

APLIKASI MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS) UNTUK MENGETAHUI FAKTOR YANG MEMPENGARUHI CURAH HUJAN DI KOTA MAKASSAR

Muhammad Reski Mattalunru¹, Suwardi Annas¹, Muhammad Kasim Aidid¹

¹Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Makassar, Indonesia

Keywords: Nonparametric regression, rainfall, MARS, GCV

Abstract:

Nature make nonparametric regression a model for the present and the future. Paying attention to natural phenomena today is increasingly difficult to predict. The rainy season is one of the natural phenomena that is increasingly leading to an erratic pattern. The month which usually marks the dry season, suddenly experienced heavy rainfall and even caused a lot of losses. So modeling is needed to find out what factors affect rainfall. The Multivariate Adaptive Regression Splines (MARS) method is one of the modern modeling methods with high estimation capabilities. In addition, MARS has flexible properties and toughness in dealing with high-dimensional data, namely data that has 3 x 20 independent variables and a sample data size of 50 n 1000. The MARS model is obtained from a combination of Basis Function (BF), Maximum Interaction (MI), and Minimum Observation (MO) with a small Generalized Cross-Validation (GCV). In this study, the number of independent variables used was 4 variables. air temperature, humidity, wind speed, and air pressure are independent variables that affect rainfall in Makassar City with contribution levels of 86.54%, 100%, 39.38%, and 54.68%, respectively. The best combination of MARS models in this study is BF=12, MI=1, and MO=1 with GCV=31.14

1. Pendahuluan

Analisis Regresi digunakan untuk mencari pengaruh variabel dependen terhadap variabel independent. Menurut Budiantara (2005), berdasarkan bentuk kurva regresinya analisis regresi terbagi menjadi tiga yaitu analisis regresi parametrik, analisis regresi semi parametrik dan analisis regresi nonparametrik. Penggunaan regresi parametrik membutuhkan asumsi yang sangat ketat seperti: nilai eror yang harus menyebar normal, kehomogenan ragam eror dan kurva regresi yang harus diketahui Sehingga dianggap tidak mampu menangani data yang berdimensi tinggi. Kekurangan regresi parametrik dapat diatasi dengan penggunaan regresi nonparametrik. Regresi nonparametrik memiliki Kemampuan mengestimasi yang tinggi dan mempunyai sifat fleksibel dalam memodelkan data (Budiantara, 2005).

MARS merupakan metode pengembangan regresi nonparametrik dengan menggabungkan *spline* dengan *recursive partitioning regression spline* (Friedman, 1991). Penggunaan MARS banyak digunakan oleh para peneliti pada dewasa ini beberapa diantaranya ada Arifin dan Mitakda (2012) menggunakan MARS Bagging untuk memodelkan

* Corresponding author.

E-mail address: reskimattalunru@gmail.com



presentase gizi buruk balita di Jawa Timur, Otok (2010) menggunakan MARS untuk pengelompokkan zona musim, Annur dkk (2015) menggunakan MARS variabel respon biner, Irmawati dkk (2019) kasus bayi dengan Berat Badan Lahir Rendah (BBLR) dan masih banyak lagi penelitian lainnya yang menggunakan MARS.

Pada kasus curah hujan pemodelan curah hujan dengan menggunakan regresi linear telah dilakukan oleh (Paradipta, Sembiring, & Bangun, 2013). Namun data curah hujan merupakan data runtut waktu dengan data yang diperoleh berdasarkan waktu (bulanan atau tahunan) dan menurut Sutikno (2002), seringkali data iklim dan cuaca melanggar asumsi terutama bentuk kurva regresinya tidak diketahui. Olehnya itu penulis menganggap bahwa penggunaan data curah hujan melanggar asumsi penggunaan regresi parametrik.

Sebagai kota besar Kota Makassar juga tidak pernah lepas dari bencana alam. Setiap tahun di Kota Makassar terjadi bencana alam akibat cuaca buruk. Banjir menjadi salah satu bencana alam yang sering melanda seiring dengan masuknya musim penghujan (Bongi, Rogi, & Sela, 2020). Curah hujan di Kota Makassar termasuk ke dalam kategori monsunial yakni terdapat curah hujan yang lebat di awal dan akhir tahun (Pabalik, Ihsan, & Arsyad, 2015).

Selain perlunya memperbaiki saluran pembuangan air dan penyediaan lahan terbuka hijau sebagai daerah serapan air menurut penulis pemerintah Kota juga mesti melakukan modifikasi cuaca. Berkaca pada bencana banjir yang melanda Kota Makassar di tahun 2000 dan tercatat di tanggal 3 Februari terjadi curah hujan yang sangat lebat mencapai 376 mm/hari (Pabalik, Ihsan, & Arsyad, 2015). Adanya anomali cuaca yang berubah-ubah maka dibutuhkan pengkajian terkait faktor-faktor apa saja yang mempengaruhi curah hujan di Kota Makassar. untuk mendeteksi faktor-faktor tersebut maka dilakukan pemodelan curah hujan dengan metode *Multivariate Adaptive Regression Spline* (MARS).

2. Tinjauan Pustaka

2.1. Analisis Regresi

Analisis regresi adalah sebuah metode yang digunakan untuk mencari pengaruh antara variabel bebas dan variabel terikat (Tiro M. A., 2011). Secara umum model regresi yang menjelaskan hubungan antara variabel terikat (y_i) dengan variabel bebas (x_i) adalah sebagai berikut (Eubank, 1999).

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

Dengan ε_i adalah error dari variabel acak yang menyebar identik dan independent dengan nilai tengah nol dan σ^2 ragam tertentu. Sementara $f(x_i)$ merupakan fungsi regresi atau curva regresi. (Eubank, 1999). Hubungan antara variabel bebas dan variabel terikat dinyatakan dalam wujud stokastik linear dan non linear (Otok B. W., 2005). Menurut Budiantara (2005) regresi jika ditinjau dari pola hubungannya maka regresi terbagi menjadi regresi parametrik, regresi semiparametrik dan nonparametrik.

