

Pengaruh Korelasi Data pada Peramalan Suhu Udara Menggunakan Backpropagation Neural Network

Dinita Rahmalia¹, Nur Aini²

¹Universitas Islam Darul Ulum Lamongan

²STKIP PGRI Bangkalan

Email: dinitarahmalia@gmail.com

DOI <https://doi.org/10.31102/zeta.2018.4.1.1-6>

ABSTRACT

Air temperature forecasting is important role in agriculture, flight, trading and so on. The method used for forecasting is Neural Network (NN). NN works as human neural system. One of NN type used for forecasting is Backpropagation where Backpropagation model is there are hidden layers between input and output. Due to forecasting result depends on data correlation, then this research will explain about the effect of data correlation on air temperature forecasting. To obtain forecasting result, Backpropagation algorithm will be used. Simulations are applied in three dataset with different structures. Based on simulation results, data which have strong correlation can result better forecasting based on smaller Mean Square Error (MSE).

Keywords: *regression, correlation, backpropagation, neural network*

ABSTRAK

Peramalan suhu udara merupakan kegiatan yang sangat penting pada bidang pertanian, penerbangan, perdagangan dan sebagainya. Salah satu metode yang digunakan untuk masalah peramalan adalah Neural Network (NN). Cara kerja NN mirip dengan sistem syaraf manusia. Salah satu jenis NN yang digunakan untuk masalah peramalan adalah Backpropagation dimana model Backpropagation adalah terdapat hidden layer di antara input dan output. Karena hasil peramalan bergantung pada nilai korelasi data, maka pada penelitian ini akan diteliti mengenai pengaruh korelasi data pada peramalan suhu udara. Untuk mendapatkan hasil peramalan, digunakan algoritma Backpropagation. Simulasi dilakukan pada tiga dataset dengan struktur yang berbeda. Berdasarkan hasil simulasi, data yang memiliki korelasi yang kuat dapat menghasilkan nilai peramalan yang lebih baik yang ditandai dengan nilai Mean Square Error (MSE) lebih kecil.

Kata Kunci: *regresi, korelasi, backpropagation, neural network*

1. PENDAHULUAN

Peramalan suhu udara merupakan kegiatan yang sangat penting pada bidang pertanian, penerbangan, perdagangan dan sebagainya. Sebagai contoh, pada bidang pertanian, suhu udara berpengaruh pada pertumbuhan tanaman. Pada bidang penerbangan, seorang pilot juga memperhatikan suhu udara sebelum melakukan penerbangan. Pada bidang perdagangan, produk yang digunakan pada musim dingin memiliki nilai jual yang tinggi jika dijual pada musim dingin. Sebaliknya, produk yang digunakan pada musim panas memiliki nilai jual yang tinggi jika dijual pada musim panas.

Salah satu metode yang digunakan untuk masalah peramalan adalah Neural Network. Neural Network (NN) diperkenalkan oleh McCulloch dan Pitts pada tahun 1943. Cara kerja NN mirip dengan sistem syaraf manusia. Salah satu jenis NN yang digunakan untuk masalah peramalan adalah Backpropagation dimana model Backpropagation adalah terdapat hidden layer di antara input dan output (Siang, 2009; Fausett, 1994).

Pada beberapa penelitian sebelumnya, telah dilakukan penelitian tentang peramalan intensitas matahari dengan input suhu udara, kelembaban udara menggunakan Backpropagation (Rahmalia dan Herlambang, 2017), peramalan cuaca menggunakan Fuzzy Logic (Matarneh, 2014), peramalan menggunakan exponential smoothing (Rahmalia, 2018). Karena hasil peramalan bergantung pada nilai korelasi data, maka pada penelitian ini akan diteliti mengenai pengaruh korelasi data pada peramalan suhu udara (Han, 2012). Pada penelitian ini menggunakan tiga dataset dengan struktur yang berbeda untuk perbandingan. Dari data yang digunakan, akan ditentukan input dan output, kemudian dihitung nilai korelasi antara input dan output. Setelah itu untuk mendapatkan hasil peramalan, digunakan algoritma Backpropagation. Sebelum Backpropagation dilakukan, partisi data ke dalam data training dan data testing (Siang, 2009). Dari hasil simulasi Backpropagation, diperoleh nilai Mean Square Error (MSE), yaitu rata-rata kuadrat selisih antara nilai target dengan hasil peramalan.

