungsi dengan meminimumkan *Average sum of Square Residual* (ASR). Pendekatan kedua yaitu pendekatan *backward stepwise* untuk memenuhi konsep Parsimoni (model sederhana) yaitu dengan memilih basis fungsi yang dihasilkan dari *forward stepwise* dengan meminimumkan nilai *Generalized Cross Validation* (GCV) (Friedman, 1991).

Prosedur *forward* akan menghasilkan model dengan banyak fungsi basis. Dalam prakteknya, biasanya jumlah maksimum fungsi basis yang akan digunakan dalam model dibatasi. Demikian juga untuk derajat interaksi, yang seringkali hanya dibatasi hanya sampai derajat tiga (Dukalang, 2017). Pada *Multivariate Adaptive Regression Splines* (MARS), model terbaik diperoleh berdasarkan nilai *Generalized Cross Validation* (GCV) minimum diantara model-model yang lain (Budiantara dkk., 2006). Fungsi GCV untuk memilih model terbaik (Friedman, 1991), sebagai berikut:

$$LOF(\hat{f}_M) = GCV(M) = \frac{\frac{1}{N}\sum_{i=1}^{N}\left[y_i - \hat{f}_M(x_i)\right]^2}{\left[1 - \frac{\tilde{C}(M)}{N}\right]^2} \tag{25}$$

dengan:

$$\tilde{C}(M) = C(M) + d.M$$
$$C(M) = tr[B(B^T B)^{-1} B^T] + 1$$

dimana:

LOF : *lack-of-fit test*
$\hat{f}_M(x_i)$: nilai taksiran variabel respon pada M fungsi basis di x_i
N : ukuran sampel
$\tilde{C}(M)$: fungsi model kompleks = $C(M) + d.M$. Nilai d menyatakan besaran *smoothing parameter*. Nilai optimumnya berada dalam interval 2 $\leq d \leq$ 4
$C(M)$: banyaknya parameter yang diestimasi
M : jumlah basis fungsi yang ditentukan
d : nilai optimal fungsi basis 2 $\leq d \leq$ 4

y_i : variabel terikat
x_i : variabel bebas

Kriteria lain yang digunakan ialah nilai MSE (Eubank, 1999) yang dihitung dengan menggunakan persamaan sebagai berikut:

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \tag{26}$$

dimana:
y_i : peubah respon ke-i
$\hat{y}_i$: nilai dugaan peubah respon ke-i
n : banyaknya data

1.6.5 Normalisasi Data

Normalisasi data merupakan proses untuk membuat skala nilai atribut ke dalam rentang yang lebih kecil dengan bobot yang sama. Normalisasi diperlukan jika data yang ada pada dataset memiliki suatu nilai dengan rentang yang tidak sama (Suryanegara dkk., 2021). Salah satu metode normalisasi data yaitu *min-max normalization*. *Min-max normalization* merupakan suatu teknik transformasi linear yang menggunakan nilai minimum dan maksimum untuk menghasilkan keseimbangan antara data satu dengan yang lainnya pada rentang yang sama (Suyanto, 2018). *Min-max normalization* dihitung menggunakan rumus berikut:

$$X_i^N = \frac{X_i - X_{min}}{X_{max} - X_{min}} \tag{27}$$

dimana:
X_i^N : hasil normalisasi data
X_i : nilai tertentu yang akan dinormalisasi
X_{min} : nilai minumum dari semua data
X_{max} : nilai maksimum dari semua data

Misal terdapat data sebagai berikut [60,30,20,25,10], kemudian kumpulan data asli akan dinormalisasi menjadi suatu nilai baru dengan range antara [0,1]. Diketahui bahwa nilai minimum yaitu 10 dan nilai maksimum yaitu 60. Maka nilai baru setelah ditransformasi yakni :

$X_1^N = \frac{60-10}{60-10} = \frac{50}{50} = 1$

$X_2^N = \frac{30-10}{60-10} = \frac{20}{50} = 0.4$

$X_3^N = \frac{20-10}{60-10} = \frac{10}{50} = 0.2$

$X_4^N = \frac{25-10}{60-10} = \frac{15}{50} = 0.3$

$X_5^N = \frac{10-10}{60-10} = \frac{0}{50} = 0$

Sehingga, diperoleh hasil data baru sebagai berikut [1,0.4,0.2,0.3,0].