Apabila kita mengasumsikan bahwa terdapat pola hubungan yang bersifat parametrik maka regresi parametrik sangat efisien digunakan mengingat kesederhanaan dan kemudahan dalam melakukan interpretasi, namun regresi parametrik memiliki asumsi-asumsi yang sangat ketat apabila asumsi ini salah satunya dilanggar maka hasil yang diperoleh dianggap bias asumsi tersebut antara lain; asumsi normalitas, homokedastisitas, autokorelasi dan multikolinearitas serta pola hubungan antara variabel (Wicaksono, suparti, & Suparti, 2014). Keterbatasan yang dimiliki analisis regresi parametrik dalam pemodelan kurva regresi yang cenderung bebas dapat diatasi dengan metode regresi nonparametrik.

2.2 Regresi Spline

Dalam pendekatan regresi nonparametrik, regresi spline merupakan salah satu jenis *piecewise polynomial* yakni jenis polynomial yang memiliki sifat tersegmen. Model yang tersegmen membuat spline lebih unggul dari *polynomial* biasa dalam fleksibilitas dan menyebabkan regresi spline dapat menyesuaikan diri dengan data (Yani, Srinadi, & Sumarjaya, 2017). Kemampuan regresi spline lainnya adalah untuk mengatasi pola data yang naik dan turun secara tajam dengan bantuan titik Knot. Secara umum spline dengan orde q dapat dinyatakan sebagai berikut (Wicaksono, suparti, & Suparti, 2014):

$$f(x) = \beta_0 + \beta_1 x + \dots + \beta_q x^q + \sum_{k=1}^K \gamma_k (x - t_k)_+^q \quad (2)$$

Dengan $q \geq 1$, $(t_1, t_2, \dots, t_K)$ adalah titik-titik knot, dan fungsi dari truncated power $(x - t_k)_+^q$ adalah :

$$(x - t_k)_+^q = \begin{cases} (x - t_k)^q & ; x - t_k > 0 \\ 0 & ; x - t_k \leq 0 \end{cases} \quad (3)$$

Sementara fungsi basis pada spline univariat dengan K knot

$$(Wicaksono, suparti, & Suparti, 2014): \left\{ x^j \right\}_1^q, \left\{ (x - t_k)_+^q \right\}_1^K \quad (4)$$

2.3. Regresi Partisi Rekursif

Perkembangan iptek utamanya dalam bidang komputasi semakin memudahkan peneliti dalam melakukan pengolahan data secara cepat dan akurat. Salah satu hasil dari perkembangan program komputasi adalah Regresi Partisi Rekursif (RPR). RPR adalah sebuah pendekatan yang menggunakan pengembangan fungsi basis untuk mengatasi permasalahan curfa regresi yang tidak diketahui. Friedmann (1991) mengutarakan bahwa tujuan dari RPR ini untuk menggunakan data sebagai estimator subregion dan parameter yang berasosiasi pada setiap subregion. Kelemahan Regresi Partisi Rekursif sendiri terletak pada diskontinuitas pada tiap knot atau *subregion*-nya. Friedmann (1991) menjabarkan persamaan RPR sebagai berikut:

$$f(x) = \sum_{m=1}^M a_m B_m(x) \quad (5)$$

Bentuk fungsi basis B_m dari RPR ialah:

$$B_m(x) = \prod_{k=1}^{K_m} H[s_{km} \cdot (x_{v(k,m)} - t_{km})] \quad (6)$$

Keterangan:

- a_m = koefisien fungsi basis ke- m
- M = maksimum fungsi basis
- K_m = derajat interaksi
- $H[\eta] = \begin{cases} 1 & ; \eta \geq 0 \\ 0 & ; \text{lainnya} \end{cases}$ = fungsi tangga (*step function*)
- s_{km} = tanda pada titik knot (bernilai ± 1)
- $x_{v(k,m)}$ = variabel bebas ke- v
- t_{km} = nilai knot dari variabel bebas $x_{v(k,m)}$

Modifikasi dilakukan Friedman pada algoritma *step function* $H[\eta]$ dengan fungsi *truncated spline*:

$$b_q^\pm(x - t) = [\pm(x - t)]_+^q \quad (7)$$

Langkah selanjutnya truncated spline yang telah dimodifikasi dimasukkan ke dalam fungsi basis sehingga diperoleh fungsi basis:

$$B_m^q(x) = \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})]_+^q \quad (8)$$

Dengan $q=1$ dan tanda + mengindikasikan nilai nol jika argument bernilai negatif.

2.4. Multivariate Adaptive Regression Splines

Multivariate Adaptive Regression Splines (MARS) merupakan suatu metode regresi nonparametric yang diperkenalkan oleh J.H. Friedman tahun 1990 metode ini merupakan pengembangan dari metode RPR yang memiliki kendala pada model yang tidak kontinu. Pengembangan ini dilakukan dengan menggunakan prinsip spline untuk menciptakan kontinuitas pada tiap segmen, model ini juga berguna untuk mengatasi data yang berdimensi tinggi dengan variabel bebas $3 \leq x \leq 20$ dan ukuran data sampel $50 \leq n \leq 1000$ (Friedman J. H., 1991). Hal-hal yang harus diperhatikan dalam membuat model MARS, yaitu (Nash & Brandford, 2001) :

1. *Knot*, adalah akhir dan awal garis region lain. Pada setiap titik knot diharapkan terjadi kontinuitas dari fungsi basis satu region dengan region lainnya. Minimum observasi (MO) antara knot adalah 0, 1, 2, dan 3 observasi.
2. *Basis Function* (BF) atau fungsi basis, yakni sebuah fungsi yang digunakan untuk menjelaskan informasi dari hubungan antara variabel bebas dan variabel terikat. Maksimum fungsi basis yang disarankan friedman adalah 2 sampai 4 kali dari jumlah variabel bebas.
3. *Interaksi*, adalah korelasi antara variabel. Maksimum interaksi (MI) yang dibolehkan adalah 1,2, atau 3. Jika MI lebih dari 3 akan diperoleh model yang sangat kompleks dan akan semakin sulit diinterpretasi.

