

# Pemodelan Faktor-Faktor yang Mempengaruhi Penyakit Tuberkulosis Menggunakan Metode *Multivariate Adaptive Regression Spline* (MARS) (Studi kasus Provinsi Jawa Barat, Jawa Timur dan Jawa Tengah tahun 2019)

Yunda Damayanti\*, Siti Sunendiari

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Islam Bandung, Indonesia.

\*damayantiyunda@gmail.com, diarisunen22@gmail.com

**Abstract.** Multivariate Adaptive Regression Spline (MARS) is a nonparametric regression approach that results from a spline with Recursive Partitioning Regression (RPR) which was first popularized by Friedman (1991). MARS is able to solve two main problems in statistics, namely prediction when the response variable is continuous and classification when the response variable is categorical. The MARS model is able to overcome the problem of high-dimensional data, namely data that has a number of predictor variables  $3 \leq p \leq 20$  and large sample data and produces accurate predictions of response variables. The MARS model is based on the Generalized Cross Validation (GCV) minimum value. Response variables used in this study using continuous data, namely the number of tuberculosis cases and the predictor variables used are the percentage of the poor (X1), population density (X2), the number of HIV sufferers (X3), the percentage of households using clean water (X4), the percentage of food processing places (TPM) that meet health requirements (X5), percentage of families with healthy basic sanitation facilities (X6) and percentage of public places that meet health requirements (X7). In this study, the best MARS model was obtained with a combination of 28, MI = 2 and MO = 2 with the minimum Generalized Cross Validation (GCV). There are 6 predictor variables that have a significant effect on the number of tuberculosis based on the level of importance of these variables.

**Keywords:** Nonparametric Regression, Multivariate Adaptive Regression Spline, Tuberculosis.

**Abstrak.** *Multivariate Adaptive Regression Spline* (MARS) merupakan pendekatan regresi nonparametrik yang dihasilkan dari *spline* dengan *Recursive Partitioning Regression* (RPR) yang pertama kali di populerkan oleh Friedman (1991). MARS mampu menyelesaikan dua permasalahan utama dalam statistika, yaitu prediksi pada saat variabel responnya kontinu dan klasifikasi pada saat variabel responnya kategorik. Model MARS mampu mengatasi permasalahan data berdimensi tinggi, yaitu data yang memiliki jumlah variabel prediktor sebesar  $3 \leq p \leq 20$  dan data sampel berukuran besar serta menghasilkan prediksi variabel respon yang akurat. Model MARS didasarkan pada nilai minimum *Generalized Cross Validation* (GCV). Variabel respon yang digunakan dalam penelitian ini menggunakan data kontinu yaitu jumlah kasus tuberkulosis dan variabel prediktor yang digunakan yaitu persentase penduduk miskin (X1), tingkat kepadatan penduduk (X2), jumlah penderita HIV (X3), persentase rumah tangga menggunakan air bersih (X4), persentase tempat pengolahan makanan (TPM) yang memenuhi syarat kesehatan (X5), persentase keluarga dengan kepemilikan sarana sanitasi dasar sehat (X6) dan persentase tempat umum yang memenuhi syarat kesehatan (X7). Pada penelitian ini diperoleh model MARS terbaik dengan kombinasi BF = 28,

MI = 2 dan MO = 2 dengan kriteria nilai *Generalized Cross Validation* (GCV) paling minimum sebesar 1696251. Terdapat 6 variabel prediktor yang berpengaruh signifikan terhadap jumlah penderita penyakit tuberkulosis.

