

BAB II

TINJAUAN PUSTAKA

1.1 Analisis Regresi

Salah satu metode statistik yang umum digunakan untuk menganalisis hubungan ataupun pengaruh antara variabel prediktor dan variabel respon. Mengatasi kurva regresi dapat dilakukan melalui dua pendekatan yaitu regresi parametrik dan regresi nonparametrik. Misalnya y adalah variabel respon dan x adalah variabel prediktor, secara umum pendekatan y dan x dapat ditulis sebagai berikut :

$$y_i = f(x_i) + \varepsilon_i, i = 1, 2, \dots, n \quad (2.1)$$

Keterangan :

ε = residual random

$f(x)$ = kurva regresi

Apabila tidak terdapat informasi apapun tentang bentuk fungsi, maka digunakan pendekatan nonparametrik (Hardle, 1990). Spline merupakan peicewise polinomial yang memiliki sifat tersegmen yang memberikan fleksibilitas lebih dibandingkan polinomial, sehingga memungkinkan $f(x_i)$ menyesuaikan diri lebih efektif terhadap karakteristik suatu fungsi atau data (Hardle, 1990).

1.2 Regresi Nonparametrik

Regresi nonparametrik merupakan metode pendekatan regresi yang sesuai untuk pola data dan tidak diketahui kurva regresi atau tidak terdapat informasi lengkap masa lalu tentang pola data (Eubank, 1999).

Eubank (1999) model regresi nonparametrik, sebagai berikut :

$$y_i = f(x_i) + \varepsilon_i, i = 1, 2, \dots, n \quad (2.2)$$

Beberapa model regresi nonparametrik yang banyak digunakan diantaranya : Spline, MARS, Kernel, Deret Fourier, Deret Orthogonal, *Neural Network* (NN), Polinomial Lokal, Histogram, *Wavelets*, k-NN, dan yang lainnya (Budiantara, 2009).

1.3 *Multivariate Adaptive Regression Splines* (MARS)

Multivariate Adaptive Regression Spline (MARS) dikembangkan oleh Friedman (1991) untuk pendekatan model regresi multivariat nonparametrik. MARS merupakan pengembangan dari pendekatan *Recursive Partitioning Regression* (RPR) yang masih memiliki kelemahan dimana model yang diperoleh tidak kontinu pada knots. Selain itu *Recursive Partitioning Regression* (RPR) tidak bisa mengidentifikasi adanya fungsi linier dan aditif. Model MARS dapat berguna untuk mengatasi data yang berdimensi tinggi seperti data yang memiliki jumlah variabel prediktor sebesar $3 \leq n \leq 20$. Selain itu MARS juga dapat menghasilkan prediksi variabel respon yang

akurat, mengurangi efek outlier pada model yang terakhir, proses pembentukan model pada MARS tidak memerlukan asumsi, menghasilkan model yang kontinu dalam knot berdasarkan *Generalized Cross Validation* (GCV) terkecil, model MARS tidak dipengaruhi oleh volume data yang hilang. Hidayat (2008) menyatakan bahwa MARS merupakan salah satu metode alternatif untuk pemodelan menggunakan regresi nonparametrik bagi data berdimensi tinggi, memiliki variabel banyak, serta ukuran sampel yang besar. MARS juga merupakan metode klasifikasi statistik modern yang sudah memanfaatkan fleksibilitas model dan menduga suatu distribusi di dalam masing-masing kelas yang pada akhirnya menyediakan suatu aturan pengelompokan (Dillon, 1978 dan Sharma, 1996).

