

Optimasi Algoritma Neural Network berbasis Fitur Seleksi Menggunakan Algoritma Genetika Untuk Prediksi Curah Hujan

Nasrul Kamal¹, Yudi Ramdhani²

^{1,2}Program Studi Teknik Informatika, Universitas Adhirajasa Reswara Sanjaya
e-mail: Nasrul.Kamal.911@gmail.com, ²yudi@ars.ac.id

Abstrak

Perubahan iklim dapat di prediksi dengan menganalisis pergerakan cuaca pada waktu tertentu, salah satunya dengan memprediksi curah hujan yang atributnya terdiri dari kelembaban dan angin sebagai acuan untuk digunakan memprediksi curah hujan di lokasi tertentu. Prediksi curah hujan dapat dicapai dengan menggunakan metode klasifikasi pada *Data Mining*. Teknik yang berbeda menghasilkan kinerja yang berbeda pula tergantung pada representasi data curah hujan termasuk representasi untuk pola jangka panjang (bulan) dan pola jangka pendek (harian). Memilih teknik yang tepat untuk durasi curah hujan tertentu merupakan tugas yang menantang. Studi ini menganalisis menggunakan *Neural Network*, setelah membandingkan dengan metode *K-NN*, *Naive Bayes*, *Decision Tree*, *Random Forest*, *Gradient Boost* untuk memprediksi curah hujan menggunakan data BMKG (Badan Meteorologi Klimatologi dan Geofisika). Dataset yang telah dikumpulkan bersumber dari Stasiun Geofisika Bandung, Indonesia. Dengan hasil 77.99% sebelum di optimasi dan 89,74% setelah di optimasi.

Kata kunci—Prediksi curah hujan, klasifikasi, perbandingan metode, optimasi

Abstract

Climate change can be predicted by analyzing weather movements at a certain time, one of which is by predicting rainfall whose attributes consist of humidity and wind as a reference to be used to predict rainfall in certain locations. Rainfall predictions can be achieved using classification methods in Data Mining. Different techniques produce different performance depending on the representation of rainfall data including representations for long-term patterns (months) and short-term patterns (daily). Selecting the appropriate technique for a given rainfall duration is a challenging task. This study analyzes using a Neural Network, after comparing it with the K-NN, Naive Bayes, Decision Tree, Random Forest, Gradient Boost methods to predict rainfall using BMKG (Meteorology, Climatology and Geophysics Agency) data. The dataset that has been collected is sourced from the Bandung Geophysical Station, Indonesia. With results of 77.99% before optimization and 89.74% after optimization.

Keywords— *Rainfall prediction, classification, method comparison, optimization*

Corresponding Author:

Yudi Ramdhani,

Email: yudi@ars.ac.id

1. PENDAHULUAN

Cuaca ekstrem adalah sifat atmosfer yang kacau sehingga cuaca yang berada di wilayah tersebut memiliki intensitas yang tak lazim [1]. Indonesia memiliki dua musim yaitu musim kemarau dan juga musim hujan, puncak musim kemarau terjadi pada bulan Juni, Juli dan Agustus sedangkan puncak musim hujan terjadi pada bulan Desember, Januari dan Februari [2]. Memprediksi curah hujan merupakan tugas yang menantang untuk melakukannya membutuhkan faktor-faktor yang terkait dengan curah hujan seperti angin, kelembaban, dan suhu [3]. Oleh karena itu pengamatan terhadap kondisi cuaca, khususnya kondisi curah hujan sangat penting untuk dilakukan. Besarnya curah hujan yang terjadi tidak dapat ditentukan secara pasti namun bisa diprediksi dengan

menggunakan data historis besarnya curah hujan beberapa waktu yang lalu, maka dapat diprediksi Berapa besarnya curah hujan yang terjadi pada masa yang akan datang. Memprediksi curah hujan merupakan tugas yang menantang untuk melakukannya membutuhkan faktor-faktor yang terkait dengan curah hujan seperti angin, kelembaban, dan suhu[3].

Untuk hasil yang maksimal diperlukan perhitungan dengan daya komputasi yang besar untuk menyelesaikan persamaan yang menggambarkan atmosfer, kesalahan dalam mengukur kondisi awal serta analisis masalah yang tidak lengkap menyebabkan hasil prediksi kurang akurat. Ini menunjukkan bahwa prakiraan cuaca menjadi kurang akurat karena perbedaan waktu saat ini dan waktu prakiraan [4].

Kegunaan teknik *Machine learning* biasanya digunakan untuk keperluan menganalisis data set besar, menemukan dan mengekstraksi sehingga mendapatkan pola dan pengetahuan dari data tersebut [5] *Data Mining* merupakan suatu proses menemukan pola dan korelasi pada dataset besar dan dituntut untuk mengambil keputusan yang tepat dalam memprediksi hasil [6].