**Kata Kunci:** Regresi Nonparametrik, *Multivariate Adaptive Regression Spline*, Tuberkulosis.

## 1. Pendahuluan

Tuberkulosis (TBC) merupakan penyakit menular yang disebabkan oleh kuman *Mycobacterium Tuberculosis* yang masuk ke dalam tubuh melalui pernafasan. Penyakit ini menyebar melalui droplet orang yang terinfeksi basil tuberkulosis yang dapat menyerang berbagai organ, terutama paru-paru (WHO, 2020). Jumlah kasus TBC di Indonesia pada tahun 2019 ditemukan sebanyak 543.287 kasus, menurun bila dibandingkan dengan semua kasus TBC pada tahun 2018 sebanyak 566.623 kasus. Jumlah kasus TBC tertinggi pada tahun 2019 di laporkan dari provinsi dengan jumlah penduduk yang besar yaitu Jawa Barat, Jawa Timur dan Jawa Tengah. Kasus TBC di tiga provinsi tersebut sebesar 45% hampir mencapai setengah dari jumlah seluruh kasus TBC di Indonesia (Kemenkes RI, 2019). Penyakit TBC dapat ditemukan di daerah pemukiman padat penduduk dengan sanitasi yang kurang bagus. Sanitasi lingkungan rumah sangat mempengaruhi keadaan bakteri *Mycobacterium Tuberculosis*. Selain faktor sanitasi lingkungan, adapun faktor ekonomi dan juga pendidikan. banyak faktor yang mempengaruhi angka kejadian penyakit TBC, sehingga perlu dilakukan analisis faktor-faktor tersebut.

Salah satu metode untuk menganalisis faktor-faktor yang mempengaruhi penyakit TBC adalah metode *Multivariate Adaptive Regression Spline* (MARS). MARS merupakan pendekatan regresi nonparametrik yang dihasilkan dari *spline* dengan *Recursive Partitioning Regression* (RPR) yang pertama kali di populerkan oleh Friedman (1991). Model MARS mampu mengatasi permasalahan data berdimensi tinggi, yaitu data yang memiliki jumlah variabel prediktor sebesar  $3 \leq n \leq 20$  dan data sampel berukuran besar serta menghasilkan prediksi variabel respon yang akurat. Model MARS didasarkan pada nilai minimum *Generalized Cross Validation* (GCV).

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut “Bagaimana pemodelan faktor-faktor yang mempengaruhi jumlah kasus penyakit TBC di provinsi Jawa Barat, Jawa Timur dan Jawa Tengah tahun 2019 menggunakan metode *Multivariate Adaptive Regression Spline* (MARS)? dan “Faktor-faktor mana saja yang berpengaruh terhadap jumlah kasus penyakit TBC di provinsi Jawa Barat, Jawa Timur dan Jawa Tengah tahun 2019?”. Selanjutnya, tujuan dalam penelitian ini diuraikan dalam pokok-pokok sbb.

1. Untuk mengetahui model pada faktor-faktor yang mempengaruhi jumlah kasus TBC di provinsi Jawa Barat, Jawa Timur dan Jawa Tengah tahun 2019 menggunakan metode *Multivariate Adaptive Regression Spline* (MARS).
2. Mengetahui faktor-faktor yang berpengaruh pada jumlah kasus TBC di provinsi Jawa Barat, Jawa Timur dan Jawa Tengah tahun 2019.

## 2. Metodologi

### Regresi Nonparametrik

Regresi nonparametrik merupakan salah satu pendekatan yang digunakan untuk mengetahui pola hubungan antara variabel prediktor dan respon yang tidak diketahui kurva regresinya atau tidak terdapat informasi masa lalu yang lengkap tentang bentuk pola data (Eubank, 1998). Kurva regresi diasumsikan menggunakan pemulusan (*smoothing*) untuk memperoleh hasil fungsi ruang tertentu. Beberapa model regresi nonparametrik yang banyak digunakan diantaranya: Spline, MARS, Kernel, Deret Fourier, Deret Orthogonal, Neural Network (NN), Polinomial Lokal dan yang lainnya.

### Regresi Spline

Dikutip dalam Rachmasari&Karyana (2020) regresi spline merupakan suatu regresi polynomial dimana terbagi menjadi beberapa segmen yang dibatasi titik-titik knot tertentu dan kontinu sehingga bersifat fleksibel dibandingkan polynomial biasa. Kurva regresi spline mengatasi permasalahan pola pada data yang menunjukkan naik atau turun yang tajam dengan bantuan titik-titik knot, dimana banyaknya titik knot dan lokasi titik knot sangat mempengaruhi bentuk kurva regresi spline.

