

PEMODELAN PENYAKIT INFEKSI SALURAN PERNAFASAN AKUT DI DAERAH SEKITAR SEMBURAN LUMPUR LAPINDO SIDOARJO DENGAN PENDEKATAN MODEL *MULTIVARIATE ADDITIVE REGRESSION SPLINE*

Mahfudhotin¹

¹ Institut Agama Islam Negeri Kediri

e-mail: mahfudhotin@iainkediri.ac.id

Abstrak

Fenomena semburan lumpur panas di Sidoarjo menarik untuk diteliti lebih lanjut. Mengenai penyebabnya, bencana tersebut terjadi karena kesalahan pengeboran sehingga mengakibatkan semburan lumpur Lapindo yang membawa dampak emisi gas menimbulkan masalah kesehatan terutama berhubungan dengan saluran pernapasan, yakni infeksi saluran pernapasan (ISPA). Faktor risiko yang dapat mempengaruhi kejadian ISPA pada umumnya adalah faktor sosio-demografi, biologis, perumahan dan kepadatan serta polusi. Oleh sebab itu, penelitian ini bertujuan untuk mendapatkan model pengklasifikasikan data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo dengan variabel-variabel yang berkontribusi dalam pengklasifikasian tersebut. Pendekatan nonparametrik *Multivariate Adaptive Regression Spline* (MARS) dipilih karena beberapa penelitian sebelumnya menyatakan bahwa metode ini menghasilkan ketepatan klasifikasi lebih tinggi dibandingkan dengan metode klasifikasi lainnya. Selain itu, MARS merupakan suatu metode klasifikasi yang mampu membentuk model dengan interaksi sebab untuk menghasilkan model MARS terbaik diperoleh dari kombinasi nilai *Maximum Interaction* (MI), *Basis Function* (BF), dan *Minimum Observation* (MO). Hasil pemodelan dengan MARS terdapat tiga variabel yang berkontribusi dalam pengelompokan, yakni persentase jarak antara rumah dan sumber luapan lumpur lapindo, banyaknya aktifitas di luar rumah, dan banyaknya ventilasi rumah. Ketepatan klasifikasi model secara keseluruhan adalah 97,4 persen dengan nilai GCV yaitu 0,096 dan R^2 sebesar 82,9 persen.

Kata Kunci: ISPA, GCV, Lumpur Lapindo, MARS

Abstract

*The phenomenon of hot mudflow in Sidoarjo is interesting to be investigated further. Regarding the cause, the disaster occurred due to drilling errors resulting in the Lapindo mudflow which resulted in gas emissions causing health problems, especially those related to the respiratory tract, namely respiratory tract infections (ARI). Risk factors that can affect the incidence of ARI in general are socio-demographic, biological, housing and density factors and pollution. Therefore, this study aims to obtain a model for classifying ARI patient data in the Jabon, Tanggulangin, and Porong sub-districts, Sidoarjo district with the variables that contribute to the classification. The nonparametric approach *Multivariate Adaptive Regression Spline* (MARS) was chosen because several previous studies stated that this method resulted in a higher classification accuracy than other classification methods. In addition, MARS is a classification method that is able to form a model with causal interactions to produce the best MARS model obtained from a combination of *Maximum Interaction* (MI), *Basis Function* (BF), and *Minimum Observation* (MO) values. The results of modeling with MARS there are three variables that contribute to the grouping, namely the percentage of the distance between the house and the source of the Lapindo mudflow, the number of activities outside the house, and the number of house ventilation. The overall model classification accuracy is 97,4 percent with a GCV value of 0,096 and an R^2 of 82,9 percent.*

Keywords: ISPA, GCV, Lapindo Mud, MARS

1. PENDAHULUAN

Kondisi bumi di Jawa Timur mengalami gejolak sejak adanya semburan lumpur lapindo yang menyengsarakan warga Porong pada 29 Mei 2006. Hal ini mengakibatkan mobilisasi pemerintah untuk menanggung risiko bencana ini atas kesalahan teknis PT Lapindo Brantas saat pengeboran sumur Banjar Panji. Penerbitan Peraturan Presiden (Perpres) tentang Badan Pengelola Lumpur Lapindo menandai awal dari kejadian ini. Semburan lumpur yang pada awal-awal kemunculannya memiliki volume mencapai kisaran antara 100.000 – 120.000 m³ per hari dengan kandungan padatan 35% bersifat kental dan kandungan kimia seperti semen yang sifat semburannya fluktuatif (Rovicky, 2006). Semburan lumpur Lapindo telah menyebabkan tergenangnya pemukiman warga, yang mengandung berbagai zat yang dapat mencemari udara sehingga desa-desa di sekitar area lumpur lapindo memiliki kualitas udara yang di bawah standar. Catatan Wahana Lingkungan Hidup (Walhi) Jawa Timur menunjukkan semburan dan gelembung di Siring Barat, Jatirejo Barat, dan Mindi menghasilkan 441.200 ppm gas hidrokarbon, 80 kali lebih tinggi dari batas biasanya 5.000 ppm. Selain itu, temuan di lapangan melebihi ambang batas udara ambien 0,24 ppm sampai dengan 55.000 ppm atau 229.100 kali ambang batas standar (Agresti, 2011).