Simulasi dilakukan pada tiga dataset dengan struktur yang berbeda. Berdasarkan hasil simulasi, data yang memiliki korelasi yang kuat (nilai korelasi mendekati 1) dapat menghasilkan nilai peramalan yang lebih baik yang ditandai dengan nilai MSE lebih kecil.

2. TINJAUAN PUSTAKA

2.1. Neural Network

Neural Network (NN) diperkenalkan oleh McCulloch dan Pitts pada tahun 1943. Cara kerja neural network mirip dengan sistem syaraf manusia yaitu :

1. Sinyal dikirim di antara neuron melalui penghubung.
2. Penghubung antar neuron memiliki bobot yang akan memperkuat atau memperlemah sinyal.
3. Untuk menentukan keluaran, setiap neuron menggunakan fungsi aktivasi yang dikenakan pada jumlah masukan yang diterima.

Dalam NN, fungsi aktivasi dipakai untuk menentukan keluaran suatu neuron. Argumen fungsi aktivasi adalah kombinasi linier antara masukan dengan bobot.

$$net = \sum_i x_i w_i \quad (1)$$

$$f(net) = f\left(\sum_i x_i w_i\right) \quad (2)$$

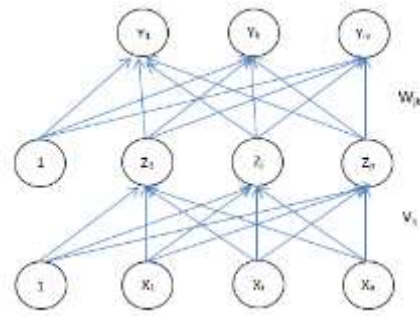
Pada umumnya, sifat fungsi aktivasi adalah : kontinu, terdiferensial, dan tidak turun. Pada penelitian ini, fungsi aktivasi yang dipakai adalah sigmoid biner dengan range (0-1) (Siang, 2009).

$$f(net) = \frac{1}{1 + e^{-net}} \quad (3)$$

$$f'(net) = f(net)(1 - f(net)) \quad (4)$$

2.2. Backpropagation

Backpropagasi adalah salah satu jenis NN yang sering digunakan dalam proses prediksi. Algoritma Backpropagasi terdiri dari beberapa unit input $x_1, x_2, \dots, x_n$, beberapa unit hidden layer $z_1, z_2, \dots, z_p$, dan beberapa unit output $y_1, y_2, \dots, y_m$. Dalam unit input dan hidden layer terdapat bias yang bernilai 1. Bobot v_{ij} menghubungkan unit input x_i menuju unit hidden layer z_j . Bobot w_{jk} menghubungkan unit hidden layer z_j menuju unit output y_k . Dalam algoritma Backpropagasi, terdapat tiga fase perhitungan yaitu : propagasi maju, propagasi mundur, dan update matriks bobot. Model Backpropagasi dapat dilihat pada Gambar 1.



Gambar 1. Model Backpropagasi

Pada propagasi maju, perhitungan menggunakan fungsi aktivasi dilakukan mulai dari input, hidden layer, dan output secara berurutan. Pada propagasi mundur, perhitungan faktor kesalahan dilakukan mulai dari output, hidden layer, dan input secara berurutan. Setelah itu, dilakukan perubahan nilai pada matriks bobot (Fausett, 1994).

Algoritma Backpropagasi menggunakan satu hidden layer dapat dinyatakan sebagai berikut :

$bppgt(V, W)$

1. Inisialisasi matriks bobot V dan W dengan bilangan acak kecil antara -0,5 sampai 0,5.

$e = 1$

while($e \leq \max \text{epoch} \ \&\& \text{MSE} \geq \min \text{MSE}$)

for($d = 1 : \text{datasize}$)

