

Implementasi Artificial Neural Network Menggunakan Backpropagation Pada Prediksi Indeks Pembangunan Manusia

I Wayan Supriana^{a1}, Gede Dikka Widya Prana^{a2}, I Ketut Adian Jayaditya^{a3}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Jl. Raya Kampus Unud Bukit Jimbaran, Badung, Indonesia

¹wayan.supriana@unud.ac.id

²dikkawidyapranagede@gmail.com

³adianjay11@gmail.com

Abstrak

Indeks Pembangunan Manusia (IPM) merupakan salah satu alat ukur yang penting bagi suatu daerah. Nilai IPM menunjukkan indikator yang digunakan untuk mengukur tingkat keberhasilan suatu daerah dalam membangun kualitas hidup manusia. IPM mencerminkan kemampuan masyarakat di daerah tersebut untuk mengakses tiga dimensi dasar pembangunan manusia, yaitu kesehatan, pendidikan, dan pendapatan. Nilai IPM ini dapat digunakan oleh pemangku kebijakan untuk merancang kebijakan pendanaan yang lebih baik ke daerah dengan IPM tertentu. Pada penelitian ini mengimplementasikan artificial neural network (ANN) dengan metode backpropagation untuk memprediksi nilai IPM di seluruh kabupaten di Indonesia guna mempermudah prediksi kualitas manusia. Hyperparameter tuning juga dilakukan untuk menentukan arsitektur jaringan optimal pada model ANN yang dibangun. Berdasarkan penelitian yang sudah dilakukan arsitektur optimal yang didapat dengan komposisi 2 hidden layer, setiap hidden memiliki neuron sebanyak 3 dan 4, learning rate sebesar 0.0001, fungsi aktivasi ReLU, dan epoch berjumlah 200. Hyperparameter optimal tersebut menghasilkan MSE sebesar 0.0005849390439398957.

Keywords: Indeks Pembangunan Manusia, Backpropagation, ANN, Prediksi, Kualitas Hidup Manusia

1. Pendahuluan

Indeks Pembangunan Manusia (IPM) atau *Human Development Index (HDI)* merupakan salah satu alat penting untuk mengukur kesejahteraan penduduk suatu negara atau daerah [1]. Indeks Pembangunan Manusia dibentuk oleh tiga aspek dasar, yaitu kesehatan, pendidikan, dan pendapatan. Ketiga aspek ini mencerminkan tingkat pembangunan manusia di suatu wilayah melalui pengukuran kondisi penduduk yang sehat, berumur panjang, teredukasi, memiliki keterampilan, dan memiliki pendapatan yang memungkinkan hidup layak [2]. Pada masing-masing aspek tersebut terdapat indikatornya, yaitu Angka Harapan Hidup, Harapan Lama Sekolah, Rata-Rata Lama Sekolah, dan Pengeluaran per Kapita [3]. Harapan Lama Sekolah dan Rata-Rata Lama Sekolah merupakan indikator pendidikan, mencerminkan tingkat melek huruf dan tingkat pendidikan dalam suatu populasi. Sedangkan, Angka Harapan Hidup mencerminkan Kesehatan masyarakat, menunjukkan berapa lama penduduk diharapkan dapat hidup [4]. IPM menjadi hal yang sangat penting untuk melihat perkembangan dari sisi manusia dan merupakan indikator pemerintah untuk merancang kebijakan pendanaan terkait pelayanan umum dan kesejahteraan masyarakat, sehingga dapat merealisasikan perencanaan pembangunan yang tepat sasaran dan berkelanjutan [5].