Kondisi udara seperti yang terjadi akibat semburan lumpur Lapindo berhubungan erat dengan masalah kesehatan, terutama yang berkaitan dengan sistem pernapasan. Salah satu hal yang dapat menyebabkan penyakit infeksi saluran pernapasan (ISPA) adalah udara yang tercemar dan perubahan cuaca, menurut data Riset Kesehatan Dasar (Riskesdas) Kementerian Kesehatan tahun 2007 (Mukono, 2008). Penyakit saluran pernapasan bahkan menempati urutan teratas dari 10 daftar penyakit menular yang paling umum di masyarakat. Penyakit infeksi akut yang menyerang organ saluran pernapasan—hidung, sinus, laring, dan alveolus—dikenal dengan ISPA atau Infeksi Saluran Pernafasan (ISK). Sistem pernapasan atas dan bawah termasuk dalam ISPA. Penyakit infeksi akut yang menyerang satu atau lebih bagian saluran pernapasan, mulai dari hidung (saluran atas) hingga paru-paru (saluran bawah) dan jaringan di dalamnya. ISPA juga termasuk infeksi saluran pernapasan yang berlangsung selama 14 hari yang seringkali dijumpai pada anak-anak hingga dewasa mulai dari ISPA ringan sampai berat (Jalil, 2018).

Faktor sosial demografi, biologis, perumahan dan kepadatan, serta polusi merupakan faktor risiko yang dapat mempengaruhi kejadian ISPA. Sedangkan, usia, jenis kelamin, pendidikan orang tua, dan pendapatan keluarga merupakan faktor sosio-demografis serta status gizi dan pemberian ASI eksklusif merupakan contoh faktor biologis. Tidak adanya cerobong asap, kebiasaan merokok ayah, dan kehadiran non-perokok merupakan faktor-faktor polusi dalam ruangan. Kondisi lantai dan dinding, jumlah orang yang tinggal di ruangan yang lebih besar dari dua, dan ventilasi rumah merupakan faktor-faktor dalam perumahan dan kepadatan (Mukono, 2008). Lebih lanjut, hasil penelitian (Hayati, 2014) meneliti faktor-faktor yang mempengaruhi kejadian ISPA di Puskesmas Pasirkaliki Kota Bandung. Faktor yang mempengaruhi kejadian ISPA antara lain pemberian ASI eksklusif, kepadatan hunian ruang tidur, ventilasi ruang tidur, keberadaan anggota keluarga yang merokok, keberadaan anggota keluarga yang menderita ISPA. Menurut Prof. Dr Mukono ahli kesehatan masyarakat Universitas Airlangga, untuk mengetahui relasi antara meningkatnya pengidap ISPA di Puskesmas Porong dan Lumpur Lapindo, harus dikaji dari 3 faktor, yakni lingkungan, manusia, dan faktor yang berada di sekitarnya. Sehingga dalam penelitian ini akan digunakan variabel respon penderita penyakit ISPA dan variabel prediktor jarak rumah dari pusat semburan lumpur, usia, jenis kelamin, dan perokok (Mukono, 2008)

Disamping itu, semenjak 29 Mei 2007 warga porong harus menghirup gas atau asap dari semburan lumpur lapindo. Dengan demikian secara umum ini menimbulkan pergerakan silia dalam hidung dapat menjadi lambat dan kaku akibat dari efek polusi udara pada saluran pernapasan hingga tidak dapat membersihkan saluran pernapasan karena telah menghirup bahan pencemar. Produksi lendir akan meningkat dan mengakibatkan penyempitan saluran pernapasan dan kerusakan sel-sel di saluran pernapasan yang membunuh bakteri. Karena itu, akan sulit untuk bernafas, yang akan menarik benda asing dan tidak memungkinkan untuk mengeluarkan bakteri lain dari saluran pernapasan. Hal ini akan memudahkan terjadinya infeksi saluran pernafasan. Hal ini berarti informasi yang dihimpun Wahana Lingkungan Hidup (Walhi) Puskesmas Porong, ISPA secara konsisten menempati urutan teratas daftar penyakit yang dikeluhkan masyarakat. Namun, sejak 2007 (setahun setelah Lapindo semburan lumpur meletus) hingga 2009, ketika mencapai jumlah tertinggi lebih dari 52 ribu, ada peningkatan tajam dalam jumlah individu yang terkena dampak. Jumlah penderita ISPA berturut-turut: 24,719 (2005); 28,640 (2006); 46,650 (2007); 42,626 (2008); 52,543 (2009) (Prasetyo, 2009). Berdasarkan data tersebut, perlu adanya sebuah penelitian terkait dengan pengaruh asap semburan lumpur lapindo terhadap resiko orang terkena penyakit ISPA di sekitar semburan lumpur lapindo dengan pendekatan model *Multivariate Additive Regression Spline* (Raditya, 2018).

