

PEMILIHAN MODEL REGRESI TERBAIK DENGAN
MENGUNAKAN METODE STEPWISE
(Studi Kasus: Data IPM Indonesia Tahun 2020)

*Selection of the Best Regression Model
Using Stepwise Method
(Case Study: Indonesia HDI Data in 2020)*

S. M. Rumanama¹, E. R. Persulesy^{2*}, Z. A. Leleury³

^{1,2,3}Program Studi Matematika, FMIPA Universitas Pattimura

Jl. Ir. M. Putuhena, Ambon, 97233, Maluku, Indonesia

e-mail: ²richardcalvin89@gmail.com

Abstrak

Analisis regresi linier berganda difokuskan pada pemilihan model regresi terbaik yang berpengaruh signifikan dan tidak terdapat multikolinieritas. Metode stepwise merupakan salah satu metode yang dapat digunakan sebab prosedur yang diterapkan adalah menentukan variabel prediktor yang dominan mempengaruhi variabel respon setelah memenuhi kriteria dan kemudian akan dimasukkan kedalam model regresi sebagai peubah. Penelitian ini bertujuan untuk mengetahui faktor-faktor yang berpengaruh signifikan terhadap IPM Indonesia tahun 2020 dan mengetahui model regresi terbaik menggunakan metode stepwise. Hasil penelitian ini menunjukkan bahwa model regresi linier berganda dengan metode stepwise diketahui Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Perkapita berpengaruh signifikan terhadap IPM Indonesia dengan nilai R-Square sebesar 93,8%.

Kata Kunci: IPM, Metode Stepwise, Regresi.

Abstract

Multiple linear regression analysis is focused on selecting the best regression model that has a significant effect and there is no multicollinerity. The stepwise method is one of the methods used because the procedure applied is to determine the dominant predictor variable influencing the response variable after meeting the criteria and then it will be entered into the regression model as a variable. This study aims to determine the factor that have a significant effect on Indonesia's HDI in 2020 and find out the best regression model using the stepwise method. The results of this study indicate that the multiple linear regression model using the stepwise method is know to have live expectancy expected length of schooling, haverage length of schooling, and percapita expenditure have a significant effect on Indonesia's HDI with an R-Square value of 93,8%.

Keywords: HDI, Stepwise Method, Regression



<https://doi.org/10.30598/parameter.v1i1pp69-78>



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

1. PENDAHULUAN

Pembangunan manusia dalam suatu negara diperlukan untuk menciptakan kesejahteraan dan membuat perkembangan dalam suatu negara sehingga perlu dilihat terlebih dahulu kualitas sumber daya manusia dalam negara tersebut. Salah satu tujuan nasional Bangsa Indonesia yakni memajukan kesejahteraan umum. Untuk mencapai tujuan tersebut maka salah satu usaha yang dilakukan adalah pembangunan nasional. Pembangunan perlu dilakukan secara menyeluruh dan berkelanjutan untuk meningkatkan kesejahteraan dan kualitas hidup masyarakat. Salah satu indikator yang dapat mengukur keberhasilan pembangunan nasional yaitu Indeks Pembangunan Manusia (IPM). IPM merupakan suatu pengukuran capaian pembangunan manusia yang didasarkan pada beberapa komponen dasar kualitas hidup, antar lain umur panjang, pengetahuan dan standar kehidupan yang layak [1]. Melalui ketiga indikator tersebut diharapkan akan terjadi peningkatan kualitas hidup manusia. Indeks Pembangunan Manusia (IPM) di Indonesia pada tahun 2020 tercatat sebesar 71,94. Angka ini menunjukkan bahwa Pembangunan Indonesia mengalami peningkatan sebesar 0,03 persen (meningkat 0,02 poin) dibandingkan tahun 2019 yaitu 71,92.