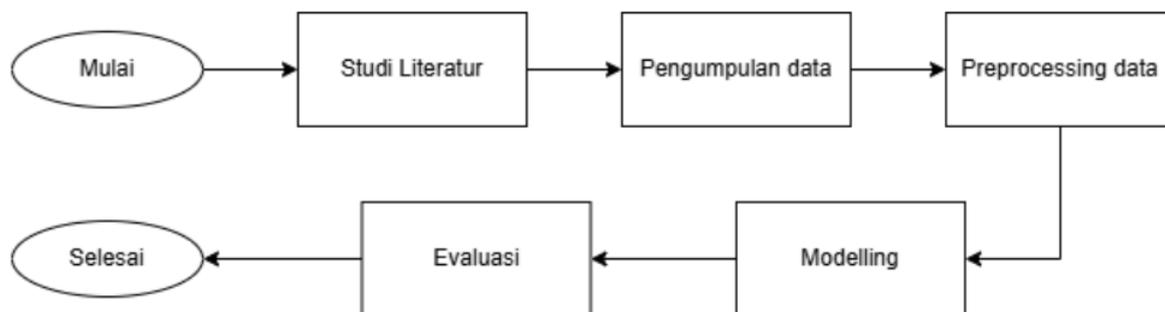
Beberapa penelitian telah dilakukan untuk memprediksi angka Indeks Pembangunan Manusia menggunakan berbagai metode *machine learning*. Penelitian yang dilakukan Atmojo et al. (2024) untuk memprediksi IPM dengan membandingkan metode KNN, Random Forest, AdaBoost, dan SVM, dengan SVM menunjukkan hasil terbaik. SVM menghasilkan MSE terendah pada data pengujian sebesar 4.419, yang menunjukkan prediksi paling akurat [1]. Penelitian lain oleh Posangi et al. (2023),

menggunakan Random Forest untuk memprediksi IPM. Penelitian ini memperoleh akurasi 80%, dengan fitur pengeluaran per kapita, harapan lama sekolah, angka harapan hidup, dan rata-rata lama sekolah sebagai variabel utama [6]. Ampang et al. (2023) melakukan penelitian untuk memprediksi angka IPM menggunakan metode Double Exponential Smoothing (DES). Hasil yang diperoleh dari penelitian tersebut ialah MAPE (Mean Absolute Percentage Error) yang diraih sebesar 0,13% untuk indeks Kesehatan, 0,55% untuk indeks pendidikan, dan 0,45% untuk indeks Daya Beli [7]. Amiroh dan Avianto (2023) juga pernah melakukan penelitian untuk memprediksi IPM menggunakan metode backpropagation. Pada penelitian yang dilakukan diperoleh MSE terbaik sebesar 2,0501 [8]. Pada penelitian yang sudah dilakukan, metode backpropagation merupakan metode terbaik dalam memprediksi IPM. Namun, beberapa penelitian tersebut terbatas pada dataset yang relatif sedikit dan memiliki ruang untuk perbaikan dalam akurasi prediksi.

Berdasarkan penelitian-penelitian sebelumnya, penulis mengusulkan penggunaan Artificial Neural Network (ANN) berbasis backpropagation dengan data lebih besar dari tahun 2010 hingga 2023 dengan fitur pengeluaran per kapita, harapan lama sekolah, angka harapan hidup, dan rata-rata lama sekolah. Data tersebut mencakup seluruh kabupaten di Indonesia dan digunakan untuk menangkap hubungan yang lebih kompleks serta menghasilkan prediksi yang lebih akurat.

2. Metode Penelitian

Penelitian ini menggunakan algoritma Artificial Neural Network berbasis backpropagation untuk memprediksi angka Indeks Pembangunan Manusia (IPM) di Indonesia. Pada penelitian ini, penulis menggunakan beberapa tahapan penelitian yang ditampilkan pada Gambar 1.



Gambar 1. Alur Penelitian

Alur penelitian pada Gambar 1 dimulai dengan tahap studi literatur untuk menggali dan mendefinisikan penelitian pendahuluan yang terkait. Tahapan selanjutnya adalah pengumpulan dan memvisualisasi data dalam mengidentifikasi kualitas dan kecukupan dataset. Tahap dilanjutkan dengan preprocessing data yaitu memperbaiki data noise, melakukan reduksi fitur, agregasi data serta validasi data. Tahap berikutnya adalah modelling yaitu membangun model prediksi menggunakan ANN. Dan tahap akhir ada evaluasi model dengan menentukan parameter input dan arsitektur jaringan ANN terbaik.