Multivariate Adaptive Regression Spline (MARS) diperkenalkan oleh (Friedman, 1991) sebagai metode baru yang lebih fleksibel yakni dengan menggunakan pendekatan MARS dapat memiliki keunggulan menggabungkan metode klasifikasi dengan pendekatan parametrik, yang dilakukan dalam dua tahap: 1) fungsi basis spline (pengaruh utama, knot, dan interaksi) ditambahkan ke model. Model akan menjadi lebih fleksibel setelah fungsi basis ditambahkan. (Rahmadhani, 2019) Prosedur ini berlanjut hingga nilai maksimum fungsi basis maksimum; (2) fungsi basis dengan kontribusi terkecil akan dikeluarkan dari model berdasarkan model yang diperoleh pada tahap pertama, dan seterusnya sampai ditemukan model terbaik (Asriani, 2016).

Kajian secara teori statistik terkait tentang penentuan variabel-variabel yang digunakan untuk pengelompokan berdasarkan tingkat pengangguran terbuka perlu dilakukan. Hal ini dimaksudkan agar pemerintah daerah mendapatkan *early warning* terkait ISPA. Pemodelan data parametrik bekerja paling baik dengan data yang bentuk model dasarnya diketahui. Asumsi struktur data juga harus dipenuhi, tetapi asumsi tersebut seringkali dilanggar. Dengan demikian, para peneliti kerap menggunakan pendekatan nonparametrik sebagai alternatif. Menurut (Otok, 2008) MARS adalah cara berbeda untuk memodelkan data berdimensi tinggi menggunakan regresi nonparametrik. Dibutuhkan perhitungan yang rumit karena memiliki banyak variabel dan ukuran sampel yang besar. Dalam penelitian ini penulis menggunakan metode MARS untuk memodelkan penyakit ISPA di daerah sekitar semburan lumpur Lapindo sehingga dapat diketahui pengaruh asap lumpur lapindo terhadap risiko orang terkena penyakit ISPA di daerah sekitar semburan.

2. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder tentang penderita ISPA yang diperoleh dari Puskesmas di Kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo. Selain itu, data diperoleh dengan cara menyebarkan kertas kuesioner dan wawancara dengan warga di Kecamatan Porong, Jabon dan Tanggulangin. Variabel respon (Y) adalah data penderita ISPA di Kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo yang terdiri atas dua kategori, yaitu tingkat seorang yang tidak menderita (0) dan tingkat seorang yang menderita (1). Sedangkan Variabel prediktor, yaitu persentase pendidikan orang tua (X_1), Usia (X_2), Jenis kelamin (X_3), persentase banyaknya jumlah yang

merokok atau tidak (X_4), banyaknya aktifitas di luar rumah (X_5), persentase penghasilan keluarga (X_6), banyaknya ventilasi rumah (X_7), dan presentase jarak antara rumah dan sumber luapan lumpur lapindo (X_8).

Penelitian ini dimaksudkan untuk memberikan informasi mengenai penyebaran jumlah penderita ISPA di daerah sekitar semburan lumpur lapindo dengan metode MARS dan untuk dunia akademik penelitian ini diharapkan dapat dijadikan salah satu referensi untuk penelitian selanjutnya yang berkaitan dengan penerapan model MARS. Dengan demikian, diperoleh model penyakit ISPA dengan metode MARS di daerah sekitar semburan lumpur lapindo dan mendapatkan secara statistik pengaruh asap lumpur lapindo terhadap resiko orang terkena penyakit ISPA di daerah sekitar semburan. Lebih lanjut, metode analisis data dilakukan dalam penelitian ini dapat dijelaskan sebagai berikut (Kishartini, 2014):

1. Melakukan studi literatur untuk mengkaji variabel-variabel prediktor yang mempengaruhi jumlah kejadian penderita ISPA, sehingga dapat digunakan untuk mengantisipasi penderita ISPA :
 - a. Jurnal ilmiah, laporan-laporan resmi dari perusahaan / instansi, disertasi / tesis yang sudah ditulis yang berhubungan dengan ISPA yang disebabkan asap lumpur lapindo menggunakan metode MARS.
 - b. Membuat summary terhadap isi sumber-sumber yang berkaitan dengan topik ISPA akibat asap lumpur lapindo menggunakan metode MARS.
 - c. Melakukan wawancara dengan warga di kecamatan Porong, Jabon dan Tanggulangin.
2. Membentuk plot matriks dari variabel respon dan variabel prediktor yang dibangun sebagai deteksi awal terkait dengan pola hubungan antara variabel prediktor dengan variabel respon.
3. Menggabungkan besarnya *Minimum Interaction* (MI), *Basis Function* (BF), dan *Maximum Observation* (MO) yang diperlukan sesuai kebutuhan dengan cara :
 - a. Menetapkan maksimal basis fungsi yang mana fungsi basis dapat dinyatakan dengan :

$$BF_m(x) = \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - u_{km})]_{\pm} \quad (1)$$

$BF_m(x)$ adalah basis fungsi ke- m , s_{km} bernilai 1 atau -1 jika data terletak di sebelah kanan titik knot atau kiri titik knot, $x_{v(k,m)}$ adalah variabel independen/respon dan nilai knots dari variabel independen $x_{v(k,m)}$ adalah u_{km} .