Untuk membuat pemodelan IPM diperlukan suatu analisis statistika, salah satu analisis statistika untuk memodelkan angka IPM yaitu regresi. Penelitian IPM tentang regresi telah dilakukan oleh Loklomin pada tahun 2019, yaitu Pemodelan Indeks Pembangunan Manusia di Kepulauan Maluku Dengan Pendekatan Estimasi Interval Parameter Model Regresi Semiparametrik *Spline Truncataed*. Hasil penelitian tersebut menunjukkan bahwa kelompok usia sekolah di Perguruan Tinggi, jumlah sarana kesehatan, persentase rumah tangga dengan akses air bersih, persentase penduduk miskin dan persentase tingkat partisipasi angkatan kerja memiliki pengaruh terhadap IPM di Kepulauan Maluku [2].

Pemodelan IPM dengan tujuan mengetahui faktor-faktor yang mempengaruhinya juga dapat dilakukan menggunakan analisis regresi linier berganda. Pada pembentukan model regresi linier berganda seringkali muncul masalah multikolinieritas, yaitu terdapat hubungan linier yang sempurna atau korelasi yang tinggi antara variabel prediktor. Dalam hal multikolinieritas ada beberapa cara untuk mendapatkan model terbaik. Pemilihan model regresi terbaik dapat dilakukan dengan menggunakan metode *stepwise*. Proses metode *stepwise* dilakukan dengan memasukan satu persatu variabel prediktor yang memiliki nilai koefisien korelasi tertinggi ke dalam model.

Penelitian tentang penggunaan metode *stepwise* telah dilakukan oleh Sanlly J. Latupeirissa dan Ronald J. Djami pada tahun 2020 dengan memodelkan IPM kota Ambon. Hasil peneltian menunjukkan bahwa pemodelan IPM kota Ambon menggunakan metode *stepwise* menghasilkan model terbaik, yaitu model regresi linear berganda yang tidak terdapat masalah multikolinieritas[3]. Metode *stepwise* ini juga telah diterapkan dalam penelitian terkait penentuan hasil jagung [4], pemodelan Impor Beras di Sulewesi Utara[5], dan pemodelan jumlah penduduk miskin di provinsi maluku[6].

Berdasarkan uraian di atas, maka peneliti tertarik untuk melakukan penelitian dengan tentang pemilihan model regresi terbaik menggunakan metode *stepwise*. Data yang digunakan pada penelitian ini adalah data IPM Indonesia tahun 2020.

2. METODE PENELITIAN

2.1 Tipe Penelitian

Tipe penelitian yang digunakan adalah studi kasus yaitu dengan menerapkan metode *stepwise* untuk membuat model regresi linier pada Indeks Pembangunan Manusia (IPM).

2.2 Data dan Sumber Penelitian

Adapun data yang digunakan pada penelitian ini berupa data sekunder yang diperoleh dari BPS Indonesia yaitu Data Indeks Pembangunan Manusia dan faktor-faktor yang diduga mempengaruhi IPM Indonesia tahun 2020.

2.3 Prosedur Penelitian

Penelitian ini dilakukan dengan prosedur atau Langkah-langkah penelitian sebagai berikut:

- Menentukan judul penelitian.
- Mencari bahan dan materi tentang regresi, IPM, Metode *Stepwise* dan aplikasinya
- Melakukan analisis deskriptif terhadap variabel penelitian
- Melakukan uji Korelasi untuk pemilihan model regresi dengan SPSS
- Implementasi Program untuk uji Asumsi klasik regresi berganda (uji autokorelasi, uji multikolinieritas, uji heteroskedastisitas dan uji normalitas) dari model regresi yang terbentuk.
- Menentukan model regresi terbaik dan menentukan nilai koefisien determinasi.
- Menyimpulkan hasil penelitian.

2.4 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari 4 variabel prediktor yaitu angka harapan hidup (X1), rata-rata lama sekolah (X2), harapan lama sekolah (X3), dan pengeluaran per kapita (X4). Sedangkan variabel respon yang digunakan dalam penelitian ini yaitu indeks prestasi manusia (Y).