2.1. Studi Literatur

Pada tahapan ini dilakukan kajian terhadap penelitian-penelitian sebelumnya yang relevan dengan konsep Indeks Pembangunan Manusia serta metode prediksi yang telah diterapkan. Studi ini bertujuan untuk memperoleh pemahaman yang mendalam mengenai teori, model, atau pendekatan yang digunakan dalam penelitian serupa. Selain itu, studi literatur ini berperan penting dalam mengidentifikasi kelebihan dan kekurangan metode yang telah ada sehingga dapat dijadikan dasar untuk mengadaptasi atau mengembangkan metode yang lebih baik.

2.2. Pengumpulan Data

Penelitian ini menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Indonesia untuk periode tahun 2010 hingga 2023. Jumlah data sebanyak 7195 yang terdiri dari atribut provinsi, kabupaten, tahun, indeks pembangunan manusia, angka harapan hidup, rata-rata lama sekolah,

harapan lama sekolah, dan pengeluaran per kapita. Data ini merupakan data dari semua kabupaten/kota yang terdapat di Indonesia. Data ini akan digunakan sebagai input dalam proses pelatihan dan pengujian ANN berbasis algoritma backpropagation.

	provinsi	kabupaten	tahun	indeks_pembangunan_manusia	angka_harapan_hidup	rerata_lama_sekolah	harapan_lama_sekolah	pengeluaran_per_kapita
0	ACEH	Simeulue	2023	68.06	65.655	9.760	14.28	7686.0
1	ACEH	Aceh Singkil	2023	70.14	67.830	8.655	14.35	9374.0
2	ACEH	Aceh Selatan	2023	68.44	64.860	8.840	14.70	8712.0
3	ACEH	Aceh Tenggara	2023	71.05	68.700	10.100	14.27	8566.0
4	ACEH	Aceh Timur	2023	69.32	69.120	8.490	13.07	9436.0
...
7191	PAPUA	Puncak	2010	33.44	64.915	0.710	2.84	4752.0
7192	PAPUA	Dogiyai	2010	47.48	64.160	3.360	7.85	4769.0
7193	PAPUA	Intan Jaya	2010	NaN	NaN	NaN	4.12	NaN
7194	PAPUA	Deiyai	2010	NaN	NaN	NaN	9.06	NaN
7195	PAPUA	Kota Jayapura	2010	76.69	69.835	10.655	13.64	13903.0

Gambar 1. Dataset Penelitian

2.3. Data Preprocessing

Pada tahap data preprocessing, dilakukan serangkaian langkah untuk memastikan data siap digunakan dalam pengembangan model prediksi.

1. Penanganan data *missing value* (data kosong)

Penanganan data missing value dilakukan dengan teknik imputasi yaitu mengganti nilai/data yang hilang (missing value; NaN; Blank) dengan nilai pengganti. Mekanisme perbaikan data yang hilang dengan teknik: (a) Missing Completely at Random (MCAR), ketika setiap sel memiliki kemungkinan hilang yang tetap dan sama. Akibatnya, nilai prediksi besar dan kecil menjadi hilang. (b) Missing at Random (MAR), ketika probabilitas yang hilang tergantung pada nilai-nilai yang diamati. (c) Missing Not at Random (MNAR), ketika kemungkinan hilang tergantung pada data yang tidak teramati atau hilang. Teknik imputasi dilakukan dengan cara menggantikan data yang hilang dengan estimasi statistik dari nilai yang hilang tersebut. Tujuan dari teknik ini adalah untuk menghasilkan dataset yang lengkap yang dapat digunakan dalam proses pelatihan model. Proses ini membutuhkan pustaka fitur, yang dapat menyederhanakan proses pengisian nilai yang hilang.

2. Transformasi Data Kategori

Encoding pada fitur kategorikal menggunakan *Label Encoding* juga dilakukan agar fitur tersebut dapat dianalisis untuk tahap selanjutnya. Kemudian, Analisis korelasi dilakukan untuk menghitung hubungan antar fitur, sehingga hanya variabel yang relevan terhadap target prediksi yang digunakan. Seleksi fitur dilakukan berdasarkan hasil analisis korelasi dan relevansi terhadap target, diikuti dengan deteksi dan penanganan outlier.

3. Data dipisahkan menjadi fitur (*variabel independen*) dan target (*variabel dependen*), lalu dilakukan normalisasi agar fitur-fitur tersebut memiliki skala nilai yang sama sehingga model prediksi dapat bekerja dengan stabil dan efisien.