- b. Menetapkan maksimal jumlah interaksi, minimal jumlah pengamatan tersebut berada diantara *knots* atau minimum pengamatan/observasi yang dilakukan dengan *trial and error* sampai diperoleh pemodelan yang optimal dengan nilai GCV yang minimum.

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

Dengan $(1/N) \sum_{i=1}^N [y_i - \hat{f}_M(x_i)]^2$ merupakan *ASR* dan $\left[1 - \frac{C(M)}{N}\right]^2$ menunjukkan suatu banyaknya parameter yang akan ditaksir pada tahap *forward stepwise*.

- c. Menentukan model terbaik yakni dengan nilai GCV minimum.
- d. Menduga koefisien model $(\alpha_1, \alpha_2, \dots, \alpha_k)$ dengan menggunakan OLS (*Ordinary Least Square*) dengan meminimumkan jumlah kuadrat residual. Model yang diperoleh adalah sebagai berikut:

$$\hat{y} = \alpha_0 + \alpha_1 * BF_1 + \alpha_2 * BF_2 + \dots + \alpha_k BF_k \quad (3)$$

y adalah variabel respon, α_0 merupakan konstanta, α_k merupakan koefisien basis fungsi ke- k , dan BF_k adalah basis fungsi ke- k .

- e. Mengkluster fungsi basis tersebut berdasarkan variabel prediktor yang sesuai dengan model.
- f. Menafsirkan nilai GCV minimum pada variabel dan tingkat kontribusi yang memiliki kepentingan dalam klasterisasi variabel respon.
- g. Mengestimasi nilai kesalahan klasifikasi dengan APER dan menentukan kestabilan klasifikasi menggunakan statistik uji *Press's Q*.

3. HASIL DAN PEMBAHASAN

Multivariate Adaptive Regression Spline (MARS)

(Friedman, 1991) mengatakan bahwa MARS merupakan cara untuk memformulasikan suatu regresi multivariat nonparametrik pada model regresi nonlinier yang menggunakan fungsi splines untuk mengestimasi model dan didasarkan pada metode *Recursive Partitioning Regression (RPR)*. Karena ukuran sampel yang besar, banyak variabel, dan dimensi yang tinggi, metode MARS memerlukan perhitungan yang rumit. Selain fakta bahwa MARS digunakan untuk mengurangi dampak *outlier* pada model akhir, model MARS memiliki kelemahan yakni model yang dihasilkan tidak kontinu pada *knot*. Menurut (Raditya A. P., 2018), ketika menggunakan model MARS, ada beberapa hal yang perlu diperhatikan. (1) *Knot*, yang berarti bahwa beberapa garis regresi berganda digunakan untuk menjelaskan semua data dari variabel independen jika satu garis regresi tidak dapat menjelaskannya. Titik sebagai wadah perubahan pola inilah yang disebut dengan *knot*. *Knot* inilah yang menandai persimpangan garis regresi akhir dan garis regresi awal dari garis regresi lainnya. Diharapkan bahwa fungsi basis akan tetap konsisten dari fungsi basis antara satu garis dengan garis lainnya pada setiap titik *knot*. (2) Fungsi Basis, yaitu fungsi yang digunakan untuk menjelaskan hubungan yang ada antara variabel bebas dan variabel terikat. Pada setiap garis didefinisikan oleh fungsi parametrik sebagai fungsi dasar. Fungsi basis tersebut populer dengan nama *truncated power*. Dengan demikian, fungsi basis yang dipilih umumnya berbentuk polinomial dengan turunan kontinu pada setiap titik *knot*.

Pada dasarnya pada metode MARS *knot* ditentukan secara otomatis oleh data dan membentuk model yang kontinu pada suatu *knot*. Penetapan *knot* pada model MARS umumnya digunakan algoritma *backward stepwise* dan *forward stepwise*. Penetapan model menggunakan *forward stepwise* dipenuhi untuk memperoleh jumlah basis fungsi dengan kriteria penetapan basis fungsi yakni dengan meminimumkan *Average Sum of Square residual (ASR)*. (Meliala, 2017) Untuk melengkapi konsep parsimony yakni model sederhana maka dilakukan pemilihan metode *backward stepwise* yakni mengeliminasi basis fungsi yang mempunyai kontribusi kecil terhadap respon dari *forward stepwise* dengan cara meminimumkan nilai *Generalized Cross Validation (GCV)* menurut (Friedman, 1991). Dengan demikian formula pada model MARS:

$$\hat{f}(x) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(k,m)} - u_{km})]_+ \quad (4)$$

Keterangan :

- a_0 : basis fungsi awal
- a_m : koefisien basis fungsi ke- m
- M : *nonconstant basis fungsi*
- K_m : derajat interaksi
- s_{km} : terletak diantara 1 atau -1 jika data di sebelah kiri titik *knot* atau kanan titik *knot*.
- $x_{v(k,m)}$: *independent variabel*
- u_{km} : nilai *knot* dari variabel $x_{v(k,m)}$

