

Pemodelan Angka Harapan Hidup di Provinsi Jawa Timur Tahun 2007 dan 2011 Berdasarkan Angka Melek Huruf, Rata-rata Lama Sekolah, dan Pengeluaran Perkapita

Ni Wayan Dewinta Ayuni

*Jurusan Statistika, Institut Teknologi Sepuluh Nopember
Jl. Prof. Dr. Soepomo, SH. Janturan Yogyakarta
e-mail: Ummunabilah67@gmail.com*

Abstrak: Angka harapan hidup merupakan salah satu indikator atau penilaian derajat kesehatan suatu negara atau wilayah dan digunakan sebagai acuan dalam perencanaan program-program kesehatan. Angka harapan hidup dipengaruhi oleh berbagai faktor misalnya angka melek huruf, rata-rata lama sekolah, dan pengeluaran per kapita masyarakat yang ada pada daerah tersebut. Penelitian ini akan membahas pemodelan angka harapan hidup di Provinsi Jawa Timur tahun 2007 dan 2011 menggunakan faktor-faktor tersebut sebagai variabel prediktor. Metode yang digunakan adalah regresi linear berganda, *feed forward neural network* (FFNN), dan *multivariate adaptive regression spline* (MARS). Hasil penelitian ini menunjukkan bahwa metode FFNN merupakan metode terbaik dalam pemodelan angka harapan hidup di Provinsi Jawa Timur dilihat dari nilai MSE prediksi yang terkecil

Keywords: Angka harapan hidup, regresi, neural network, MARS.

1. Pendahuluan

Angka harapan hidup merupakan salah satu indikator atau penilaian derajat kesehatan suatu negara dan digunakan sebagai acuan dalam perencanaan program-program kesehatan. Angka Harapan Hidup disebut juga lama hidup manusia didunia. Angka harapan hidup pada suatu umur tertentu didefinisikan sebagai rata-rata tahun hidup yang akan dijalani oleh seseorang yang telah berhasil mencapai umur tersebut dalam situasi kematian yang berlaku di lingkungan masyarakatnya. Angka harapan hidup juga merupakan indikator kesejahteraan masyarakat pada suatu wilayah. Penelitian mengenai angka harapan hidup di suatu wilayah sangat penting guna mengetahui faktor-faktor yang berpengaruh terhadap angka harapan hidup sehingga pemerintah di daerah tersebut dapat mencanangkan program untuk meningkatkan angka harapan hidup masyarakat di wilayahnya. Beberapa faktor yang diduga berpengaruh terhadap angka harapan hidup adalah angka melek huruf, rata-rata lama sekolah, dan pengeluaran per kapita masyarakat.

Secara geografis Jawa Timur merupakan wilayah dengan populasi kedua terbesar di Indonesia. Angka harapan hidup Provinsi Jawa Timur menduduki posisi 11 secara nasional. Pada penelitian ini dilakukan pemodelan angka harapan hidup Provinsi Jawa Timur Pada Tahun 2007 dan 2011 berdasarkan angka melek huruf, rata-rata lama sekolah, dan pengeluaran per kapita masyarakat.

2. Tinjauan Pustaka

2.1. Regresi Linear Berganda

Model regresi bertujuan untuk membentuk model hubungan antara variabel prediktor dan variabel respon (Rawling, Pentula, & Dickey, 1932). Hubungan antara variabel prediktor dan variabel respon dikatakan linear jika dapat dinyatakan dalam model berikut:

$$y_t = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} + \varepsilon_i \quad (1)$$

dimana $i = 1, 2, \dots, n$ dan error-nya diasumsikan identik, independen dan berdistribusi normal dengan mean nol dan varians konstan (Myers, 1990). Bentuk penaksir dari parameter tersebut menggunakan ordinary least square adalah:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (2)$$

dengan $\hat{\beta}$: vektor parameter yang ditaksir ($(p+1) \times 1$)

X : matriks variabel bebas berukuran ($n \times (p+1)$)

y : vektor observasi dari variabel respon berukuran ($n \times 1$)

k : banyaknya variabel bebas ($k = 1, 2, \dots, p$)

Pada analisis regresi, signifikansi parameter diuji melalui dua tahap yaitu secara serentak (simultan) dan secara individu (parsial). Pengujian parameter secara serentak dilakukan dengan menggunakan uji F dan hipotesis yang diuji adalah sebagai berikut:

$$H_0 : \beta_1, \beta_2, \dots, \beta_p = 0$$

$$H_1 : \text{paling sedikit ada satu } \beta_k \neq 0$$

Dengan statistik uji F dijabarkan pada persamaan (3)

$$F_{hit} = \frac{MSR}{MSE} \quad (3)$$

Daerah penolakan H_0 adalah jika nilai $F_{hit} > F_{\alpha, v1, v2}$ dimana $v1 = p$ dan $v2 = n - p - 1$ atau nilai $p - value < \alpha$.