1.6.6 Algoritma Differential Evolution

Differential Evolution (DE) merupakan suatu teknik optimasi yang dikembangkan oleh Kenneth Price yang dipublikasikan dalam majalah Dr. Dobb's Journal pada Oktober 1994 (Price dkk., 2005). Metode ini merupakan suatu pendekatan optimasi matematis untuk fungsi-fungsi multidimensi dan termasuk dalam kelompok algoritma evolusioner (Julianto, 2016). Kemudian, seiring dengan kemajuannya, dalam ICEO (*International Contest on Evolutionary Optimization*) yang pertama, *Differential Evolution* (DE) menjadi salah satu algoritma genetika terbaik dan dapat menemukan global optimum yang multi dimensi dengan probabilitas yang baik (Arif & Mauludiyanto, 2015).

Differential Evolution (DE) memperbaiki kekurangan algoritma evolusi lainnya dengan menggunakan strategi optimasi yang sederhana untuk proses optimasi yang cepat (waktu perhitungan yang cepat dengan sedikit iterasi untuk menemukan solusi global optimal). Metode ini adalah evolusi dari *Genetic Algorithm* (GA), dengan mengganti operator logika dengan operator matematis (Fajarwati & Anggraeni, 2012). Menurut Storn & Price (1997), beberapa kelebihan dari algoritma DE dibandingkan dengan algoritma lain yaitu kemampuan untuk menyelesaikan persamaan (non-differentiable, nonlinear, dan multimodal), dapat bekerja secara terpisah dengan program utama sehingga dapat mempercepat proses komputasi secara keseluruhan, kemudahan penggunaan misalnya dengan menggunakan variabel kontrol yang sedikit namun cukup berpengaruh pada algoritma tersebut, serta mempunyai parameter konvergensi yang baik (Jason & Prayogo, 2023). Ada empat tahapan pemrosesan dalam algoritma *Differential Evolution* (DE) (Sholihin & Donoriyanto, 2023) dan (Abdulelah Al-Sudani et al., 2019), yaitu:

a. Inisialisasi

Pada tahap inisialisasi akan dilakukan pembangkitan bilangan *random* sebanyak ukuran populasi antara (0,1). Pada tahap ini, populasi awal (x) (populasi *parents*) akan dihasilkan. Proses inisialisasi dinyatakan dengan persamaan berikut:

$$x_{i,g} = (Ub - Lb) \times rand + Lb \tag{28}$$

Bilangan *random* tersebut dihasilkan berdasarkan distribusi uniform pada rentang (0,1) atau dalam kata lain 0 ≤ $rand$ < 1. Dimana Ub merupakan *upper bound* dan Lb merupakan *lower bound* yaitu batas bawah dan atas yang mengatur ruang pencarian parameter dalam algoritma Differential Evolution.

b. Mutasi

Tahap selanjutnya yaitu mutasi, dimana akan menghasilkan populasi mutan (V). Pembentukan individu mutan dilakukan dengan menggabungkan tiga individu yang dipilih secara acak dari populasi awal dengan F sebagai faktor skala pembeda. Nilai individu pertama dipilih sekali untuk semua individu dipopulasi yang sama, dan nilai individu kedua dan ketiga dipilih untuk setiap individu yang akan dibentuk. Pembentukan dilakukan sebanyak jumlah populasi. Adapun persamaan yang dibentuk untuk vektor mutan adalah:

$$v_{i,g+1} = x_{r1,g} + F \times (x_{r2,g} - x_{r3,g}) \tag{29}$$

Bilangan bulat $r_1, r_2, r_3 \in \{1,2,3,\dots,NP\}$ dipilih secara acak dan akan berubah ketika terjadi generasi/iterasi (i) baru, sehingga $r_1 \neq r_2 \neq r_3 \neq i$.

c. *Crossover*

Nilai individu dari populasi awal akan dipertukarkan dengan nilai individu dari populasi mutan dengan menggunakan perbandingan nilai CR dan bilangan *random* untuk membentuk populasi trial.