Estimator Friedman (1991) model MARS, sebagai berikut :

$$\hat{f}(x) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{K_m} [S_{km}(x_{v(k,m)} - t_{km})]_+ \quad (2.3)$$

Keterangan :

a_0 = koefisien konstanta dari basis fungsi B_0

a_m = koefisien dari basis fungsi ke- m

M = banyaknya fungsi basis

K_m = banyaknya iteraksi pada fungsi basis ke- m

$x_{v(k,m)}$ = variabel independen

t_{km} = nilai knot dari variabel independen $x_{v(k,m)}$

S_{km} = nilainya 1 atau -1 jika data berada di sebelah kanan atau kiri titik knot

v = banyaknya variabel prediktor

k = banyaknya iteraksi

Pembentukan model MARS yang pertama yaitu menentukan titik-titik perubahan pola perilaku data atau titik knot. Pemilihan knot pada MARS menggunakan algoritma *forward* dan *backward*. *Forward* digunakan untuk mendapatkan subregion-subregion supaya dapat menentukan basis fungsi. Untuk memenuhi konsep persemoni (model sederhana) dilakukan tahap *backward* yaitu mengeluarkan suku model (basis fungsi) yang kontribusinya kecil terhadap nilai dugaan respon. Penentuan titik knots dan koefisien sangat penting untuk mendapatkan model terbaik.

Ukuran kontribusi yang digunakan yaitu modifikasi kriteria *generalized cross-validation* (GCV). Beberapa hal yang perlu diperhatikan dalam pemilihan model paling optimum (terbaik) adalah jika nilai *generalized cross-validation* (GCV) dari model tersebut mempunyai nilai yang paling rendah (minimum) diantara model-model yang lain. Pada model *Multivariate Adaptive Regression Spline* (MARS), model terbaik diperoleh berdasarkan nilai *generalized cross-validation* (GCV) dari model tersebut mempunyai nilai yang paling rendah (minimum) diantara model-model yang lain (Otok et al, 2012).

$$GCV = \frac{MSE}{\left[1 - \frac{C(M)}{N}\right]} \frac{1/N \sum_{i=1}^n [y_i - f_M(x_i)]^2}{\left[1 - \frac{C(M)}{N}\right]^2} \quad ((2.5))$$

Keterangan :

MSE = kuadrat tengah error

Y_i = variabel respon

X_i = variabel prediktor

N = banyaknya pengamatan

$f_{iM}(X_i)$ = nilai taksiran variabel respon pada M fungsi di X_i

M = maksimal jumlah fungsi basis

$C(M)$ = $CM + DM$

$C(M)$ = $Trace [B(B^T B)^{-1} B^T] + 1$; adalah matriks dari M fungsi basis

d = nilai ketika setiap fungsi basis mencapai nilai optimal ($2 \leq d \leq 4$)

Klasifikasi model MARS, berdasarkan pada pendekatan analisa regresi. Jika klasifikasi pada variabel respon terdiri dari dua nilai, maka menurut (Cox dan Snell, 1989) dikatakan sebagai regresi *binary response*, sehingga dapat digunakan model probabilitas dengan persamaan sebagai berikut :

$$\text{Logit } \pi(x) = \frac{e^{f(x)}}{1 + e^{f(x)}} \text{ dan } \{1 - \pi(x)\} = \frac{1}{1 + e^{f(x)}} \quad (2.6)$$

Dengan,

Probabilitas ($Y = 1$) = $\pi(x)$

Probabilitas ($Y = 0$) = $1 - \pi(x)$

Y merupakan variabel respon biner (0 dan 1) dengan banyaknya m variabel predicktor $x(x_1, \dots, x_m)$ model MARS untuk klasifikasi dinyatakan sebagai berikut (Otok, 2008) :

$$\pi(x) = \ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) = a_0 + \sum_{m=1}^M a_m \prod_{k=1}^{km} [S_{km}(x_{v(1,m)} - t_{tm})] \quad ((2.7)$$

dalam bentuk matriks dapat dinyatakan sebagai berikut :

$$\text{logit } \pi(x) = B_a$$

sedangkan dalam bentuk fungsi probabilitas dapat dinyatakan sebagai berikut:

$$\pi(x) = \sigma_L(B^T a)$$

Metode klasifikasi yang baik akan menghasilkan kesalahan klasifikasi kecil atau peluang kesalahan klasifikasi (alokasi) yang kecil (Agresti, 1990).