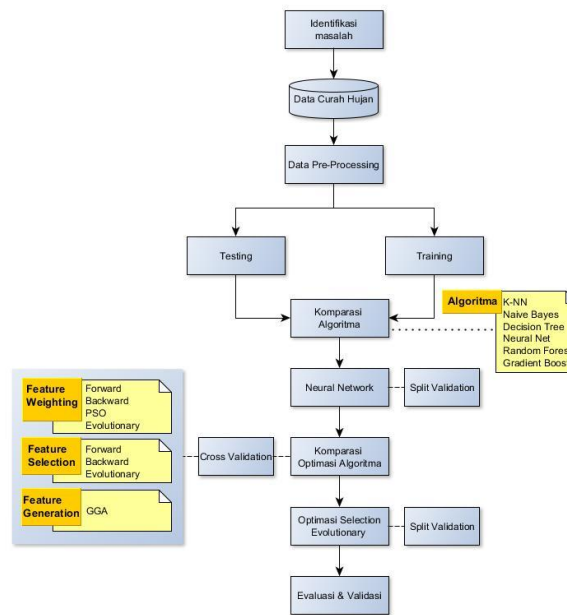
Berbagai teknik seperti *K-NN*, *Naive Bayes*, *Decision Tree*, *Random Forest* dan *Gradient Boost Tree* telah dianalisis untuk prediksi curah hujan, dan hasilnya *Neural Network* memiliki akurasi yang paling tinggi karena kekuatan dan keunggulan dari *Neural Network* terletak pada kemampuan untuk mewakili hubungan linier dan non linier langsung dari data yang dimodelkan, sedangkan model linier tradisional sama sekali tidak memadai jika menyangkut data pemodelan sejati yang mengandung karakteristik non linier, *Neural Network* merupakan sistem pemrosesan informasi yang meniru cara kerja jaringan manusia yang terdiri dari beberapa neuron dan ada hubungan antara neuron neuron tersebut.

Model *Neural Network* adalah struktur yang dapat disesuaikan untuk menghasilkan pemetaan dari kumpulan data tertentu ke fitur atau hubungan antar data. Model disesuaikan, atau dilatih, menggunakan kumpulan data dari sumber tertentu sebagai input, biasanya disebut sebagai set pelatihan. Setelah pelatihan berhasil, *Neural Network* akan mampu melakukan klasifikasi, estimasi, prediksi, atau simulasi pada data baru dari sumber yang sama atau serupa.

2. METODE PENELITIAN

Menurut Hidayat dan Sedarmayanti merupakan pembahasan tentang konsep teoritik berbagai metode dengan kelebihan dan juga kekurangan yang dalam karya ilmiah dilanjutkan dengan pemilihan metode yang digunakan [18]. Pra-pemrosesan data adalah salah satu fase utama dalam proses penemuan pengetahuan yang melibatkan lebih banyak usaha dan waktu dalam keseluruhan proses analisis data. Ini dapat dianggap sebagai salah satu teknik data mining [19].

Dalam penelitian ini akan memilih dan menerapkan teknik yang tepat untuk klasifikasi curah hujan. Tahap pertama dalam penelitian ini adalah membagi dataset curah hujan dengan membagi dua yaitu, data training dan data testing. Langkah selanjutnya adalah melakukan seleksi fitur terbaik pada dataset curah hujan menggunakan algoritma *Neural Network*. Seleksi fitur yang digunakan dalam penelitian menggunakan Algoritma Evolutionary membuat populasi terdiri dari banyak individu yang dipilih dengan paling banyak nilai yang relevan dengan klasifikasi [20]. Kemudian fitur yang telah diseleksi dilakukan klasifikasi menggunakan algoritma *Evolutionary*. Tahapan penelitian yang dilakukan ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

Pada tahap ini ditampilkan hasil eksperimen klasifikasi dataset curah hujan. Tahap awal yang dilakukan yaitu mengidentifikasi masalah. Diketahui bahwa dalam mengklasifikasikan curah hujan diperlukan suatu metode atau algoritma dengan model terbaik. Berdasarkan hal tersebut maka dilakukan penelitian mengenai klasifikasi curah hujan.

Penelitian ini menggunakan data penelitian BMKG Stasiun Bandung dataset yang didapatkan dari website BMKG. Dataset dalam penelitian ini memiliki 396 data record Curah hujan yang terdiri dari 9 atribut dan 1 label yang terdiri dari 3 *class* yaitu hujan, tidak hujan dan hujan ekstrem.