Sedangkan untuk pengujian secara individu dilakukan dengan menggunakan uji t dan hipotesis yang diuji adalah sebagai berikut:

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0, k = 1, 2, \dots, p$$

Statistik uji t dijabarkan pada persamaan (4):

$$t_{hit} = \frac{\beta_k}{SE\beta_k} \quad (4)$$

Keputusan menolak H_0 diambil apabila nilai $|t_{hit}| > t_{df; (1-\alpha)/2}$ dimana $df = n - 2 - k$.

2.2. Feed Forward Neural Network (FFNN)

FFNN disebut juga Multi Layer Perceptron (MLP). Jaringan pada model FFNN menggunakan algoritma backpropagation yang meliputi tiga tahap yaitu umpan maju (*feed*

forward) dari pola input, penghitungan dan propagasi balik dari error dan penyesuaian bobot. Pada tahap umpan maju setiap unit input menerima sinyal input (x_i) dan menyebarkannya ke unit tersembunyi z_1, \dots, z_p . Setiap unit tersembunyi menghitung aktivasinya dan jumlah terboboti dari input-inputnya dalam bentuk:

$$z_{in_j} = \sum_i w_{ji}x_i + w_{bj} \quad (5)$$

dimana x_i adalah aktivasi dari unit input ke- i yang mengirimkan sinyal ke unit hidden ke j , w_j adalah bobot dari sinyal yang terkirim dan $j = 1, 2, \dots, q$ adalah jumlah unit hidden. Hasil penjumlahan ditransformasi dengan fungsi aktivasi nonlinear:

$$Z_j = f(z_{in_j}) \quad (6)$$

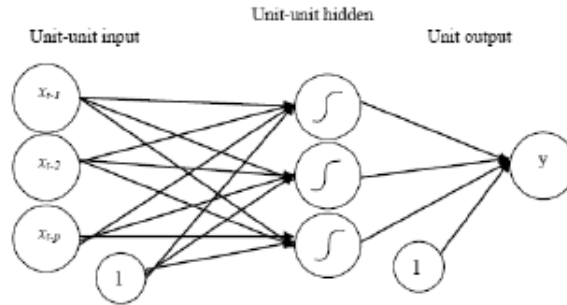
Setelah semua unit tersembunyi menghitung aktivasinya kemudian mengirimkan sinyal (z_j) ke unit output. Kemudian unit output menghitung aktivasinya dalam bentuk:

$$g(w, z) = \sum_j w_{j0}Z_j + w_{b0} \quad (7)$$

merupakan nilai output dari jaringan yaitu:

$$y = \sum_i w_{j0}f(a_j) + w_{b0} \quad (8)$$

dimana w_{b0} adalah bobot dari bias ke unit output. Arsitektur model FFNN dengan unit input lag 1 sampai p dan unit konstan, satu *hidden layer* dengan 3 neuron dan 1 unit output diilustrasikan pada Gambar 1.



Gambar 1. Arsitektur model FFNN

Fungsi aktivasi sigmoid pada neural network adalah:

$$f(x) = \frac{1}{1 + \exp(\beta x)} \quad (9)$$

Model FFNN dengan satu hidden dan input x_{t-1}, \dots, x_{t-p} ditulis dalam bentuk:

$$\hat{y} = \Psi_0 \left\{ w_{c0} + \sum_n w_{n0} \Psi_n \left(w_{cn} + \sum_t w_{in} x_t - J_t \right) \right\} \quad (10)$$

dimana $\{w_{cn}\}$ adalah bobot antara unit konstan dan neuron dan w_{co} adalah bobot antara unit konstan dan output. $\{w_{in}\}$ dan $\{w_{no}\}$ masing-masing menyatakan bobot koneksi input dengan neuron dan antara neuron dengan output. Kedua fungsi Ψ_n dan Ψ_0 masing-masing fungsi aktivasi yang digunakan pada neuron dan output. Notasi untuk model FFNN adalah NN (j_1, \dots, j_k, n) yang menyatakan NN dengan input lag j_1, \dots, j_k dan n neuron (Warsito, 2006).

