

# IMPLEMENTASI METODE EXTREME LEARNING MACHINE (ELM) UNTUK MEMPREDIKSI INDEKS PEMBANGUNAN MANUSIA PROVINSI DKI JAKARTA

Bagus Arief Aditiya<sup>1)</sup>, Didit Widiyanto<sup>\*2)</sup>, Noor Falih<sup>\*3)</sup>

Informatika, Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

Bagusariefaditiya@gmail.com<sup>1)</sup>, Didit.widiyanto89@gmail.com<sup>\*2)</sup>, falih@upnvj.ac.id<sup>\*3)</sup>

**Abstrak.** Indeks Pembangunan Manusia (IPM) adalah tolok ukur untuk menghitung pencapaian pembangunan masyarakat di wilayah tertentu. IPM dibentuk berdasarkan 3 indikator dasar yaitu, pengetahuan, standar hidup layak dan kesehatan. Hal ini menyebabkan perhitungan IPM oleh Badan Pusat Statistik (BPS) membutuhkan waktu yang cukup lama ditambah masih harus mempublikasikan hasil perhitungan IPM tiap tahunnya, sehingga dibutuhkan prediksi nilai IPM tersebut untuk mempersingkat waktu. Dari permasalahan tersebut penulis ingin membuat model yang dapat digunakan untuk memprediksi nilai IPM dengan mengimplementasikan metode *Extreme Learning Machine* (ELM). Data yang digunakan akan melalui tahap pra-proses mean substitution dan normalisasi menggunakan metode z-score. Selanjutnya akan dicari performa terbaik dari 8 skenario yang telah dirumuskan, performa terbaik ditentukan berdasarkan nilai RMSE yang didapatkan tiap skenario. Hasil terbaik yang didapatkan dari 8 skenario penelitian yang dibuat adalah hasil dari skenario Y3. RMSE yang didapatkan sebesar 0,025355 pada proses pengujian.

**Kata kunci:** Indeks Pembangunan Manusia, Jaringan Saraf Tiruan, ELM

## 1. Pendahuluan

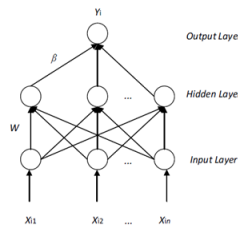
IPM adalah indeks yang digunakan sebagai alat ukur untuk menghitung tingkat kesuksesan pemerintah membangun SDM di suatu daerah, diterbitkan oleh *United Nations Development Programme* (UNDP) di tahun 1990 dan dipublikasikan dalam laporan tahunan *Human Development Report* (HDR) secara berkala. IPM terdiri dari 3 indikator yaitu kesehatan, pengetahuan, dan standar hidup layak. Di DKI Jakarta sendiri IPM merupakan data penting karena selain sebagai taraf ukur kinerja pemerintah, IPM juga digunakan sebagai salah satu untuk menentukan Dana Alokasi Umum (DAU). Untuk proses perhitungan IPM memakan waktu cukup lama, dikarenakan IPM terdiri dari 3 indikator dasar yang berbeda sehingga Badan Pusat Statistik membutuhkan waktu yang lama untuk melakukan perhitungan dan mempublikasikan hasil perhitungan. Untuk membantu untuk mempersingkat waktu dalam perhitungan IPM di masa depan dapat melakukan prediksi (*forecast*). Memperkirakan apa yang akan terjadi di masa mendatang adalah prediksi, artinya prediksi bukan hanya sekedar perkiraan akan tetapi menggunakan metode-metode tertentu hingga menjadi lebih dari sekedar perkiraan. Berdasarkan hal yang disampaikan diatas, penulis melihat suatu masalah yaitu bagaimana memprediksi nilai IPM. Oleh sebab itu penulis ingin membuat model prediksi yang baik dengan mengimplementasikan metode ELM. Menggunakan metode ELM karena metode ini memiliki akurasi dan *learning speed* yang baik berdasarkan penelitian sebelumnya yang serupa dengan penelitian ini. Adapun tujuan yang ingin dicapai dalam penelitian ini adalah menghasilkan model prediksi dengan performa yang baik dari beberapa skenario penelitian yang dijalankan menggunakan pembelajaran Jaringan Syaraf Tiruan (JST) dengan Metode Extreme Learning Machine yang diaplikasikan pada Indeks Pembangunan Manusia Provinsi DKI Jakarta.