2. Tiap unit input menerima sinyal dan meneruskan ke unit hidden layer.

Propagasi Maju

3. Hitung semua keluaran di unit hidden layer $z_j, j = 1, 2, \dots, p$

$$z_net_j = v_{0j} + \sum_{i=1}^n x_i v_{ij} \quad (5)$$

$$z_j = f(z_net_j) = \frac{1}{1 + e^{-z_net_j}} \quad (6)$$

4. Hitung semua keluaran di unit output $y_k, k = 1, 2, \dots, m$

$$y_net_k = w_{0k} + \sum_{j=1}^p z_j w_{jk} \quad (7)$$

$$y_k = f(y_net_k) = \frac{1}{1 + e^{-y_net_k}} \quad (8)$$

Propagasi Mundur

5. Hitung faktor δ unit output berdasarkan kesalahan di setiap unit output $y_k, k = 1, 2, \dots, m$

$$\delta_k = (t_k - y_k) f'(y_net_k) \quad (9)$$

6. Hitung perubahan bobot

$$\Delta w_{jk} = \alpha \delta_j z_j$$

$$k = 1, 2, \dots, m \quad j = 0, 1, 2, \dots, p$$
(10)

7. Hitung faktor δ unit hidden layer berdasarkan kesalahan di setiap unit hidden layer $z_j, j = 1, 2, \dots, p$

$$\delta_{net_j} = \sum_{k=1}^m \delta_k w_{jk}$$

$$\delta_j = \delta_{net_j} f'(z_{net_j})$$
(11)

8. Hitung perubahan bobot

$$\Delta v_{ij} = \alpha \delta_j x_i$$

$$j = 1, 2, \dots, p \quad i = 0, 1, 2, \dots, n$$
(12)

Update Matriks Bobot

9. Hitung matriks bobot yang baru

$$w_{jk} \leftarrow w_{jk} + \Delta w_{jk}$$
(13)

$$v_{ij} \leftarrow v_{ij} + \Delta v_{ij}$$
(14)

end

10. Hitung nilai Mean Square Error (MSE)

$$MSE = \frac{1}{datasize} \sum_{d=1}^{datasize} \frac{1}{m} \sum_{k=1}^m (T_{dk} - Y_{dk})^2$$
(15)

11. $e \leftarrow e + 1$

end

2.3. Korelasi Data

Korelasi adalah hubungan linear antara dua variabel. Koefisien korelasi dapat dihitung menggunakan persamaan (16). Nilai dari koefisien korelasi berkisar antara $-1 \leq r \leq 1$. Jika korelasi bernilai +1, maka hubungan dua variabel sempurna bernilai positif. Jika korelasi bernilai -1, maka hubungan dua variabel sempurna bernilai negatif. Jika korelasi bernilai 0, maka tidak ada hubungan dua variabel. Jadi korelasi yang kuat bernilai mendekati +1 atau mendekati -1 (Walpole, 2012).

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$
(16)

Dengan :

$$S_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

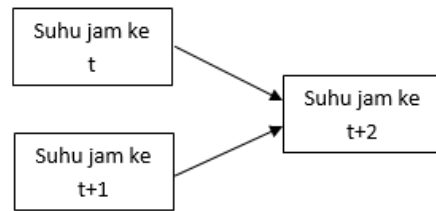
$$S_{xx} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$S_{yy} = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

3. METODE PENELITIAN

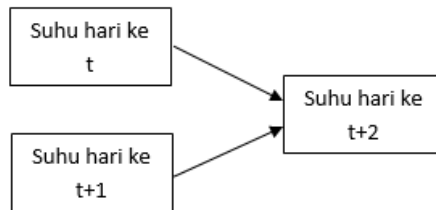
Data yang digunakan dalam penelitian ini terdiri dari tiga dataset dengan masing-masing dataset terdapat input dan output dengan konstruksi seperti Gambar 2, Gambar 3, dan Gambar 4.

Gambar 2 adalah konstruksi dari dataset 1 dengan input suhu udara pada jam ke t dan suhu udara pada jam ke t+1 dan output suhu udara pada jam ke t+2.



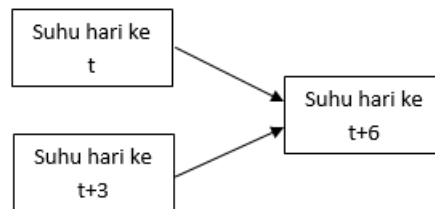
Gambar 2. Konstruksi pada Dataset 1

Gambar 3 adalah konstruksi dari dataset 2 dengan input suhu udara pada hari ke t dan suhu udara pada hari ke t+1 dan output suhu udara pada hari ke t+2.