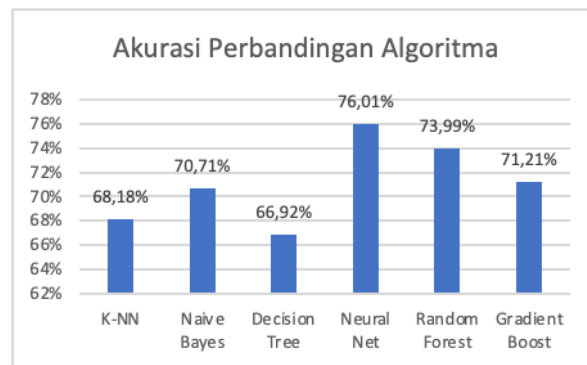
Setelah data dikumpulkan, peneliti melakukan *pre-processing* data. Pada tahap ini untuk dilakukan pengecekan missing value terhadap data tujuannya agar terlihat apakah terdapat data yang tidak sesuai. Selanjutnya melakukan remove duplicate agar tidak ada data yang sama, dan melakukan normalisasi dengan *method Z-Transformation*. Data yang dinormalisasi tersebut akan memiliki rentang nilai yang sama yaitu antara 0 sampai 1. Berikut ini adalah hasil dari normalisasi yang telah dilakukan.

Setelah proses *preprocessing* data selesai, maka langkah selanjutnya adalah melakukan komparasi algoritma. Komparasi algoritma dilakukan untuk membandingkan 6 algoritma yang diuji pada penelitian ini. Algoritma yang digunakan yaitu; *K-NN*, *Naive Bayes*, *Decision Tree*, *Random Forest* dan *Gradient Boost Tree*. Untuk mengetahui performa kinerja yang dihasilkan dari 6 algoritma tersebut, maka proses Validasi data dilakukan dengan menggunakan *method 10-Fold validation* yang menghasilkan nilai *Accuracy*, *Precision* dan *Recall*. Berikut ini adalah nilai *Accuracy* yang dihasilkan oleh masing-masing algoritma.

Tabel 1. Komparasi Algoritma

Algoritma	Validasi	Ratio	Accuracy
<i>K-NN</i>	<i>Cross Validation</i>	-	68.18%
<i>Naive Bayes</i>	<i>Cross Validation</i>	-	70.71%
<i>Decision Tree</i>	<i>Cross Validation</i>	-	66.92%
Neural Net	<i>Cross Validation</i>	-	76.01%
<i>Random Forest</i>	<i>Cross Validation</i>	-	73.99%
Gradient Boost	<i>Cross Validation</i>	-	71.21%
Rata2			71,17%

Untuk mempermudah dalam memahami perbedaan akurasi dari Performa dalam komparasi algoritma maka perlu dibuatkan grafik. Berikut adalah grafik 6 algoritma dalam proses komparasi algoritma.



Gambar 2. Komparasi Algoritma

Berdasarkan komparasi algoritma tersebut diketahui bahwa algoritma *Neural Network* memiliki nilai performa algoritma tertinggi dibandingkan algoritma yang lainnya yaitu sebesar 76,01% untuk akurasi. Berikut tabel *Confusion Matrix* yang dihasilkan klasifikasi algoritma *Neural Network* dapat dilihat pada Tabel 2.

Tabel 2. *Confusion Matrix Neural Network*

	true Hujan	true Hujan Ekstrem	true Tidak Hujan	class precision
pred. Hujan	125	18	33	71.02%
pred. Hujan Ekstrem	0	0	0	0.00%
pred. Tidak Hujan	29	15	176	80.00%
class recall	81.17%	0.00%	84.21%	

Accuracy

Dari hasil pengujian nilai akurasi yaitu 76,01%. Berikut perhitungan manual akurasi.

$$\begin{aligned}
 \text{Accuracy} &= \frac{\text{THujan} + \text{THujan Ekstrem} + \text{TTidak Hujan}}{\text{THujan} + \text{THujan Ekstrem} + \text{TTidak Hujan} + \text{FHujan} + \text{FHujan Ekstrem} + \text{FTidak Hujan}} * 100\% \\
 &= \frac{125 + 0 + 176}{125 + 176 + 18 + 33 + 29} * 100\% \\
 &= \frac{301}{396} * 100\%
 \end{aligned}$$

$$= 0,7601 * 100\%$$

$$= 76,01\%$$

Precision

Dari hasil pengujian nilai presisi yang diperoleh yaitu 85,08% untuk *class* Hujan dan 89,39% untuk *class* Tidak Hujan. Berikut perhitungan presisinya.