$$u_{ji,g+1} = \begin{cases} v_{ji,g+1} \; jika \; r(j) \leq CR \; atau \; j = rn(i) \\ x_{ji,g} \; untuk \; lainnya \end{cases} \tag{30}$$

Probabilitas *crossover*, $CR \in (0,1)$ adalah nilai yang digunakan untuk mengendalikan fraksi nilai variabel yang disalin dari mutan. Nilai CR memberi pengaruh terhadap individu yang akan mengalami *crossover*, semakin tinggi nilainya maka semakin banyak individu yang mengalami *crossover.*

d. Seleksi

Seleksi dilakukan diantara populasi hasil *crossover* dengan populasi awal, yaitu Jika hasil *crossover* (u) memiliki nilai fungsi tujuan lebih besar dari populasi awal (x) maka populasi awal (x) tetap pada posisinya. Jika nilai fungsi tujuan populasi hasil crossover (u) lebih rendah dari populasi awal (x), populasi hasil crossover (u) akan digunakan untuk generasi berikutnya. Menurut Price dkk (2005), pada dasarnya ada dua tahapan dalam proses evolusi yang menggunakan seleksi yaitu *parent selection* dan *survivor selection* (Julianto, 2016). Pada *parent selection*, vektor dipilih berdasarkan nilai fungsi terbaik dan probabilitas seleksi yang tinggi. Adapun pada *survivor selection*, dilakukan replikasi untuk menentukan vektor mana yang akan bertahan dalam generasi selanjutnya. Untuk mengetahui apakah vector menjadi anggota generasi $g + 1$, maka vector uji $u_{i,g+1}$ dibandingkan dengan vector target $x_{i,g}$. Apabila vektor $u_{i,g+1}$ menghasilkan nilai fungsi yang lebih kecil daripada $x_{i,g}$ maka $x_{i,g+1}$ akan diatur menjadi $u_{i,g+1}$, dan jika sebaliknya maka nilai $x_{i,g}$ yang lama dipertahankan. Penjelasan tersebut ditunjukan dalam persamaan berikut :

$$x_{i,g+1} = \begin{cases} u_{i,g+1} \; jika \; f(u_{i,g+1}) < f(x_{i,g}) \\ x_{i,g} \; yang \; lainnya \end{cases} \tag{31}$$

Setelah melalui tahapan yang sudah dijelaskan sebelumnya, penghentian iterasi dilakukan hingga kondisi optimal tercapai.

1.6.7 Indeks Harga Saham Gabungan

Indeks Harga Saham Gabungan (IHSG) merupakan suatu indeks yang menjelaskan mengenai pergerakan nilai harga saham yang tercatat pada Bursa Efek Indonesia (BEI) (Mochlasen dkk., 2023). Indeks inilah yang paling banyak digunakan sebagai acuan dalam perkembangan kegiatan di pasar modal. IHSG melibatkan semua harga saham yang tercatat di bursa dan dapat digunakan untuk menilai situasi pasar secara keseluruhan atau mengukur apakah harga saham mengalami kenaikan atau penurunan (Dewi, 2020). Indeks Harga Saham Gabungan (IHSG) ini sangatlah dibutuhkan untuk memahami pergerakan harga saham di indonesia dan menggunakannya sebagai acuan untuk portofolio saham. Beberapa faktor makro dan mikro ekonomi yang mempengaruhi Indeks Harga

Saham Gabungan (IHSG) yaitu nilai tukar (kurs) rupiah bulanan terhadap dolar Amerika, inflasi, tingkat suku bunga di Indonesia, indeks saham Dow Jones, indeks saham Nikkei 225, dan indeks saham Shanghai Stock Exchange Composite.