Gambar 3. Konstruksi pada Dataset 2

Gambar 4 adalah konstruksi dari dataset 3 dengan input suhu udara pada jam ke t dan suhu udara pada jam ke t+3 dan output suhu udara pada jam ke t+6.



Gambar 4. Konstruksi pada Dataset 3

Persiapan Data : Dari data cuaca yang diperoleh, tentukan input dan output menggunakan korelasi.

1. Berdasarkan hasil korelasi pada dataset 1, suhu udara jam ke t+2 dipengaruhi oleh suhu udara jam ke t dan suhu udara jam ke t+1, maka input yaitu : suhu udara pada jam ke t dan suhu udara

- pada jam ke $t+1$ serta output yaitu suhu udara pada jam ke $t+2$.
2. Berdasarkan hasil korelasi pada dataset 2, suhu udara hari ke $t+2$ dipengaruhi oleh suhu udara hari ke t dan suhu udara hari ke $t+1$, maka input yaitu : suhu udara pada hari ke t dan suhu udara pada hari ke $t+1$ serta output yaitu suhu udara pada hari ke $t+2$.
 3. Berdasarkan hasil korelasi pada dataset 3, suhu udara hari ke $t+6$ dipengaruhi oleh suhu udara hari ke t dan suhu udara hari ke $t+3$, maka input yaitu : suhu udara pada hari ke t dan suhu udara pada hari ke $t+3$ serta output yaitu suhu udara pada hari ke $t+6$.

Pada penelitian ini partisi data yang akan digunakan untuk data training sebesar 80% dari seluruh data dan data testing sebesar 20% dari seluruh data.

Transformasi Input dan Output ke dalam Range [0.1-0.9] : Fungsi aktivasi yang digunakan dalam proses prediksi adalah fungsi sigmoid biner yang memiliki nilai minimum 0 dan maksimum 1. Namun karena fungsi sigmoid biner tidak pernah menyentuh nilai 0 maupun 1, maka input dan output harus ditransformasikan ke dalam range [0.1-0.9] sebelum proses training dimulai.

Inisialisasi Matriks Bobot : Sebelum proses training dimulai, tentukan matriks bobot awal yang menghubungkan input, hidden layer dan output dengan bilangan acak kecil antara -0,5 sampai 0,5.

Proses Training : Proses training menggunakan data training sebesar 80% dari data keseluruhan. Aturan pemberhentian dalam proses training dapat berupa maksimum epoch atau nilai MSE terkecil telah dicapai.

Matriks Bobot Baru : Proses training pada data training menghasilkan nilai matriks bobot yang baru. Matriks bobot inilah yang akan digunakan pada proses training data testing.

Proses Testing : Proses training menggunakan data testing yang telah ditentukan sebelumnya. Aturan pemberhentian adalah 1 epoch untuk menghitung nilai MSE. Nilai MSE adalah nilai rata-rata dari kuadrat selisih nilai target dengan nilai prediksi. Nilai MSE dapat dihitung dari persamaan (15).

De-normalisasi Data : Setelah nilai MSE terkecil tercapai, lakukan de-normalisasi pada data testing untuk mengembalikan data ke dalam bentuk asalnya sehingga dapat diinterpretasikan (Rahmalia dan Herlambang, 2017).

4. HASIL PENELITIAN

Data yang digunakan dalam penelitian ini adalah data suhu udara (Celcius) di daerah Jakarta selama periode Januari-Juli 2012 sebanyak 200 data. Dari data tersebut akan dihitung korelasi data dan dilakukan simulasi Backpropagation $bppg(V, W)$.

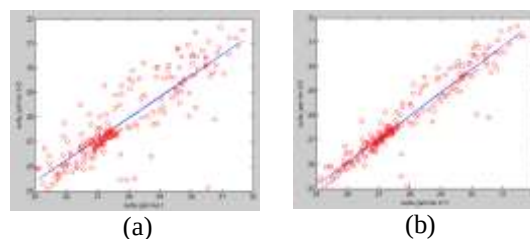
Parameter yang digunakan pada Backpropagation dapat dilihat pada Tabel 1.