Estimator model MARS dituliskan sebagai berikut (Friedman J. H., 1991) :

$$f(x) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - t_{km})]_+ \quad (9)$$

keterangan:

a_0 = koefisien fungsi basis *parent*

a_m = koefisien fungsi basis ke – m
 M = maksimum fungsi basis
 K_m = maksimum derajat interaksi
 s_{km} = bernilai ± 1
 $x_{v(k,m)}$ = variabel bebas ke- v
 t_{km} = nilai knot dari variabel bebas

Dari persamaan (2.9) dapat dituliskan kembali dalam bentuk :

$$f(x) = a_0 + \sum_{K_m=1} f_1(x_1) + \sum_{K_m=2} f_{12}(x_1, x_2) + \dots + \sum_{K_m=k} f_{12\dots p}(x_1, x_2, \dots, x_p) \quad (10)$$

Penjumlahan pertama pada persamaan (10) meliputi semua fungsi basis untuk satu variabel, penjumlahan kedua meliputi semua fungsi basis yang melibatkan dua variabel untuk interaksi dengan cara yang sama penjumlahan ketiga meliputi semua fungsi basis untuk interaksi tiga variabel dan seterusnya. Contoh jika $V(m) = \{v(k, m)\}_1^{K_m}$ sebuah himpunan dari variabel yang dihubungkan dengan fungsi basis ke m , (B_m), maka seluruh penjumlahan pertama pada persamaan (2.10) dapat dituliskan sebagai berikut (Friedman J. H., 1991) :

$$f_1(x_1) = \sum_{K_m=1} a_m B_m(x_1) \quad (11)$$

Fungsi (2.11) merupakan penjumlahan semua basis fungsi untuk satu variabel yang mewakilkan fungsi univariat. Seluruh fungsi bivariat pada penjumlahan kedua dan fungsi multivariat pada penjumlahan ketiga dalam persamaan (2.10) dapat dituliskan sebagai berikut (Friedman J. H., 1991):

$$f_{12}(x_1, x_2) = \sum_{K_m=2} a_m B_m(x_1, x_2) \quad (12)$$

$$f_{12\dots p}(x_1, x_2, \dots, x_p) = \sum_{K_m=k} a_m B_m(x_1, x_2, \dots, x_p) \quad (13)$$

Memperhatikan fungsi (12) dan (13) Penambahan fungsi univariat dan bivariat dengan variabel terkait memiliki interaksi dengan variabel respon dan dijabarkan dalam bentuk fungsi berikut (Friedman J. H., 1991):

$$f_{12}^*(x_1, x_2) = f_1(x_1) + f_2(x_2) + f_{12}(x_1, x_2) \quad (14)$$

$$f_{12\dots p}^*(x_1, x_2, \dots, x_p) = f_1(x_1) + f_2(x_2) + \dots + f_p(x_p) + f_{12}(x_1, x_2) + \dots + f_{1\dots p}(x_1, x_2, \dots, x_p) \quad (15)$$

Dalam bentuk matrix:

$$\hat{f}(x) = B\hat{a} \quad (16)$$

Dengan :

$$\begin{aligned}
 \hat{a} &= (a_0, a_1, a_2, \dots, a_M)^T \\
 B &= \begin{bmatrix} 1 & \prod_{k=1}^{K_1} [s_{1m'}(x_{v1(1,m)} - t_{1m})]_+ & \dots & \prod_{k=1}^{K_M} [s_{Mm'}(x_{v1(M,m)} - t_{Mm})]_+ \\ 1 & \prod_{k=1}^{K_1} [s_{1m'}(x_{v2(1,m)} - t_{1m})]_+ & \dots & \prod_{k=1}^{K_M} [s_{Mm'}(x_{v2(M,m)} - t_{Mm})]_+ \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1} [s_{1m'}(x_{vn(1,m)} - t_{1m})]_+ & \dots & \prod_{k=1}^{K_M} [s_{Mm'}(x_{vn(M,m)} - t_{Mm})]_+ \end{bmatrix}
 \end{aligned}$$

(Otok B. W., 2010)

$$\underline{Y} = \underline{f}(x) + \underline{\varepsilon}$$

$$\hat{\underline{Y}} = \hat{\underline{f}}(x) \quad (17)$$

$$\hat{\underline{Y}} = B\hat{\underline{a}}$$

Dengan:

$\underline{Y} = (y_1, y_2, y_3, \dots, y_n)^T$ dan $\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_n)^T$ serta $\hat{\underline{a}}$ merupakan estimasi atau pendugaan nilai $\underline{a}$ (Ampulembang, Otok, Rumiat, & Budiasih, 2015).

2.5. Estimasi Parameter MARS Data Kontinu

Persamaan (17) dapat dituangkan ulang Ke dalam bentuk:

$$\underline{Y} = B \underline{a} + \underline{\varepsilon} \quad (18)$$

dengan ε adalah eror dari variabel acak yang diasumsikan identik, independen dan menyebar normal dengan nilai tengah nol dengan ragam tertentu σ^2 atau digambarkan dengan $IIDN(0, \sigma^2)$.

untuk menaksir parameter $\underline{a}$, metode estimasi parameter yang digunakan adalah metode *Least Square* (Ampulembang, Otok, Rumiati, & Budiasih, 2015):

$$\hat{\underline{a}} = (B^T B)^{-1} B^T \underline{Y} \quad (19)$$

Apabila persamaan (19) disubstitusikan ke persamaan (17) maka akan menghasilkan persamaan seperti berikut (Ampulembang, Otok, Rumiati, & Budiasih, 2015):

$$\hat{\underline{Y}} = B(B^T B)^{-1} B^T \underline{Y} \quad (20)$$

Dengan memisalkan $H = B(B^T B)^{-1} B^T$ maka akan menjadi:

$$\hat{\underline{Y}} = H \underline{Y} \quad (21)$$

(Ampulembang, Otok, Rumiati, & Budiasih, 2015).

2.6. Pemilihan Model Terbaik MARS

Pemilihan model terbaik MARS dilakukan dengan dua pendekatan yang pertama menggunakan *forward stepwise* untuk memperoleh jumlah fungsi basis pemilihan fungsi basis dengan cara meminimumkan *Average sum of Square Residual* (ASR) (Friedman J. H., 1991). Selanjutnya dilakukan pendekatan *backward stepwise* untuk memenuhi konsep Parsimoni (model yang sederhana) dengan cara menghilangkan fungsi basis yang memiliki kontribusi kecil terhadap respon dari forward stepwise dengan meminimumkan *Generalized Cross Validation* (GCV) (Otok B. W., 2005). Hal yang perlu diperhatikan dalam penentuan model terbaik pada MARS adalah melihat nilai GCV yang paling kecil.

