

# Pemodelan Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia Menggunakan Regresi Nonparametrik *Spline* di Jawa Tengah

Ni Putu Dera Yanthi<sup>1</sup> dan I Nyoman Budiantara<sup>2</sup>

Jurusan Statistika, Fakultas MIPA, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

e-mail: <sup>2</sup>i\_nyoman\_b@statistika.its.ac.id dan <sup>1</sup>derayanthi@gmail.com

**Abstrak**—Manusia merupakan salah satu modal dasar yang dimiliki oleh suatu bangsa. Pada tahun 1990, Badan Perserikatan Bangsa-Bangsa (PBB) menetapkan suatu ukuran standar pembangunan manusia yaitu Indeks Pembangunan Manusia (IPM) atau *Human Development Index* (HDI). Pada tahun 2013, Jawa Tengah memiliki nilai IPM sebesar 68,02 yang termasuk dalam kategori menengah atas belum mencapai kategori tinggi yang ditetapkan oleh pemerintah Indonesia. Hal ini menunjukkan bahwa perkembangan pembangunan di provinsi Jawa Tengah belum bisa mencapai target yang ditetapkan pemerintah. Tindakan yang dapat dilakukan pemerintah untuk meningkatkan nilai IPM adalah dengan mengetahui faktor-faktor yang memengaruhi IPM di Jawa Tengah. Metode statistika yang digunakan untuk mengetahui hubungan antara variabel respon dengan variabel prediktor adalah analisis regresi. Pada penelitian ini, pola data IPM dan faktor-faktor yang memengaruhi IPM di provinsi Jawa Tengah dilihat dari *scatterplot* memiliki pola data yang tidak diketahui bentuk polanya sehingga metode yang dapat digunakan adalah regresi nonparametrik *spline*. Model terbaik didapatkan dari titik knot optimal berdasarkan nilai *Generalized Cross Validation* (GCV) terkecil. Berdasarkan penelitian ini, model regresi nonparametrik *spline* terbaik adalah dengan menggunakan kombinasi knot (3,3,2,1,2) dan lima variabel signifikan yaitu tingkat partisipasi angkatan kerja, rasio sekolah-siswa, kepadatan penduduk, angka kesakitan, dan PDRB/1juta setiap kabupaten/kota di Jawa Tengah. Model tersebut memiliki  $R^2$  sebesar 93,14% dan MSE sebesar 6,45564.

**Kata Kunci**—GCV, Indeks Pembangunan Manusia, Regresi Nonparametrik *Spline*, Titik Knot

## I. PENDAHULUAN

Manusia merupakan salah satu kekayaan yang dimiliki oleh suatu bangsa, sehingga kesejahteraan manusia harus diperhitungkan. Pada tahun 1990, UNDP memperkenalkan konsep *Human Development* atau pembangunan manusia sebagai paradigma baru model pembangunan. Untuk dapat ikut berpartisipasi dalam proses pembangunan, tentunya dibutuhkan masyarakat Indonesia yang tidak hanya unggul dari segi kuantitas, tetapi juga unggul dari segi kualitas [1]. Oleh karena itu, pada tahun 1990 Badan PBB menetapkan suatu ukuran standar pembangunan manusia yaitu Indeks Pembangunan Manusia. IPM dibangun melalui tiga dimensi dasar, yaitu kesehatan dengan indikator angka harapan hidup, pengetahuan dengan indikator angka melek huruf dan rata-rata lama sekolah, dan yang terakhir ialah dimensi hidup layak dengan indikator kemampuan daya beli [2].

IPM merupakan relasi antara manusia dengan pembangunan disekitarnya. Sehingga semakin padat penduduk dalam suatu wilayah maka angka IPM harus diperhitungkan. Di Indonesia Pulau Jawa merupakan pulau dengan penduduk yang paling padat, salah satunya adalah Jawa Tengah. Pada tahun 2013 salah satu provinsi yang berada di Pulau Jawa yaitu Jawa Tengah memiliki nilai IPM sebesar 68,02. Angka IPM di Jawa Tengah masih dibawah dari provinsi-provinsi lain yang ada di Pulau Jawa. Hal ini menunjukkan bahwa perkembangan pembangunan di Provinsi Jawa Tengah masih tertinggal jauh. Oleh karena itu perlu diketahui faktor-faktor apa sajakah yang memengaruhi IPM di Jawa Tengah sehingga pemerintah provinsi setempat bisa lebih memerhatikan dan mengupayakan program-program pembangunan manusia guna meningkatkan angka IPM sebagai usaha perbaikan kesejahteraan manusia.

Faktor-faktor pengaruh tersebut dapat diketahui dengan menggunakan metode pemodelan. Salah satu metode yang dapat digunakan untuk menyelesaikan masalah pemodelan adalah analisis regresi. Analisis regresi merupakan suatu metode statistika yang digunakan untuk mengetahui pola hubungan antara variabel respon dengan prediktor [3]. Pendekatan dalam metode analisis regresi ada tiga yaitu pendekatan parametrik, pendekatan nonparametrik dan pendekatan semiparametrik [4]. Metode yang dapat digunakan untuk mengetahui hubungan antara variabel respon dengan prediktor jika kurva regresinya tidak diketahui bentuk polanya adalah regresi nonparametrik [5].