Tabel 1. Parameter Backpropagasi

Parameter	Nilai
Maximum epoch	1000
Laju pembelajaran α	0,2
Jumlah hidden layer	2
Data training	80% (160 data)
Data testing	20% (40 data)

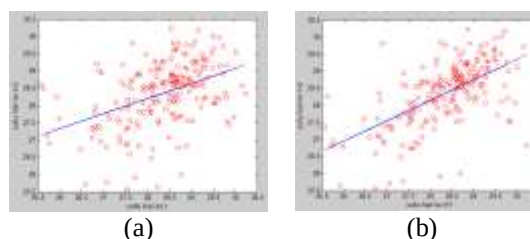
4.1. Korelasi Data

Hasil korelasi pada dataset 1 dapat dilihat pada Gambar 5(a) dan Gambar 5(b). Gambar 5(a) adalah korelasi suhu udara pada jam ke t terhadap suhu udara pada jam ke $t+2$. Nilai korelasi dihitung berdasarkan persamaan (16) sehingga diperoleh nilai korelasi antara suhu udara pada jam ke t terhadap suhu udara pada jam ke $t+2$ adalah 0,8530. Gambar 5(b) adalah korelasi suhu udara pada jam ke $t+1$ terhadap suhu udara pada jam ke $t+2$. Nilai korelasi dihitung berdasarkan persamaan (16) sehingga diperoleh nilai korelasi antara suhu udara pada jam ke $t+1$ terhadap suhu udara pada jam ke $t+2$ adalah 0,9477.



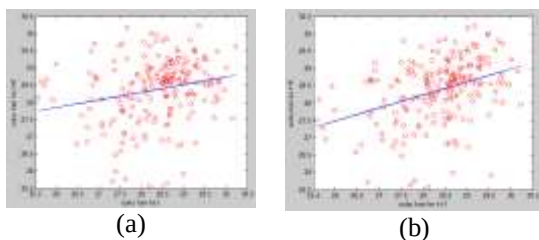
Gambar 5. Korelasi pada Dataset 1

Hasil korelasi pada dataset 2 dapat dilihat pada Gambar 6(a) dan Gambar 6(b). Gambar 6(a) adalah korelasi suhu udara pada hari ke t terhadap suhu udara pada hari ke $t+2$. Nilai korelasi dihitung berdasarkan persamaan (16) sehingga diperoleh nilai korelasi antara suhu udara pada hari ke t terhadap suhu udara pada hari ke $t+2$ adalah 0,4354. Gambar 6(b) adalah korelasi suhu udara pada hari ke $t+1$ terhadap suhu udara pada hari ke $t+2$. Nilai korelasi dihitung berdasarkan persamaan (16) sehingga diperoleh nilai korelasi antara suhu udara pada hari ke $t+1$ terhadap suhu udara pada hari ke $t+2$ adalah 0,5991.



Gambar 6. Korelasi pada Dataset 2

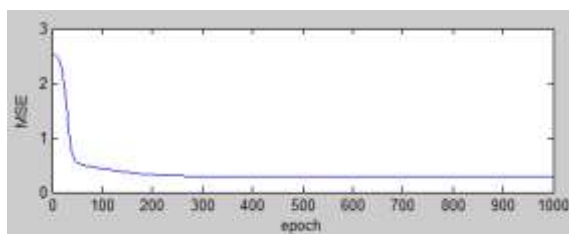
Hasil korelasi pada dataset 3 dapat dilihat pada Gambar 7(a) dan Gambar 7(b). Gambar 7(a) adalah korelasi suhu udara pada hari ke t terhadap suhu udara pada hari ke $t+6$. Nilai korelasi dihitung berdasarkan persamaan (16) sehingga diperoleh nilai korelasi antara suhu udara pada hari ke t terhadap suhu udara pada hari ke $t+6$ adalah 0,2234. Gambar 7(b) adalah korelasi suhu udara pada hari ke $t+3$ terhadap suhu udara pada hari ke $t+6$. Nilai korelasi dihitung berdasarkan persamaan (16) sehingga diperoleh nilai korelasi antara suhu udara pada hari ke $t+3$ terhadap suhu udara pada hari ke $t+6$ adalah 0,3767.