Fungsi GCV untuk memilih model terbaik (Friedman & Silverman, 1989).

$$LOF(\hat{f}_M) = GCV(M) = \frac{\frac{1}{N} \sum_{i=1}^N [y_i - \hat{f}_M(x_i)]^2}{\left[1 - \frac{\tilde{C}(M)}{N}\right]^2} \quad (22)$$

Dimana:

$$\tilde{C}(M) = C(M) + d \cdot M \quad (23)$$

$$C(M) = tr[B(B^T B)^{-1} B^T] + 1 \quad (24)$$

keterangan:

LOF = *lack-of-fit test*

$\hat{f}_M$ = penduga fungsi f dengan M fungsi basis

N = ukuran sampel

$C(M)$ = banyaknya parameter yang diestimasi

d = nilai optimal fungsi basis, $2 \leq d \leq 4$

y_i = Variabel terikat.

2.7. Hujan

Hujan merupakan wujud persipitasi uap air yang terdapat di troposfer sehingga jumlahnya dipengaruhi oleh suhu tekanan udara dan angin. Hujan adalah peristiwa sampainya air dalam bentuk cair maupun padat yang di curahkan dari atmosfer ke permukaan bumi. Bentuk lain dari persipitasi uap air seperti salju dan es. Uap air naik ke atmosfer tepatnya di troposfer sehingga mendingin dan terjadi kondensasi menjadi butir-butir air dan kristal-kristal es yang akhirnya jatuh sebagai hujan (Triatmojo, 1998). Curah hujan merupakan ketinggian air hujan yang terkumpul dala tempat yang datar, tidak menguap, tidak meresap dan tidak mengalir (Ramli, 2010). Jumlah air yang jatuh ke permukaan bumi dapat diukur dengan alat penakar hujan. Satuan yang digunakan untuk mengukur curah hujan adalah millimeter (mm) intensitas hujan adalah jumlah curah hujan persatuan jangka waktu tertentu. Hujan terbentuk di daerah trofosfer dan pada lapisan ini terjadi gejala klimatologi yang akan mempengaruhi jumlah curah hujan yang turun, unsur-unsur klimatologi tersebut antara lain:

- a) Suhu udara, merupakan salah satu unsur paling penting untuk diamati dalam pemantauan cuaca maupun iklim. Suhu udara sangat bergantung pada keadaan radiasi matahari oleh sebab itu suhu udara sering mengalami perubahan. Suhu akan mengalami fluktuasi secara nyata selama 24 jam (Ramli, 2010). Fluktuasi ini berkaitan

erat dengan proses pertukaran energi yang berlangsung di atmosfer. Pada siang hari sebagian dari radiasi matahari akan diserap oleh gas-gas dan partikel-partikel yang melayang dalam atmosfer sehingga mengakibatkan suhu udara meningkat (Pabalik, Ihsan, & Arsyad, 2015). suhu udara diukur menggunakan thermometer air raksa. Suhu udara harian rata-rata didefinisikan sebagai rata-rata pengamatan selama 4 jam dalam sehari. (Ramli, 2010).

- b) Kelembaban udara, adalah jumlah kandungan air keseluruhan seperti uap, tetes air, dan kristal es yang terdapat diudara dalam suatu waktu dan tempat tertentu (lakitan, 2002). Tingginya kelembaban suatu wilayah menjadi faktor yang dapat menstimulus hujan. Data kelembaban udara yang diamati dan dilaporkan pada umumnya merupakan data kelembaban relatif, kelembaban relative yang dimaksud adalah perbandingan antara tekanan uap air yang terukur dengan tekanan uap air pada kondisi jenuh (Ramli, 2010). Diukur dengan menggunakan hygrometer dinyatakan dengan satuan persen.
- c) Angin, adalah pergerakan molekul udara yang terjadi di atas permukaan bumi, yang disebabkan oleh perbedaan tekanan udara pada dua arah yang berbeda. Perbedaan tekanan udara pada daerah tropis diakibatkan oleh perbedaan tingkat pemanasan matahari di permukaan bumi (Ramli, 2010). Udara akan bergerak dari tempat yang bertekanan tinggi ke tempat yang bertekanan rendah selain itu perputaran bumi juga mempengaruhi arah, kecepatan dan kekuatan angin. Kecepatan angin diukur dengan anemometer dengan satuan meter per detik (m/s).

3. Metode Penelitian

3.1. Jenis Penelitian

Penelitian ini adalah penelitian terapan dengan tinjauan kuantitatif, dengan mengambil data yang diperlukan selanjutnya melakukan analisis dengan *Multivariate Adaptive Regression Spline* (MARS) untuk mengetahui faktor-faktor yang mempengaruhi curah hujan di Kota Makassar.

3.2. Sumber Data

Data yang digunakan diperoleh dari publikasi Badan Meteorologi Klimatologi dan Geofisika (BMKG) *stasiun meteorology maritime Paotere* melalui website www.bmkg.go.id dan publikasi BPS Kota Makassar melalui website www.makassarkota_bps.go.id, variabel terkait yang akan digunakan diambil dari rentang waktu tahun 2015-2020. Variabel bebas yang digunakan adalah suhu udara, kelembaban udara, kecepatan angin, dan tekanan udara.

3.3. Definisi Operasional Variabel

Variabel yang digunakan pada penelitian ini adalah:

- a) Curah Hujan (Y) milimeter (mm).
- b) Suhu Udara (X_1) derajat celcius ($^{\circ}\text{C}$).
- c) Kelembaban Udara (X_2) satuan persen (%).
- d) Kecepatan Angin (X_3) meter per detik (m/s).
- e) Tekanan Udara (X_4) milibar (mb).

3.4. Prosedur Penelitian

Adapun prosedur yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Mencari serta mengumpulkan referensi dan informasi terkait yang dibutuhkan dalam penelitian.
2. Mengambil data terkait yang dibutuhkan sesuai dengan peubah yang digunakan.
3. Mengolah data menggunakan teknik analisis yang sesuai dengan tujuan penelitian.
4. Menyusun laporan penelitian.
5. Membuat kesimpulan berdasarkan hasil penelitian.