1.4 *Apparent Error Rate (APER) dan Press's Q*

Apparent Error Rate (APER) digunakan untuk menghitung kesalahan pada suatu klasifikasi. Menurut Johnson dan Wichern (1992) *Apparent Error Rate (APER)* adalah prosedur evaluasi yang digunakan untuk melihat kesalahan klasifikasi yang

dilakukan oleh suatu fungsi klasifikasi. Penentuan kesalahan klasifikasi MARS, dapat dilihat pada tabel klasifikasi bentuk matrik konfigurasi sebagai berikut :

Tabel 2.1 Tabel Klasifikasi MARS Respon Biner

Hasil Observasi	Taksiran Observasi	
	y_1	y_2
y_1	n_{11}	n_{12}
y_2	n_{21}	n_{22}

Keterangan :

y_1 = variabel respon kategori 1

y_2 = variabel respon kategori 2

n_{11} = jumlah observasi kelompok dari y_1 yang tepat diklasifikasikan sebagai kelompok y_1

n_{22} = jumlah observasi kelompok y_2 yang tepat diklasifikasikan sebagai kelompok y_2

n_{21} = jumlah observasi kelompok y_2 yang salah diklasifikasikan sebagai kelompok y_1

n_{12} = jumlah observasi kelompok y_1 yang salah diklasifikasikan sebagai kelompok y_1

n = jumlah observasi

Nilai APER merupakan nilai jumlah prediksi yang salah diklasifikasikan. Nilai APER didapat dengan perhitungan menggunakan rumus sebagai berikut :

$$\begin{aligned} \text{APER}(\%) &= \frac{\text{Jumlah Prediksi Salah}}{\text{Jumlah total Prediksi}} \\ &= \frac{n_{21} + n_{12}}{n} \times 100\% \end{aligned} \quad (2.8)$$

Press's Q yaitu untuk mengetahui kestabilan dalam pengklasifikasian. Nilai yang dihasilkan dibandingkan dengan sebuah nilai kritis (tabel chi-square dengan

derajat bebas 1), Menurut Hair J.F, Rolf E.J.(2006) jika nilai *Press's Q* melebihi nilai kritis maka klasifikasi dianggap sudah stabil atau konsisten secara statistik. Nilai *Press's Q* dapat dihitung menggunakan

$$PRESS'SQ = \frac{[N - n_i K]^2}{N(K - 1)} \quad (2.9)$$

Keterangan :

N = ukuran total sampel

n_i = jumlah individu yang tepat diklasifikasikan ($n_{11} + n_{22}$)

K = jumlah kelompok (kategori)

2.5 *Bootstrap Aggregating (Bagging)*

Bootstrap pertama kali digunakan oleh Efron pada tahun 1979. *Bootstrap* merupakan suatu metode penaksiran nonparametrik yang dapat menaksir parameter-parameter dari suatu distribusi, variansi dari sampel median, dan dapat menaksir error (Efron dan Tibshirani, 1993). *Bootstrap* mampu memberikan estimasi terbaik dari hasil estimasinya yang memiliki bias sedikit.

Metode *bagging* pertama kali digunakan oleh Breiman (1994). *Bagging* digunakan untuk memperbaiki stabilitas dan kekuatan prediksi dengan cara mereduksi variansi dari suatu prediktor. *Bagging* juga dapat meningkatkan akurasi prediksi. *Bagging* prediktor adalah metode untuk membangkitkan multiple version dari prediktor dan menggunakannya untuk aggregate prediktor.