## 2. Landasan Teori

Jaringan Saraf Tiruan (JST) adalah sistem pemroses informasi yang memiliki karakteristik mirip dengan jaringan saraf biologi[5]. JST dibentuk sebagai generalisasi model matematika dari jaringan saraf biologi, dengan asumsi bahwa pemrosesan informasi terjadi pada banyak elemen sederhana (neuron), sinyal dikirimkan diantara neuron-neuron melalui penghubung-penghubung, penghubung antar neuron memiliki bobot yang akan memperkuat atau memperlemah sinyal, untuk menentukan output, setiap neuron menggunakan fungsi aktivasi (biasanya bukan fungsi linear) yang dikenakan pada jumlah input yang diterima.

### 2.1. Extreme Learning Machine

Extreme Learning Machine (ELM) adalah bagian dari Jaringan Saraf Tiruan (JST). ELM termasuk kedalam Jaringan Saraf Tiruan *feedforward* dengan *single hidden layer* atau disebut dengan Single Hidden Layer Feedforward Networks [1]. Untuk asitektur ELM dapat dilihat pada Gambar 1, input layer, output layer dan hidden layer merupakan 3 layer yang dimiliki ELM. Satu buah hidden layer merupakan ciri-ciri dari algoritma ELM.



Gambar 97 Arsitektur

Proses prediksi dengan ELM melalui beberapa tahap yaitu data di normalisasi kemudian dilakukan proses *training* dan *testing* selanjutnya dilakukan denormalisasi untuk membangkitkan nilai asli dari data yang sudah di normalisasi, setelah itu nilai error dihitung dengan metode *Root Mean Square Error (RMSE)*.

Normalisasi adalah proses penskalaan nilai dari atribut data sehingga dapat memiliki range yang sama. Berikut persamaan untuk menghitung *Z-score Normalization*.

$$X' = ((x - \text{mean}) / (\text{std.dev})) \quad (1)$$

$X'$  merupakan hasil dari normalisasi, sedangkan  $x$  adalah nilai asli data yang digunakan,  $\text{mean}$  adalah nilai rata-rata dari  $x$  dan  $\text{std.dev}$  adalah nilai standar deviasi dari  $x$ .

Langkah-langkah yang dilakukan dalam proses training menggunakan metode elm adalah sebagai berikut:

1. Menginisialisasi  $W$  dan bias secara *random*.
2. Mencari keluaran dari *hidden layer* dengan persamaan berikut.

$$Y_{\text{in train}} = X_{\text{train}} W^T + b \quad (2)$$

Dimana  $X$  adalah matriks data training sedangkan  $W^T$  adalah transpose matriks bobot.

3. Mengaktivasi keluaran hidden layer.

$$Y = \frac{1}{1 + \exp(-H_{\text{in train}})} \quad (3)$$

4. Hitung matriks Moore-Penrose Generalized Inverse menggunakan persamaan berikut.

$$Y^+ = (Y^T \cdot Y)^{-1} Y^T \quad (4)$$

$Y^+$  merupakan matriks Moore-Penrose Generalized Inverse dari matriks  $Y$ , sedangkan matriks  $Y$  merupakan matriks keluaran hidden layer.

5. Hitung output weight dengan persamaan berikut.

$$\beta = Y^+ \cdot Z \quad (5)$$

dimana  $Y^+$  merupakan matriks *Moore-Penrose Generalized Inverse* dan  $Z$  adalah matriks target.