2.4. Modelling

Pada tahap implementasi, dilakukan penerapan model ANN dengan model pembelajaran dengan algoritma backpropagation untuk memprediksi nilai IPM. Proses ini melibatkan sekenario dengan berbagai kombinasi hyperparameter untuk mendapatkan arsitektur model yang optimal. Hyperparameter yang diuji meliputi jumlah hidden layers, jumlah neuron di setiap hidden layers, fungsi aktivasi, nilai learning rate, dan jumlah epoch pelatihan. Tahapan implementasi ini dilakukan secara iteratif, di mana setiap kombinasi hyperparameter diuji untuk memahami pengaruhnya terhadap performa model. Berikut merupakan tahapan-tahapan dari proses pelatihan menggunakan ANN berbasis *backpropagation* [8].

1. Pemberian nilai awal pada bobot dilakukan secara acak dengan memerhatikan batasan learning rate dan epoch yang sesuai.
2. Setiap data yang ada dilakukan proses langkah ke-3 hingga 8

Proses Feed Forward

3. Inputan akan menerima isyarat masukan x_i yang akan dilanjutkan ke node-node pada hidden layers.

4. Node hidden layers (z) akan dijumlahkan bobot sinyal inputnya. Pada proses ini parameter n merupakan banyaknya iterasi yang dilakukan.

$$Z_{in_{jk}} = v_{0j} + \sum_{i=1}^n x_i v_{ij} \quad (1)$$

sementara fungsi aktivasi dihitung dengan:

$$Z_j = f(Z_{in_j}) \quad (2)$$

5. Setiap node keluaran dari hidden layers diteruskan ke output layers yang akan dijumlahkan bobot dengan syarat masukkan, dengan persamaan berikut:

$$Y_{in_k} = w_{0j} + \sum_{i=1}^n z_j v_{jk} \quad (3)$$

fungsi aktivasi dihitung dengan:

$$Y_j = f(Y_{in_k}) \quad (4)$$

Proses Backpropagation

6. Pada tiap node output mendapatkan suatu pola training dari node input yang sudah ada. Perhitungan error dilakukan dengan persamaan sebagai berikut:

$$\delta_k = (t_k - y_k) f'(y_{in_k}) \quad (5)$$

Sementara perhitungan dan koreksi bobot serta bias, menggunakan persamaan berikut:

$$\Delta w_{jk} = a \delta_k x_j \quad (6)$$

$$\Delta w_{0k} = a \delta_k \quad (7)$$

7. Masing-masing node hidden akan menjumlahkan input delta (dari unit yang berada pada lapisan atasnya)

$$\delta_{in_j} = \sum_{i=1}^m \delta_k w_{jk} \quad (8)$$

pada perhitungan node hidden, faktor δ menggunakan persamaan berikut:

$$\delta_j = \delta_{in_j} f'(Z_{in_{jk}}) = \delta_{in_j} f'(Z_j) \quad (9)$$

selanjutnya dilakukan perubahan bobot dan nilai biasnya, dengan persamaan berikut:

$$\Delta v_{ij} = a \delta_j x_i \quad (10)$$

$$\Delta v_{0j} = a \delta_j \quad (11)$$

Proses Pembobotan

8. Menghitung semua bobot dan bias yang berubah ke node output

$$w_{jk}(\text{baru}) = w_{jk}(\text{lama}) + \Delta w_{jk} \quad (12)$$

$$w_{0k}(\text{baru}) = w_{0k}(\text{lama}) + \Delta w_{0k} \quad (13)$$

Bobot yang sudah berubah lalu diteruskan ke hidden layers

$$v_{ij}(\text{baru}) = v_{ij}(\text{lama}) + \Delta v_{ij} \quad (14)$$

$$v_{0j}(\text{baru}) = v_{0j}(\text{lama}) + \Delta v_{0j} \quad (15)$$

Hasil perhitungan dari hidden layer dan output layer akan dimasukkan ke suatu fungsi yang bernama fungsi aktivasi. Fungsi aktivasi adalah elemen penting dalam ANN, bertugas memberikan kemampuan non-linear kepada jaringan dan memungkinkan pemodelan hubungan yang kompleks dalam data. Pada penelitian ini menggunakan tiga fungsi aktivasi yaitu ReLU, Sigmoid, dan Tanh. Fungsi ini juga memerlukan turunan untuk memperbarui bobot jaringan.