Multiple version dibentuk dengan replikasi *bootstrap* dari sebuah data set. Sebuah data set \mathcal{L} terdiri dari $\{(y_i, x_i)\}, i = 1, 2, \dots, n\}$. Dilakukan replikasi *bootstrap* sehingga mendapatkan $\mathcal{L}^* = (y^*_i, x^*_i), i = 1, 2, \dots, n$. replikasi *bootstrap* dilakukan sebanyak B kali, sehingga didapatkan $\{\mathcal{L}^{(B)}\}$ dari \mathcal{L} . $\{\mathcal{L}^{(B)}\}$ adalah resampling dengan pengembalian (Otok et al, 2012). Penentuan batas replikasi *bootstrap* dilakukan apabila hasil replikasi *bootstrap* sudah menghasilkan nilai akurasi yang tinggi. peningkatan akurasi dan prediksi akan terjadi jika banyaknya replikasi ditingkatkan dari 50 ke 100 kali dan jika replikasinya ditingkatkan menjadi lebih dari 100 kali akan menghasilkan akurasi yang tidak lebih besar dibandingkan replikasi 100 kali (Hestie, Tibshirani dan Friedman, 2001). Untuk memperoleh hasil yang lebih baik, maka replikasi *bootstrap* dilakukan sebanyak 50 sampai 200 kali (Efron dan Tibshirani, 1993). Algoritma pada bagging merujuk pada algoritma pada *bootstrap* yaitu pengambilan sampel dengan pengembalian pada data training sehingga terbentuk variasi data baru.

Algoritma bagging MARS menurut (Buhlmann, P. dan Yu, B. (2002) sebagai berikut :

1. Sebuah data set \mathcal{L} terdiri dari $\{(y_i, x_i), i = 1, 2, \dots, n\}$ Melakukan replikasi *bootstrap* pada data tersebut sehingga didapatkan $\mathcal{L}^*_i = (y^*_i, x^*_i), i = 1, 2, \dots, n$.
2. Replikasi *bootstrap* dilakukan sebanyak B kali, sehingga didapatkan, sehingga didapatkan $\{\mathcal{L}^{(B)}\}$ dari \mathcal{L} . $\{\mathcal{L}^{(B)}\}$ adalah *resampling* dengan pengembalian.

Menurut Buhlmann, P. dan Yu, B. (2002) bagging memiliki potensi untuk menurunkan kuadrat *error* permalan pada berbagai proses.

2.6 Pembangunan Manusia

Manusia pada dasarnya adalah kekayaan yang di miliki suatu negara. Tujuan dari sebuah pembangunan adalah menciptakan lingkungan yang memungkinkan bagi rakyatnya untuk menikmati umur panjang, sehat, dan menjalankan kehidupan produktif. Tetapi hal yang seringkali terlupakan dalam diri manusia adalah kesibukan dalam memperkaya diri, mengumpulkan harta dan uang.

Dalam UNDP (*United Nations Development Programme*), pembangunan manusia adalah suatu proses untuk memperbesar pilihan-pilihan bagi manusia. Pada saat yang sama pembangunan manusia dapat dilihat juga sebagai pembangunan (formation) kemampuan manusia melalui perbaikan taraf kesehatan, pengetahuan, dan ketrampilan, sekaligus pemanfaatan kemampuan/ketrampilan mereka.

Menurut UNDP dalam *Human Development Report* (HDR) yang menekankan bahwa untuk memperluas pilihan-pilihan manusia, konsep pembangunan manusia harus dibangun dari empat dimensi yang tidak terpisahkan. Berdasarkan konsep di atas maka untuk menjamin tercapainya tujuan pembangunan manusia, ada empat unsur pokok yang perlu diperhatikan (UNDP: 1995) yaitu:

a. Produktivitas (*Productivity*)

Masyarakat harus mampu untuk meningkatkan produktifitas mereka dan berpartisipasi penuh dalam proses mencari penghasilan dan lapangan pekerjaan. Oleh karena itu, pembangunan ekonomi merupakan bagian dari model pembangunan manusia.

b. Pemerataan (*equity*)

Masyarakat harus mempunyai akses untuk memperoleh kesempatan yang adil. Semua hambatan terhadap peluang ekonomi dan politik harus dihapuskan sehingga masyarakat dapat berpartisipasi di dalam dan memperoleh manfaat dari peluang-peluang yang ada.

c. Kestinambungan (*Sustainability*)

Akses untuk memperoleh kesempatan harus dipastikan bahwa tidak hanya untuk generasi sekarang tetapi juga untuk generasi yang akan datang. Semua jenis pemodalan baik itu fisik, manusia, dan lingkungan hidup harus dilengkapi.

d. Pemberdayaan (*Empowerment*)

Pembangunan harus dilakukan oleh masyarakat, dan bukan hanya untuk mereka. Masyarakat harus berpartisipasi penuh dalam mengambil keputusan dan proses-proses yang memengaruhi kehidupan mereka.