Penelitian sebelumnya tentang IPM sudah banyak dilakukan, beberapa diantaranya yaitu penelitian yang dilakukan oleh Melliana [6] yang menghasilkan faktor-faktor yang berpengaruh terhadap IPM di provinsi Jawa Timur menggunakan regresi panel. Penelitian Awal [7] yang menghasilkan faktor-faktor yang berpengaruh terhadap IPM di provinsi papua menggunakan metode nonparametrik. Penelitian oleh Retno [8] yang menghasilkan faktor-faktor yang berpengaruh terhadap IPM di Jawa Timur.

## II. TINJAUAN PUSTAKA

### A. Regresi Nonparametrik *Spline*

Pendekatan model regresi nonparametrik digunakan jika bentuk kurva regresi tidak diketahui bentuk polanya seperti pada regresi parametrik atau informasi tentang bentuk pola data dimasa lalu tidak lengkap [9]. Bentuk model regresi nonparametrik secara umum disajikan sebagai berikut.

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

dengan  $y_i$  adalah variabel respon, fungsi  $f$  yang tidak diketahui bentuk kurva regresinya dengan  $x_i$  sebagai variabel prediktor dan  $\varepsilon_i$  adalah *error* random yang diasumsikan berdistribusi  $N(0, \sigma^2)$ . Pada model regresi nonparametrik *spline*, kurva regresi dihipotesis dengan fungsi *spline* berorde  $p$  dengan titik knot  $K_1, K_2, \dots, K_r$  yang disajikan dalam bentuk.

$$f(x_i) = \sum_{j=0}^p \gamma_j x_i^j + \sum_{k=1}^r \gamma_{p+k} (x_i - K_k)_+^p \quad (2)$$

dimana  $\gamma$  adalah parameter-parameter model,  $p$  adalah orde polinomial dan  $k$  adalah banyak knot *spline*. Dari persamaan (1) dan (2) dapat diperoleh persamaan regresi seperti berikut.

$$y_i = \sum_{j=0}^p \gamma_j x_i^j + \sum_{k=1}^r \gamma_{p+k} (x_i - K_k)_+^p + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (3)$$

dengan fungsi *truncated* (potongan-potongan) disajikan dalam bentuk.

$$(x_i - K_k)_+^p = \begin{cases} (x_i - K_k)_+^p, & x_i \geq K_k \\ 0, & x_i < K_k \end{cases} \quad (4)$$

Kelebihan regresi nonparametrik *spline* bisa memodelkan data yang perilakunya berubah pada selang tertentu yang tidak mampu dimodelkan oleh parametrik.

### B. Pemilihan Titik Knot Optimal

Banyak metode yang dapat digunakan untuk memilih titik knot optimal dalam estimator *spline*. Salah satu metode untuk memilih titik knot yang optimal yaitu dengan metode GCV (*Generalized Cross Validation*) [10]. Titik knot yang optimal diperoleh dari nilai GCV minimum. Fungsi GCV diberikan oleh.

$$GCV(K_1, K_2, \dots, K_r) = \frac{MSE(K_1, K_2, \dots, K_r)}{(n^{-1} \text{trace}[I - A(K_1, K_2, \dots, K_r)])^2} \quad (5)$$

dengan

$$MSE(K_1, K_2, \dots, K_r) = n^{-1} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2 \quad (6)$$

### C. Estimasi Parameter

Estimasi merupakan proses yang menggunakan sampel statistik untuk menduga atau menaksir hubungan parameter populasi [11]. Untuk mencari estimasi parameter pada penelitian ini menggunakan metode *Ordinary Least Square* (OLS) yang dituliskan dengan rumus sebagai berikut

$$\hat{\gamma} = (X'X)^{-1}X'y \quad (7)$$

### D. Pengujian Parameter Model

#### 1) Pengujian Secara Serentak

Uji serentak dilakukan untuk mengetahui signifikansi parameter model regresi secara bersama-sama. Hipotesis adalah sebagai berikut.

$$H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_{p+r} = 0$$

$$H_1 : \text{minimal ada satu } \gamma_j \neq 0, \quad j = 1, 2, 3, \dots, p+r$$

dimana nilai  $p+r$  adalah jumlah parameter dalam model regresi *spline*,  $p$  adalah derajat polinomial dan  $r$  adalah jumlah knot. Statistik uji yang digunakan adalah sebagai berikut.

$$F_{hitung} = \frac{MS_{regresi}}{MS_{residual}} \quad (8)$$

Daerah penolakan yang digunakan adalah dengan membandingkan nilai  $F_{hitung}$  dengan  $F_{\alpha, (p+r), n-(p+r)-1}$ . Keputusan  $H_0$  ditolak jika  $F_{hitung}$  lebih besar  $F_{tabel}$ . Artinya dapat disimpulkan bahwa minimal terdapat satu parameter pada model regresi *spline* yang signifikan.