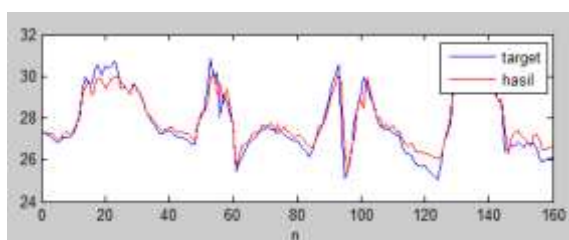
Gambar 7. Korelasi pada Dataset 3

4.2. Simulasi Backpropagation

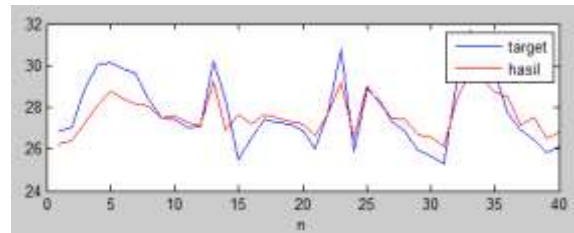
Hasil simulasi Backpropagation pada dataset 1 dapat dilihat pada Gambar 7, Gambar 8, dan Gambar 9. Gambar 7 adalah grafik konvergensi pada proses training sebanyak 1000 epoch pada data training dan diperoleh nilai MSE sebesar 0,0025. Gambar 8 adalah hasil perbandingan antara nilai target dengan hasil perhitungan pada data training. Setelah proses training dilakukan, diperoleh bobot optimal yang akan digunakan pada data testing. Gambar 9 adalah hasil perbandingan antara nilai target dengan hasil perhitungan pada data testing. Pada data testing, diperoleh nilai MSE sebesar 0,0107.



Gambar 7. Konvergensi Proses Training pada Dataset 1

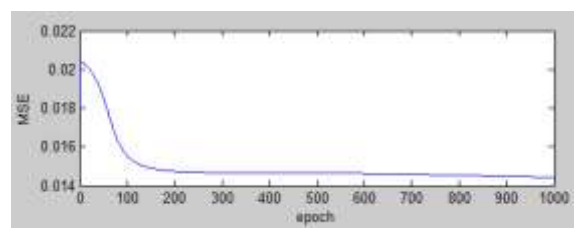


Gambar 8. Perbandingan Nilai Target dan Hasil Perhitungan Data Training pada Dataset 1

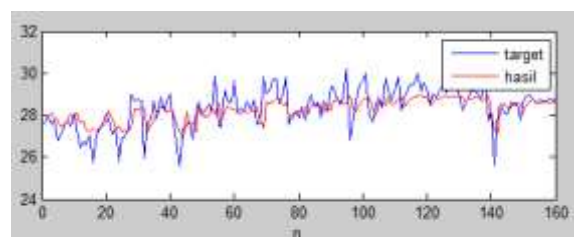


Gambar 9. Perbandingan Nilai Target dan Hasil Perhitungan Data Testing pada Dataset 1

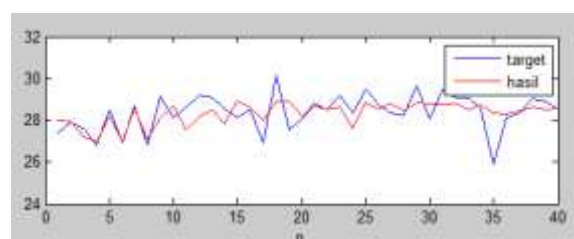
Hasil simulasi Backpropagation pada dataset 2 dapat dilihat pada Gambar 10, Gambar 11, dan Gambar 12. Gambar 10 adalah grafik konvergensi pada proses training sebanyak 1000 epoch pada data training dan diperoleh nilai MSE sebesar 0,0144. Gambar 11 adalah hasil perbandingan antara nilai target dengan hasil perhitungan pada data training. Setelah proses training dilakukan, diperoleh bobot optimal yang akan digunakan pada data testing. Gambar 12 adalah hasil perbandingan antara nilai target dengan hasil perhitungan pada data testing. Pada data testing, diperoleh nilai MSE sebesar 0,0156.