Berdasarkan pengertian-pengertian di atas maka dapat disimpulkan bahwa pelayanan merupakan suatu bentuk sistem, procedure atau metode tertentu diberikan kepada orang lain. Dalam hal ini, kebutuhan pelanggan tersebut dapat terpenuhi sesuai dengan harapan atau keinginan pelanggan dengan tingkat persepsi mereka.

2.7 Indeks Pembangunan Manusia(IPM)

Secara khusus, Indeks Pembangunan Manusia (IPM) adalah mengukur capaian pembangunan manusia manusia berbasis sejumlah komponen dasar kualitas hidup (BPS 2008:9). Indeks pembangunan manusia mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup. Sebagai ukuran kualitas hidup, IPM dibangun melalui tiga dimensi dasar yaitu:

- a. Dimensi umur panjang dan hidup sehat.
- b. Dimensi pengetahuan.
- c. Dimensi standar hidup layak (BPS:2014)

Untuk mengukur dimensi umur panjang dan hidup sehat digunakan angka harapan hidup waktu lahir. Selanjutnya untuk mengukur dimensi pengetahuan digunakan gabungan indikator harapan lama sekolah dan rata-rata lama sekolah. Adapun untuk mengukur dimensi hidup layak digunakan indikator pengeluaran perkapita disesuaikan sebagai pendekatan pendapatan yang mewakili capaian pembangunan untuk hidup layak.

Sebelum penghitungan IPM, setiap komponen IPM harus dihitung indeksnya. Formula yang digunakan dalam penghitungan indeks komponen IPM adalah sebagai berikut:

$$\text{Dimensi Kesehatan: } I_{\text{kesehatan}} = \frac{AHH - AHH_{\min}}{AHH_{\text{maks}} - AHH_{\min}} \quad (2.10)$$

$$\text{Dimensi Pendidikan: } I_{\text{pendidikan}} = \frac{HLS - HLS_{\min}}{HLS_{\text{maks}} - HLS_{\min}} \quad (2.11)$$

$$I_{\text{RLS}} = \frac{RLS - RLS_{\min}}{RLS_{\text{maks}} - RLS_{\min}} \quad (2.13)$$

$$\text{Dimensi Pengeluaran ln} I_{\text{pengeluaran}} = \frac{\ln(\text{pengeluaran}) - \ln(\text{pengeluaran}_{\min})}{\ln(\text{pengeluaran}_{\text{maks}}) - \ln(\text{pengeluaran}_{\min})} \quad (2.14)$$

Untuk menghitung indeks masing-masing komponen IPM digunakan batas maksimum dan minimum seperti terlihat dalam Tabel berikut:

Tabel 2.2 batas maksimum dan minimum

No	Komponen IPM	Maksimum	Minimum	Keterangan
1	Angka Harapan Hidup Saat Lahir(Tahun)	85	20	Standar UNDP
2	Angka harapan Lama Sekolah(Tahun)	18	0	Standar UNDP
3	Rata-rata Lama Sekolah (Tahun)	15	0	Standar UNDP
4	Pengeluaran Perkapita Disesuaikan	26.572.352**(Rp)	1.007.436*(Rp)	Pengeluaran per Kapita Disesuaikan

Keterangan:

- * Daya beli minimum merupakan garis kemiskinan terendah kabupaten tahun 2010(data empiris yaitu di Tolikara-Papua)
- ** Daya beli maksimum merupakan nilai tertinggi kabupaten yang diproyeksikan hingga 2025 (akhir RPJPN) yaitu perkiraan pengeluaran per kapita Jakarta Selatan tahun 2025Sumber: BPS, 2014