$$\text{Hujan Precision} = \frac{\text{THujan}}{\text{THujan} + \text{FHujan}} * 100\%$$

$$= \frac{125}{125 + 18 + 33} * 100\%$$

$$= \frac{125}{176} * 100\%$$

$$= 0,7102 * 100\%$$

$$= 71,02\%$$

$$\text{Tidak Hujan Precision} = \frac{\text{TTidakHujan}}{\text{TTidakHujan} + \text{FTidakHujan}} * 100\%$$

$$= \frac{176}{176 + 15 + 17} * 100\%$$

$$= \frac{176}{208} * 100\%$$

$$= 0,8 * 100\%$$

$$= 80,00\%$$

Recall

Dari hasil pengujian nilai *Recall* yang diperoleh yaitu 90,00% untuk *class* Hujan dan 84,22% untuk *class* Tidak Hujan. Berikut perhitungan *Recall*nya.

$$\text{Hujan Recall} = \frac{\text{THujan}}{\text{THujan} + \text{FTidakHujan}} * 100\%$$

$$= \frac{125}{125 + 0 + 20} * 100\%$$

$$= \frac{125}{154} * 100\%$$

$$= 0,8117 * 100\%$$

$$= 81,17\%$$

$$\text{Tidak Hujan Recall} = \frac{\text{TTidakHujan}}{\text{TTidakHujan} + \text{FHujan}} * 100\%$$

$$= \frac{176}{176 + 0 + 33} * 100\%$$

$$= \frac{176}{209} * 100\%$$

$$= 0,8421 * 100\%$$

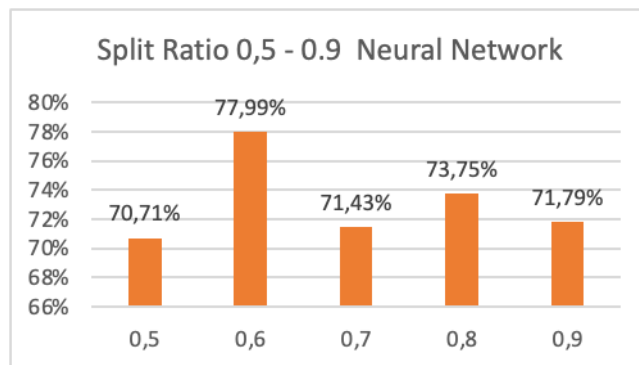
$$= 84,21\%$$

Setelah diketahui performa kinerja terbaik algoritma *Neural Network* dalam melakukan klasifikasi Curah hujan, dilakukan validasi data menggunakan *Split Validation* untuk menguji algoritma tersebut. Berikut ini adalah hasil dari validasi data menggunakan *Split Validation* dengan split ratio 0,5 hingga 0,9 yang terdapat pada Tabel 3.

Tabel 3. Split Ratio 0,5 - 0.9 *Neural Network*

Algoritma	Validasi	Ratio	Accuracy
<i>Neural Network</i>	<i>Split Validation</i>	0.5	70.71%
<i>Neural Network</i>	<i>Split Validation</i>	0.6	77.99%
<i>Neural Network</i>	<i>Split Validation</i>	0.7	71.43%
<i>Neural Network</i>	<i>Split Validation</i>	0.8	73.75%
<i>Neural Network</i>	<i>Split Validation</i>	0.9	71.79%
Average			73.13%

Berikut ini adalah grafik pengujian algoritma *Neural Network* dengan split ratio 0,5 hingga 0,9.



Gambar 3. Split Ratio 0,5-0,9 *Neural Network*

Berdasarkan Tabel IV.4 dapat diketahui bahwa algoritma *Neural Network* dengan split ratio 0,5 hingga 0,9 memiliki nilai average 73.13% untuk akurasi. Validasi dengan split ratio 0,6 memiliki nilai *Accuracy* tertinggi sebesar 77,99%. Berikut tabel *Confusion Matrix* yang dihasilkan dapat dilihat pada Tabel 4.