2.3. Multivariate Adaptive Regression Spline

Model MARS berguna untuk mengatasi permasalahan data berdimensi tinggi dan menghasilkan prediksi variabel respon yang akurat, serta menghasilkan model yang kontinu dalam knot berdasarkan nilai GCV terkecil (Friedman, 1991). MARS dikembangkan oleh Friedman (1991) untuk pendekatan model regresi multivariate nonparametrik antara variabel respon dan beberapa variabel prediktor pada *piecewise* regresi. *Piecewise* regresi merupakan regresi yang memiliki sifat tersegmen. Apabila suatu garis regresi tidak bisa menjelaskan keseluruhan data maka beberapa garis regresi digunakan untuk menjelaskan seluruh data yang ada dari variabel yang independen. Tempat perubahan pola itulah yang dinamakan knot. Perlu dilakukan pengujian untuk mengontrol jumlah derajat bebas (DF) untuk optimalisasi knot. Sedangkan basis fungsi merupakan suatu fungsi yang digunakan untuk menjelaskan hubungan antara variabel respon dan variabel prediktor. Pada umumnya basis fungsi yang dipilih adalah berbentuk polinomial dengan turunan yang kontinu pada setiap titik knot. Friedman (1991) menyarankan jumlah maksimum basis fungsi (BF) dua sampai empat kali jumlah variabel prediktor. Model MARS digunakan untuk mengatasi kelemahan *Recursive Partitioning Regression* (RPR) yaitu menghasilkan model yang kontinu pada knot dan dapat mengidentifikasi adanya fungsi linear dan aditif. Perbaikan yang dilakukan untuk mengatasi keterbatasan RPR, antara lain menghasilkan fungsi basis menjadi:

$$B_k(x) = \prod_{t=1}^{L_k} \left[s_{kl} \left(x_{i(k,l)} - x_{i(k,l)}^* \right) \right]_+^m \quad (11)$$

dengan $[x]_+$ merupakan bagian bernilai positif dari x , dimana x adalah $x_{i(k,l)} - x_{i(k,l)}^*$. $[x]_+ = x$ jika $x > 0$ dan $[x]_+ = 0$ jika $x \leq 0$, m adalah orde dari spline, L_k adalah banyaknya interaksi pada fungsi basis K , $x_{i(k,l)}^*$ merupakan knot dari variabel prediktor $x_{i(k,l)}$.

Modifikasi Friedman dalam mengatasi kelemahan RPR menghasilkan persamaan untuk model MARS seperti dijelaskan pada persamaan selanjutnya (Friedman, 1991).

$$\begin{aligned} f(x) &= c_0 + \sum_{k=1}^K c_k \prod_{l=1}^{L_k} \left[\left(x_{i(k,l)} - x_{i(k,l)}^* \right) \right]_+^m + \varepsilon_i \\ &= c_0 + \sum_{k=1}^K c_k B_k(x) \end{aligned} \quad (12)$$

dengan fungsi

$$[x_{i(k,l)} - x_{i(k,l)}^*]_+ = \begin{cases} (x_{i(k,l)} - x_{i(k,l)}^*), & x_{i(k,l)} \geq x_{i(k,l)}^* \\ 0 & , x_{i(k,l)} < x_{i(k,l)}^* \end{cases} \quad (13)$$

Jika dalam bentuk matriks:

$$\mathbf{Y} = \mathbf{cB} + \varepsilon \quad (14)$$

dengan,

$$\mathbf{Y} = (Y_1, \dots, Y_b)^T, \quad \mathbf{c} = (c_1, \dots, c_k)^T, \quad \varepsilon = (\varepsilon_1, \dots, \varepsilon_k)^T \quad (15)$$

Atau dapat dijabarkan sebagai berikut:

$$f(x) = c_0 + \sum_{i=1}^d f_i(x_i) + \sum_{i,j=1}^d f_{ij}(x_i, x_j) + \sum_{i,j,k=1}^d f_{ijk}(x_i, x_j, x_k) + \dots \quad (16)$$

Ukuran kontribusi yang digunakan dalam tahap ini adalah modifikasi kriteria GCV sebagai berikut (Lewis, 2000).

$$GCV^*(K) = \frac{(1/n) \sum_{i=1}^n [y_i - \hat{f}_k(x_i)]^2}{\left[1 - \frac{C(K)^*}{n}\right]^2} \quad (17)$$

Keterangan:

K : jumlah basis fungsi (nonconstant basis fungsi) yang ditentukan pada tahap forward

$C(K)^*$: $C(K)^* = C(K) + dK$, $2 \leq d \leq 4$

$C(K)$: jumlah parameter dalam model = trace $(\mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T) + 1$