#### 2) Pengujian Secara Individu

Pengujian ini bertujuan untuk mengetahui mana variabel yang berpengaruh secara signifikan terhadap model. Hipotesisnya sebagai berikut.

$$H_0 : \gamma_j = 0$$

$$H_1 : \gamma_j \neq 0, \quad j = 1, 2, 3, \dots, p+r$$

Statistik uji sebagai berikut.

$$t_{hitung} = \frac{\hat{\gamma}_j}{SE(\hat{\gamma}_j)}, \quad j = 1, 2, \dots, p+r \quad (9)$$

Sedangkan daerah penolakan  $H_0$  adalah  $|t_{hitung}| \geq t_{(\frac{\alpha}{2}, n-(p+r)-1)}$  atau  $p\text{-value} < \alpha$  yang menunjukkan bahwa parameter signifikan terhadap model.

### E. Pengujian Asumsi Residual

#### 1) Uji Asumsi Identik

Pengujian ini biasanya dilakukan dengan menggunakan uji *glejser*. Uji *glejser* dilakukan dengan meregresikan absolut dari residual dengan variabel prediktornya [12]. Hipotesis yang digunakan adalah sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{p+r} = 0$$

$$H_1 : \text{minimal ada satu } \beta_i \neq 0, \quad i = 1, 2, \dots, p+r$$

Statistik uji yang digunakan adalah sebagai berikut.

$$F_{hitung} = \frac{[\sum_{i=1}^n (|\hat{\varepsilon}_i| - |\bar{\varepsilon}|)^2] / (p+r-1)}{[\sum_{i=1}^n (|\varepsilon_i| - |\bar{\varepsilon}|)^2] / (n-p+r)} \quad (10)$$

Tolak  $H_0$  apabila  $F_{hitung} \geq F_{\alpha, (p+r-1), n-p+r}$ . Apabila gagal tolak  $H_0$  maka nilai absolut residual sama dengan nilai  $\beta_0$ , dimana nilai absolut residual merupakan variansi nilai  $y$  dengan  $\hat{y}$ . Jika nilai variansi sama dengan  $\beta_0$  atau konstan, maka variansi di setiap titik  $x$  sama. Artinya dapat disimpulkan bahwa tidak terdapat indikasi adanya kasus heteroskedastisitas pada residual.

#### 2) Uji Asumsi Independen

Pengujian autokorelasi dilakukan melalui plot *Autocorrelation Function* (ACF). Berikut ini merupakan formula untuk menghitung ACF [13].

$$\hat{\rho}_l = \frac{\hat{\gamma}_l}{\hat{\gamma}_0} = \frac{\sum_{i=1}^{n-l} (e_i - \bar{e})(e_{i+l} - \bar{e})}{\sum_{i=1}^n (e_i - \bar{e})^2}, \quad l = 1, 2, 3 \dots \quad (11)$$

Interval konfidensi  $(1 - \alpha)$  100% untuk autokorelasi  $\rho_l$  diberikan oleh.

$$-t_{(1-\frac{\alpha}{2}, n-1)} SE(\hat{\rho}_l) \leq \rho_l \leq t_{(\frac{\alpha}{2}, n-1)} SE(\hat{\rho}_l) \quad (12)$$

Apabila tidak ada autokorelasi ( $\rho_l$ ) yang keluar dari batas atas maupun batas bawah interval konfidensi maka dapat disimpulkan bahwa asumsi independen telah terpenuhi atau dengan kata lain, tidak ada autokorelasi antar residual. Begitu sebaliknya, bila terdapat autokorelasi ( $\rho_l$ ) yang keluar dari batas interval konfidensi maka dapat disimpulkan asumsi independen tidak terpenuhi.

#### 3) Uji Asumsi Distribusi Normal

Untuk mendeteksi apakah residual telah berdistribusi normal dilakukan dengan menggunakan uji *Kolmogorov-Smirnov* [14]. Hipotesis yang digunakan adalah sebagai berikut.

$$H_0 : F_0(x) = F(x)$$

$$H_1 : F_0(x) \neq F(x)$$

Statistik uji yang digunakan adalah

$$D = \max |F_0(x) - S_N(x)| \quad (13)$$

Daerah penolakan yang digunakan adalah  $H_0$  ditolak apabila  $|D| > q_{(1-\alpha)}$  dengan nilai  $q_{(1-\alpha)}$  didapatkan dari tabel Kolmogorov-Smirnov, sehingga residual tidak berdistribusi normal

### F. Kriteria Pemilihan Model Terbaik

Beberapa kriteria yang digunakan untuk menentukan model regresi terbaik salah satunya ialah menggunakan koefisien determinasi  $R^2$ . Berikut ini merupakan rumus untuk menghitung  $R^2$ .

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \times 100\% \quad (14)$$

Selain dengan  $R^2$ , pemilihan model terbaik juga bisa menggunakan MSE. MSE adalah rata-rata dari kesalahan peramalan dikuadratkan [15].