### Multivariate Adaptive Regression Spline (MARS)

*Multivariate Adaptive Regression Splines* (MARS) merupakan salah satu metode regresi nonparametrik yang diperkenalkan oleh Friedman tahun 1991. Model MARS berguna sebagai metode baru yang mengakomodasi pembangunan model-model-model prediksi yang lebih akurat untuk variabel-variabel kontinu maupun kategorik biner. Model MARS mampu mengatasi permasalahan data yang berdimensi tinggi yaitu yang memiliki variabel predictor sebesar  $3 \leq n \leq 20$  dan jumlah sampel yang diharuskan untuk pendekatan MARS adalah  $50 \leq n \leq 1000$ . MARS merupakan pengembangan dari pendekatan *Recursive Partitioning Regression* (RPR) dan *Generalized Cross Validation* (CGV). Metode MARS juga mengatasi kelemahan *Recursive Partitioning Regression* (RPR) yaitu menghasilkan model yang tidak kontinu pada titik *knot* dan dapat mengidentifikasi adanya fungsi linier aditif. Model MARS dapat di definisikan melalui fungsi basis dalam bentuk persamaan:

$$\hat{f}(x) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [S_{km} \cdot (x_{v(k,m)} - t_{km})]_+ \quad \dots (2.1)$$

dengan fungsi basis MARS

$$B_m(x) = \prod_{k=1}^{K_m} H[S_{km} \cdot (x_{v(k,m)} - t_{km})]_+ \quad \dots (2.2)$$

dimana:

- $a_0$  : konstanta fungsi basis
- $a_m$  : koefisien fungsi basis ke-m
- $M$  : maksimum fungsi basis (nonconstant fungsi basis)
- $K_m$  : derajat interaksi
- $S_{km}$  : tanda pada titik knotnya (bernilai  $\pm 1$ )
- $x_{v(k,m)}$  : variabel prediktor
- $t_{km}$  : nilai knot dari variabel prediktor

### Pemilihan Model MARS Terbaik

Model terbaik pada MARS ditentukan berdasarkan nilai *Generalized Cross Validation* (GCV), dari model tersebut mempunyai nilai yang paling minimum. Pemilihan model MARS terbaik menggunakan algoritma *stepwise (forward dan backward)*. Algoritma *stepwise* bertujuan membangkitkan model dan kemudian memangkas model kembali ke model yang lebih efektif (Knafl & Ding, 2016). Adapun rumus GCV sebagai berikut:

$$G(M) = \frac{ASR}{\left[1 - \frac{\tilde{C}(M)}{N}\right]^2} = \frac{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{f}_M(x_i))^2}{\left(1 - \frac{\tilde{C}(M)}{N}\right)^2} \\ = \frac{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{f}_M(x_i))^2}{\left(1 - \frac{M + d_x(M-1)/2}{N}\right)^2} \quad \dots (2.2)$$

dimana:

- $x_i$  : Variabel prediktor
- $y_i$  : Variabel respon
- $\hat{f}_M(x_i)$  : Taksiran/prediksi  $y_i$
- $M$  : Banyaknya basis fungsi
- $C(M)$  :  $C(M) + d \cdot M$
- $C(M)$  :  $trace \left( \mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \right) + 1$  adalah banyaknya parameter yang diestimasi
- $d$  : Nilai Ketika basis fungsi mencapai kondisi optimal (antara  $2 \leq d \leq 4$ )

Menurut Friedman (1991) ada beberapa istilah yang perlu diperhatikan dalam metode dan pemodelan MARS adalah sebagai berikut:

1. Knot  
*Knot* adalah nilai variabel prediktor ketika *slope* (kemiringan) suatu garis regresi mengalami peubah yang dapat di definisikan sebagai akhir dari sebuah garis regresi dan awal dari garis regresi lainnya. Kontinuitas dari fungsi basis antar tiap region diharapkan pada setiap titik *knot* yang disarankan 0,1,2 dan 3. Berlebihannya MO akan semakin mengurangi fleksibilitas model dan akurasinya.
3. Basis Function (BF)  
 Fungsi basis merupakan fungsi parametrik yang menjelaskan hubungan antara variabel respon dan variabel prediktor yang dapat didefinisikan pada tiap region. Fungsi basis terdiri dari satu atau lebih variabel prediktor. Fungsi basis ini berupa selang antara *knot* yang berurutan. Pada umumnya fungsi basis yang dipilih berbentuk polynomial dengan turunan yang kontinu pada setiap titik *knot*. Maksimum fungsi basis yang disarankan adalah 2 hingga 4 kali jumlah variabel prediktor (Friedman, 1991).
4. Interkasi  
 Interaksi adalah perkalian silang antar variabel yang saling berkorelasi. Jumlah maksimum interaksi (MI) yang diperoleh yaitu 1,2,3. Ketika jumlah MI melebihi 3, maka akan menyebabkan model semakin kompleks dan sulit untuk di interpretasikan serta nilai CGV akan meningkat.

**Data**

Data yang digunakan dalam penelitian ini -yaitu data sekunder yang di peroleh dari Badan Pusat Statistika (BPS) Jawa Barat, Jawa Timur dan Jawa Tengah pada tahun 2019 dan dari Profil Kesehatan Jawa Barat, Jawa Timur dan Jawa Tengah pada tahun 2019. Data yang digunakan mengenai jumlah kasus penyakit tuberkulosis dan faktor-faktor yang mempengaruhi jumlah kasus penyakit tuberkulosis meliputi persentase penduduk miskin ( $X_1$ ), tingkat kepadatan penduduk ( $X_2$ ), jumlah penderita HIV ( $X_3$ ), persentase rumah tangga menggunakan air bersih ( $X_4$ ), persentase tempat pengolah makanan yang memenuhi syarat kesehatan ( $X_5$ ), persentase keluarga dengan kepemilikan sanitasi dasar sehat ( $X_6$ ), persentase tempat umum yang memenuhi syarat kesehatan ( $X_7$ ).

**3. Pembahasan dan Diskusi**

**Pemodelan Multivariate Adaptive Regression Spline (MARS)**

Model MARS terbaik dipilih berdasarkan nilai GCV terkecil yaitu 1696251. Nilai GCV diperoleh dengan kombinasi BF = 28, MI = 2 dan MO = 2. Pemodelan menggunakan metode MARS pada penyakit tuberkulosis yaitu:

$$\hat{f}(x) = 40675.6 - 7,411*BF_1 + 3,196*BF_2 - 4,207*BF_3 + 5,863*BF_4 - 64,333*BF_5 + 4,254*BF_6 - 176,647*BF_7 - 44,336*BF_8 - 41,253*BF_9 + 0,013*BF_{10} + 0,011*BF_{11} - 0,016*BF_{12} - 0,007*BF_{13} + 4,980*BF_{14} + 1,948*BF_{15} + 15,770*BF_{17}$$

dengan fungsi basis sebagai berikut:

$BF_1 = \max (0, X_2 - 2201)$	$BF_9 = \max (0, X_6 - 71,84)$
$BF_2 = \max (0, X_2 - 4186)$	$BF_{10} = \max (0, X_2 - 1644) * BF_5$
$BF_3 = \max (0,9385 - X_2)$	$BF_{11} = \max (0,6787 - X_2) * BF_5$
$BF_4 = \max (0, X_2 - 9835)$	$BF_{12} = \max (0, X_2 - 6787) * BF_5$
$BF_5 = \max (0,443 - X_3)$	$BF_{13} = \max (0, X_4 - 50,67) * BF_3$
$BF_6 = \max (0, X_3 - 443)$	$BF_{14} = \max (0,64.67 - X_4) * BF_8$
$BF_7 = \max (0,54,3 - X_4)$	$BF_{15} = \max (0, X_4 - 64,67) * BF_8$
$BF_8 = \max (0,79,6 - X_5)$	$BF_{17} = \max (0,68,8 - X_7) \max (0, X_5 - 79,6)$