Tabel 4. *Confusion Matrix* Split Ratio 0,6 *Neural Network*

	true Hujan	true Hujan Ekstrem	true Tidak Hujan	class precision
pred. Hujan	54	9	14	70.13%
pred. Hujan Ekstrem	0	0	0	0.00%
pred. Tidak Hujan	8	4	70	85.37%
class recall	87.10%	0.00%	83.33%	

Accuracy

Dari hasil pengujian nilai akurasi yaitu 77,99%. Berikut perhitungan manual akurasinya.

$$\begin{aligned}
 Accuracy &= \frac{THujan+TTidakHujan+THujanEkstrem}{THujan+TTidakHujan+THujanEkstrem+FHujan+FTidakHujan+FHujanEkstrem} * 100\% \\
 &= \frac{54+0+70}{54+0+70+9+0+0+8+4+14} * 100\% \\
 &= \frac{124}{158} * 100\%
 \end{aligned}$$

$$= 0,7798 * 100\%$$

$$= 77,98\%$$

Precision

Dari hasil pengujian nilai presisi yang diperoleh yaitu 70,13% untuk *class* Hujan dan 85,37% untuk *class* Tidak Hujan. Berikut perhitungan presisinya.

$$\text{Hujan Precision} = \frac{\text{THujan}}{\text{THujan}+\text{FHujan}} * 100\%$$

$$= \frac{54}{54+9+14} * 100\%$$

$$= \frac{54}{77} * 100\%$$

$$= 0,7013 * 100\%$$

$$= 70,13\%$$

$$\text{Tidak Hujan Precision} = \frac{\text{TTidakHujan}}{\text{TTidakHujan}+\text{FTidakHujan}} * 100\%$$

$$= \frac{70}{70+8+4} * 100\%$$

$$= \frac{70}{82} * 100\%$$

$$= 0,8537 * 100\%$$

$$= 85,37\%$$

Recall

Dari hasil pengujian nilai *Recall* yang diperoleh yaitu 87,10% untuk *class* Hujan dan 83,33% untuk *class* Tidak Hujan. Berikut perhitungan *Recall*nya.

$$\text{Hujan Recall} = \frac{\text{THujan}}{\text{THujan}+\text{FTidakHujan}} * 100\%$$

$$= \frac{54}{54+0+8} * 100\%$$

$$= \frac{54}{62} * 100\%$$

$$= 0,8710 * 100\%$$

$$= 87,10\%$$

$$\text{Tidak Hujan Recall} = \frac{\text{TTidakHujan}}{\text{TTidakHujan}+\text{FHujan}} * 100\%$$

$$= \frac{70}{70+0+84} * 100\%$$

$$= \frac{70}{84} * 100\%$$

$$= 0,8333 * 100\%$$

$$= 83,33\%$$

Untuk meningkatkan nilai akurasi pada algoritma *Neural Network* maka digunakan fitur optimasi, fitur *Weight* dan fitur generation. Pada penelitian ini melakukan komparasi fitur optimasi yaitu *Optimize Selection*. Validasi yang dilakukan menggunakan *method* 10-Fold validation atau *Cross validation*. Berikut ini adalah nilai akurasi dari setiap fitur optimasi yang terdapat pada Tabel 5.

Tabel 5. *Neural Network + Feature Optimize Selection*

Algoritma	Metode	Validasi	Accuracy
Neural Network	Forward Selection	Cross Validation	75.25%
Neural Network	Backward Selection	Cross Validation	74.49%
Neural Network	Evolutionary	Cross Validation	81.13%

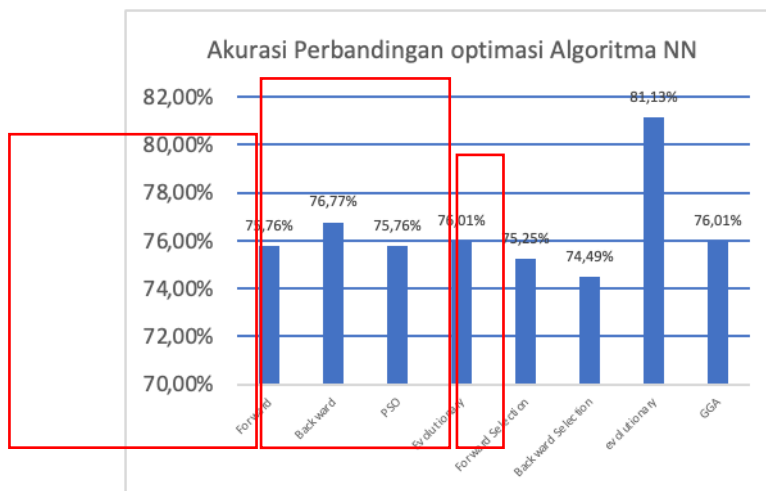
Tabel 6. *Neural Network + Feature Optimize Weight*

Algoritma	Metode	Validasi	Accuracy
Neural Network	Forward	Cross Validation	75.76%
Neural Network	Backward	Cross Validation	76.77%
Neural Network	PSO	Cross Validation	75.76%
Neural Network	Evolutionary	Cross Validation	76.01%

Tabel 7. *Neural Network + Feature Optimize Generation*

Algoritma	Metode	Validasi	Accuracy
Neural Network	GGA	Cross Validation	76.01%

Berikut ini adalah grafik Accuracy dari setiap fitur optimasi.