3. HASIL DAN PEMBAHASAN

3.1 Statistik Deskriptif

Tabel 1 menyajikan statistik deskriptif terhadap data Indeks Pembangunan Manusia di Indonesia pada tahun 2020 serta faktor-faktor yang mempengaruhinya [7]. Berdasarkan pada Tabel 1 berikut terlihat bahwa nilai rata-rata IPM Indonesia pada tahun 2020 adalah sebesar 71,08 dengan nilai IPM terendah sebesar 60,44 dan IPM terbesar adalah 80,77.

Tabel 1. Statistik Deskriptif Data Indeks Pembangunan Manusia

	N	Minimum	Maximum	Mean	Std. Error	Std. Deviation	Variance
	Statistic	Statistic	Statistic	Statistic		Statistic	Statistic
Y	34	60,44	80,77	71,08	0,669	3,902	15,225
X1	34	65,06	74,99	70,57	0,642	3,746	14,029
X2	34	11,08	15,59	13,14	0,128	0,746	0,557
X3	34	6,69	11,13	8,65	0,159	0,928	0,862
X4	34	6,95	18,23	10,69	0,375	2,186	4,780

3.2 Pemodelan IPM Dengan Metode *Stepwise*

a. Seleksi Variabel Prediktor Pertama

Pada [Tabel 2](#) terlihat bahwa dengan menggunakan metode korelasi pearson diperoleh variabel prediktor pertama yaitu pendapatan perkapita (X_4) karena memiliki nilai korelasi tertinggi yaitu sebesar 0,866 dan signifikan dengan nilai p-value 0,000 yang kurang dari nilai kritis 0,05. Selanjutnya adalah meregresikan variabel respon (Y) dengan variabel prediktor dengan hasil seperti yang disajikan pada [Tabel 3](#).

Tabel 2. Ringkasan Pengujian Korelasi Pearson

Y Dengan X_n	Nilai Korelasi	p-value
Y Dengan X_1	0,424	0,013
Y Dengan X_2	0,497	0,003
Y Dengan X_3	0,781	0,000
Y Dengan X_4	0,866	0,000

Tabel 3. Regresi Variabel Respon Dengan Variabel Prediktor Pertama

	Model	Koefisien (B)	Std. Error	t	p-value
1	(Constant)	54,575	1,722	31,684	0,000
	X4	1,545	0,158	9,775	0,000

Pada [Tabel 3](#) terlihat bahwa hasil regresi menunjukkan variabel prediktor (X_4) signifikan dengan nilai p-value 0,000 sehingga model yang diperoleh untuk hasil seleksi variabel pertama adalah :

$$Y = 54,575 + 1,545 X_4$$

b. Seleksi Variabel Prediktor Kedua

Variabel prediktor kedua yang dimasukkan adalah variabel prediktor dengan nilai korelasi parsial tertinggi dan masih signifikan terhadap variabel respon. Hasil pengujian korelasi parsial disajikan pada [Tabel 4](#) berikut.

Tabel 4. Korelasi Parsial

Variabel Control			Y	X_1	X_2	X_3
X4	Y	Correlation	1,000	0,500	0,720	0,693
		Significance (2-tailed)		0,003	0,000	0,000
		Df	0	31	31	31
X1		Correlation	0,500	1,000	0,234	0,277
		Significance (2-tailed)	0,003		0,190	0,118
		Df	31	0	31	31
X2		Correlation	0,720	0,234	1,000	0,469
		Significance (2-tailed)	0,000	0,190		0,006
		Df	31	31	0	31

X3	Correlation	0,693	0,277	0,469	1,000
	Significance (2-tailed)	0,000	0,118	0,006	
	Df	31	31	31	0

Berdasarkan [Tabel 4](#) terlihat bahwa hasil pengujian korelasi parsial dengan X_4 sebagai variabel kontrol menunjukkan bahwa variabel prediktor X_2 memiliki nilai korelasi tertinggi diantara variabel prediktor lainnya. Setelah didapatkan variabel prediktor kedua selanjutnya dilakukan evaluasi dengan meregresikan variabel respon (Y) dengan variabel prediktor (X_4) dan variabel prediktor yang baru masuk yaitu X_2 .