Langkah-langkah yang dilakukan dalam proses testing menggunakan algoritma ELM sebagai berikut:

1. Bobot  $W$  dan bias serta  $\beta$  diperoleh berdasarkan hasil training.
2. Hitung matriks keluaran hidden layer dengan persamaan berikut.

$$Y_{in\ test} = X_{test} W^T + b \quad (6)$$

$X_{test}$  adalah data *testing* dan  $W^T$  adalah *transpose* matriks bobot

3. Mengaktivasi keluaran *hidden layer*.

$$Y = \frac{1}{1 + \exp(-H_{in\ train})} \quad (7)$$

4. Menghitung nilai output layer dengan persamaan berikut.

$$H = Y \cdot \beta \quad (8)$$

Dimana  $H$  adalah hasil output layer sedangkan  $Y$  adalah matriks keluaran hidden layer dan  $\beta$  adalah matriks keluaran hidden layer pada proses training.

5. Melakukan denormalisasi untuk mendapatkan nilai asli dengan persamaan berikut.

$$X = (x' \cdot \text{std.dev}) + \text{mean} \quad (9)$$

Dimana  $X$  adalah nilai asli dan  $x'$  adalah hasil output layer.

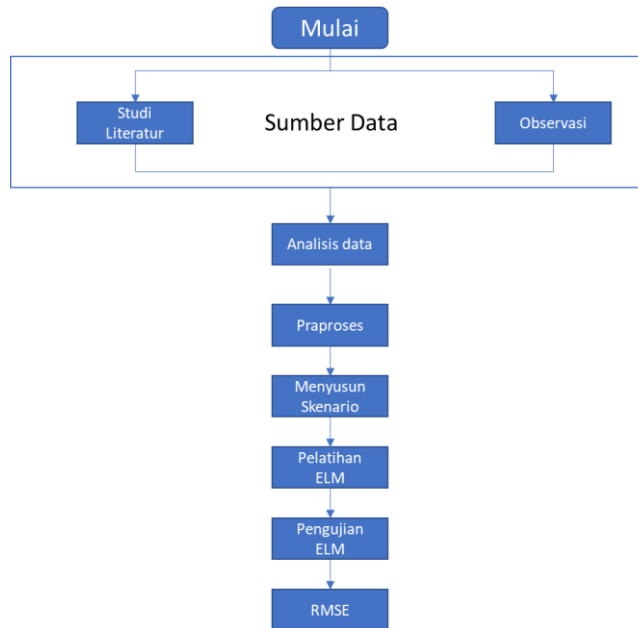
Evaluasi hasil prediksi dengan RMSE, berikut persamaan yang digunakan.

$$RMSE = \left( \frac{\sum (y_i - x_i)}{n} \right)^{1/2} \quad (10)$$

dimana  $n$  adalah jumlah data yang diprediksi dan  $y$  adalah nilai hasil prediksi.

### 3. Metodologi Penelitian

Alur penelitian yang akan dilakukan dalam penelitian ini dapat dilihat dalam gambar:



#### 3.1. Pengumpulan data

Pengumpulan data merupakan sebuah metode atau cara untuk mendapatkan sebuah informasi yang akan digunakan untuk pembangunan sebuah sistem. Pada tahap pengumpulan data ini terdapat beberapa hal yang harus dilakukan untuk membangun sebuah sistem, diantaranya adalah sebagai berikut:

##### a. Observasi

Observasi merupakan kegiatan melakukan pengamatan pada suatu objek atau bidang yang diteliti. Kegiatan observasi dilakukan dengan mengunjungi website resmi BPS dengan alamat situs <http://ipm.bps.go.id>. Observasi dilakukan untuk mengumpulkan data maupun informasi melalui naskah publikasi sebagai acuan penelitian yang dilakukan penulis.

##### b. Studi literatur

Studi pustaka merupakan kegiatan mencari materi-materi yang dibutuhkan sebagai acuan untuk keperluan penelitian. Studi pustaka dapat diperoleh dari buku, *e-book*, jurnal serta website yang mempunyai keterkaitan dengan Indeks Pembangunan Manusia dan Jaringan Saraf Tiruan.

### 3.2. Analisis dan perancangan

#### a. Preprocessing

*Preprocessing* merupakan proses yang dilakukan untuk membuat data mentah hingga menjadi data yang berkualitas dan siap untuk diproses. Berikut beberapa langkah *preprocessing* yang dilakukan.