1. ReLU

ReLU adalah fungsi aktivasi sederhana yang didefinisikan sebagai berikut.

$$f(x) = \max(0, x) \quad (16)$$

Nilai Output adalah nol untuk nilai input negatif dan sama dengan input untuk nilai input positif.

Turunannya didefinisikan sebagai berikut.

$$f'(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (17)$$

Turunan ini sangat efisien untuk perhitungan. Namun, jika neuron menghasilkan nilai nol secara terus-menerus (*dying* ReLU), gradien menjadi nol, sehingga bobot tidak diperbarui selama pelatihan.

2. Sigmoid

Fungsi sigmoid memiliki bentuk S (*S-shaped curve*) dan didefinisikan sebagai berikut.

$$f(x) = \frac{1}{1+e^{-x}} \quad (18)$$

Fungsi ini mengubah nilai input menjadi rentang (0, 1), yang cocok untuk tugas biner atau menghitung probabilitas. Turunannya didefinisikan sebagai berikut.

$$f'(x) = f(x)(1 - f(x)) \quad (19)$$

Karena $f(x)$ selalu berada di antara 0 dan 1, nilai turunan sigmoid menjadi sangat kecil pada input yang ekstrem (sangat besar atau kecil). Hal ini menyebabkan masalah *vanishing gradient*.

3. Tanh

Fungsi Tanh (*tangens* hiperbolik) didefinisikan sebagai berikut.

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (20)$$

Output Tanh berada dalam rentang (-1, 1), sehingga lebih informatif dibandingkan sigmoid karena mencakup nilai negatif dan positif. Turunannya didefinisikan sebagai berikut.

$$f'(x) = 1 - f(x)^2 \quad (21)$$

Fungsi Tanh juga rentan terhadap masalah vanishing gradient pada input dengan nilai besar atau kecil.

2.5. Evaluasi

Model yang telah dibangun akan dievaluasi menggunakan metrik Mean Squared Error (MSE) untuk mengukur seberapa baik model memprediksi angka Indeks Pembangunan Manusia. MSE menghitung rata-rata selisih kuadrat antara nilai sebenarnya dan nilai prediksi, di mana semakin kecil nilai MSE, semakin baik kinerja model. Hasil evaluasi ini akan digunakan untuk menilai performa model dan menentukan apakah diperlukan perbaikan lebih lanjut.

3. Hasil dan Pembahasan

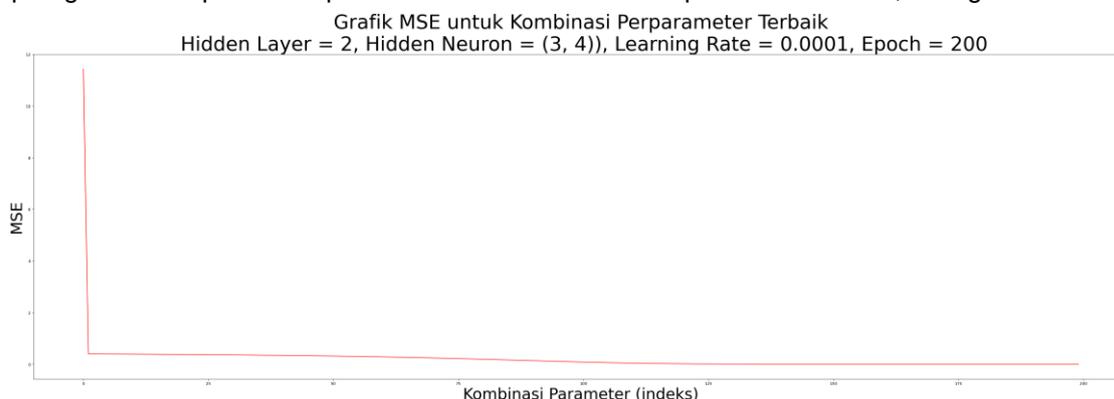
3.1. Hasil Training.