Berikut ini merupakan rumus untuk menghitung MSE.

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (15)$$

MSE didapatkan dengan menggunakan  $k$ -fold,  $k$ -fold merupakan salah satu metode *Cross Validation*.  $K$ -fold dilakukan untuk membagi data menjadi data training dan data testing, dengan mengulang sebanyak  $k$  untuk membagi sebuah himpunan secara acak menjadi  $k$  subset yang saling bebas.

### G. Indeks Pembangunan Manusia

Menurut UNDP, Indeks Pembangunan Manusia mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup. Sebagai ukuran kualitas hidup, IPM dibangun melalui pendekatan tiga dimensi yaitu, dimensi umur panjang dan sehat, dimensi pengetahuan, dan dimensi kehidupan yang layak [1]. Ketiga dimensi tersebut memiliki pengertian sangat luas karena terkait banyak faktor. Untuk mengukur dimensi kesehatan, digunakan angka harapan hidup waktu lahir. Selanjutnya untuk mengukur dimensi pengetahuan digunakan gabungan indikator angka melek huruf dan rata-rata lama sekolah. Adapun untuk mengukur dimensi hidup layak digunakan indikator kemampuan daya beli.

## III. METODOLOGI PENELITIAN

### A. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari publikasi Statistika Daerah Jawa Tengah 2015, Jawa Tengah dalam angka 2015 dan website Badan Pusat Statistika Jawa Tengah. Data yang digunakan ialah data pada tahun 2013 dengan banyaknya observasi yang digunakan adalah 35 kabupaten/kota di Jawa Tengah.

### B. Variabel Penelitian

Berikut ini adalah variabel yang digunakan pada penelitian ini.

TABEL 1. VARIABEL PENELITIAN

Variabel	Keterangan	Skala
y	Indeks Pembangunan Manusia	Rasio
$x_1$	Tingkat Partisipasi Angkatan Kerja	Rasio
$x_2$	Rasio Sekolah-Siswa	Rasio
$x_3$	Kepadatan Penduduk	Rasio
$x_4$	Angka Kesakitan	Rasio
$x_5$	PDRB/1jt	Rasio

Berdasarkan Tabel 1 yang merupakan variabel yang termasuk pada dimensi pengetahuan ialah rasio sekolah-siswa dimana variabel ini rata-rata jumlah siswa yang dapat belajar di suatu sekolah SMA. Variabel angka kesakitan masuk pada dimensi kesehatan dimana pengertian dari variabel ini adalah persentase penduduk yang mengalami gangguan kesehatan dan terganggu

aktifitasnya sehari-hari. Variabel yang termasuk pada dimensi standar hidup layak ialah tingkat partisipasi angkatan kerja, kepadatan penduduk dan PDRB/1jt.

### C. Langkah-Langkah Penelitian

Langkah-langkah dalam penelitian ini adalah sebagai berikut.

1. Mengumpulkan data Indeks Pembangunan Manusia di Jawa Tengah beserta variabel-variabel yang berpengaruh.
2. Mendeskripsikan karakteristik dari data IPM di Jawa Tengah beserta variabel-variabel yang berpengaruh, menggunakan *scatterplot*. Apabila pola data tidak mengikuti pola tertentu maka metode yang digunakan adalah regresi nonparametrik *spline*.
3. Memodelkan data dengan pendekatan *spline* dengan satu, dua, tiga titik knot dan kombinasi knot.
4. Memilih titik knot optimal berdasarkan nilai *Generalized Cross Validation* (GCV) yang paling minimum.
5. Mendapatkan model regresi *spline* dengan titik knot optimal.
6. Menguji signifikansi parameter regresi *spline* secara serentak dan parsial.
7. Mendeteksi asumsi residual model *spline* terbaik. Apabila residual model *spline* tidak memenuhi asumsi, maka dilakukan transformasi data. Kemudian, memulai kembali dari langkah (2).
8. Memilih model terbaik
9. Menarik kesimpulan.

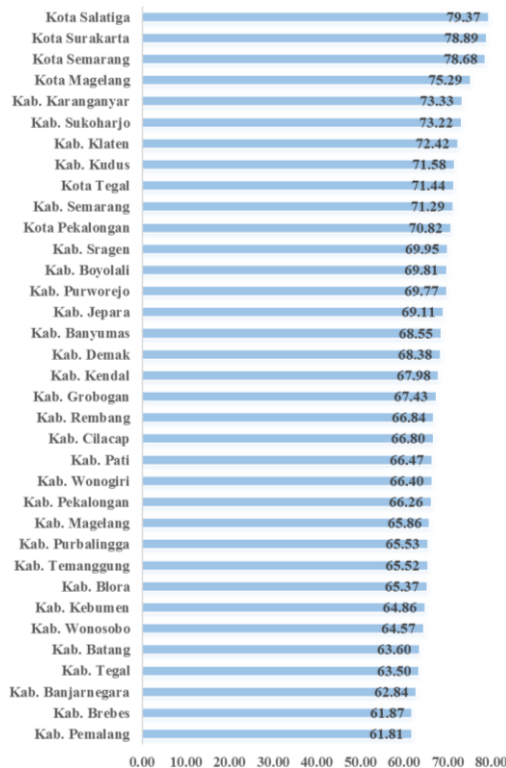
## IV. ANALISIS DAN PEMBAHASAN

### A. Karakteristik Indeks Pembangunan Manusia di Jawa Tengah

Menurut UNDP, IPM mengukur pencapaian hasil pembangunan dari suatu daerah/wilayah dalam tiga dimensi dasar pembangunan yaitu kesehatan, pendidikan, dan standar hidup layak. Setiap tahunnya nilai IPM Jawa Tengah mengalami peningkatan, pada tahun 2013 Jawa Tengah memiliki nilai IPM sebesar 68,02.