Tabel 5. Regresi antara Variabel Respon, Variabel kontrol & Variabel Prediktor Kedua

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	30,616	4,327		7,076	0,000
	X4	1,438	0,113	0,806	12,731	0,000
	X2	1,910	0,331	0,365	5,770	0,000

Berdasarkan [Tabel 5](#), terlihat bahwa variabel prediktor X_4 dan variabel prediktor X_2 dinyatakan signifikan mempengaruhi variabel respon (Y) karena nilai p-value 0,000 kurang dari nilai kritis 0,05 sehingga variabel X_2 tetap dimasukkan ke dalam model. Adapun model regresi yang diperoleh untuk hasil seleksi variabel prediktor kedua adalah:

$$Y = 30,616 + 1,438 X_4 + 1,910 X_2$$

c. Seleksi Variabel Prediktor Ketiga

Variabel prediktor ketiga yang dimasukkan kedalam model regresi adalah variabel prediktor dengan nilai korelasi tertinggi dan masih signifikan terhadap variabel respon, dimana sebagai variabel kontrol. Hasil uji korelasi pearson diperoleh hasil seperti yang disajikan pada [Tabel 6](#).

Tabel 6. Seleksi Variabel Prediktor Ketiga Dengan X_4 dan X_2 Sebagai Variabel kontrol

Control Variables		Y	X_1	X_3
X4 & X2	Y	1,000	0,492	0,579
	Correlation		0,004	0,001
	Significance (2-tailed)			
	Df	0	30	30
X1	Y	0,492	1,000	0,195
	Correlation	0,004		0,284
	Significance (2-tailed)			
	Df	30	0	30
X3	Y	0,579	0,195	1,000
	Correlation	0,001	0,284	
	Significance (2-tailed)			
	Df	30	30	0

Berdasarkan **Tabel 6** terlihat bahwa hasil pengujian korelasi parsial dengan variabel prediktor X_4 dan X_2 sebagai variabel kontrol menunjukkan bahwa variabel prediktor X_3 memiliki nilai korelasi tertinggi yaitu sebesar 0,759 dan signifikan terhadap variabel respon dengan nilai signifikansi 0,001. Setelah variabel prediktor ketiga diperoleh, maka selanjutnya dilakukan evaluasi dengan meregresikan variabel respon Y dengan variabel kontrol X_4 dan X_2 serta variabel yang baru masuk yaitu X_3 .

Berdasarkan **Tabel 7** terlihat bahwa hasil regresi menunjukkan variabel X_2 , X_3 dan X_4 signifikan mempengaruhi variabel respon Y karena memiliki nilai p-value kurang dari 0,05 sehingga model regresi yang diperoleh untuk hasil seleksi variabel prediktor ketiga adalah:

$$Y = 30,649 + 1,343X_2 + 1,171X_3 + 1,184X_4$$

**Tabel 7. Regresi Variabel Respon Dengan Variabel kontrol X_4 dan X_2
Dengan Variabel Ketiga X_3**

Model		Unstandardized		t	p-value
		Coefficients	Std. Error		
		B			
1	(Constant)	30,649	3,585	8,549	0,000
	X2	1,343	0,311	4,325	0,000
	X3	1,171	0,301	3,892	0,001
	X4	1,184	0,114	10,372	0,000

d. Seleksi Variabel Prediktor Keempat

Variabel prediktor keempat yang dimasukkan kedalam model regresi adalah variabel prediktor tersisa yakni X_1 . Berdasarkan **Tabel 8** terlihat bahwa hasil pengujian korelasi parsial dengan variabel prediktor X_4 , X_2 dan X_3 sebagai variabel kontrol menunjukkan bahwa variabel prediktor X_1 memiliki nilai signifikansi 0,007 yang kurang dari 0,05 sehingga variabel prediktor X_1 juga dimasukkan kedalam model regresi.