Fungsi dari sebuah basis fungsi adalah akhir dari algoritma *forward stepwise* dengan ketetapan basis fungsi yang telah dideterminasi sebelumnya. Menurut Friedman (1991) jumlah maksimum pada basis fungsi adalah 2 sampai 4 kali jumlah variabel prediktor. Lebih lanjut, maksimum interaksi 1,2 atau 3 dengan memperhitungkan apabila lebih dari 3 maka dihasilkan model yang sangat rumit. Menurut (Arifin, 2015) minimum jarak antar *knot* atau minimum observasi antar *knot* yakni bernilai 10, 20, 50 dan 100. Sehingga persamaan (4) dapat diformulasikan dengan (Kasse, 2017):

$$\hat{f}(x) = a_0 + \sum_{m=1}^M a_m [s_{1m} \cdot (x_{v(1,m)} - u_{1m})] + \sum_{m=1}^M a_m [s_{1m} \cdot (x_{v(1,m)} - u_{1m})][s_{2m} \cdot (x_{v(2,m)} - u_{2m})] + \sum_{m=1}^M a_m [s_{1m} \cdot (x_{v(1,m)} - u_{1m})][s_{2m} \cdot (x_{v(2,m)} - u_{2m})][s_{3m} \cdot (x_{v(3,m)} - u_{3m})] + \dots \quad (5)$$

selanjutnya Persamaan (5) dapat didefinisikan sebagai:

$$\hat{f}(x) = a_0 + \sum_{K_m=1} f_i(x_i) + \sum_{K_m=2} f_{ij}(x_i, x_j) + \sum_{K_m=3} f_{ijk}(x_i, x_j, x_k) + \dots \quad (6)$$

Pada model MARS, selain pemilihan *knot* dilakukan secara otomatis dari data dan diperoleh model yang kontinu pada *knot*, penentuan model pada MARS menggunakan metode *stepwise (forward dan backward)*. Lebih lanjut, menurut (Friedman, 1991) *Forward stepwise* ditetapkan karena untuk mendapatkan fungsi dengan jumlah fungsi basis maksimum. Kriteria penetapan fungsi basis pada *forward stepwise* yakni dengan meminimumkan *Mean Squared Error (MSE)*. Untuk menetapkan konsep parsimony yakni model yang sederhana perlu dilakukan *backward stepwise*, yakni menetapkan fungsi basis yang dihasilkan dari *forward stepwise* dengan meminimumkan nilai *Generalized Cross-Validation/GCV*. Dengan demikian pada model MARS, beberapa hal yang perlu diperhatikan dalam pemilihan model yang paling optimal/terbaik yakni jika nilai GCV dari model tersebut mempunyai nilai yang paling minimum diantara model-model yang lainnya. Fungsi GCV minimum dijabarkan (Respita, 2017):

$$GCV(M) = \frac{MSE}{[1 - \frac{C(\hat{M})}{n}]^2} = \frac{\frac{1}{n} \sum_{i=1}^n [y_i - \hat{f}_M(x_i)]^2}{[1 - \frac{C(\hat{M})}{n}]^2} \quad (7)$$

dengan:

$$C(\hat{M}) = C(M) + dM$$

$$C(M) = Trace [\mathbf{B}(\mathbf{B}^T \mathbf{B}) - \mathbf{1B}^T] + 1$$

keterangan:

- x_i : variabel independen/prediktor
- y_i : variabel dependen/respon

- n : banyaknya pengamatan
- d : nilai saat setiap fungsi basis mencapai optimasi yakni terletak pada $2 \leq d \leq 4$

Analisis Deskriptif

Untuk mengetahui karakteristik data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo yang telah dikelompokkan pada tingkat seorang yang tidak menderita (0) dan tingkat seorang yang menderita (1) maka dilakukan analisis deskriptif. Karakteristik yang dimaksud meliputi eksplorasi variabel prediktor pada masing-masing kelompok sehingga diperoleh informasi yang lebih luas.

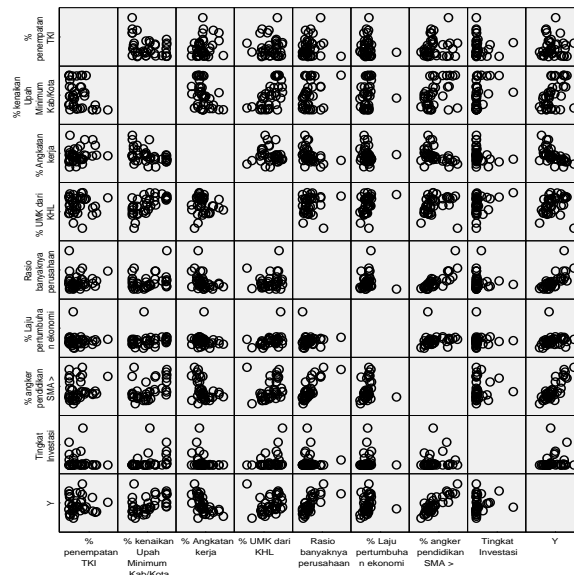
Tabel 1 Persentase Kelompok data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo

| Kelompok | Jumlah | Persentase |
|-----------------|--------|------------|
| Tidak menderita | 16 | 42,11 |
| Menderita | 22 | 57,89 |

Berdasarkan Tabel 1 dapat diketahui bahwa 57,89 persen data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo tergolong dalam wilayah yang memiliki tingkat seorang yang menderita ISPA. Terdapat banyak kemungkinan penyebab kondisi tersebut, diantaranya permasalahan lingkungan, manusia, dan faktor yang berada di sekitarnya sehingga berakibat pada tingginya penderita ISPA di sekitar lumpur Lapindo.