Model MARS terbaik memiliki 16 basis fungsi yang menunjukkan adanya interaksi antar variabel prediktor dan variabel respon. Berikut merupakan interpretasi dari model MARS:

1. Koefisien fungsi basis model MARS dengan satu variabel:
  - a.  $BF_1 = \max (0, X_2 - 2201)$  dengan koefisien - 7,411, artinya kenaikan  $BF_1$  satu

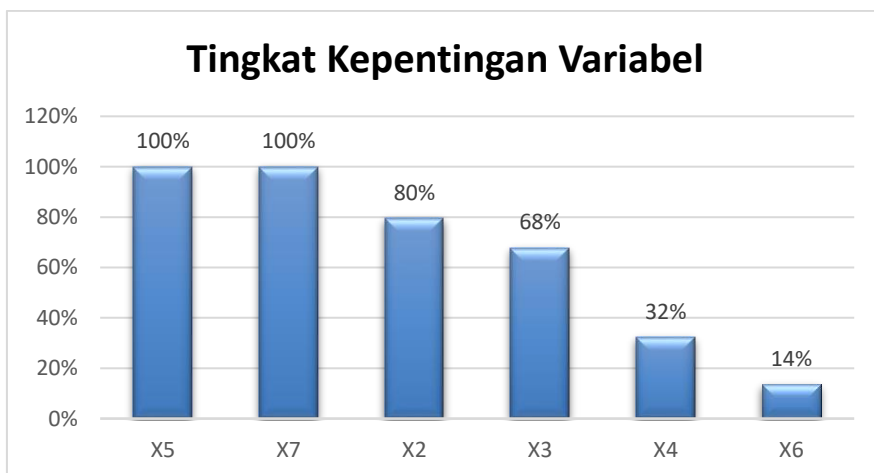
satuan akan menurunkan jumlah penyakit TBC sebesar 7,411 atau jika nilai tingkat kepadatan penduduk ( $X_2$ ) lebih besar dari 2201, maka jumlah penyakit TBC menurun sebesar 7,411.