#### Missing value

Data Indeks Pembangunan Manusia (IPM) yang digunakan sebagai objek penelitian berasal dari website resmi Badan Pusat Statistik (BPS). Data yang telah didapatkan memiliki beberapa nilai yang hilang, data yang hilang dilakukan imputasi menggunakan metode mean *substitution*.

#### Time series

Setelah dilakukan proses pada data yang hilang, selanjutnya data disusun menggunakan metode time series dengan rentang waktu lima tahun dan tiga tahun

#### Normalisasi data

Sebelum dilakukan proses perhitungan jaringan saraf tiruan data ditransformasikan terlebih dahulu. Untuk mentransformasikan data dapat menggunakan metode *scaling z-score*. Hasil dari normalisasi tersebut yang akan disimpan untuk dilakukan proses perhitungan ELM.

#### b. Penyusunan skenario penelitian

Dalam penelitian ini dibuat 8 skenario penelitian sebagai berikut:

1. SkenarioX1: Data disusun historis dengan rentang waktu 3 tahun, membagi data latih 70% dan data uji 30%.
2. SkenarioX2: Data disusun historis dengan rentang waktu 3 tahun, membagi data latih 30% dan data uji 70%.
3. SkenarioX3: Data disusun historis dengan rentang waktu 5 tahun, membagi data latih 70% dan data uji 30%.
4. SkenarioX4: Data disusun historis dengan rentang waktu 5 tahun, membagi data latih 30% dan data uji 70%.
5. SkenarioY1: Data disusun historis dengan rentang waktu 3 tahun, membagi data latih 70% dan data uji 30%, dan ditambahkan variabel Umur Harapan Hidup, Rata-rata Lama Sekolah..
6. SkenarioY2: Data disusun historis dengan rentang waktu 3 tahun, membagi data latih 30% dan data uji 70%, dan ditambahkan variabel Umur Harapan Hidup, Rata-rata Lama Sekolah.
7. SkenarioY3: Data disusun historis dengan rentang waktu 5 tahun, membagi data latih 70% dan data uji 30%, dan ditambahkan variabel Umur Harapan Hidup, Rata-rata Lama Sekolah.
8. SkenarioY4: Data disusun historis dengan rentang waktu 5 tahun, membagi data latih 30% dan data uji 70%, dan ditambahkan variabel Umur Harapan Hidup, Rata-rata Lama Sekolah.

## 4. Hasil dan Pembahasan

#### a. Pengumpulan Data

Tahap pertama penelitian ini adalah pengumpulan data. Data yang digunakan didapatkan dari website BPS berupa nilai IPM Provinsi DKI Jakarta 1996-2019, Angka Harapan Hidup Provinsi DKI Jakarta 1996-2019, Rata-rata Lama Sekolah Provinsi DKI Jakarta 1996-2019. Berikut data yang digunakan dapat dilihat pada **Tabel 1**.

*Tabel 3 Data yang digunakan*

DATA IPM		DATA RATA-RATA sekolah		DATA HARAPAN HIDUP	
tahun	ipm	tahun	rata-rata sekolah	tahun	harapan hidup
1996	76,10	1996	9,50	1996	70,20
1997		1997		1997	
1998		1998		1998	
1999	72,50	1999	9,70	1999	71,10
2000		2000		2000	
2001		2001		2001	
2002	75,60	2002	10,40	2002	72,30
2003		2003		2003	
2004	75,80	2004	10,40	2004	72,40
2005	76,07	2005	10,60	2005	72,50
2006	76,33	2006	10,80	2006	72,60
2007	76,59	2007	10,80	2007	72,80
2008	77,03	2008	10,80	2008	72,90
2009	77,36	2009	10,90	2009	73,05
2010	76,31	2010	10,37	2010	71,71
2011	76,98	2011	10,40	2011	71,87
2012	77,53	2012	10,43	2012	72,03
2013	78,08	2013	10,47	2013	72,19
2014	78,39	2014	10,54	2014	72,27
2015	78,99	2015	10,70	2015	72,43
2016	79,60	2016	10,88	2016	72,49
2017	80,06	2017	11,02	2017	72,55
2018	80,47	2018	11,05	2018	72,67
2019	80,76	2019	11,08	2019	72,79

## b. Praproses Data

### a. Missing value

Tahap kedua dalam penelitian ini adalah praproses data, dimana data yang didapatkan masih memiliki nilai yang hilang, maka selanjutnya dilakukan mean substitution untuk mengisi nilai yang hilang dari data tersebut. Mean substitution adalah metode untuk mengatasi missing value. Mean substitution dapat dilakukan dengan cara mencari nilai rata-rata dari data sesudah dan sebelum data yang hilang tersebut. Data hasil mean substitution dapat dilihat pada **Tabel 2**.