Model MARS

Sebelum membangun model MARS terlebih dahulu dilakukan identifikasi pola bentuk data dengan menggunakan matrik plot. Berdasarkan Gambar 1 terlihat bahwa data tidak memiliki pola yang jelas sehingga digunakan metode MARS sebagai pendekatan nonparametrik.



Gambar 1. Matrik Plot Antar Variabel

Selanjutnya dilakukan kombinasi nilai maksimum BF, MI, dan MO sampai mendapatkan nilai GCV minimum. BF yang dipilih adalah 2 sampai 3 kali variabel prediktor, yaitu 16 karena lebih dari itu (24 dan 32) nilai GCV sama dengan jika nilai BF = 16.

Kemudian nilai MI yang digunakan adalah 1, 2, dan 3 sedangkan MO yaitu 0, 1, dan 2. Hasil kombinasi disajikan pada Tabel 2 berikut.

Tabel 2 Hasil Kombinasi nilai BF, MI, dan MO

| BF | MI | MO | GCV | R ² | Jumlah Variabel |
|-------|----|----|-------|----------------|-----------------|
| 16 | 1 | 0 | 0,146 | 63,9 % | 2 |
| 16 | 2 | 0 | 0,132 | 71,7 % | 3 |
| 16 | 3 | 0 | 0,111 | 71,3 % | 3 |
| 16 | 1 | 1 | 0,118 | 82,8% | 4 |
| 16 | 2 | 1 | 0,110 | 68,5 % | 3 |
| 16 | 3 | 1 | 0,100 | 81,8 % | 4 |
| 16 *) | 1 | 2 | 0,096 | 82,9 % | 3 |
| 16 | 2 | 2 | 0,118 | 68,5 % | 3 |
| 16 | 3 | 2 | 0,110 | 71,4 % | 3 |

Keterangan: *) Model terbaik

Fungsi pengelompokan untuk membedakan penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo dengan tingkat pada tingkat seorang yang tidak menderita dan tingkat seorang yang menderita adalah sebagai berikut.

$$\hat{f}(x) = 0,649 - 0,073BF_7 - 6,439BF_8 + 4,819BF_{10} - 0,086BF_{12} + 0,063BF_{14} \quad (8)$$

dengan

$$\begin{aligned} BF_7 &= (10 - X_8)_+ \\ BF_8 &= (X_5 - 0,330)_+ \\ BF_{10} &= (X_5 - 0,120)_+ \\ BF_{12} &= (X_7 - 27,230)_+ \\ BF_{14} &= (X_7 - 18,710)_+ \end{aligned}$$

Pada persamaan (8) terdapat lima fungsi basis tanpa interaksi yang memenuhi model. Lebih lanjut, interpretasi masing-masing fungsi basis dapat dituliskan sebagai berikut.

$$BF_7 = (10 - X_8)_+ = \begin{cases} (10 - X_8) & , \text{jika } X_8 < 10 \\ 0 & , \text{lainnya} \end{cases}$$

Dengan demikian dapat diartikan koefisien BF_7 akan bermakna jika nilai X_8 lebih kecil dari 10, tetapi jika nilai X_8 lebih besar dari 10 maka koefisien BF_7 tidak bermakna sehingga nilainya adalah 0.

$$BF_8 = (X_5 - 0,330)_+ = \begin{cases} (X_5 - 0,330) & , \text{jika } X_5 > 0,330 \\ 0 & , \text{lainnya} \end{cases}$$

Dengan demikian dapat diartikan koefisien BF_8 akan bermakna jika nilai X_5 lebih besar dari 0,330, tetapi jika nilai X_5 lebih kecil dari 0,330 maka koefisien BF_8 tidak bermakna sehingga nilainya adalah 0.

Interpretasi Model MARS

Penafsiran model MARS pada Tabel 3 memperlihatkan bahwa tingkat kepentingan variabel prediktor pada fungsi pengelompokan, yang diestimasi oleh kenaikan nilai GCV dikarenakan beralihnya variabel-variabel yang dipertimbangkan dari suatu model. Dapat dilihat bahwa persentase jarak antara rumah dan sumber luapan lumpur lapindo (X_8) adalah variabel terpenting pada model MARS dengan tingkat kepentingannya 100 persen, kemudian diikuti oleh banyaknya aktifitas di luar rumah (X_5) dengan tingkat kepentingan 36,709 persen. Pada urutan ketiga yaitu variabel banyaknya ventilasi rumah (X_7) dengan tingkat kepentingan 16,728 persen. Sedangkan lima variabel terakhir tidak memiliki tingkat kepentingan (0 persen) karena sudah terwakili oleh tiga variabel sebelumnya.