Gambar 4. Accuracy Neural Network + Feature Optimize

Berdasarkan gambar 4 dapat diketahui bahwa *Optimize Selection* dan *Optimize Selection* telah berhasil untuk meningkatkan nilai akurasi algoritma *Neural Network* dalam klasifikasi Curah hujan. Fitur *Optimize Selection* dan *Optimize Weight* dengan *Neural Network* memiliki nilai akurasi yang tertinggi dibandingkan dengan fitur optimasi metode lainnya, selain itu nilai akurasi yang dihasilkan memiliki nilai yang sama besar yaitu 81,13%. Berdasarkan hal

tersebut, peneliti memilih untuk untuk menguji algoritma *Neural Network* berbasis *Optimize Selection Evolutionary* dalam mengklasifikasi Curah hujan.

Dataset curah hujan dibagi menjadi data training dan data testing menggunakan *Split Validation* dengan split ratio 0,5 hingga 0,9. Untuk mengetahui nilai akurasi, *Precision* dan *Recall*, model seleksi yang dihasilkan dari algoritma *Neural Network* berbasis *Optimize Selection Evolutionary* dilakukan menggunakan Rapidminer.

1. Split Ratio 0,5

Dalam pembagian dataset dengan parameter split ratio 0,5 menghasilkan nilai akurasi sebesar 78.79%. Nilai tersebut termasuk klasifikasi kurang bagus. Berikut tabel *Confusion Matrix* yang dihasilkan dapat dilihat pada Tabel 8.

Tabel 8. *Confusion Matrix Split Ratio 0,5
Neural Network + Optimize Selection Evolutionary*

accuracy: 78.79%

	true Hujan	true Hujan Ekstrem	true Tidak Hujan	class precision
pred. Hujan	66	7	15	75.00%
pred. Hujan Ekstrem	0	0	0	0.00%
pred. Tidak Hujan	11	9	90	81.82%
class recall	85.71%	0.00%	85.71%	

2. Split Ratio 0,6

Dalam pembagian dataset dengan parameter split ratio 0,6 menghasilkan nilai akurasi sebesar 80.50%. Nilai tersebut termasuk klasifikasi bagus. Berikut tabel *Confusion Matrix* yang dihasilkan dapat dilihat pada Tabel 9.

Tabel 9. *Confusion Matrix Split Ratio 0,6
Neural Network + Optimize Selection GA*

	true Hujan	true Hujan Ekstrem	true Tidak Hujan	class precision
pred. Hujan	55	6	11	76.39%
pred. Hujan Ekstrem	0	0	0	0.00%
pred. Tidak Hujan	7	7	73	83.91%
class recall	88.71%	0.00%	86.90%	

3. Split Ratio 0,7

Dalam pembagian dataset dengan parameter split ratio 0,7 menghasilkan nilai akurasi sebesar 84,03%. Nilai tersebut termasuk klasifikasi bagus sekali. Berikut tabel *Confusion Matrix* yang dihasilkan dapat dilihat pada Tabel 10.

Tabel 10. *Confusion Matrix Split Ratio 0,7
Neural Network + Optimize Selection GA*

	true Hujan	true Hujan Ekstrem	true Tidak Hujan	class precision
pred. Hujan	42	6	5	79.25%
pred. Hujan Ekstrem	0	0	0	0.00%
pred. Tidak Hujan	4	4	58	87.88%
class recall	91.30%	0.00%	92.06%	

4. Split Ratio 0,8

Dalam pembagian dataset dengan parameter split ratio 0,8 menghasilkan nilai akurasi sebesar 85,00%. Nilai tersebut termasuk klasifikasi bagus sekali. Berikut tabel *Confusion Matrix* yang dihasilkan dapat dilihat pada Tabel 11.

Tabel 11. *Confusion Matrix* Split Ratio 0,8
Neural Network + Optimize Selection GA

	true Hujan	true Hujan Ekstream	true Tidak Hujan	class precision
pred. Hujan	55	6	11	76.39%
pred. Hujan Ekstream	0	0	0	0.00%
pred. Tidak Hujan	7	7	73	83.91%
class recall	88.71%	0.00%	86.90%	

5. Split Ratio 0,9

Dalam pembagian dataset dengan parameter split ratio 0,9 menghasilkan nilai akurasi sebesar 89,74%. Nilai tersebut termasuk klasifikasi sangat bagus sekali. Berikut tabel *Confusion Matrix* yang dihasilkan dapat dilihat pada Tabel 12.