*Tabel 4 Hasil mean substitution*

Tahun	Ipm	rata-rata sekolah	harapan hidup
1996	76,10	9,50	70,20
1997	74,30	9,60	70,65
1998	74,30	9,60	70,65
1999	72,50	9,70	71,10
2000	74,05	10,05	71,70
2001	74,05	10,05	71,70
2002	75,60	10,40	72,30
2003	75,70	10,40	72,35
2004	75,80	10,40	72,40
2005	76,07	10,60	72,50
2006	76,33	10,80	72,60
2007	76,59	10,80	72,80
2008	77,03	10,80	72,90
2009	77,36	10,90	73,05
2010	76,31	10,37	71,71
2011	76,98	10,40	71,87
2012	77,53	10,43	72,03
2013	78,08	10,47	72,19

2014	78,39	10,54	72,27
2015	78,99	10,70	72,43
2016	79,60	10,88	72,49
2017	80,06	11,02	72,55
2018	80,47	11,05	72,67
2019	80,76	11,08	72,79

**b. Normalisasi**

Setelah nilai yang hilang sudah terisi kemudian dilakukan normalisasi dengan menggunakan metode z-score, normalisasi dilakukan untuk membuat data pada dataset memiliki rentang nilai yang sama. Berikut data hasil normalisasi pada **tabel 3**.

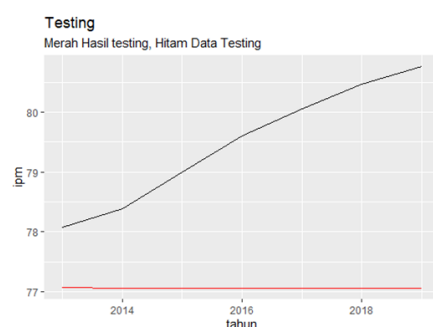
*Tabel 5 Hasil Normalisasi*

	data.tahun	ipm_normal	hrpnhdnp_normal	rt2sklh_normal
1	1996	1.44164427	-1.3244152	-1.0399620
2	1997	-0.09773859	-0.7197909	-0.7366397
3	1998	-0.09773859	-0.7197909	-0.7366397
4	1999	-1.63712146	-0.1151665	-0.4333175
5	2000	-0.31154177	0.6909993	0.6283104
6	2001	-0.31154177	0.6909993	0.6283104
7	2002	1.01403792	1.4971650	1.6899382
8	2003	1.09955919	1.5643455	1.6899382
9	2004	1.18508046	1.6315260	1.6899382
10	2005	1.41598789	1.7658870	2.2965827
11	2006	1.63834319	1.9002479	2.9032272
12	2007	1.86069850	2.1689699	2.9032272
13	2008	2.23699209	2.3033308	2.9032272
14	2009	2.51921228	2.5048723	3.2065494
15	2010	1.62123894	0.7044353	1.5989415
16	2011	2.19423145	0.9194129	1.6899382
17	2012	2.66459844	1.1343904	1.7809349
18	2013	3.13496542	1.3493680	1.9022638
19	2014	3.40008136	1.4568568	2.1145893
20	2015	3.91320898	1.6718343	2.5999049
21	2016	4.43488873	1.7524509	3.1458849
22	2017	4.82828658	1.8330675	3.5705361
23	2018	5.17892378	1.9943006	3.6615328
24	2019	5.42693547	2.1555338	3.7525294