\mathbf{B} : matriks basis fungsi (*nonconstant* basis fungsi)

n : jumlah data

y_i : nilai variabel respon

$\hat{f}_k(x_i)$: nilai taksiran variabel respon pada K basis fungsi

3. Data dan Metodologi

Data yang digunakan pada penelitian ini adalah data angka harapan hidup (AHH) Provinsi Jawa Timur sebagai variabel respon sedangkan variabel prediktor yang digunakan adalah angka melek huruf (AMH), rata-rata lama sekolah (RLS) dan pengeluaran per kapita (PP) masyarakat. Periode data yang digunakan adalah data tahun 2007 dan tahun 2011. Data ini akan diolah dengan 3 metode yaitu regresi linear berganda, feed forward neural network, dan multivariate adaptive regression spline. Analisis dibagi menjadi analisis eksplanasi yaitu tanpa data missing dan analisis prediksi yaitu dengan data missing 1, missing 2, dan missing 3. Model terbaik dinilai dari nilai MSEF terkecil yang dihasilkan.

4. Hasil dan Pembahasan

4.1. Analisis Eksplanasi

4.1.1. Regresi Linear Berganda

Dengan menggunakan metode regresi linier berganda, diperoleh model untuk angka harapan hidup (AHH) untuk data tahun 2007 sebagai berikut:

$$AHH = 42,3 + 0,378 AMH - 0,101 RLS - 0,0121 PP \quad (18)$$

Interpretasi dari model tersebut adalah apabila angka melek huruf (AMH) naik satu satuan maka angka harapan hidup (AHH) akan meningkat sebesar 0,378 kali sedangkan apabila rata-rata lama sekolah (RLS) naik satu satuan maka angka harapan hidup (AHH) akan menurun sebesar 0,101 kali dan apabila pengeluaran perkapita (PP) naik satu satuan maka angka harapan hidup (AHH) akan menurun sebesar 0,0121 kali. Model tersebut menghasilkan koefisien determinasi R^2 sebesar 66% artinya angka harapan hidup (AHH) dapat dijelaskan oleh variabel angka melek huruf (AMH), rata-rata lama sekolah (RLS) dan pengeluaran perkapita (PP) sebesar 66% sedangkan sisanya sebesar 34% dijelaskan oleh variabel lain yang tidak diikutsertakan dalam penelitian.

Pengujian secara serentak terhadap parameter model regresi dilakukan dengan menggunakan uji F. Hipotesis yang akan diuji adalah:

$$H_0 : \beta_1, \beta_2, \dots, \beta_p = 0$$

$$H_1 : \text{paling sedikit ada satu } \beta_k \neq 0$$

Hasil analisis data dari output minitab diperoleh F_{hitung} sebesar 24,9 dengan p-value = 0,0 sehingga untuk $\alpha = 5\%$ terlihat bahwa p-value $< \alpha$ maka keputusan yang diambil adalah tolak H_0 dan kesimpulannya model tersebut secara keseluruhan sudah sesuai terhadap data. Pengujian secara individu terhadap parameter regresi dilakukan dengan menggunakan uji t dengan menggunakan hipotesis sebagai berikut:

$$H_0 : \beta_k = 0 \text{ (Variabel prediktor ke-} k \text{ tidak berpengaruh signifikan terhadap respon)}$$

$$H_1 : \beta_k \neq 0 \text{ (Variabel prediktor ke-} k \text{ berpengaruh signifikan terhadap respon)}$$

dengan $k = 1, 2, 3$. Tabel 1 menyajikan hasil pengujian parameter regresi secara individu untuk model regresi angka harapan hidup (AHH) data tahun 2007.

Tabel 1. Hasil Uji Signifikansi Parameter Regresi Linear Berganda Tahun 2007

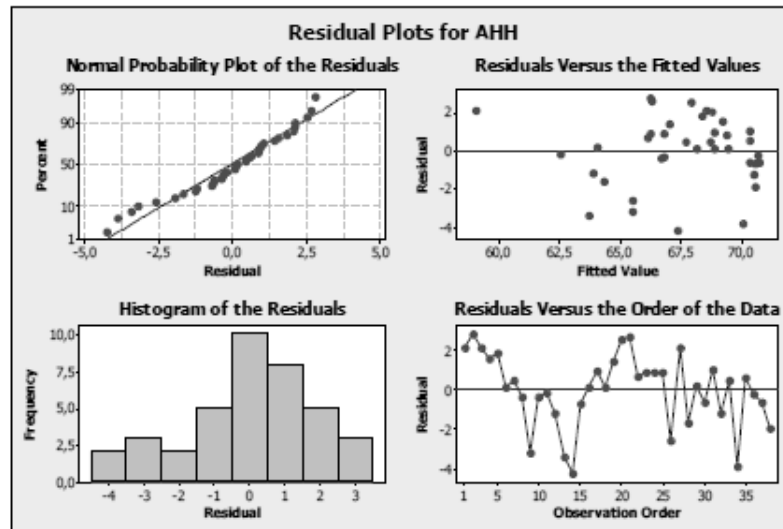
Prediktor	Koef.	SE Koef.	T hitung	p-value	Keterangan
Konstan	42,33	28,8300	1,47	0,151	
AMH	0,378	0,0928	4,07	0,000	signifikan
RLS	-0,101	0,4855	-0,21	0,837	tidak signifikan
PP	-0,0121	0,0467	-0,26	0,789	tidak signifikan