Pada Gambar 1 diketahui bahwa Provinsi Jawa Tengah nilai IPM terendah yaitu Kabupaten Pemalang dengan nilai IPM sebesar 61,81. Dilihat dari angka IPM tersebut, berdasarkan kategori IPM yang dikeluarkan oleh PBB, Kabupaten Pemalang termasuk ke dalam IPM menengah bawah (50,0 sampai 65,9). Nilai IPM tertinggi di Jawa Tengah berada pada Kota Salatiga dengan nilai sebesar 79,37. Kota Salatiga termasuk ke dalam IPM menengah atas (66,0 sampai 79,9).

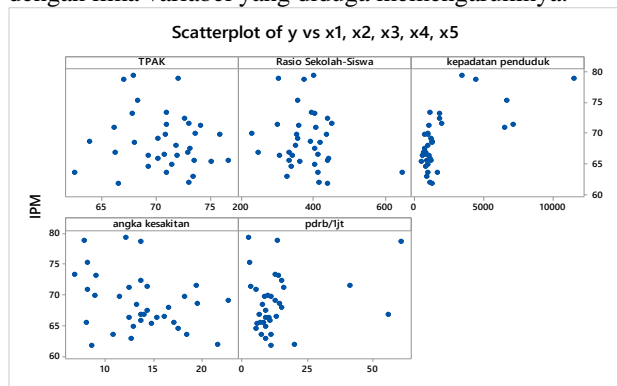
Bila dilihat dari kategori IPM terdapat 11 kabupaten yang masih berada di kategori menengah bawah yaitu Kab. Pemalang, Kab. Brebes, Kab. Banjarnegara, Kab. Tegal, Kab. Batang, Kab. Wonosobo, Kab. Kebumen, Kab. Blora, Kab. Temanggung, Kab. Purbalingga, Kab. Magelang. Nilai IPM Kab. Magelang sebesar 65,86. Sisanya ialah kategori IPM menengah atas dengan 18 kabupaten dan 5 kota. Walaupun sebagian besar kabupaten/kota provinsi Jawa Tengah sudah berada dalam kategori menengah atas, namun belum ada yang mampu menembus kategori tinggi. Dikarenakan pemerintah Indonesia menargetkan nilai IPM berada pada kategori tinggi di setiap provinsi yang ada di Indonesia, maka IPM di Jawa Tengah perlu ditingkatkan.



Gambar 1. IPM di Kabupaten/Kota Provinsi Jawa Tengah

B. Analisis Faktor-faktor yang Diduga Memengaruhi Indeks Pembangunan Manusia

Selanjutnya, dilakukan analisis deskriptif dari variabel-variabel yang diduga berpengaruh terhadap IPM di provinsi Jawa Tengah. Salah satunya ialah membuat *scatterplot* antara variabel respon dengan variabel prediktor agar mengetahui pola hubungan IPM dengan masing-masing variabel prediktor. Gambar 1 adalah *scatterplot* yang menunjukkan pola hubungan IPM (*y*) dengan lima variabel yang diduga memengaruhinya.



Gambar 2 Scatterplot variabel prediktor terhadap respon

Berdasarkan Gambar 2 diketahui bahwa kelima pola data tidak ada yang membentuk suatu pola tertentu. Dengan demikian, dalam pemodelan regresi akan digunakan pendekatan regresi nonparametrik *spline*. Berikut ini adalah karakteristik dari kelima variabel yang diduga memengaruhi IPM provinsi Jawa Tengah

TABEL 2. STATISTIKA DESKRIPTIF VARIABEL PENELITIAN

Variabel	Mean	St. Deviasi	Minimum	Maximum
X <sub>1</sub>	70,672	3,264	62,56	76,7
X <sub>2</sub>	380,4	72,1	227,1	646,9
X <sub>3</sub>	1985	2403	471	11534
X <sub>4</sub>	13,719	4,053	6,880	22,790
X <sub>5</sub>	13,55	13,05	2,28	61,09

Variabel X<sub>1</sub> merupakan variabel tingkat partisipasi angkatan kerja. Pada Tabel 2 diketahui bahwa nilai tertinggi variabel tingkat partisipasi angkatan kerja terdapat di Kabupaten Temanggung sedangkan terendah terdapat di Kabupaten Tegal. Variabel X<sub>2</sub> merupakan variabel rasio sekolah-siswa, diketahui bahwa nilai tertinggi variabel rasio sekolah-siswa terdapat di Kabupaten Batang sedangkan nilai terendah terdapat di Kabupaten Sragen.