**Tabel 8. Seleksi Variabel Prediktor Keempat dengan X_4 , X_2 dan X_3
Sebagai Variabel kontrol**

X4 X2 & X3	Y	Correlation	1,000	0,474
		Significance (2-tailed)		0,007
		Df	0	29
X1		Correlation	0,474	1,000
		Significance (2-tailed)	0,007	
		Df	29	0

Setelah variabel prediktor keempat didapatkan yaitu variabel X_1 , maka selanjutnya dilakukan evaluasi dengan meregresikan variabel respon Y dengan variabel kontrol X_4 , X_2 , dan X_3 serta variabel yang baru masuk yaitu X_1 .

Tabel 9. Regresi Variabel Respon Dengan Variabel kontrol (X_4 , X_2 dan X_3)
Dengan Variabel prediktor Keempat X_1

Model		Unstandardized		t	p-value.
		Coefficients			
		B	Std. Error		
1	(Constant)	22,819	4,197	5,436	0,000
	X2	1,244	0,280	4,437	0,000
	X3	1,016	0,275	3,696	0,001
	X4	1,175	0,102	11,481	0,000
	X1	0,150	0,052	2,897	0,007

Berdasarkan **Tabel 9** terlihat bahwa hasil regresi menunjukkan variabel X_2 , X_3 , X_4 dan X_1 signifikan mempengaruhi variabel respon Y karena memiliki nilai p-value kurang dari 0,05 sehingga model regresi yang diperoleh untuk hasil seleksi variabel prediktor keempat adalah:

$$Y = 22,819 + 0,150X_1 + 1,244X_2 + 1,016X_3 + 1,175X_4$$

Karena koefisien regresi pada semua variabel prediktor bernilai positif atau memiliki hubungan yang searah mengakibatkan kondisi bahwa jika persentase umur harapan hidup saat lahir (X_1), harapan lama sekolah (X_2), rata-rata lama sekolah (X_3), dan pendapatan perkapita (X_4) mengalami peningkatan maka persentase Indeks Pembangunan Manusia di Indonesia juga ikut meningkat. Setelah model regresi didapatkan maka dilakukan analisis variansi dari model yang diperoleh.

Tabel 10. Analisis Variansi

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	471,092	4	117,773	109,041	.000 ^b
	Residual	31,322	29	1,080		
	Total	502,414	33			

Berdasarkan **Tabel 10** dapat dilihat bahwa nilai F hitung sebesar 109,041 dan p-value sebesar 0,000. Dengan demikian dapat dikatakan bahwa model yang diperoleh sesuai karena nilai p-value kurang dari 0,05. Adapun nilai koefisien determinasi yang diperoleh sebesar 93,8%. Hal ini menunjukkan bahwa besarnya pengaruh variabel prediktor X_1 , X_2 , X_3 dan X_4 terhadap variabel Y adalah 93,8% sisanya 6,2% dipengaruhi oleh variabel lainnya.

e. Pengujian Asumsi Residual

Model regresi akan dijadikan alat estimasi yang baik dan tidak bias jika telah dilakukan beberapa pengujian asumsi residual.

a. Uji Heteroskedastisitas.

Pengujian Heteroskedastisitas dilakukan dengan menggunakan uji glejser

Tabel 11. Uji Glejser

Model	t	Sig.
(Constant)	-0,674	0,505
X1	1,800	0,082
X2	-1,265	0,216
X3	0,671	0,508
X4	-1,271	0,214

Berdasarkan Tabel 11 terlihat bahwa nilai sig. dari variabel prediktor lebih dari taraf signifikansi 0,05. Hal ini menunjukkan bahwa variabel prediktor tidak signifikan mempengaruhi residual. Sehingga tidak terdapat kasus heteroskedastisitas.

b. Uji Autokorelasi

Pengujian ini bertujuan untuk mengetahui korelasi antara residual periode saat ini pada periode selanjutnya. Model regresi yang baik adalah yang tidak terdapat autokorelasi, nilai d_L (batas atas) dan nilai d_U (batas bawah) dapat diperoleh dari tabel *Durbin-Watson*, dengan nilai $n = 34$ dan $k = 4$ didapat $d_L = 1,01$ dan $d_U = 1,51$ sehingga nilai $4-d_U = 2,49$ dan $4-d_L = 2,99$.