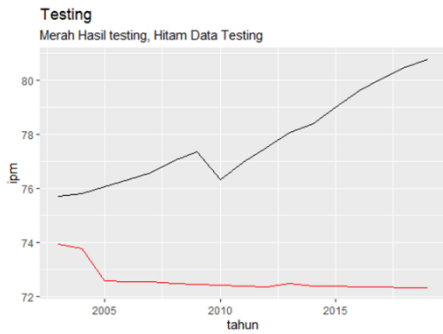
**c. Training dan Testing**

Pelatihan dilakukan pada masing-masing skenario penelitian dengan menggunakan 3 *hidden neuron*, matriks bobot W dan matriks bias di acak dengan range -4 sampai 4. Dari hasil pelatihan akan didapatkan bobot terbaik yang akan digunakan untuk proses pengujian.

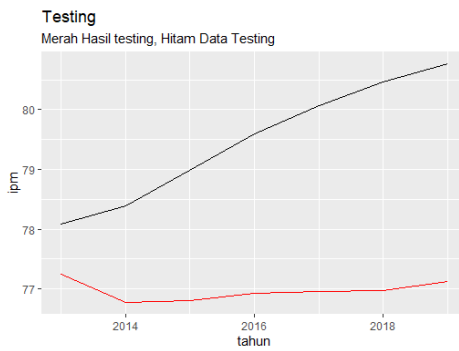
Proses selanjutnya adalah pengujian data. Hasil pengujian data untuk setiap skenario dapat dilihat pada Gambar sampai Gambar.