Variabel X<sub>3</sub> merupakan variabel kepadatan penduduk. Dalam Tabel 2 diketahui bahwa nilai tertinggi variabel kepadatan penduduk terdapat di Kota Surakarta sedangkan terendah terdapat di Kabupaten Blora. Variabel X<sub>4</sub> merupakan variabel angka kesakitan, diketahui bahwa nilai tertinggi variabel angka kesakitan terdapat di Kabupaten Jepara sedangkan terendah di Kabupaten Karanganyar. Variabel X<sub>5</sub> merupakan variabel PDRB/ljuta, diketahui bahwa nilai tertinggi variabel PDRB/ljuta terdapat di Kota Semarang sedangkan terendah terdapat di Kota Salatiga.

C. Pemilihan Titik Knot Optimal

Titik knot merupakan titik perpaduan dimana terjadi perubahan pola data. Model regresi nonparametrik *spline* terbaik didapatkan dari titik knot yang optimal. Untuk menentukan titik knot yang optimal digunakan metode GCV. Nilai GCV minimum pada pemilihan titik knot optimal dengan satu titik knot, dua titik knot, tiga titik knot, dan kombinasi titik knot ditampilkan sebagai berikut.

TABEL 3. PERBANDINGAN NILAI GCV MINIMUM

Model	GCV Minimum
1 Titik Knot	12,51
2 Titik Knot	9,68
3 Titik Knot	6,36
Kombinasi Knot (3,3,2,1,2)	4,80

Berdasarkan Tabel 3 diketahui bahwa pemodelan yang menghasilkan nilai GCV paling minimum merupakan pemodelan regresi nonparametrik *spline* dengan nilai GCV sebesar 4,80 yaitu kombinasi knot (3,3,2,1,2). Oleh karena itu, diputuskan bahwa model terbaik yang akan dipilih adalah model dengan menggunakan kombinasi knot.

Pemodelan regresi nonparametrik *spline* menggunakan titik knot optimal diperoleh dari nilai GCV paling minimum. Model regresi nonparametrik *spline* terbaik dituliskan sebagai berikut.

$$\hat{y} = 3,699 + 2,561 X_1 - 14,949 (X_1 - 65,73)_+^1 + 39,517 (X_1 - 66,60)_+^1 - 27,005 (X_1 - 66,89)_+^1 - 0,057 X_2 + 0,279(X_2 - 321,31)_+^1 - 0,347(X_2 - 347,01)_+^1 + 0,107(X_2 - 355,58)_+^1 + 0,005 X_3 - 0,047(X_3 - 3631,86)_+^1 + 0,043 (X_3 - 3857,63)_+^1 - 12,918 X_4 + 12,812 (X_4 - 7,20)_+^1 + 0,153 X_5 - 6,385 (X_5 - 19,09)_+^1 + 6,524(X_5 - 20,29)_+^1$$

D. Pengujian Signifikansi Parameter Model Regresi Nonparametrik Spline

1) Uji Serentak

Uji serentak dilakukan pada parameter model regresi terhadap variabel IPM secara bersama-sama atau serentak. Nilai  $\alpha$  yang digunakan sebesar 0,05. Berikut ini adalah hasil uji serentak model regresi nonparametrik *spline*.

TABEL 4. HASIL UJI SERENTAK

Sumber	df	SS	MS	Fhit	p-value
Regresi	16	672,7788	42,04868	15,28829	2,448358e-07
Error	18	49,50693	2,750385		
Total	34	722,2857			

Berdasarkan Tabel 4 dapat diketahui bahwa nilai  $F_{hitung} > F_{(0,05;16,18)}$  yaitu  $15,28829 > 2,249587$  sehingga tolak  $H_0$  yang artinya minimal terdapat satu parameter yang signifikan dalam model. Untuk mengetahui parameter mana yang signifikan dalam model maka dilakukan pengujian secara individu

2) Uji Individu

Uji individu dilakukan untuk mengetahui parameter mana saja yang berpengaruh signifikan terhadap model regresi. Berikut ini adalah hasil pengujian signifikansi parameter model secara individu.