3.5. Teknik Analisis

Untuk mencapai tujuan penelitian maka langkah-langkah analisisnya disusun sebagai berikut:

1. Membuat statistika deskriptif dan plot data variabel terikat dengan masing-masing variabel bebasnya.
2. Apabila kurva regresi dari plot antara variabel terikat dan masing-masing variabel bebasnya tidak diketahui maka dilanjutkan dengan metode regresi nonparametrik yakni MARS (*Multivariate Adaptive Regression Splines*).
3. Menentukan maksimum *Basis Function* (BF) yang telah ditentukan yakni 2-4 kali dari jumlah variabel bebasnya. Variabel bebas yang digunakan dalam penelitian ini sebanyak 4 variabel maka maksimum fungsi basis yg terbentuk adalah 8, 12, dan 16.
4. Menentukan jumlah maksimum interaksi (MI) yaitu 1, 2, dan 3 interaksi. Apabila $MI < 3$ maka akan menyebabkan interpretasi model yang sangat kompleks.

5. Menentukan model terbaik dengan memperhatikan nilai *Generalized Cross Validation* (GCV). Jika nilai GCV Semakin kecil, maka semakin baik model tersebut dibandingkan model lainnya.
6. Mendapatkan variabel-variabel yang berpengaruh signifikan dari pembentukan model MARS.
7. Melakukan Interpretasi hasil.

4. Hasil dan Pembahasan

4.1. Deskripsi Variabel

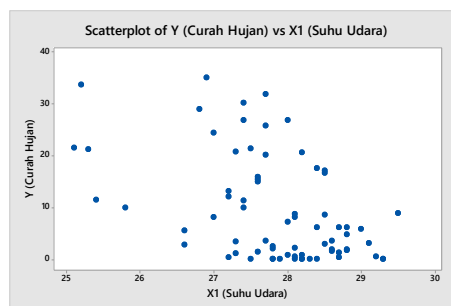
Data yang digunakan pada penelitian ini adalah data curah hujan bulanan dari tahun 2015-2020 yang telah melalui tahapan estimasi data hilang dengan metode rerata aljabar akibat terdapat beberapa data yang kurang lengkap. Untuk menjawab pertanyaan-pertanyaan penelitian yang telah diajukan oleh peneliti maka pada bab ini dijabarkan langkah-langkah pengujian yang terdiri dari beberapa tahapan. Tahap yang pertama dilakukan statistik deskriptif terhadap semua variabel-variabel terkait serta membuat plot antara variabel terikat (Y) Curah hujan terhadap variabel-variabel bebas yakni (X1) Suhu udara, (X2) Kelembaban udara, (X3) Kecepatan angin, (X4) Tekanan udara. Variabel-variabel bebas yang merupakan data kontinu dideskripsikan dengan nilai maksimum, minimum, rata-rata, serta tabel sebaran frekuensi dan rata-rata curah hujan pada kelompok-kelompok tertentu. Langkah-langkah yang dilakukan dalam pembuatan sebaran frekuensi menggunakan aturan *Struges* (Tiro M. A., 2008)

4.1.1. Curah Hujan (Y)

Curah hujan di Kota Makassar sepanjang tahun 2015 -2020 memiliki titik curah hujan tertinggi yakni di bulan Januari 2015 dimana pada bulan tersebut memiliki rata rata hujan sebesar 35.135 mm/hari dan kondisi terendah dengan curah hujan 0 mm. terjadi pada bulan Juli, Agustus, September dan Oktober 2015 dan bulan Agustus 2016 serta bulan Agustus, September, dan Oktober 2019. Dari data tersebut menunjukkan bahwa tahun dengan cuaca paling ekstrim terjadi di tahun 2015 dimana terdapat 4 bulan tanpa adanya hujan dan juga terdapat bulan Januari dengan curah hujan yang lebat 35.135 mm. curah hujan di Kota Makassar sangat lebat di awal dan di akhir tahun yang dikenal dengan pola monsun.

4.1.2. Suhu Udara (X1)

suhu udara di Kota Makassar dari Januari 2015-Desember 2020 memiliki suhu paling rendah pada bulan Februari 2020 dengan suhu 25,1°C dan suhu udara paling tinggi pada bulan November 2015 dengan suhu udara rata-rata 29,5°C dan rata-rata suhu udara pada periode 2015-2020 adalah 27,82°C. Untuk mengetahui kurva regresi antara curah hujan dan suhu udara, berikut ini diberikan plot antara curah hujan dengan suhu udara di Kota Makassar pada Gambar 1

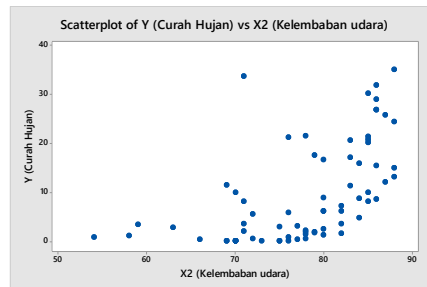


Gambar 1. Plot curah hujan dengan suhu udara.

Pada Gambar 1 menunjukkan bahwa antara curah hujan dan suhu udara tidak memiliki sebuah pola hubungan yang jelas

4.1.3. Kelembaban Udara (X2)

kelembaban udara paling tinggi selama 2015-2020 di Kota Makassar sebesar 88% dan yang paling rendah 54% yang terjadi pada bulan September 2020. Rata-rata kelembaban udara di kota Makassar sepanjang Januari 2015-Desember 2020 sebesar 77,76%. Untuk mengetahui kurva regresi antara curah hujan dan kelembaban udara, berikut ini diberikan plot antara curah hujan dengan suhu udara di Kota Makassar pada Gambar 2

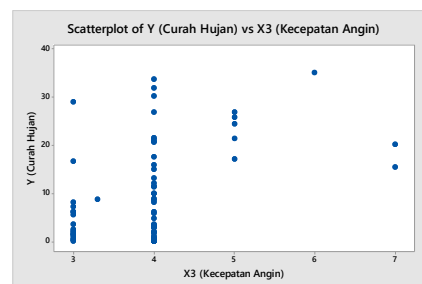


Gambar 2. Plot Data Curah Hujan dengan Kelembaban Udara.