a. Nilai Tukar (Kurs) Rupiah

Nilai tukar rupiah merupakan harga rupiah terhadap mata uang negara lain. Jadi, Nilai tukar rupiah adalah nilai mata uang rupiah terhadap mata uang asing, seperti nilai tukar rupiah terhadap dolar Amerika, nilai tukar rupiah terhadap Euro, dan sebagainya (Savira & Hidayat, 2021). Nilai tukar menjadi sangat penting bagi suatu perekonomian, hal tersebut karena pengaruh dari nilai kurs yang besar didalam suatu neraca dari transaksi yang sedang berjalan maupun untuk variabel makro di dalam suatu perekonomian. Kurs mata uang menjadi acuan bagi investor. Investor dapat memanfaatkan kurs untuk bertransaksi di valuta asing atau juga menanamkan modal di sebuah perusahaan. Apabila mata uang sedang lemah, investor tidak mungkin akan menanamkan modalnya di Negara yang kurs mata uangnya melemah (Sutandi et al., 2021).

b. Inflasi

Inflasi merupakan kecenderungan kenaikan harga barang dan jasa termasuk faktor–faktor produksi secara umum dan berlangsung terus menerus. Inflasi merupakan suatu faktor makro ekonomi yang dapat sekaligus menguntungkan atau merugikan suatu perusahaan. Para pelaku pasar modal tidak menyukai inflasi yang tinggi karena akan meningkatkan biaya produksi. Kenaikan biaya produksi perusahaan menyebabkan kenaikan harga barang dalam negeri yang akan berdampak pada kinerja perusahaan dan hal ini dapat terlihat dari harga sahamnya (Savira & Hidayat, 2021). Kenaikan harga dari satu atau dua barang saja tidak dapat dianggap sebagi inflasi kecuali jika kenaikan tersebut berdampak pada kenaikan harga sebagian besar barang barang lain (Dewi, 2020).

c. Tingkat Suku Bunga

Suku bunga SBI merupakan tingkat suku bunga acuan di Indonesia. Jika suku bunga suatu negara meningkat, masyarakat akan memilih untuk menabung dalam bentuk deposito, yang akan menyebabkan investasi saham menurun sehingga harga saham dan pergerakan IHSG juga menurun. Apabila suku bunga SBI mengalami peningkatan maka akan mengakibatkan penurunan pergerakan IHSG, namun jika suku bunga SBI menurun hal tersebut akan mengakibatkan peningkatan IHSG (Zabidi & Asandimitra, 2018).

d. Indeks saham Dow Jones

Indeks Dow Jones merupakan salah satu dari tiga indeks utama yang ada di Amerika Serikat (*New York Stock Exchange*). Petumbuhan ekonomi negara maju pada dasarnya mempunyai hubungan terhadap perekonomian negara berkembang. Indeks Dow Jones sebagai cerminan dari kinerja perekonomian Amerika Serikat, apabila indeks Dow Jones mengalami peningkatan maka perekonomian Amerika juga akan membaik. Amerika Serikat merupakan salah satu negara tujuan ekspor bagi negara Indonesia, jika ekonomi Amerika Serikat terus berkembang, pertumbuhannya dapat mendorong pertumbuhan ekonomi Indonesia baik melalui kegiatan ekspor, investasi langsung, dan investasi pada pasar modal Indonesia (Jayanti et al., 2014).