TABEL 5. HASIL UJI INDIVIDU

Var	Parameter	p-value	Keputusan	Kesimpulan
Const	$\gamma_0$	0,001	Tolak	Signifikan
	$\gamma_1$	0,000	Tolak	Signifikan
$X_1$	$\gamma_2$	0,000	Tolak	
	$\gamma_3$	0,000	Tolak	
	$\gamma_4$	0,000	Tolak	
	$\gamma_5$	0,018	Tolak	Signifikan
$X_2$	$\gamma_6$	0,012	Tolak	
	$\gamma_7$	0,211	Gagal	
	$\gamma_8$	0,594	Gagal	
$X_3$	$\gamma_9$	0,000	Tolak	Signifikan
	$\gamma_{10}$	0,000	Tolak	
	$\gamma_{11}$	0,001	Tolak	
$X_4$	$\gamma_{12}$	0,023	Tolak	Signifikan
	$\gamma_{13}$	0,025	Tolak	
$X_5$	$\gamma_{14}$	0,149	Gagal	Signifikan
	$\gamma_{15}$	0,003	Tolak	
	$\gamma_{16}$	0,002	Tolak	

Berdasarkan Tabel 5 diketahui bahwa semua parameter signifikan terhadap model yaitu variabel tingkat partisipasi angkatan kerja ( $X_1$ ), rasio sekolah-siswa ( $X_2$ ), kepadatan penduduk ( $X_3$ ), angka kesakitan ( $X_4$ ), dan PDRB/1juta ( $X_5$ ).

E. Pengujian Asumsi Residual

1) Asumsi Identik

Uji *glejser* dilakukan dengan meregresikan absolut dari residual dengan variabel prediktornya. Berikut ini adalah hasil uji Glejser.

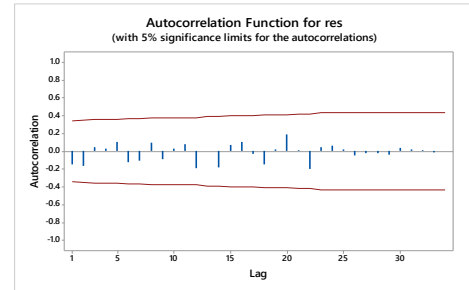
TABEL 6. HASIL UJI GLEJSER

Sumber	df	SS	MS	Fhit	p-value
Regresi	16	8,872724	0,5545452	0,7607888	0,7062689
Error	18	13,12035	0,7289083		
Total	34	21,99307			

Berdasarkan Tabel 4.9 diketahui bahwa nilai  $F_{hitung} < F_{(0,05,16,18)}$  yaitu  $0,7607888 < 2,249587$  maka gagal tolak  $H_0$ , maka disimpulkan bahwa tidak ada heteroskedastisitas atau dengan kata lain variansi antar residual sama.

2) Asumsi Independen

Plot *Autocorrelation Function* (ACF) dapat digunakan untuk mengetahui ada atau tidaknya autokorelasi antar residual

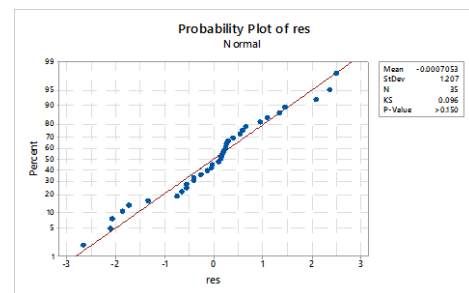


Gambar 3. Plot ACF Residual

Gambar 3 menunjukkan bahwa tidak terlihat adanya autokorelasi yang keluar batas interval konfidensi. Hal ini mengindikasikan bahwa asumsi independen pada residual model telah terpenuhi.

3) Asumsi Distribusi Normal

Pengujian dapat dilakukan dengan uji *Kolmogorov-Smirnov* yang ditampilkan pada gambar sebagai berikut.



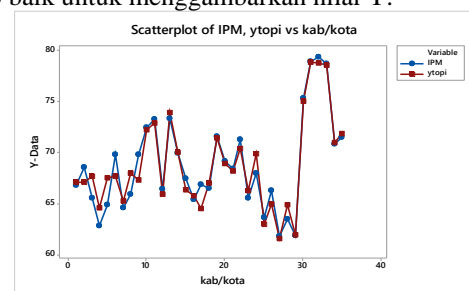
Gambar 4. Hasil Uji Kolmogorov-Smirnov

Berdasarkan Gambar 4 diketahui nilai *Kolmogorov-Smirnov* sebesar 0,096 dimana nilai ini lebih kecil dibandingkan  $q_{(1-\alpha)}$  sehingga gagal tolak  $H_0$ . Hal ini menunjukkan bahwa residual model regresi nonparametrik *spline* telah memenuhi asumsi distribusi normal.

F. Pemilihan Model Terbaik

Dari perhitungan didapatkan  $R^2$  sebesar 93,1458%. Hal ini menunjukkan bahwa model tersebut mampu menjelaskan keragaman nilai IPM Provinsi Jawa Tengah sebesar 93,1458% sedangkan sisanya dijelaskan oleh variabel lain. Selain dengan  $R^2$  cara lain untuk pemilihan model terbaik yaitu MSE, didapatkan dengan menggunakan metode *k-fold*. Dalam perhitungan *k-fold* digunakan  $k=10$ , sehingga didapatkan nilai MSE sebesar 6,45564. Nilai MSE termasuk kecil sehingga dikatakan bahwa model yang diperoleh sudah baik.

Untuk melihat apakah model yang didapatkan sudah cukup menggambarkan nilai sebenarnya, dapat dari Gambar 5, yang menunjukkan nilai variabel respon (Y) dengan nilai yang didapatkan dari model penduga ( $\hat{Y}$ ). Terlihat bahwa nilai  $\hat{Y}_i$  dan  $Y_i$  cenderung berhimpit, hal ini menunjukkan bahwa model yang didapatkan sudah cukup baik untuk menggambarkan nilai Y.