Pada Gambar 2. menunjukkan bahwa masing-masing variabel bebas memiliki pola hubungan yang cenderung nonlinear

4.1.4. Kecepatan Angin (X3)

Kecepatan angin pada Januari 2015-Desember 2020 di Kota Makassar memiliki kecepatan tertinggi sebesar 7 m/s terjadi pada bulan Desember 2016 dan bulan Februari 2017 sedangkan kecepatan angin yang terendah 3 m/s. dan rata-rata kecepatan angin sebesar 3,92 m/s. Untuk mengetahui kurva regresi antara curah hujan dan kecepatan angin, berikut ini diberikan plot antara curah hujan dengan suhu udara di Kota Makassar pada Gambar 3

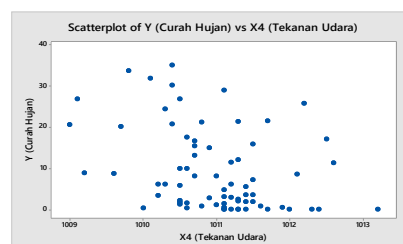


Gambar 3. Plot Data Hubungan Curah Hujan dan Kecepatan Angin.

Pada Gambar 3. menunjukkan bahwa semakin tinggi kecepatan angin tidak menjamin bahwa curah hujan akan semakin tinggi.

4.1.5. Tekanan Udara (X4)

tekanan udara di Kota makassar selama Januari 2015-Desember 2020 memiliki nilai terendah sebesar 1009 mb. terjadi pada bulan Nopember 2017. dan tekanan udara tertinggi sebesar 1013,2 mb. pada bulan September 2019 dengan rata-rata 1010,9 mb. Untuk mengetahui kurva regresi antara curah hujan dan tekanan udara , berikut ini diberikan plot antara curah hujan dengan suhu udara di Kota Makassar pada Gambar 4



Gambar 4. Plot Data Curah Hujan (Y) dengan Tekanan udara (X4)

Dari Gambar 4. diketahui bahwa data curah hujan dan tekanan udara tidak menunjukkan pola hubungan yang jelas. Dari scatterplot antara Variabel curah hujan (Y) dengan variabel variabel bebas yang diasumsikan berpengaruh dapat ditarik kesimpulan bahwa ada hubungan antara variabel Curah hujan dengan Variabel bebas Suhu udara, Kelembaban udara, Kecepatan angin dan Tekanan udara namun tidak memiliki pola data yang jelas sehingga menjadi pertimbangan utama untuk menggunakan regresi nonparametric dengan metode MARS yang dianggap mampu mengatasi permasalahan dimensi data yang tinggi dan diskontinuitas pada data.

4.2. Pemodelan Curah Hujan di Kota Makassar menggunakan MARS

Pada penelitian ini untuk memodelkan curah hujan peneliti menggunakan *Software Salford Predictive Modeler* (SPM) versi 8.3. dalam pemodelan Curah hujan di Kota Makassar digunakan kombinasi Basis Function (BF),

Maksimum Observasi (MO) dan Minimum Interaksi (MI). BF diperoleh berdasarkan kriteria yakni 2-4 kali kelipatan jumlah Variabel bebas, pada penelitian ini menggunakan 4 variabel bebas maka BF yang digunakan adalah 8, 12 dan 16, Maksimum Interaksi yang di anjurkan adalah 1, 2, dan 3 interaksi jika lebih dari tiga maka akan menghasilkan model yang kompleks. Nilai 1 pada maksimum interaksi menunjukkan bahwa variabel variabel bebas akan memiliki hubungan langsung dengan Variabel terikat, jika MI nilainya 2 maka terdapat interaksi antara dua variabel bebas yang berpengaruh terhadap variabel terikat, serta jika nilai MI adalah 3 maka terdapat interaksi antara 3 variabel bebas yang berpengaruh terhadap Variabel terikat. Pada Minimum Observasi (MO) yang dianjurkan adalah 0, 1, 2 dan 3 observasi sehingga memberi kontinuitas antara satu region dengan region lainnya. Pada tahap selanjutnya dilakukan metode Forward stepwise dan Backward stepwise untuk membentuk model terbaik pada setiap kombinasi BF, MI dan MO.

1. *Forward stepwise*
 - a. Menentukan BF ke 0 = 1
 - b. Membentuk BF hingga BF ke M berdasarkan kombinasi Variabel Bebas dan titik knot.
 - c. Membentuk interaksi antar BF sehingga menghasilkan GCV minimum
 - d. Mengulangi langkah c sehingga diperoleh model MARS sebanyak Maksimum BF
2. *Backward stepwise*
 - a. Menghapus BF hasil Forward Stepwise yang memiliki Kontribusi yang kecil terhadap model untuk nilai GCV yang kecil
 - b. Mengulang langkah 1 untuk mendapatkan BF yang konstan dan optimal.

Berdasarkan kombinasi BF, MI dan MO yang telah melewati forward stepwise dan backward stepwise maka diperoleh 36 model terbaik pada masing-masing kombinasi dan selanjutnya akan dipilih model terbaik berdasarkan kriteria nilai GCV terkecil apabila terdapat nilai GCV yang sama maka dipilih melalui *Mean Square Error* (MSE) terkecil apabila terdapat MSE yang sama maka kriteria yang terakhir adalah nilai R^2 atau koefisien determinasi tertinggi. Berikut diberikan hasil model MARS terbaik berdasarkan kombinasi BF, MI dan MO kombinasi BF= 12, MI= 1 dan Mo= 1 sebagai model terbaik dengan nilai GCV minimum 31,14 dan R^2 Maksimum 81,20% dan bentuk modelnya sebagai berikut:

$$\begin{aligned} Y = & -77,4917 + 4,2764 * BF1 + 33,6979 * BF4 \\ & + 5,70488 * BF6 - 58,2981 * BF7 \\ & + 26,7361 * BF8 - 7,58973 * BF9 + 36,0657 * BF11; \end{aligned} \quad (25)$$

Dengan:

$$BF1 = \max(0; X2 - 82) \quad BF4 = \max(0; 25,8 - X1) \quad BF6 = \max(0; 1011,1 - X4)$$

$$BF7 = \max(0; X3 - 6) \quad BF8 = \max(0; 6 - X3) \quad BF9 = \max(0; X2 - 86) \quad BF11 = \max(0; X3 - 3,3)$$

Dari persamaan (25) semua variabel bebas mempengaruhi variabel terikat yakni suhu udara(X_1), kelembaban udara(X_2), kecepatan angin(X_3) dan tekanan udara (X_4). Sementara itu terdapat 7 basis fungsi yang berkontribusi dalam pembentukan model (25) yakni BF1, BF4, BF6, BF7, BF8, BF9, dan BF11. Pada model (25) tidak terdapat interaksi antara variabel-variabel bebas yang berpengaruh terhadap variabel terikat.