- b.  $BF_2 = \max(0, X_2 - 4186)$  dengan koefisien 3,196, artinya kenaikan  $BF_2$  satu satuan akan meningkatkan jumlah penyakit TBC sebesar 3,196 atau jika nilai tingkat kepadatan penduduk ( $X_2$ ) lebih besar dari 4186 maka jumlah penyakit TBC meningkat sebesar 3,196.
  - c.  $BF_3 = \max(0, 9385 - X_2)$  dengan koefisien  $-4,207$ , artinya kenaikan  $BF_3$  satu satuan akan menurunkan jumlah pnyakit TBC sebesar 4,207 atau jika nilai tingkat kepadatan penduduk ( $X_2$ ) lebih kecil dari 9385, maka jumlah penyakit TBC menurun sebesar 4,207.
  - d.  $BF_4 = \max(0, X_2 - 9835)$  dengan koefisien 5,863, artinya kenaikan  $BF_4$  satu satuan akan meningkatkan jumlah penyakit TBC sebesar 5,863 atau jika nilai tingkat kepadatan penduduk ( $X_2$ ) lebih besar dari 9835, maka jumlah penyakit TBC meningkat sebesar 5,863.
  - e.  $BF_5 = \max(0, 443 - X_3)$  dengan koefisien  $-64,333$ , artinya kenaikan  $BF_5$  satu satuan akan menurunkan jumlah penyakit TBC sebesar 64,33 atau jika nilai jumlah penderita HIV ( $X_3$ ) lebih kecil dari 443, maka jumlah penyakit TBC menurun sebesar 64,333.
  - f.  $BF_6 = \max(0, X_3 - 443)$  dengan koefisien 4,254, artinya kenaikan  $BF_6$  satu satuan akan meningkatkan jumlah penyakit TBC sebesar 4,254 au jika nilai jumlah penderita HIV ( $X_3$ ) lebih besar dari 443 maka jumlah penyakit TBC meningkat sebesar 4,254.
  - g.  $BF_7 = \max(0, 54,3 - X_4)$  dengan koefisien  $-176,647$ , artinya kenaikan  $BF_7$  satu satuan akan menurunkan jumlah penyakit TBC sebesar 176,547 atau jika nilai persentase rumah tangga menggunakan air bersih ( $X_4$ ) lebih kecil dari 54,3 maka jumlah penyakit TBC menurun sebesar 176,647.
  - h.  $BF_8 = \max(0, 79,6 - X_5)$  dengan koefisien  $-44,336$ , artinya kenaikan  $BF_8$  satu atuan menurunkan jumlah penyakit TBC sebesar 44,36 atau jika nilai persentase tempat pengolahan makanan (TPM) yang memenuhi syarat kesehatan ( $X_5$ ) lebih kecil dari 79,6 maka jumlah penyakit TBC menurun sebesar 44,336.
  - i.  $BF_9 = \max(0, X_6 - 71,84)$  dengan koefisien  $-41,253$ , artinya kenaikan  $BF_9$  satu satuan akan menurunkan jumlah penyakit TBC sebesar 41,253 atau jika nilai persentase keluarga yang memiliki sarana sanitasi dasar sehat ( $X_6$ ) lebih besar dari 71,84, aka jumlah penyakit TBC menurun sebesar 41,253.
2. Koefisien fungsi basis model MARS dengan dua variabel:
    - a.  $BF_{10} = \max(0, X_2 - 1644) * BF_5$   
 $BF_5 = \max(0, 443 - X_3)$   
 dengan koefisien 0,013, artinya kenaikan  $BF_{10}$  satu satuan akan meningkatkan jumlah penyakit TBC sebesar 0,013 atau jika nilai tingkat kepadatan penduduk lebih besar dari 1664 dan nilai jumlah penderita HIV lebih kecil dari 443, maka jumlah penyakit TBC meningkat sebesar 0,013.
    - b.  $BF_{11} = \max(0, 6787 - X_2) * BF_5$   
 $BF_5 = \max(0, 443 - X_3)$   
 dengan koefisien 0,011, artinya kenaikan  $BF_{11}$  satu satuan akan meningkatkan jumlah penyakit TBC sebesar 0,011 atau jika nilai tingkat kepadatan penduduk lebih kecil dari 6787 dan nilai jumlah penderita HIV lebih kecil dari 443 maka jumlah penyakit TBC meningkat sebesar 0,011.
    - c.  $BF_{12} = \max(0, X_2 - 6787) * BF_5$   
 $BF_5 = \max(0, 443 - X_3)$   
 dengan koefisien  $-0,016$ , artinya kenaikan  $BF_{12}$  satu satuan akan menurunkan jumlah penyakit TBC sebesar 0,016 atau jika nilai tingkat kepadatan penduduk lebih besar dari 6787 dan nilai jumlah penderita HIV lebih kecil dari 443 maka jumlah