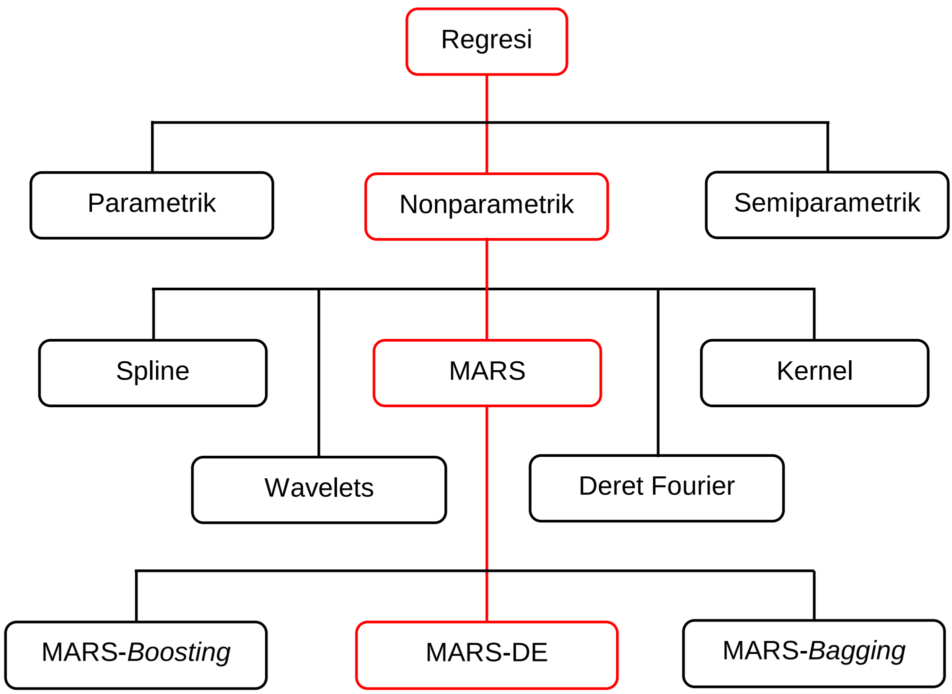
e. Indeks saham Nikkei 225

Indeks Nikkei 225 merupakan sebuah Indeks besar milik Jepang. Peningkatan terhadap indeks Nikkei 225 menunjukkan bahwa perekonomian Jepang sedang mengalami peningkatan. Meningkatnya perekonomian Jepang akan mengakibatkan peningkatan ekspor dari Indonesia ke Jepang, sehingga investasi di Indonesia akan meningkat. Meningkatnya investasi akan membuat harga saham naik dan Indeks Harga Saham Gabungan (IHSG) juga akan mengalami peningkatan (Zabidi & Asandimitra, 2018).

f. Indeks Shanghai Stock Exchange Composite (SSEC)

Indeks pasar saham SSEC merupakan indeks pasar saham utama China dan umumnya digunakan sebagai indikator utama untuk mengevaluasi kinerja pasar saham di Bursa Efek China. Perubahan perekonomian China akan berdampak pada Indonesia, karena Indonesia dan China memiliki kesamaan geografis yaitu terletak di kawasan Asia (Putra & Nurmatias, 2024). Oleh karenanya, kenaikan IHSG dapat disebabkan oleh peningkatan indeks Shanghai Stock Exchange Composite, yang merupakan salah satu indikator ekonomi yang mengukur kinerja perekonomian China (Beureukat & Andriani, 2021).

1.6.8 Kerangka Konseptual



Gambar 1. Kerangka Konseptual

BAB II

METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder, yaitu data yang diperoleh dari sumber yang telah dipublikasikan. Data IHSG (Indeks Harga Saham Gabungan) bulanan, data indeks saham Dow Jones, indeks saham Nikkei 225, dan indeks Shanghai Stock Exchange Composite diperoleh dari situs finance.yahoo.com. Data kurs dan inflasi diperoleh dari situs www.bi.go.id. Data suku bunga di Indonesia bulanan diperoleh dari situs Badan Pusat Statistik diakses melalui https://www.bps.go.id/indicator/13/379/1/bi-rate.html. Adapun data yang digunakan adalah data Indeks Harga Saham Gabungan periode 2017 sampai dengan 2022.

2.2 Variabel Penelitian

Dalam penelitian ini terdapat satu variabel respon (Y) yaitu Indeks Harga Saham Gabungan (IHSG) dan terdapat enam variabel prediktor (X) yaitu Nilai Tukar (Kurs) Rupiah terhadap Dolar Amerika bulanan (X_1), Inflasi bulanan (X_2), Suku Bunga di Indonesia bulanan (X_3), Indeks saham Dow Jones (X_4), Indeks saham Nikkei 225 (X_5), dan Indeks Shanghai Stock Exchange Composite (X_6). Adapun definisi operasional variabel dalam penelitian ini disajikan dalam Tabel 1 berikut.