Interpretasi Model MARS

Interpretasi hasil MARS dengan BF=12, MI= 1 dan MO= 1 diperoleh model MARS sebagaimana yang dijabarkan pada persamaan (25) dan interpretasi dari masing-masing *Basis Function* (BF) sebagai berikut :

1. BF1 = max(0, X2 - 82)

Setiap peningkatan satu satuan BF1 maka akan menyebabkan curah hujan di Kota Makassar meningkat 4,2764 mm dengan fungsi basis lainnya bernilai konstan. Dan BF1 akan bernilai ($X_2 - 82$) jika kelembaban udara lebih dari 82% dan akan bernilai 0 jika kurang dari atau sama dengan 82%.

2. BF4 = max(0, 25,8 - X1)

Setiap peningkatan satu satuan BF4 maka akan meningkatkan curah hujan di Kota Makassar sebesar 33,6979 mm dengan fungsi basis lainnya bernilai konstan, BF4 akan bernilai ($25,8 - X_1$) jika suhu udara lebih rendah dari 25,8 °C dan akan bernilai 0 jika sama atau lebih tinggi dari 25,8 °C

3. BF6 = max(0, 1011,1 - X4)

Setiap peningkatan satu satuan BF6 maka akan meningkatkan 5,70489 mm curah hujan di Kota Makassar dengan fungsi basis lainnya bernilai konstan, BF6 akan bernilai ($1011,1 - X_4$) jika tekanan udara lebih rendah dari 1011,1 mb. dan akan bernilai 0 jika lebih atau sama dengan 1011,1 mb.

4. BF7 = max(0, X3 - 6)

Setiap peningkatan BF7 satu satuan akan mengurangi curah hujan di Kota Makassar sebesar 58,29894 mm dengan fungsi basis lainnya bernilai konstan. BF7 akan bernilai (X3-6) jika kecepatan angin lebih dari 6 m/s dan BF7 akan bernilai 0 jika sama dengan atau lebih rendah dari 6 m/s.

5. BF8 = max(0, 6 - X3)

Setiap peningkatan satu satuan BF8 akan meningkatkan curah hujan di Kota Makassar sebesar 26,73661 mm dengan fungsi basis lainnya bernilai konstan. BF8 akan bernilai (6-X3) jika kecepatan angin lebih lambat dari 6 m/s apabila kecepatan angin sama atau lebih cepat dari 6 m/s maka BF8 akan bernilai 0

6. BF9 = max(0, X2 - 86)

Setiap peningkatan satu satuan BF9 akan mengurangi curah hujan di Kota Makassar sebesar 7,58977 mm dengan fungsi basis lainnya dianggap konstan, BF9 akan bernilai (X2-86) jika kelembaban udara lebih dari 86% dan akan bernilai 0 jika kelembaban udara sama dengan atau lebih rendah dari 86%.

7. BF 11= max(0, X3- 3,3)

Setiap peningkatan satu satuan BF11 akan meningkatkan curah hujan sebesar 36.06637 mm dengan BF lainnya bernilai Konstan. BF11 akan bernilai (X3-3,3) jika kecepatan angina lebih dari 3,3 m/s dan akan bernilai 0 jika sama dengan atau lebih rendah dari 3,3 m/s.

4.3. Uji Signifikansi Model MARS

Pengujian ini dilakukan untuk melihat signifikansi parameter dan mengevaluasi kecocokan model, dengan cara pengecekan koefisien regresi secara simultan dan secara parsial

1. Uji Simultan.

Uji simultan dilakukan dengan menggunakan uji F untuk mengetahui pengaruh variabel bebas secara bersamaan terhadap variabel terikat berdasarkan koefisien fungsi basis (α)

Keputusan : Berdasarkan hasil pengolahan data menggunakan SPM 8.3 diperoleh P-value sebesar 0,00000 dan nilai F hitung sebesar 39,664 sedangkan F tabel ($F_{(0,05;18;64)} = 2,15$). Karena $F_{hit} > F_{(0,05;7;64)}$ dan $p - value \leq \alpha$ maka H_0 ditolak. Dan dapat disimpulkan bahwa variabel bebas secara Bersama-sama signifikan mempengaruhi variabel terikat sehingga sehingga layak digunakan untuk memodelkan curah hujan

2. Uji Parsial

Setelah dilakukan uji simultan yang menunjukkan bahwa secara bersama-sama variabel bebas mempengaruhi variabel terikat maka dalam uji parsial ini kita akan melihat Variabel bebas apa saja yang berpengaruh terhadap variabel terikat. Statistic uji yang akan digunakan adalah uji t.

Keputusan : t_{hit} menyatakan nilai hitung uji t dan $t_{(0,025;64)} = 1,98$ menyatakan nilai tabel dari statistik uji t pada taraf signifikansi yang digunakan dengan dk adalah 64. Berdasarkan statistik uji yang digunakan, diketahui bahwa seluruh parameter BF memiliki nilai $p - value < \alpha$, sehingga H_0 ditolak. Hal ini menunjukkan bahwa setiap parameter BF yang mewakili variabel bebas secara parsial berpengaruh signifikan terhadap variabel terikat.