Proses pelatihan untuk memprediksi nilai Indeks Pembangunan Manusia (IPM) pada data training telah dilakukan beberapa kali menggunakan hyperparameter dari arsitektur yang berbeda-beda. Percobaan dilakukan dengan menggunakan semua kombinasi yang mungkin dari hidden layer sebanyak 1 dan 2; neuron pada hidden layer sebanyak 2, 3, dan 4; learning rate sebesar 0.0001, 0.001; epoch sebesar 100 dan 200. Pada Tabel 1 menunjukkan *hyperparameter* arsitektur optimal yang didapatkan setelah melakukan proses *hyperparameter tuning*.

Tabel 1. *Hyperparameter Tuning*

<i>Hyperparameter</i>	Nilai
Hidden Layer	2
Hidden Neuron	(3, 4)
Learning Rate	0.0001
Fungsi Aktivasi	ReLU
Epochs	200
MSE	0.0005849390439398957

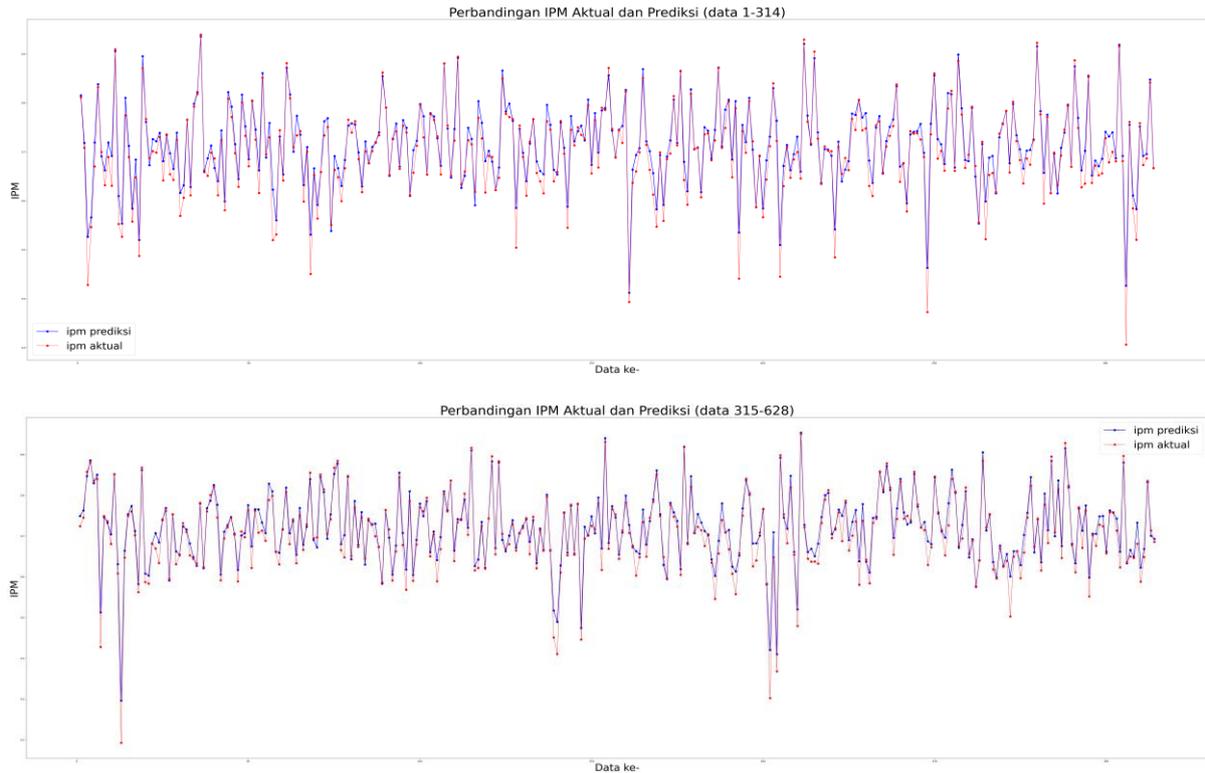
Adapun grafik MSE pada setiap iterasi untuk kombinasi nilai parameter terbaik, sebagai berikut.