Tabel 1. Definisi Operasional Variabel

Variabel	Keterangan	Satuan	Definisi Operasional
Y	IHSG	Rupiah	Indeks harga saham gabungan yang dimiliki perusahaan Bursa Efek Indonesia pada bulan Januari tahun 2017 – bulan Desember 2022.
X_1	Kurs	Rupiah	Tingkat perubahan nilai tukar jual rupiah terhadap dolar Amerika yang diterbitkan oleh Bank Indonesia setiap bulan selama periode Januari tahun 2017 – bulan Desember 2022.
X_2	Inflasi	Persen	Tingkat perubahan inflasi indeks harga konsumen nasional yang dipublikasikan oleh Bank Indonesia selama bulan Januari 2017 – bulan Desember 2022.
X_3	Suku Bunga	Persen	Tingkat perubahan pada suku bunga BI Rate yang dipublikasikan oleh Bank Indonesia selama bulan Januari 2017

			– bulan Desember 2022.
X_4	Indeks Saham Dow Jones	Rupiah	Indeks Dow Jones merupakan indeks pasar saham yang ada di Bursa Efek Amerika Serikat.
X_5	Indeks Saham Nikkei 225	Rupiah	Indeks Nikkei 225 merupakan indeks pasar saham yang ada di Bursa Efek Jepang.
X_6	Indeks Saham Shanghai Stock Exchange Composite	Rupiah	Indeks Shanghai Stock Exchange Composite merupakan indeks pasar saham yang ada di Bursa Efek China.

2.3 Metode Analisis

Adapun tahapan-tahapan analisis dalam pemodelan *Multivariate Adaptive Regression Spline* dengan Algoritma *Differential Evolution* adalah sebagai berikut:

1. Melakukan eksplorasi data pada setiap variabel terikat dengan masing-masing variabel bebasnya untuk mendapatkan statistik deskriptif data.
2. Melakukan normalisasi data terhadap setiap variabel yang terlibat (variabel terikat dan variabel bebas) dengan metode *min-max normalization*
$$X_i^N = \frac{X_i - X_{min}}{X_{max} - X_{min}}$$
Kemudian dilanjutkan dengan membuat plot antar variabel setiap variabel terikat dengan masing-masing variabel bebasnya.
3. Melakukan pemodelan Indeks Harga Saham Gabungan (IHSG) menggunakan MARS-DE dengan menentukan jumlah maksimum *Basis Function* (BF), Maksimum Interaksi (MI), dan Minimum Observasi (MO).
   a. Jumlah maksimum interaksi (MI), jumlah maksimum interaksi yaitu 1, 2, dan 3. Apabila terdapat lebih dari 3 interaksi, maka akan menimbulkan interpretasi model yang sangat kompleks.
   b. Menentukan minimum observasi (MO) yaitu 0,1,2, dan 3.