Terlihat bahwa variabel prediktor yang signifikan adalah angka melek huruf (AMH) sedangkan variabel lainnya yaitu rata-rata lama sekolah (RLS) dan pengeluaran perkapita (PP) tidak signifikan pada $\alpha = 5\%$. Setelah melakukan pemilihan model regresi terbaik dengan menggunakan metode stepwise, diperoleh model regresi sebagai berikut:

$$AHH = 36,6 + 0,349 AMH \quad (19)$$

Model regresi tersebut menjelaskan bahwa apabila angka melek huruf (AMH) naik satu satuan maka angka harapan hidup (AMH) akan meningkat sebesar 0,349 kali. Model tersebut menghasilkan R^2 sebesar 68,5% artinya model tersebut dapat menjelaskan keragaman data sebesar 68,5%. Residual (sisaan)

Residual (sisaan) pada analisis regresi harus memenuhi beberapa asumsi yaitu asumsi identik, independen dan berdistribusi normal. Pemeriksaan terhadap asumsi tersebut dapat dilakukan dengan meninjau plot residual (sisaan). Gambar 2 menyajikan plot residual (sisaan) dari model regresi linier yang dihasilkan untuk data tahun 2007.



Gambar 2. Plot Residual Regresi Tahun 2007

Dari plot residual tersebut dapat dikatakan bahwa residual memenuhi asumsi kenormalan dengan nilai K-S sebesar 0,11 dan p-value $> 0,15$. Residual diduga tidak memenuhi asumsi identik karena plot residual terhadap nilai dugaan (*fitted value*) membentuk pola seperti corong (tidak acak) sedangkan asumsi independen diduga terpenuhi karena plot residual terhadap order acak (tidak membentuk suatu pola tertentu).

Selanjutnya dengan menggunakan metode yang sama, diperoleh model regresi linear berganda untuk data tahun 2011 sebagai berikut:

$$AHH = 56,8 + 0,329 AMH + 0,196 RLS - 0,0303 PP \quad (20)$$

Interpretasi dari model tersebut adalah apabila angka melek huruf (AMH) naik satu satuan maka angka harapan hidup (AHH) akan meningkat sebesar 0,329 kali sedangkan apabila rata-rata lama sekolah (RLS) naik satu satuan maka angka harapan hidup (AHH) akan meningkat sebesar 0,196 kali dan apabila pengeluaran perkapita (PP) naik satu satuan maka angka harapan hidup (AHH) akan menurun sebesar 0,0303 kali. Model tersebut menghasilkan koefisien determinasi R^2 sebesar 61,8% artinya model tersebut dapat menjelaskan keragaman data sebesar 61,8%. Pengujian secara serentak dilakukan dengan menggunakan uji F dengan hipotesis yang akan diuji adalah:

$$H_0 : \beta_1, \beta_2, \dots, \beta_p = 0$$

$$H_1 : \text{paling sedikit ada satu } \beta_k \neq 0$$

Hasil analisis data dari output minitab diperoleh Fhitung sebesar 20,93 dengan p-value = 0,0 sehingga untuk $\alpha = 5\%$ terlihat bahwa p-value $< \alpha$ maka keputusan yang diambil adalah tolak H_0 . Kesimpulannya model tersebut sudah sesuai terhadap data. Pengujian secara individu dilakukan dengan menggunakan uji t. Hipotesisnya adalah:

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

dengan $k = 1, 2, 3$

Hasil uji t pengujian secara individu terhadap parameter regresi disajikan pada Tabel 2.