- penyakit TBC menurun sebesar 0,016.
- d.  $BF_{13} = \max(0, X_4 - 50,67) * BF_3$   
 $BF_3 = \max(0,9385 - X_2)$   
 dengan koefisien -0,007 artinya kenaikan  $BF_{13}$  satu satuan akan menurunkan jumlah penyakit TBC sebesar 0,007 atau jika nilai persentase rumah tangga menggunakan air bersih lebih besar dari 50,67 dan nilai tingkat kepadatan penduduk lebih kecil dari 9385 maka jumlah penyakit TBC menurun sebesar 0,007.
  - e.  $BF_{14} = \max(0,64,67 - X_4) * BF_8$   
 $BF_8 = \max(0,79,6 - X_5)$   
 dengan koefisien 4,980, artinya kenaikan  $BF_{14}$  satu satuan akan meningkatkan jumlah penyakit TBC sebesar 4,980 atau jika nilai persentase rumah tangga menggunakan air bersih lebih kecil dari 64,67 dan nilai persentase tempat pengolahan makanan (TPM) yang memenuhi syarat kesehatan lebih kecil dari 79,6 maka setiap kenaikan satu satuan  $BF_{14}$  dapat meningkatkan jumlah penyakit tuberkulosis sebesar 4,980
  - f.  $BF_{15} = \max(0, X_4 - 64,67) * BF_8$   
 $BF_8 = \max(0,79,6 - X_5)$   
 dengan koefisien 1,948, artinya kenaikan  $BF_{15}$  satu satuan akan meningkatkan jumlah penyakit TBC sebesar 1,948 atau jika nilai persentase rumah tangga menggunakan air bersih lebih besar dari 64,67 dan nilai persentase tempat pengolahan makanan (TPM) yang memenuhi syarat kesehatan lebih kecil dari 79,6 maka jumlah penyakit TBC meningkat sebesar 1,948.
  - g.  $BF_{17} = \max(0,68,8 - X_7) * \max(0, X_5 - 79,6)$   
 dengan koefisien 15,770, artinya kenaikan  $BF_{17}$  satu satuan akan meningkatkan jumlah penyakit TBC sebesar 15,770 atau jika nilai persentase tempat umum yang memenuhi syarat kesehatan lebih kecil dari 68,8 dan nilai persentase tempat pengolahan makanan (TPM) yang memenuhi syarat kesehatan lebih besar dari 79,6 maka jumlah penyakit TBC meningkat sebesar 15,770.

**Tingkat Kepentingan Variabel**

Tingkat kepentingan berada pada rentang 0-100% berdasarkan tiap variabel respon. Model MARS terbaik menghasilkan 6 variabel prediktor yang berpengaruh signifikan terhadap jumlah penyakit tuberkulosis. Tingkat kepentingan masing-masing variabel prediktor dilihat berdasarkan rentangnya dari yang paling tinggi sampai paling rendah pada penyajian Gambar 1.



**Gambar 1.** Tingkat Kepentingan Variabel

Berdasarkan tabel diatas, menunjukkan tingkat kepentingan dari setiap variabel yang digunakan dalam model MARS. Variabel prediktor yang berpengaruh signifikan terhadap jumlah penyakit tuberkulosis diantaranya, persentase tempat pengolahan makanan (TPM) yang memenuhi syarat

kesehatan (X5) dengan tingkat kepentingan 100%, persentase tempat umum yang memenuhi syarat kesehatan (X7) dengan tingkat kepentingan 100%, tingkat kepadatan penduduk (X2) dengan tingkat kepentingan 80%, jumlah penderita HIV (X3) dengan tingkat kepentingan 68%, persentase rumah tangga menggunakan air bersih (X4) dengan tingkat kepentingan 32% dan persentase keluarga dengan kepemilikan sarana sanitasi dasar sehat (X6) dengan tingkat kepentingan sebesar 14%.

#### 4. Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, maka diperoleh kesimpulan penelitian sebagai berikut.

1. Model MARS terbaik untuk memprediksi jumlah penderita penyakit TBC di provinsi Jawa Barat, Jawa Timur dan Jawa Tengah tahun 2019 adalah model MARS dengan kombinasi  $BF = 28$ ,  $MI = 2$  dan  $MO = 2$ . Pemodelan menggunakan metode MARS pada penyakit tuberkulosis yaitu:

$$\begin{aligned} \hat{f}(x) = & 40675.6 - 7,411*BF_1 + 3,196*BF_2 - 4,207*BF_3 + 5,863*BF_4 \\ & - 64,333*BF_5 + 4,254*BF_6 - 176,647*BF_7 - 44,336*BF_8 \\ & - 41,253*BF_9 + 0,013*BF_{10} + 0,011*BF_{11} - 0,016*BF_{12} \\ & - 0,007*BF_{13} + 4,980*BF_{14} + 1,948*BF_{15} + 15,770*BF_{17} \end{aligned}$$