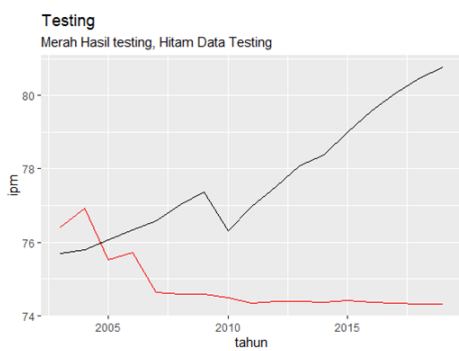
**Gambar 3** Plot testing skenarioX1



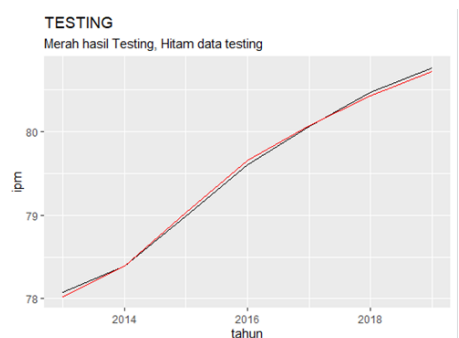
**Gambar 4** Plot testing SkenarioX2



**Gambar 5** Plot testing SkenarioX3

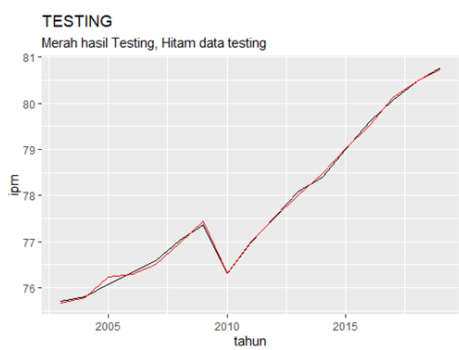


**Gambar 6** Plot testing SkenarioX4

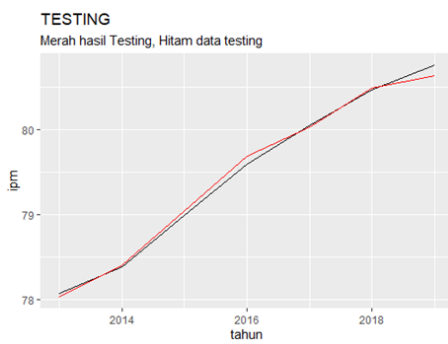




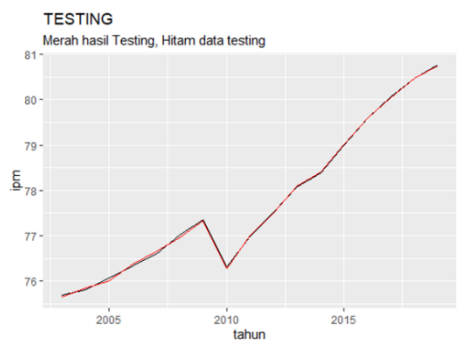
**Gambar 7** Plot testing SkenarioY1



**Gambar 8** Plot testing SkenarioY2



**Gambar 9** Plot testing SkenarioY3



**Gambar 10** Plot testing SkenarioY4

Gambar-gambar diatas adalah plot hasil pelatihan dari setiap skenario dimana garis merah menunjukkan hasil prediksi model dan garis hitam menunjukkan data uji asli. Berdasarkan gambar-gambar diatas dapat dilihat ada kemiripan yang cukup signifikan dari garis merah dan garis hitam pada plot skenarioY.

**d. Evaluasi RMSE**

Hasil yang dihasilkan dari pelatihan kemudian dilakukan perhitungan nilai *error* dengan RMSE. Hasil perhitungan RMSE untuk setiap skenario dapat dilihat pada **Tabel 4**:

**Tabel 4** Hasil perhitungan RMSE

no	SKENARIO PENELITIAN	RMSE
1	SkenarioX1	2,605955
2	SkenarioX2	5,523923
3	SkenarioX3	2,681295
4	SkenarioX4	3,681259
5	SkenarioY1	0,10454
6	SkenarioY2	0,13707
7	SkenarioY3	0,025355
8	SkenarioY4	0,049349

Berdasarkan **Tabel 4** diatas SkenarioY3 adalah skenario terbaik dengan nilai RMSE sebesar 0,025355, SkenarioY3 dengan data latih 70%, data uji 30%, historis IPM 5 tahun, dan 2 variabel pembentuk IPM yaitu Umur Harapan Hidup dan Rata-rata Lama Sekolah.

## 5. Kesimpulan dan Saran

### 5.7 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan penulis, Extreme Learning Machine memiliki performa yang baik dalam proses pengujian. Dengan nilai eror paling rendah yang didapatkan sebesar 0,025355 dihitung dengan RMSE pada proses pengujian data uji. Skenario terbaik yang didapatkan dari 8 skenario yang dibuat adalah skenarioY3 dengan susunan data latih 70%, data uji 30%, data historis 5 tahun, dan ditambah 2 variabel pembentuk IPM.

### 5.8 Saran

Berdasarkan kesimpulan mengenai implementasi Extreme Learning Machine (ELM) untuk memprediksi nilai Indeks Pembangunan Manusia (IPM) Provinsi DKI Jakarta, dapat diberikan saran berikut:

- Penelitian berikutnya dapat memprediksi nilai IPM dari provinsi-provinsi lainnya
- Metode saat ini dikombinasikan dengan metode lain untuk meingkatkan akurasi.
- Penelitian dapat dilakukan dengan parameter yang lebih bervariasi seperti pembagian data latih dan data uji yang berbeda dari penelitian saat ini,
- Penelitian selanjutnya dapat menambahkan beberapa variabel lain yang mempengaruhi nilai IPM.

## Referensi

- [1] Badan Pusat Statistik, 2018. Indeks Pembangunan Manusia 2017, Jakarta: s.n.
- [2] Sun, Z. L., Choi, T. M., Au, K. F. & Yu, Y., 2008. Sales Forecasting Using Extreme Learning Machine With Applications In Fashion Retailing. *Decision Support Systems*, Volume 46, pp. 411-419.
- [3] Rachmad Hidayat, S., 2012. Meminimalisasi Nilai Error Peramalan Dengan Algoritma Extreme Learning Machine. *Optimasi Sistem Industri*, Volume 11, pp. 187-192.
- [4] Ashar, N. M., Cholissodin, I. & Dewi, C., 2018. Penerapan Metode Extreme Learning Machine (ELM) Untuk Memprediksi Jumlah Produksi Pipa Yang Layak. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, II(11), pp. 4621-4628.
- [5] Siang, J. J., (2009), Jaringan Syaraf Tiruan dan Pemrogramannya menggunakan Matlab, Yogyakarta: CV. Andi Offset.