Gambar 3. Grafik Elbow Method

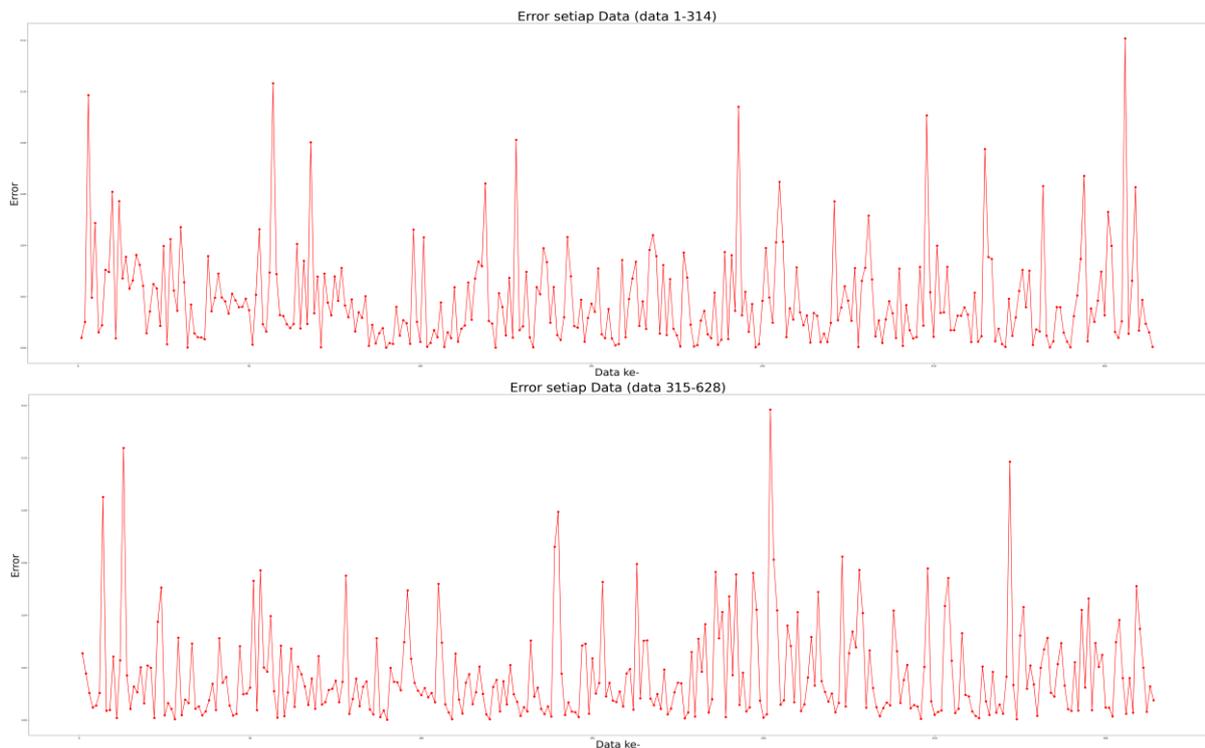
3.2. Hasil Testing

Pengujian dilakukan menggunakan data *testing* yang berjumlah 628 data. Berikut pada Gambar 3 dapat dilihat grafik perbandingan antara hasil prediksi dan aktual. Data tersebut dipecah menjadi 2 bagian agar dapat melihat grafik lebih baik.



Gambar 3. Grafik perbedaan nilai prediksi dan actual

Adapun grafik error (selisih antara prediksi dan actual) pada setiap data *testing* dapat dilihat pada Gambar 4.



Gambar 4. Grafik error pada setiap data

Model yang telah berhasil dibangun akan diintegrasikan dengan antarmuka berbasis web, sehingga memudahkan untuk melakukan prediksi IPM. Tampilan dari web yang dibangun dapat dilihat pada Gambar 5.