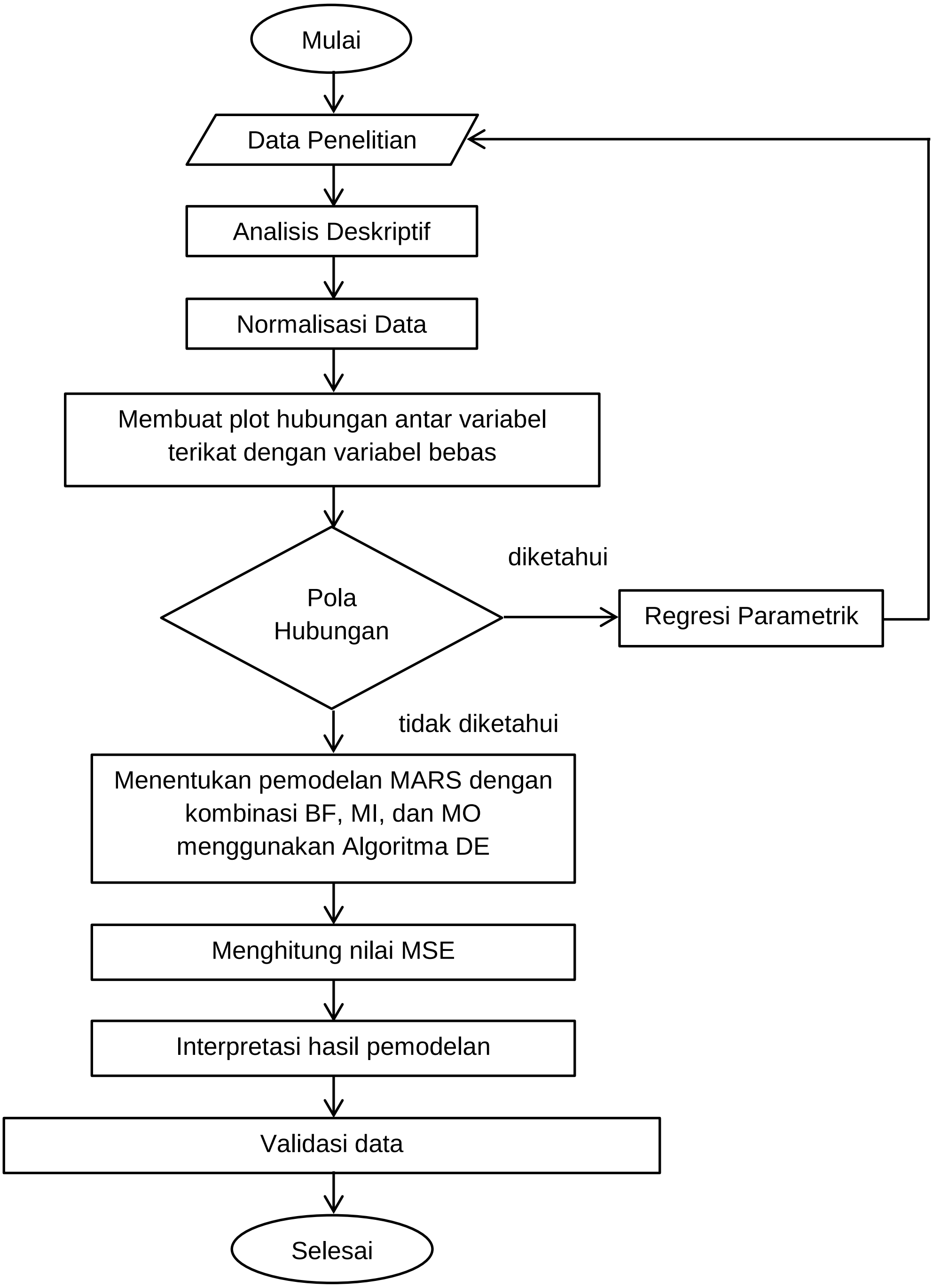
   Pemodelan *Multivariate Adaptive Regression Spline* menggunakan algoritma *differential evolution* dengan langkah berikut:
   a. Inisialisasi nilai parameter (BF, MI, MO) dari MARS secara acak serta dilakukan pembangkitan bilangan *random* berdasarkan distribusi *uniform.* Proses inisialisasi seperti yang diberikan dalam persamaan:
$$x_{i,g} = (Ub - Lb) \times rand + Lb$$
   b. Melakukan proses mutasi
$$v_{i,g+1} = x_{r1,g} + F \times (x_{r2,g} - x_{r3,g})$$
   c. Melakukan proses *crossover*
$$u_{ji,g+1} = \begin{cases} v_{ji,g+1} \; jika \; r(j) \leq CR \; atau \; j = rn(i) \\ x_{ji,g} \; untuk \; lainnya \end{cases}$$

d. Melakukan proses seleksi. Jika trial vector mempunyai nilai fungsi tujuan yang lebih kecil dari fungsi tujuan vector targetnya. Maka, trial vector akan menggantikan posisi fungsi tujuan vector targetnya dalam populasi pada generasi berikutnya. Jika terjadi sebaliknya, vector target akan tetap pada posisinya dalam populasi.

4. Menentukan model *Multivariate Adaptive Regression Spline* (MARS) terbaik berdasarkan pada nilai *Generalized Cross Validation* (GCV). Semakin kecil nilai GCV, maka semakin baik model tersebut dibandingkan model lainnya.
5. Mendapatkan variabel-variabel yang berpengaruh signifikan dari pembentukan model MARS.
6. Menghitung nilai *Mean Square Error* (MSE) dari hasil prediksi model yang didapatkan.
7. Membuat kesimpulan penelitian.

2.4 Diagram Alir

Diagram alir (*flow chart*) dalam penelitian ini adalah sebagai berikut:



Gambar 2. Diagram Alir