Diketahui dari Tabel 12 nilai *Durbin-Watson* sebesar 1,625, terletak diantara $d_U < DW < 4-d_U$ atau $1,51 < 1,625 < 2,49$, maka hipotesis nol diterima yang berarti tidak terjadi autokorelasi.

Tabel 12. Hasil Uji Durbin-Watson

	Unstandardized Residual
Test Statistic	0,106
Asymp. Sig. (2-tailed)	.200

c. Uji Normalitas

Uji normalitas dalam suatu model regresi merupakan salah satu uji asumsi klasik yang bertujuan untuk mendeteksi residual berdistribusi normal atau tidak.

Tabel 13. Uji Kolmogorov-Smirnov

Model	R	R Square	Durbin-Watson
1	.971 ^a	0,943	1,625

Berdasarkan Tabel 13 terlihat bahwa nilai Asymp.sig.(2-tailed) dari uji kolmogorov-smirnov lebih dari taraf signifikan 0,05. Hal ini menunjukkan bahwa residual berdistribusi normal.

d. Uji Multikolinieritas

Untuk mendeteksi ada atau tidaknya masalah multikolinieritas pada suatu model dapat digunakan nilai *Variance Inflation Vector* (VIF).

Tabel 14. Nilai VIF Untuk Deteksi Multikolinearitas

Variabel	VIF
X_1	1,149
X_2	1,337
X_3	1,990
X_4	1,529

Berdasarkan Tabel 14 terlihat bahwa nilai VIF dari semua variabel prediktor kurang dari 10 maka dapat dikatakan bahwa tidak terjadi multikolinearitas antara variabel prediktor.

Dari hasil pengujian asumsi terlihat bahwa model yang dihasilkan telah memenuhi asumsi-asumsi yang ada, sehingga dapat disimpulkan bahwa model yang diperoleh dapat digunakan untuk menggambarkan persentase indeks pembangunan manusia di Indonesia.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan maka diperoleh kesimpulan yaitu pemodelan regresi terbaik terhadap Indeks Pembangunan Manusia (IPM) Indonesia dengan menerapkan metode *stepwise* melibatkan umur harapan hidup saat lahir (X_1), harapan lama sekolah (X_2), rata-rata lama sekolah (X_3), dan pendapatan perkapita (X_4) Model regresi linier adalah $Y = 22,819 + 0,150X_1 + 1,244X_2 + 1,016X_3 + 1,175X_4$ dengan nilai R-Square sebesar 93,8%.

DAFTAR PUSTAKA

- [1] Pratiwi, W. 2020. Kuliah Pengantar : Indeks Pembangunan Sub Bidang Pembangunan Pedesaan Di Program Studi Arsitektur, ITB.
- [2] Loklomin, S. B. 2019. *Pemodelan Indeks Pembangunan Manusia di Kepulauan Maluku Dengan Pendekatan Estimasi Interval Parameter Model Regresi Semiparametrik Sline Truncated* Barekeng, Vol 13, No 2.
- [3] Latupeirissa, S. J., & Djami, R. J. 2020. Pemodelan Indeks Pembangunan Manusia (IPM) Kota Ambon Menggunakan Metode Stepwise. *VARIANCE: Journal of Statistics and Its Applications*, 2(1), 45-52.
- [4] Nining N. dkk . 2016. Aplikasi Model Regresi Stepwise Dalam Penentuan Hasil Jagung Putih. *Informatika Pertanian*, Vol. 25, No.1, Juni 2016 : 21-28.
- [5] Selfina. C. W et al. 2017. Penentuan Model Regresi Terbaik Dengan Menggunakan Metode Stepwise (Studi kasus: Impor Beras di Sulawesi Utara). *Jurnal Ilmiah Sains*, Vol.17, N0.2.
- [6] Lalin, D. 2018. Metode Stepwise Untuk Memodelkan Jumlah Penduduk Miskin di Provinsi Maluku. Skripsi Jurusan Matematika FMIPA Unpatti, Ambon.
- [7]. BPS. 2021. *Indeks Pembangunan Manusia*. aJkarta : Badan Pusat Statistik