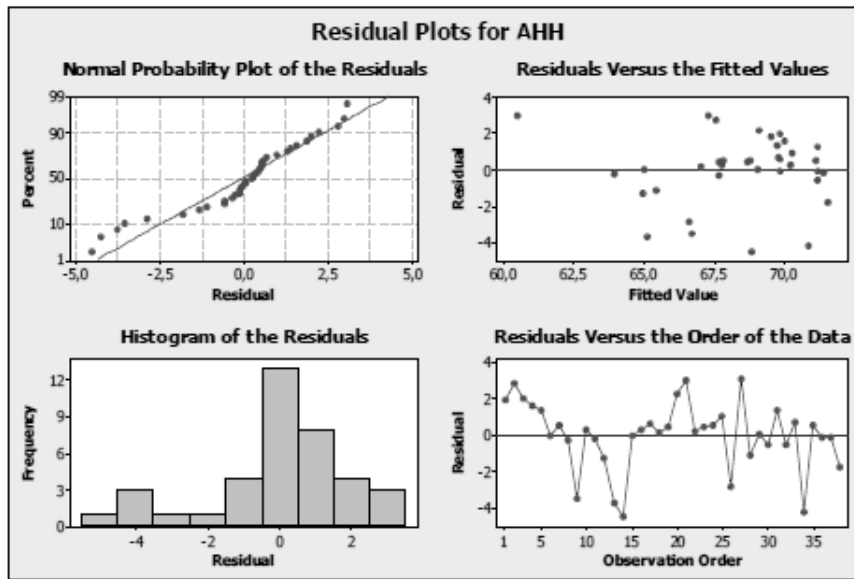
Tabel 2. Hasil Uji Signifikansi Parameter Regresi Linear Berganda Tahun 2011

Prediktor	Koef.	SE Koef.	T hitung	p-value	Keterangan
Konstan	56,76	31,21	1,82	0,078	
AMH	0,329	0,0911	3,61	0,001	signifikan
RLS	0,196	0,4480	0,44	0,664	tidak signifikan
PP	-0,0303	0,0494	-0,61	0,554	tidak signifikan

Tabel 2 menunjukkan bahwa variabel prediktor yang signifikan adalah angka melek huruf (AMH) sedangkan variabel lainnya yaitu rata-rata lama sekolah (RLS) dan pengeluaran perkapita (PP) tidak signifikan pada $\alpha = 5\%$. Setelah melakukan pemilihan model terbaik dengan menggunakan regresi stepwise, diperoleh model linier sebagai berikut:

$$HH = 37,2 + 0,348 AMH \quad (21)$$

Artinya apabila angka melek huruf (AMH) naik satu satuan maka angka harapan hidup (AMH) akan meningkat sebesar 0,348 kali. R^2 yang dihasilkan sebesar 64,4% yang menunjukkan bahwa model tersebut dapat menjelaskan keragaman data sebesar 64,4%. Gambar 3 menyajikan plot residual untuk memeriksa asumsi yang harus dipenuhi untuk data tahun 2011.



Gambar 3. Plot Residual Regresi Tahun 2011

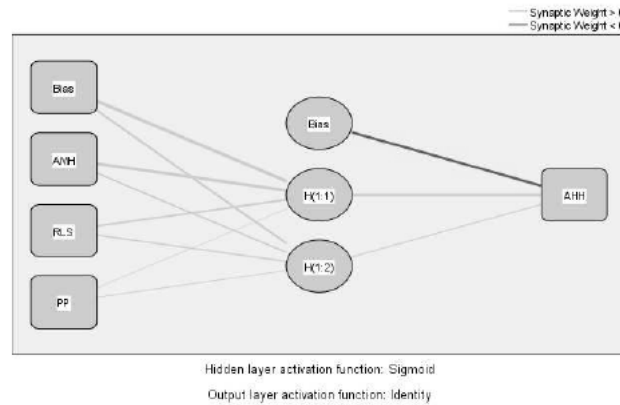
Dari plot residual tersebut dapat dikatakan bahwa residual tidak memenuhi asumsi kenormalan dengan nilai K-S sebesar 0,17 dan p-value $< 0,01$. Residual diduga tidak memenuhi asumsi identik karena plot residual terhadap nilai dugaan (*fitted value*) membentuk pola seperti corong (tidak acak) sedangkan asumsi independen diduga terpenuhi karena plot residual terhadap order acak (tidak membentuk suatu pola tertentu).

Oleh karena tidak semua asumsi residual terpenuhi pada metode regresi linier ini baik untuk data tahun 2007 maupun 2011, maka dapat dikatakan bahwa metode ini kurang layak untuk memodelkan angka harapan hidup (AMH) berdasarkan angka melek huruf (AMH), rata-rata lama sekolah (RLS) dan pengeluaran per kapita (PP).

4.1.2. Feed Forward Neural Network (FFNN)

Metode FFNN merupakan salah satu metode pada *neural network* dengan menggunakan fungsi aktivasi. Pada analisis ini fungsi aktivasi yang digunakan adalah fungsi sigmoid. Fungsi ini merupakan fungsi aktivasi yang sangat umum digunakan untuk membangun jaringan dan secara matematik dapat diturunkan menjadi fungsi naik.