Tabel 12. *Confusion Matrix* Split Ratio 0,9
Neural Network + Optimize Selection GA

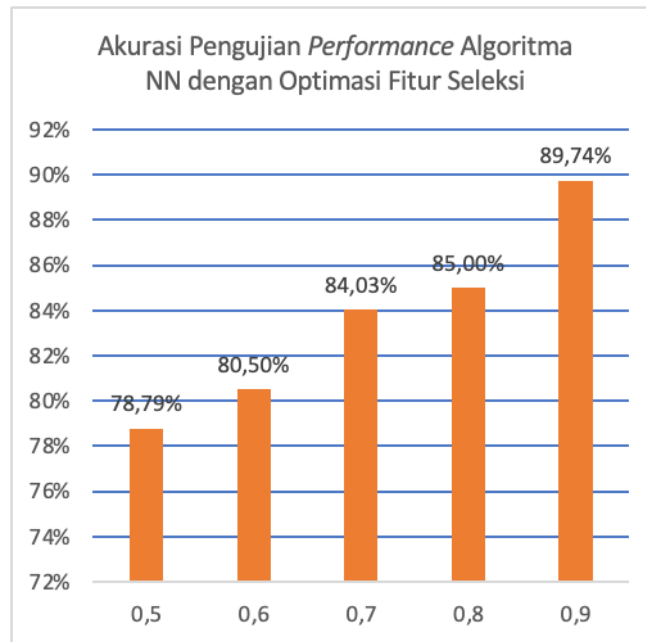
	true Hujan	true Hujan Ekstream	true Tidak Hujan	class precision
pred. Hujan	15	2	1	83.33%
pred. Hujan Ekstream	0	0	0	0.00%
pred. Tidak Hujan	0	1	20	95.24%
class recall	100.00%	0.00%	95.24%	

Berdasarkan atribut yang dihasilkan dari split ratio 0,9 terdapat 5 atribut yang berpengaruh tinggi terhadap *Accuracy* klasifikasi curah hujan. yaitu atribut . Temperatur Rata-rata, Kelembapan Rata-rata, Lamanya penyinaran matahari, Kecepatan angin rata-rata

Berikut ini adalah tabel keseluruhan nilai Akurasi dari setiap parameter split ratio 0,5 hingga 0,9 algoritma *Neural Network* berbasis *Optimize Selection Neural Network* dapat dilihat pada Tabel 13.

Tabel 13. Akurasi *Neural Network + Optimize Selection Evolutionary*

Algoritma	Metode	Ratio	Accuracy
<i>Neural Network</i>	<i>Evolutionary</i>	0.5	78.79%
<i>Neural Network</i>	<i>Evolutionary</i>	0.6	80.50%
<i>Neural Network</i>	<i>Evolutionary</i>	0.7	84.03%
<i>Neural Network</i>	<i>Evolutionary</i>	0.8	85.00%
<i>Neural Network</i>	<i>Evolutionary</i>	0.9	89.74%
Average			83.61%



Gambar 5. Split ratio 0,5 – 0,9
Neural Network + Optimize Selection Evolutionary

Berikut ini adalah grafik pengujian algoritma *Neural Network* berbasis *Optimize Selection Neural Network* dengan split ratio 0,5 hingga 0,9.

Berdasarkan Tabel 14 diketahui bahwa algoritma *Neural Network* berbasis *Optimize Selection Neural Network* dengan split ratio 0,5 hingga 0,9 memiliki nilai rata-rata 91,56%, Validasi dengan split ratio 0,9 memiliki hasil akhir yang tinggi dibandingkan split ratio lainnya. Berikut ini adalah perhitungan manualnya.

1. Accuracy

Dari hasil pengujian nilai akurasi yaitu 89,74%. Berikut perhitungan akurasinya.

$$\begin{aligned}
 Accuracy &= \frac{THujan+TTidakHujan}{THujan+TTidakHujan+FHujan+FTidakHujan} * 100\% \\
 &= \frac{15+20}{15+2+1+0+0+0+0+1+20} * 100\% \\
 &= \frac{35}{39} * 100\% \\
 &= 0,8974 * 100\% \\
 &= 89,74\%
 \end{aligned}$$