2. Terdapat 6 variabel prediktor yang berpengaruh signifikan terhadap jumlah penderita penyakit TBC di Provinsi Jawa Barat, Jawa Timur dan Jawa Tengah tahun 2019 yaitu persentase tempat pengolahan makanan (TPM) yang memenuhi syarat kesehatan (X5) persentase tempat umum yang memenuhi syarat kesehatan (X7), tingkat kepadatan penduduk (X2), jumlah penderita penyakit HIV (X3), persentase rumah tangga menggunakan air bersih (X4) dan persentase keluarga dengan kepemilikan sarana sanitasi dasar sehat (X6).

#### Acknowledge

Penulis mengucapkan terima kasih kepada Allah SWT, ibu dan kakak-kakak saya yang selalu mendo'akan dan memberi semangat kepada penulis. Kepada ibu Siti Sunendiari, DRA., M.Si yang telah memberikan bimbingan kepada penulis hingga penelitian bisa terselesaikan dengan baik. Dosen dan Staf Prodi Statistika Unisba Universitas Islam Bandung Bandung yang telah banyak memberikan ilmu pengetahuan dan motivasi selama perkuliahan berlangsung. Sahabat dan semua pihak yang terlibat dalam penelitian ini telah membantu hingga penelitian ini terselesaikan.

#### Daftar Pustaka

- [1] [BPS] Badan Pusat Statistika. (2019). Statistik Kesejahteraan Rakyat Provinsi Jawa Barat 2019. Jawa Barat: Badan Pusat Statistika Jawa Barat.
- [2] [BPS] Badan Pusat Statistika. (2019). Statistik Kesejahteraan Rakyat Provinsi Jawa Tengah 2019. Jawa Tengah: Badan Pusat Statistika Jawa Tengah.
- [3] [BPS] Badan Pusat Statistika. (2019). Statistik Kesejahteraan Rakyat Provinsi Jawa Timur 2019. Jawa Timur: Badan Pusat Statistika Jawa Timur.
- [4] Dinas Kesehatan Provinsi Jawa Barat. (2019). *Profil Kesehatan Provinsi Jawa Barat 2019*. Jawa Barat.
- [5] Dinas Kesehatan Provinsi Jawa Timur. (2019). *Profil Kesehatan Provinsi Jawa Timur 2019*. Jawa Timur.
- [6] Dinas Kesehatan Provinsi Jawa Tengah. (2019). *Profil Kesehatan Provinsi Jawa Tengah 2019*. Jawa Tengah.
- [7] Eubank, R. L. (1998). *Nonparametric Regression and Spline Smothing, second edition*. New York: Marcel Dekker.
- [8] Friedman, J. (1991). Multivariate Adaptive Regression Spline. *The Analysis of Statistics*, 19(1), 1-141.
- [9] Kemenkes RI. (2019). *Profil Kesehatan Indonesia 2019*. Jakarta.

- [10] Knafl, G., & Ding, K. (2016). *Adaptive Regression for Modeling Nonlinear Relationships*. Switzerland (CHE): Springer.
- [11] Nash, M. S., & Bradford, D. F. (2001). *Parametric and Non Parametric Logistic Regression for Prediction of Precense Absence of an Amphibian*. Las Vegas: Nevada.
- [12] Rachmasari, R., & Karyana, Y. (2020). Pemodelan Regresi Nonparametrik Spline Truncated Untuk Mengetahui Faktor-Faktor yang Mempengaruhi Penggunaan Konstrasepsi. 128-134.
- [13] WHO. (2020). *Global Tuberculosis Report 2020*. Switzerland.
- [14] Utama Muhammad Bangkit Riksa, Hajarisman Nusar. (2021). *Metode Pemilihan Variabel pada Model Regresi Poisson Menggunakan Metode Nordberg*. Jurnal Riset Statistika, 1(1), 35-42.