Gambar 4 menyajikan arsitektur jaringan yang digunakan untuk pemodelan angka harapan hidup di Jawa Timur tahun 2007 dengan menggunakan satu hidden layer dan 2 nodes.



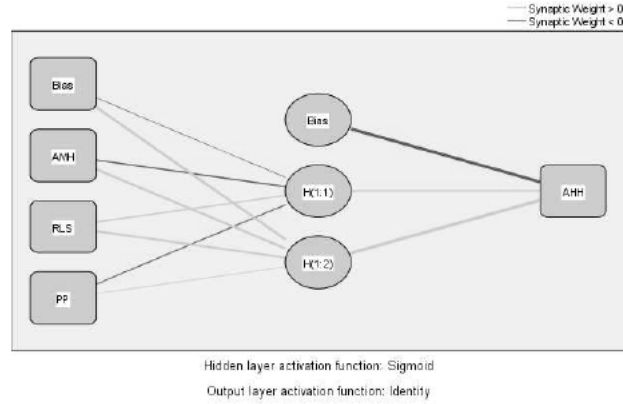
Gambar 4. Arsitektur Jaringan FFNN Data Tahun 2007

Model yang terbentuk untuk arsitektur tersebut adalah:

$$y = -1,921 + \frac{2,053}{1 + \exp(-(2,243 + 2,204AMH + 1,411RLS + 0,013PP))} + \frac{0,609}{1 + \exp(-(1,665 + 1,331AMH + 0,965RLS + 0,327PP))} \quad (22)$$

MSE yang dihasilkan oleh model ini cukup kecil yaitu sebesar 2,1882.

Selanjutnya Gambar 5 menyajikan arsitektur jaringan yang digunakan untuk pemodelan angka harapan hidup di Jawa Timur tahun 2011 dengan menggunakan satu hidden layer dan 2 nodes:



Gambar 5. Arsitektur Jaringan FFNN Data Tahun 2011

Model yang terbentuk untuk arsitektur tersebut adalah:

$$y = -2,601 + \frac{1,543}{1 + \exp(-(-0,053 - 0,247AMH + 0,27RLS - 0,221PP))} + \frac{2,633}{1 + \exp(-(1,978 + 1,676AMH + 1,674RLS + 0,165PP))} \quad (23)$$

MSE yang dihasilkan oleh model ini adalah sebesar 2,3813.

4.1.3. Multivariate Adaptive Regression Spline (MARS)

Pemodelan angka harapan hidup (AHH) di Provinsi Jawa Timur pada tahun 2007 menggunakan metode MARS menghasilkan nilai GCV terkecil yaitu 3,583 dengan jumlah *basis function* (BF) yang masuk dalam model yaitu 3 *basis function*. Variabel yang masuk ke dalam model berjumlah 2 yaitu angka melek huruf (AMH) dan rata-rata lama sekolah (RLS). Terdapat interaksi antara RLS dan AMH. Model MARS yang terbentuk adalah:

$$y = 69.614 - 0.80682 * BF1 - 0.92143 * BF2 + 1.0582 * BF3 \quad (24)$$

dengan

$$BF1 = \max(0.745 - RLS) * \max(0, 89.8 - AMH)$$

$$BF2 = \max(0, 85.1 - AMH)$$

$$BF3 = \max(0, 7.45 - RLS) * \max(0, 86.6 - AMH)$$

Model ini menghasilkan MSE sebesar 2,16.

Sedangkan pada tahun 2011 pemodelan angka harapan hidup (AHH) di Provinsi Jawa Timur menggunakan metode MARS menghasilkan nilai GCV terkecil yaitu 3,915 dengan jumlah *basis function* yang masuk dalam model yaitu 3 *basis function*. Variabel yang masuk ke dalam model berjumlah 2 yaitu angka melek huruf (AMH) dan rata-rata lama sekolah (RLS). Terdapat interaksi antara kedua variabel yang masuk ke dalam model. MSE yang dihasilkan adalah sebesar 2,36. Model MARS yang terbentuk adalah:

$$y = 70.921 - 0.31973 * BF1 + 0.17006 * BF2 - 3.2157 * BF3 \quad (25)$$

dengan :

$$BF1 = \max(0, 94.98 - AMH)$$

$$BF2 = BF1 * \max(0, 6.33 - RLS)$$

$$BF3 = \max(0, 6.96 - RLS)$$

4.2. Analisis Prediksi

Seperti langkah-langkah pada analisis eksplanasi, model dibentuk dengan menggunakan ketiga metode yang telah disebutkan kemudian model yang terbentuk digunakan untuk memprediksi data yang telah dimissingkan. Banyaknya data yang dimissingkan pada analisis ini adalah 1 data, 2 data, dan 3 data.