Gambar 5. Tampilan antarmuka *website*

3. Kesimpulan

Model ANN berbasis algoritma backpropagation berhasil dirancang untuk memprediksi IPM di Indonesia. Proses perancangan meliputi pengumpulan data sekunder dari BPS, preprocessing data (termasuk penghapusan data kosong, seleksi fitur, deteksi outlier, normalisasi, dan pembagian data menjadi set pelatihan dan pengujian), serta hyperparameter tuning untuk menentukan arsitektur jaringan terbaik, seperti jumlah hidden layer, jumlah neuron pada hidden layer, learning rate, dan jumlah epoch. Model yang dirancang juga diintegrasikan dengan antarmuka berbasis web untuk memudahkan pengguna dalam melakukan prediksi IPM. Evaluasi performa model menggunakan metrik MSE menunjukkan bahwa model mampu memprediksi IPM dengan akurasi yang sangat baik. Arsitektur jaringan ANN terbaik adalah 2 hidden layer, 3 dan 4 neuron, fungsi aktivasi ReLU, learning rate 0.0001, dan 200 epoch, menghasilkan MSE sebesar 0.0005849390439398957.

References

- [1] F. W. Atmojo, C. I. Nurlita, and N. Nurchim, "ANALISIS PEMANFAATAN MACHINE LEARNING GUNA PREDIKSI INDEKS PEMBANGUNAN MANUSIA," *Simtek : jurnal sistem informasi dan teknik komputer*, vol. 9, no. 2, pp. 89–96, Oct. 2024, doi: 10.51876/simtek.v9i2.390.
- [2] A. M. Kahar, "Analisis Angka Harapan Lama Sekolah di Indonesia Timur Menggunakan Weighted Least Squares Regression," *Jurnal Matematika "MANTIK,"* vol. 4, no. 1, pp. 32–41, May 2018, doi: 10.15642/mantik.2018.4.1.32-41.
- [3] M. Y. Alang, K. Suryowati, and F. Astuti, "PREDIKSI INDEKS PEMBANGUNAN MANUSIA PADA KABUPATEN/KOTA DI PROVINSI NUSA TENGGARA TIMUR MENGGUNAKAN BACKPROPAGATION DENGAN KOMBINASI LEARNING RATE DAN EPOCH," *Jurnal Statistika Industri dan Komputasi*, vol. 9, no. 1, pp. 10–20, Jan. 2024, doi: 10.34151/statistika.v9i1.4821.
- [4] D. I. Ginting, I. Lubis, I. Lubis, and I. Lubis, "PENGARUH ANGKA HARAPAN HIDUP DAN HARAPAN LAMA SEKOLAH TERHADAP INDEKS PEMBANGUNAN MANUSIA," *Bisnis-Net Jurnal Ekonomi dan Bisnis*, vol. 6, no. 2, pp. 519–528, Dec. 2023, doi: 10.46576/bn.v6i2.3884.
- [5] I. Kemala and A. W. Wijayanto, "Perbandingan Kinerja Metode Bagging dan Non-Ensemble Machine Learning pada Klasifikasi Wilayah di Indonesia menurut Indeks Pembangunan Manusia," *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 9, no. 2, p. 269, Apr. 2021, doi: 10.26418/justin.v9i2.44166.

- [6] T. Posangi, L. Yahya, and D. Wungguli, "Implementasi Algoritma Random Forest dengan Forward Selection untuk Klasifikasi Indeks Pembangunan Manusia," *Jambura Journal of Probability and Statistics*, vol. 4, no. 2, pp. 85–91, Nov. 2023, doi: 10.37905/jjps.v4i2.18460.
- [7] A. C. Ampang, B. E. W. Asrul, and M. A. Nur, "Implementasi Metode Double Exponential Smoothing Untuk Memprediksi Indeks Pembangunan Manusia (IPM) Di Kabupaten Toraja Utara," *Jurnal Fokus Elektroda : Energi Listrik, Telekomunikasi, Komputer, Elektronika Dan Kendali*, vol. 8, no. 3, pp. 197–207, 2023.
- [8] N. N. Amiroh and D. Avianto, "Prediksi Indeks Pembangunan Manusia di Kabupaten Wonosobo Menggunakan Algoritma Backpropagation," *Techno.Com*, vol. 22, no. 2, pp. 388–399, May 2023, doi: 10.33633/tc.v22i2.7980.