Tabel 3 Kepentingan Variabel Prediktor

| Variabel | Tingkat Kepentingan | Nilai GCV Minimum |
|----------|---------------------|-------------------|
| X_8 | 100,00 % | 0,263 |
| X_5 | 36,709 % | 0,118 |
| X_7 | 16,728 % | 0,101 |
| X_1 | 0,00 | 0,096 |
| X_2 | 0,00 | 0,096 |
| X_3 | 0,00 | 0,096 |
| X_4 | 0,00 | 0,096 |
| X_6 | 0,00 | 0,096 |

Ketepatan Klasifikasi Model MARS

Untuk mengetahui besarnya peluang kesalahan di dalam pengklasifikasian data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo dengan seorang yang menderita ISPA dihitung dengan menggunakan APER (*Apparent Error Rate*). Nilai APER ini menyatakan nilai proporsi sampel yang salah diklasifikasikan oleh fungsi klasifikasi. Berdasarkan Tabel 4, terlihat bahwa besarnya kesalahan klasifikasi dalam pengklasifikasian data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo adalah 2,6 persen. Karena terdapat kesalahan dalam pengklasifikasian data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo yang tidak menderita (0), yaitu dari 15 data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo yang tidak menderita, diprediksi satu data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo masuk ke dalam kelompok menderita (1).

Tabel 4 Kesalahan Klasifikasi data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo

| Hasil Observasi (<i>Actual Class</i>) | Taksiran (<i>Predicted Class</i>) | | A P E R (%) |
|--|--|---------------|----------------|
| | Tidak Menderita (0) | Menderita (1) | |
| Tidak Menderita (0) | 15 | 1 | 2,6 |
| Menderita (1) | 0 | 22 | |

Untuk membuktikan kestabilan klasifikasi, maka digunakan suatu uji statistik yang dinamakan dengan *Press's Q*. Tabel 5 memperlihatkan besarnya kestabilan klasifikasi data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo berdasarkan nilai dari *Press's Q* adalah 34,105. Jika dibandingkan dengan $\chi^2_{(1;0,05)} = 3,841$ maka nilai dari *Press's Q* untuk kedua model jauh lebih besar daripada nilai $\chi^2_{(1;0,05)}$ sehingga keakuratan pengklasifikasian data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo dengan pendekatan MARS telah konsisten secara statistik.

Tabel 5 Ketepatan dan Kestabilan

| Ketepatan Klasifikasi (%) | Ketepatan Klasifikasi | | <i>Press's Q</i> |
|---------------------------|-----------------------|---------------|------------------|
| | Tidak Menderita (0) | Menderita (1) | |
| 97,4 | 15 (93,8 %) | 22 (100 %) | 34,105 |

Evaluasi Pengklasifikasian

Untuk memastikan bahwa model MARS yang terbentuk telah sesuai untuk pengelompokan, maka dilakukan evaluasi pengklasifikasian dengan menggunakan data observasi. Kemudian untuk mengevaluasi pengklasifikasian digunakan persamaan (9) berikut. Pengambilan keputusan dilakukan dengan menentukan nilai *cut point*, yaitu 0,5 karena diasumsikan probabilitas pengamatan masuk ke masing-masing kelompok adalah 0,5. Apabila nilai $P(Y = 1|X = x)$ lebih besar dari nilai *cut point* (0,5) maka, pengamatan tersebut masuk ke dalam kelompok data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo dengan tingkat seorang yang menderita dan Apabila nilai $P(Y = 0|X = x)$ lebih besar dari nilai *cut point* (0,5) maka, pengamatan tersebut masuk ke dalam kelompok data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo dengan tingkat seorang yang tidak menderita. Sehingga diperoleh hasil seperti pada Tabel 6.

$$P(Y = 1|X = x) = \frac{e^{\hat{f}(x)}}{1+e^{\hat{f}(x)}} \text{ dan } P(Y = 0|X = x) = \frac{1}{1+e^{\hat{f}(x)}} \quad (9)$$

Tabel 6 Evaluasi Pengklasifikasian

| Hasil Observasi (<i>Actual Class</i>) | Taksiran (<i>Predicted Class</i>) | | Jumlah |
|--|--|---------------|--------|
| | Tidak Menderita (0) | Menderita (1) | |
| Tidak Menderita (0) | 9 | 7 | 16 |
| Menderita (1) | 0 | 22 | 22 |
| Jumlah | 9 | 29 | 38 |

Evaluasi pengklasifikasian dilakukan pada semua pengamatan, yaitu 38 data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo. Hasil evaluasi menunjukkan bahwa terdapat kesalahan klasifikasi, yaitu pada kelompok data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo dengan tingkat seorang tidak menderita. Ketepatan klasifikasi untuk evaluasi adalah 81,57 persen. Sebanyak tujuh dari 16 data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo yang ditaksir berubah kelompok klasifikasi menjadi data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo dengan tingkat seorang yang menderita.