Missing 1 : Data Kabupaten Mojokerto

Missing 2 : Data Kabupaten Mojokerto dan Pamekasan

Missing 3 : Data Kabupaten Mojokerto, Pamekasan, dan Kabupaten Malang.

Tabel 3 menyajikan hasil prediksi angka harapan hidup kota yang missing yang diduga dengan menggunakan masing-masing metode.

Tabel 3. Nilai Aktual Angka Harapan Hidup dan Prediksi Masing-Masing Model

Tahun	Missing	Kab/Kota	Aktual	Prediksi		
				Linier	FFNN	MARS
2007	Missing 1	Mojokerto	69,58	69,49	69,59	69,96
	Missing 2	Mojokerto	69,58	69,45	69,78	69,55
		Pamekasan	62,70	64,46	62,95	65,13
	Missing 3	Mojokerto	69,58	69,44	69,78	69,37
		Pamekasan	62,70	64,45	63,01	65,09
		Malang	68,22	67,77	68,40	68,12
2007	Missing 1	Mojokerto	70,56	70,25	70,55	70,92
	Missing 2	Mojokerto	70,56	70,21	70,56	70,42
		Pamekasan	64,33	65,48	64,34	66,03
	Missing 3	Mojokerto	70,56	70,21	70,51	70,37
		Pamekasan	64,33	65,48	64,34	66,66
	Malang	69,35	68,81	69,39	68,51	

Ketepatan hasil prediksi diukur dengan kriteriaan MSEP (Mean Square Error of Prediction). Semakin kecil nilai MSEP maka semakin tepat hasil prediksi dan model yang digunakan semakin baik. Tabel berikut menyantumkan nilai MSEP masing-masing metode.

Tabel 4. Nilai MSEP Masing-Masing Model

Tahun	Data	MSEP		
		Linear	FNN	MARS
2007	Missing 1	0,0074	0,0001	0.1430
	Missing 2	1,5606	0,0513	2.9439
	Missing 3	1,0975	0,0562	1.9140
2007	Missing 1	0,0981	0,0001	0.1561
	Missing 2	0,7193	0,0001	1.4612
	Missing 3	0,5785	0,0014	2.0505

Tabel 4 menunjukkan bahwa metode FFNN memberikan nilai prediksi yang memiliki ketepatan yang sangat tinggi. Hal ini dilihat dari nilai MSEP yang dihasilkan metode ini sangat kecil dan jika dibandingkan dengan metode lain, MSEP yang dihasilkan oleh metode FFNN ini adalah yang paling kecil.

5. Kesimpulan

Dari hasil analisis yang dilakukan, diperoleh bahwa metode FFNN menghasilkan nilai MSEP terkecil sehingga dapat disimpulkan bahwa pemodelan angka harapan hidup (AHH) di Provinsi Jawa Timur berdasarkan angka melek huruf (AMH), rata-rata lama sekolah (RLS), dan Pengeluaran Perkapita (PP) paling baik menggunakan metode FFNN.

Daftar Pustaka

- [1] Drapper, N., R. dan Smith, H.1996. Applied Regression Analysis, 2nd edition. New York: John Wiley & Sons. Chapman and Hall.
- [2] Friedman, J.H. (1991). Multivariate Adaptive Regression Splines (with discussion), Ann Statist. 19, 1-141.
- [3] Lewis, R.J.: An Introduction to Classification And Regression Trees (CART) Analysis. Departement of Emergency Medicine Harbor-UCLA Medical Center, Torrance, California (2000)
- [4] Rawling, J. O., Pentula, S. G., & Dickey, D. A. (1932). Applied Regression Analysis: A Research Tool-2nd ed. New York: Springer-Verlag New York, Inc.
- [5] Warsito, B. dan Ispriyanti, D. 2004. Uji Linearitas Data Time Series dengan Reset Test. Jurnal Matematika dan Komputer, Vol. 7. No. 3, 36 - 44, Desember 2004, ISSN : 1410-8518.
- [6] Warsito, B. 2006. Perbandingan Model Feed Forward Neural Network dan Generalized Regression Neural Network pada Data Nilai Tukar Yen terhadap Dolar AS. Prosiding SPMIPA. pp. 127-131.
- [7] Walpole, Ronald E. 1995. Pengantar Statistika Edisi ke-3. Jakarta: Gramedia Pustaka Utama.