Selanjutnya nilai IPM dapat dihitung sebagai:

$$IPM = \sqrt[3]{I_{Kesehatan} \times I_{Pendidikan} \times I_{pengeluaran}} \times 100 \quad (2.15)$$

2.8 Angka Harapan Hidup

Angka Harapan Hidup(AHH) merupakan alat untuk mengevaluasi kinerja pemerintah dalam meningkatkan kesejahteraan penduduk pada umumnya, dan meningkatkan derajat kesehatan pada khususnya. Angka Harapan Hidup yang rendah di suatu daerah harus diikuti dengan program pembangunan kesehatan, dan program sosial lainnya termasuk kesehatan lingkungan, kecukupan gisi dan kalori termasuk program pemberantasan kemiskinan. Angka harapan hidup mencerminkan drajat kesehatan suatu masyarakat.

2.9 Indeks Pendidikan

Salah satu komponen pembentuk IPM adalah dari dimensi pengetahuan yang diukur melalui tingkat pendidikan. Dalam hal ini, indikator yang digunakan adalah

rata-rata lama sekolah (*mean years of schooling*) dan angka harapan lama sekolah. Rata-rata lama sekolah didefinisikan sebagai jumlah yang digunakan oleh penduduk dalam menjalani pendidikan formal. Diasumsikan bahwa dalam kondisi normal rata-rata lama sekolah suatu wilayah tidak akan turun. Cakupan penduduk yang dihitung dalam perhitungan rata-rata lama sekolah adalah penduduk berusia 25 tahun ke atas.

Angka Harapan Lama Sekolah didefinisikan lamanya sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak pada umur tertentu di masa mendatang. Diasumsikan bahwa peluang anak tersebut akan tetap bersekolah pada umur-umur berikutnya sama dengan peluang penduduk yang bersekolah per jumlah penduduk untuk umur yang sama saat ini.

Angka Harapan Lama Sekolah dihitung untuk penduduk berusia 7 tahun ke atas. HLS dapat digunakan untuk mengetahui kondisi pembangunan sistem pendidikan di berbagai jenjang yang ditunjukkan dalam bentuk lamanya pendidikan (dalam tahun) yang diharapkan dapat dicapai oleh setiap anak.

2.10 Indeks Standar Hidup Layak

Dimensi lain dari ukuran kualitas hidup manusia adalah standar hidup layak. Dalam cakupan lebih luas, standar hidup layak menggambarkan tingkat kesejahteraan yang dinikmati oleh penduduk sebagai dampak semakin membaiknya ekonomi. UNDP mengukur standar hidup layak menggunakan Produk Domestik Bruto (PDRB) riil yang

disesuaikan, sedangkan BPS dalam menghitung standar hidup layak menggunakan Pengeluaran rata-rata disesuaikan.

Pengeluaran per kapita yang disesuaikan ditentukan dari nilai pengeluaran per kapita dan paritas daya beli *Purchasing Power Parity* (PPP). Rata-rata pengeluaran per kapita setahun diperoleh dari Susenas, dihitung dari level provinsi hingga level kab/kota. Rata-rata pengeluaran per kapita dibuat konstan/riil dengan tahun dasar 2012=100. Perhitungan paritas daya beli pada metode baru menggunakan 96 komoditas dimana 66 komoditas merupakan makanan dan sisanya merupakan komoditas nonmakanan. Metode penghitungan paritas daya beli menggunakan Metode Rao. Perhitungan paritas daya beli dihitung dari bundle komoditas makanan dan non makanan.

Rumus Perhitungan Paritas Daya Beli (PPP):

$$PPP_j = \prod_{i=1}^m \left(\frac{p_{ij}}{p_{ik}} \right)^{\frac{1}{m}} \quad (2.16)$$

Keterangan :

p_{ij} = Harga komoditas i di kab atau kota j

p_{ik} = Harga Komoditas i di Jakarta Selatan

m = Jumlah Komoditas

Pengelompokan indeks pembangunan manusia (BPS, 2014):

IPM < 60 : IPM rendah

60 ≤ IPM < 70 : IPM sedang

70 ≤ IPM < 85 : IPM tinggi