Gambar 5. Scatterplot  $\hat{Y}_i$  dan  $Y_i$

## V. KESIMPULAN DAN SARAN

Pada tahun 2013 Jawa Tengah memiliki nilai IPM sebesar 63,02. Nilai IPM terendah yaitu Kabupaten Pemalang yaitu sebesar 61,81. Nilai IPM tertinggi yaitu Kota Salatiga dengan nilai sebesar 79,37. Berdasarkan kategori IPM yang dikeluarkan oleh PBB, terdapat 11 kabupaten yang masih berada di kategori menengah bawah, dan sisanya berada kategori menengah atas. Belum ada satu kabupaten/kota yang bisa memasuki kategori tinggi, nilai IPM tertinggi di Jawa Tengah pun belum memasuki kategori tinggi.

Model regresi nonparametrik *spline* terbaik untuk pemodelan Indeks Pembangunan manusia di Jawa Tengah adalah dengan menggunakan kombinasi knot 3,3,2,1,2. Model ini mempunyai lima variabel yang berpengaruh secara signifikan yaitu tingkat partisipasi angkatan kerja, rasio sekolah-siswa, kepadatan penduduk, angka kesakitan, PDRB/1juta dengan nilai 6,45564 dan nilai koefisien determinasi ( $R^2$ ) 93,1458%. Hal ini menunjukkan bahwa model tersebut mampu menjelaskan keragaman nilai IPM Provinsi Jawa Tengah sebesar 93,1458% sedangkan sisanya dijelaskan oleh variabel lain.

Beberapa saran yang dapat diberikan oleh penulis yaitu untuk penelitian selanjutnya, sebaiknya menambah jumlah variabel yang digunakan dan diduga berpengaruh terhadap Indeks Pembangunan Manusia (IPM) di Jawa Tengah, sehingga diharapkan akan mendapatkan model yang lebih sesuai. Bagi pemerintah, sebaiknya lebih memerhatikan ketiga dimensi yang terdapat pada Indeks Pembangunan Manusia (IPM). Membuat program-program yang menunjang nilai dari ketiga dimensi yang sekiranya bisa meningkatkan angka Indeks Pembangunan Manusia (IPM) terutama di provinsi Jawa Tengah.

## DAFTAR PUSTAKA

- [1] BPS, Indeks Pembangunan Manusia 2010-2011, Jakarta: Badan Pusat Statistik Indonesia, 2011.
- [2] BPS, Indeks Pembangunan Manusia 2006-2007, Jakarta: Badan Pusat Statistik Indonesia, 2007.
- [3] N. R. Drapper and H. Smith, Analisis Regresi Terapan, Jakarta: Gramedia Pustaka Utama, 1992.
- [4] I. N. Budiantara, "Model Keluarga Spline Polinomial Truncated Dalam Regresi Semiparametrik," *Makalah Seminar Nasional Matematika*, pp. Jurusan Matematika Universitas Diponegoro, Semarang, 2005.
- [5] R. Eubank, Nonparametric Regression and Spline Smoothing, New York: Marcel Dekker Inc, 1999.
- [6] A. Melliana, Analisis Statistika Faktor yang Mempengaruhi IPM di kabupaten/kota provinsi Jawa Timur dengan Menggunakan regresi panel, Surabaya: Institut Teknologi Sepuluh Nopember, 2013.
- [7] C. N. P. Awal, Pendekatan Regresi Nonparametrik Spline untuk Memodelkan Indeks Pembangunan Manusia (IPM) di Papua, Surabaya: Institut Teknologi Sepuluh Nopember, 2014.
- [8] A. T. Retno, Faktor-Faktor yang mempengaruhi Indeks Pembangunan Manusia di Jawa Timur dengan Pendekatan Regresi Semiparametrik Spline, Surabaya: Institut Teknologi Sepuluh Nopember, 2014.
- [9] R. Eubank, Spline Smoothing and Nonparametric Regression, New York: Marcel Dekker Inc, 1988.
- [10] G. Wahba, Spline Models for Observation Data, SIAM Pennsylvania, 1990.
- [11] I. Hasan, Pokok-Pokok Materi Teori Pengambilan Keputusan, Jakarta: Ghalia Indonesia, 2002.
- [12] D. N. Gujarati, Basic Econometrics 4th edition, New York: McGraw-Hill Inc, 2003.
- [13] W. W. Wei, Time Series Analysis Univariate and Multivariate Methods, United States: Pearson Education, 2006.
- [14] W. W. Daniel, Statistika Nonparametrik Terapan, Jakarta: PT. Gramedia, 1989.
- [15] Subagyo, Forecasting Konsep dan Aplikasi, Yogyakarta: BPFE, 1986.
- [16] L. Fu, Neural Networks In Computer Intelligence, Singapore: McGraww-Hill, 1994.