Gambar 10. Konvergensi Proses Training pada Dataset 2



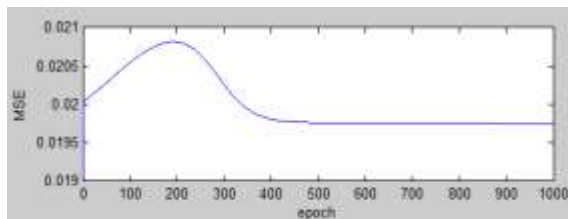
Gambar 11. Perbandingan Nilai Target dan Hasil Perhitungan Data Training pada Dataset 2



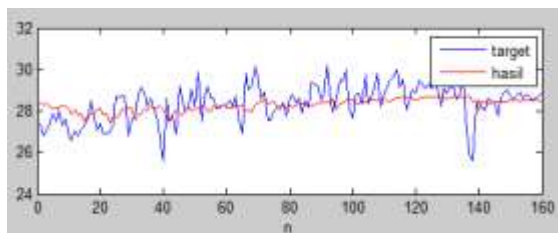
Gambar 12. Perbandingan Nilai Target dan Hasil Perhitungan Data Testing pada Dataset 2

Hasil simulasi Backpropagation pada dataset 3 dapat dilihat pada Gambar 13, Gambar 14, dan

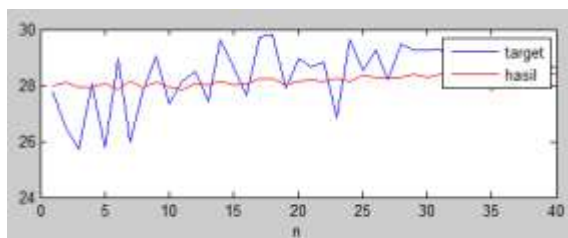
Gambar 15. Gambar 13 adalah grafik konvergensi pada proses training sebanyak 1000 epoch pada data training dan diperoleh nilai MSE sebesar 0,0198. Gambar 14 adalah hasil perbandingan antara nilai target dengan hasil perhitungan pada data training. Setelah proses training dilakukan, diperoleh bobot optimal yang akan digunakan pada data testing. Gambar 15 adalah hasil perbandingan antara nilai target dengan hasil perhitungan pada data testing. Pada data testing, diperoleh nilai MSE sebesar 0,02806.



Gambar 13. Konvergensi Proses Training pada Dataset 3



Gambar 14. Perbandingan Nilai Target dan Hasil Perhitungan Data Training pada Dataset 3



Gambar 15. Perbandingan Nilai Target dan Hasil Perhitungan Data Testing pada Dataset 3

5. KESIMPULAN

Backpropagation merupakan salah satu metode yang digunakan dalam proses peramalan. Cara kerja Backpropagation menyerupai sistem syaraf manusia dimana pada model Backpropagation adalah terdapat hidden layer di antara input dan output. Nilai korelasi memberikan pengaruh pada hasil peramalan menggunakan Backpropagation. Simulasi dilakukan pada tiga dataset dengan struktur yang berbeda menggunakan Backpropagation. Berdasarkan hasil simulasi, data yang memiliki korelasi yang kuat (nilai korelasi mendekati 1) dapat menghasilkan nilai peramalan yang lebih baik yang ditandai dengan nilai MSE lebih kecil.

DAFTAR PUSTAKA

- Matarneh, L.A., Sheta, A., Bani-Ahmad, S., Alshaer, J., dan Al-oqily, I., (2014) *Development of Temperature-based Weather Forecasting Models Using Neural Network and Fuzzy Logic*, International Journal of Multimedia and Obiquitous Engineering Vol. 9 No. 12 pp.343-366.
- Rahmalia, D., (2018) *Estimation of Exponential Smoothing Parameter on Pesticide Characteristic Forecast using Ant Colony Optimization (ACO)*, Eksakta : Jurnal Ilmu-Ilmu MIPA Vol. 18 No. 1 pp.56-63.
- Fausett, L., (1994) *Fundamentals of Neural Networks*, Prentice Hall, USA.
- Han, J., Kamber, M., Pei, J., (2012) *Data Mining Concepts and Techniques*, Elsevier, USA.
- Siang, J.J., (2009) *Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan Matlab*, Penerbit ANDI, Yogyakarta.
- Walpole, R.E., Myers, R.H., Myers, S.L., (2012) *Probability & Statistics for Engineers & Scientists*, Prentice Hall, USA.
- Rahmalia, D., Herlambang, T., (2017) *Prediksi Cuaca Menggunakan Algoritma Particle Swarm Optimization-Neural Network (PSO-NN)*, Prosiding Seminar Nasional Matematika dan Aplikasinya (pp 41-48), Surabaya, Universitas Airlangga