5. Kesimpulan

5.1. Kesimpulan

- Dari kombinasi BF = 8, 12 dan 16 , MI= 1,2 dan 3, serta MO= 0,1,2,dan 3 terbentuk 36 model MARS dan model MARS yang terpilih adalah Model BF=12, MI=1 dan MO=1 dengan kriteria model terbaik berdasarkan nilai GCV terkecil = 31,14 dan nilai R^2 sebesar 81,20% sehingga model tersebut yang terpilih dalam memodelkan curah hujan di Kota Makassar dengan persamaan sebagaimana yang dijelaskan pada Persamaan (25).
- Berdasarkan model terbaik yang diperoleh menggunakan MARS telah diketahui bahwa seluruh Variabel bebas berpengaruh secara signifikan yaitu suhu udara (X1), kelembaban udara (X2), kecepatan angin (X3) dan tekanan udara (X4) dengan variabel kelembaban udara (X2) sebagai variabel bebas yang memiliki kontribusi terbesar dalam meminimumkan GCV pada Model MARS.

5.2. Saran

- Bagi yang ingin menggunakan metode MARS pada Kasus curah Hujan di tempat lain sebaiknya perhatikan kehomogenan letak geografis wilayah tersebut jika tidak homogen silahkan gunakan MARS dengan data biner.
- Bagi yang akan menggunakan metode MARS ada baiknya jika dikembangkan pada software R

References

- Ampulembang, A. P., Otok, B. W., Rumiati, A. T., & Budiasih. (2015). Bi-Responses Nonparametric Regression Model Using MARS and Its Propertis. *Applied Mathematical Sciences* 9(29), 1417-1427.
- Annur, M., Dahlan, J. A., & Agustina, F. (2015). Penerapan Metode Multivariate Adaptive Regression Spline (MARS) untuk Menentukan Faktor yang Mempengaruhi Masa Studi Mahasiswa FPMIPA UPI. *Eurekamatika* 3(1), 135-155.
- Arisandi, R., & Annas, S. (2016). Metode MARS Bagging dalam Memodelkan Berat Badan Balita Sulawesi-Selatan. *Seminar Nasional VARIANSI*, 160-182.
- BMKG. (2021, 8 25). *BMKG Database*. Diambil kembali dari Badan Meteorologi Klimatologi dan Geofisika: <http://www.bmkg.go.id/cuaca/probabilistik-curah-hujan.bmkg>
- Bongi, A., Rogi, O. H., & Sela, R. E. (2020). Mitigasi Risiko Banjir di Kota Makassar. *SABUA* 9(1), 1-12.
- Budiantara, I. N. (2005). Model Keluarga Spline Polinomial Truncated Dalam Regresi Semiparametrik. *Berkala Mipa* 15(3), 55-61.
- Eubank, R. L. (1999). *Nonparametrik Regression and Spline Smoothing (2end ed.)*. New York: Marcell Dekker.
- Friedman, J. H. (1991). Multivariat Adaptive Regression Splines. *The Annals of Statistics* 19(1), 1-67.
- Friedman, J., & Silverman, B. (1989). Flexible Parsimony Smoothing and Additive Modeling. *Techonometrics* 31, 3-39.
- Irmawati, Bustan, M. N., & Annas, S. (2019). Aplikasi Multivariate Adaptive Regression Spline (MARS) Terhadap Pemodelan Risiko Kesehatan Bayi dengan Berat Badan Lahir Rendah (BBLR). *VARIANSI: Journal of Statistics and Its appplication on Teaching and Research* 1(2), 1- 6.
- Mahardy, A. I. (2014). *Analisis dan Pemetaan Daerah Rawan Banjir di Kota Makassar Berbasis Spatial*. Makassar: skripsi Teknik Sipil, Universitas Hasanuddin.
- Nash, M., & Brandford, D. (2001). Parametric and Nonparametric logistic Regression for Prediction of Presence/Absence of An Amphibian. *U.S. Enviromental rotection Agency*.
- Otok, B. W. (2005). Klasifikasi Perbangkan dengan Pendekatan CART dan MARS. *Widya Management dan Akutansi* 5 (1), 50-82.
- Otok, B. W. (2010). Pendekatan Multivariate Adaptive Regression Spline (MARS) Pada Pengelompokkan Zona Musim Suatu Wilayah. *Statistika* 10(2), 107-120.
- Pabalik, I., Ihsan, N., & Arsyad, M. (2015). Analisis Fenomena Perubahan Iklim dan Karakteristik Curah Hujan Ekstrim di Kota Makassar. *Jurnal Sains dan Pendidikan Fisika* 11(1), 88-92.
- Paradipta, N. S., Sembiring, P., & Bangun, P. (2013). Analisis Pengaruh Curah Hujan di Kota Medan. *Saintia Matematika* 1(5), 459-468.
- Poerwanto, B., & Budiantara, I. N. (2014). Estimasi Kurva Regresi Semiparametrik Spline untuk Data Longitudinal. *Prosiding Seminar Nasional Matematika, Universitas Udayana*.
- Ramli, U. (2010). *Meteorologi dan Kliatologi*. Makassar: Badan Penerbit UNM.
- Salford System. (t.thn.). MARS (Multivariate Adaptive Regression Splines). <http://www.salford-system.com>.
- Saputro, D. (2011). penduga data tidak lengkap curah hujan di kabupaten indramayu(berdasarkan data tahun 1980-2000). *Sains IPB Press*.
- Sutikno. (2002). *Penggunaan Regresi Spline Adaptive Berganda untuk Peramalan Indeks ENSO dan Hujan Bulanan*. BOGOR: Tesis jurusan Statistika, Institut ertanian Bogor.
- Tiro, M. A. (2008). *Dasar-Dasar Statistika*. Makassar: Andira Publisier.
- Tiro, M. A. (2011). *Analisis Regresi dengan Data Kategori (Edisi Ketiga)*. Makassar: Andhira Publisher.
- Triatmojo, B. (1998). Studi Keseimbangan Air di Pulau Jawa. *Media Teknik* 20(1), 32-38.
- Wicaksono, W., suparti, & Suparti. (2014). Pemodelan Multivariate Adaptive Regression Spline (MARS) Pada Faktor-Faktor Resiko Kesakitan Diare. *Jurnal Gaussian* 3(2), 253-262.
- Yani, N. M., Srinadi, I. A., & Sumarjaya, I. (2017). Aplikasi Model Regresi Semiparametrik Studi kasus DBD di Rumah Sakit Puri Raharja. *E-Journal Matematika* 6(1), 65-73.