4. KESIMPULAN

Penelitian ini bertujuan untuk mendapatkan model pengklasifikasikan data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo akibat semburan lumpur Lapindo yang menimbulkan masalah kesehatan, terutama berhubungan dengan saluran pernapasan, yakni infeksi saluran pernapasan (ISPA) terhadap variabel-variabel yang berkontribusi dalam pengklasifikasiannya. Analisis statistik dengan pendekatan nonparametrik MARS menyatakan sebanyak tiga variabel prediktor yang memberikan kontribusi dalam pengelompokan objek. Ketiga variabel tersebut diantaranya persentase jarak antara rumah dan sumber luapan lumpur lapindo (X_8), banyaknya aktifitas di luar rumah (X_5), dan banyaknya ventilasi rumah (X_7). Lebih lanjut, model pengelompokan MARS terbaik untuk data penderita ISPA di Kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo dihasilkan dari kombinasi $BF = 16$, $MI = 1$, dan $MO = 2$, dengan persamaan sebagai berikut:

$$\hat{f}(x) = 0,649 - 0,073BF_7 - 6,439BF_8 + 4,819BF_{10} - 0,086BF_{12} + 0,063BF_{14} \quad (10)$$

dengan

$$BF_7 = (10 - X_8)_+$$

$$BF_8 = (X_5 - 0,330)_+$$

$$BF_{10} = (X_5 - 0,120)_+$$

$$BF_{12} = (X_7 - 27,230)_+$$

$$BF_{14} = (X_7 - 18,710)_+$$

Klasifikasi data penderita ISPA di kecamatan Jabon, Tanggulangin, dan Porong kabupaten Sidoarjo dengan pendekatan MARS menghasilkan ketepatan 97,4 persen dengan nilai GCV yakni 0,096 dan R^2 sebesar 82,9 persen dan konsistensi secara statistik.

DAFTAR PUSTAKA

- Agresti, A. (2011). *Categorical Data Analysis*. New York: John Wiley and Sons.
- Arifin, M. Z. (2015). *Metode Pendekatan Multivariate Adaptive Regression Spline (MARS) Bagging Dalam Memodelkan Presentase Gizi Buruk Balita di Jawa Timur*. Malang: Universitas Brawijaya.
- Asriani, E. D. (2016). *Estimasi Multivarite Adaptive Regression Spline (MARS) Pada Indeks Harga Saham (IHSG)*. Semarang: Universitas Negeri Semarang.
- Friedman, J. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, Vol. 19 No. 1.
- Hayati. (2014). Gambaran Faktor Penyebab Infeksi Saluran Pernafasan Akut (ISPA) Pada Balita di Puskesmas Pasirkaliki Kota Bandung. *Jurnal Ilmu Keperawatan*, 11 (1), 62–67.
- Jalil, R. (2018). *Faktor-Faktor Yang Berhubungan Dengan Kejadian Ispa Pada Balita Di Wilayah Kerja Puskesmas Kabangka Kecamatan Kabangka Kabupaten Muna*. Kendari: Universitas Halu Oleo.
- Kasse, I. (2017). Memodelkan Faktor-Faktor Yang Mempengaruhi Gizi Buruk Balita Dengan Metode Multivariate Adaptive Regression Spline (MARS). *Matematika dan Statistika serta Aplikasinya Vol 5, No 1*.
- Kishartini, S. I. (2014). *Multivariate Adaptive Regression Spline (MARS) untuk Klasifikasi Status Kerja di Kabupaten Demak*. Universitas Diponegoro, Semarang: *Jurnal Gaussian*, 3(4):711-718.
- Meliala, B. (2017). *Regresi Spline Multivariat Adaptive (Multivariate Adaptive Regression Spline)*. Yogyakarta: UIN Sunan Kalijaga.
- Mukono. (2008). *Pencemaran udara dan pengaruhnya terhadap gangguan saluran pernapasan*. Surabaya: Universitas Airlangga Press.
- Otok, B. S. (2008). *Pengembangan Model Klasifikasi Berbasis Machine Learning Untuk Prediksi Kapailitan Bank Umum di Indonesia*. Surabaya: Lembaga Penelitian dan Pengabdian Kepada Masyarakat Institut Teknologi Sepuluh Nopember.
- Prasetyo, G. C. (2009). *Klasifikasi Deteksi Intrusi Menggunakan Pendekatan Classification And Regression Trees (CART) Dan Multivariate Adaptive Regression Spline (MARS)*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Raditya, A. P. (2018). *Penerapan metode Multivariate Adaptive Regression Spline (MARS) dengan estimator nadaraya-watson fungsi kernel gaussian*. Yogyakarta: Fakultas MIPA UII.
- Raditya, A. P. (2018). *Penerapan Metode Multivariate Adaptive Regression Spline (MARS) Dengan Estimator Nadaraya-Watson Fungsi Kernel Gaussian*. Yogyakarta: Universitas Islam Indonesia.
- Rahmadhani, N. F. (2019). *Multivariate Adaptive Regression Spline (MARS) dalam Menentukan Faktor-Faktor Kecelakaan Lalu Lintas di Kota Yogyakarta*. Yogyakarta: UIN Sunan Kalijaga Yogyakarta.
- Respita, R. D. (2017). *Perbandingan Model Regresi Spline dan Multivariate Adaptive Regression Spline untuk Analisis Survival pada Pasien Kanker Serviks di RSUD Dr. Soetomo Surabaya*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Rovicky. (2006). *Hasil Pengujian Toksikologis Lumpur Porong Sidoarjo*. Jakarta: Sucofindo, Corelab, dan Bogorlab.