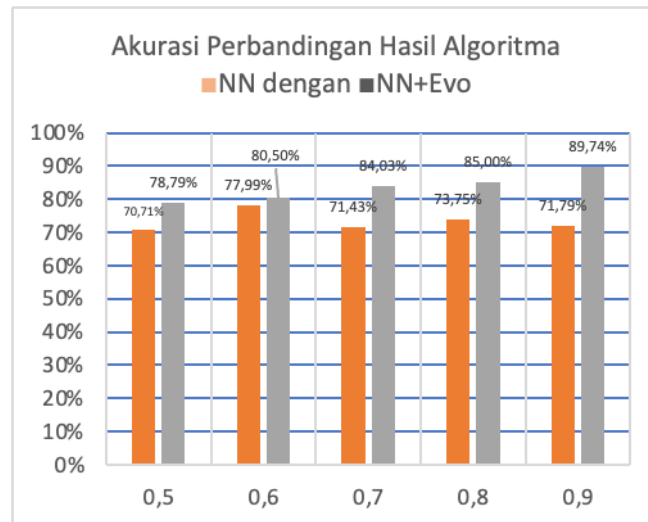
Tabel 14. Accuracy Neural Network dan Neural Network + Optimize Selection Evolutionary

Validasi	Akurasi	
	algoritma NN	Algoritma NN +Evolutionary
0.5	70.71%	78,79%
0.6	77.99%	80,50%
0.7	71.43%	84,03%
0.8	73.75%	85,00%

0,9	71,79%	89,74%
-----	--------	--------

Tabel 14. adalah tabel perbandingan algoritma *Neural Network* dan *Neural Network* berbasis *Optimize Selection Evolutionary* dengan split ratio 0,5 hingga 0,9.

Untuk mempermudah dalam memahami perbedaan akurasi dari hasil perhitungan *Confusion Matrix* dan Performa metode *Neural Network* Berbasis *Optimize Selection Evolutionary* split ratio 0,9 maka perlu dibuatkan grafik. Berikut adalah Gambar grafik perbandingan nilai akurasinya.



Gambar 6. Accuracy Split ratio 0,5 – 0,9 *Neural Network* dan *Neural Network + Optimize Selection Evolutionary*

Setelah menguji algoritma *Neural Network* dengan *Neural Network* Berbasis *Optimize Selection Evolutionary* maka langkah terakhir yang akan dilakukan yaitu melakukan *T-Test Paired Two Samples*. *T-Test* tersebut dilakukan peneliti untuk mengetahui apakah terdapat perbedaan rata-rata nilai Curah hujan sebelum dilakukan optimasi dan setelah dilakukan optimasi. Berikut ini adalah hasil dari *T-Test Paired Two Samples* menggunakan *Microsoft Excel*.

Tabel 15. T-Test Paired Two Samples

	Variable 1	Variable 2
Mean	0.73134	0.83612
Variance	0.000863648	0.001814747
Observations	5	5
Pearson Correlation	-0.22997963	
Hypothesized Mean Difference	0	
df	4	
t Stat	-4.107139691	
P(T<=t) one-tail	0.007383303	
t Critical one-tail	2.131846786	
P(T<=t) two-tail	0.014766606	
t Critical two-tail	2.776445105	

Berdasarkan *T-Test* yang terdapat pada Tabel 15 dapat dibandingkan antara nilai signifikansi dengan 0,05. Nilai signifikansi yang dihasilkan pada *T-Test* tersebut adalah 0,014766606 lebih kecil dari pada 0,05 artinya terdapat perbedaan antara sebelum *Optimize* dan setelah *optimize*.

Pembahasan terhadap hasil penelitian dan pengujian yang diperoleh disajikan dalam bentuk uraian teoritik, baik secara kualitatif maupun kuantitatif. Hasil dapat disajikan dalam angka, grafik, tabel, dan lain-lain yang membuat pembaca memahami dengan mudah. Pada bagian ini ditekankan nilai baru dari penelitian yang memuat inovasi, serta implikasinya. Pembahasan dapat dibuat dalam beberapa sub-bab.

4. KESIMPULAN

Penelitian ini dilakukan menggunakan algoritma *Neural Network* berbasis *Optimize Selection Evolutionary* dengan menggunakan dataset curah hujan yang didapatkan dari web BMKG. Validasi data dengan 10-Fold-Validation atau *Cross validation* pada Algoritma *Neural Network* memiliki tingkat performa terbaik diantara 5 algoritma lainnya seperti *K-NN*, *Naive Bayes*, *Decision Tree*, *Random Forest* dan *Gradient Boost Tree*. Algoritma *Neural Network* menghasilkan nilai akurasi sebesar 76.01% dari rata-rata akurasi 71,17%. Untuk menguji algoritma *Neural Network* menggunakan operator *Split Validation* dengan ratio 0,5 hingga 0,9. Berdasarkan pengujian yang telah dilakukan, pada nilai rata-rata akurasi menghasilkan 73,13%, dan data dengan rasion 0,6 memiliki tingkat akurasi lebih tinggi yaitu 77.99%. Untuk mengoptimalkan nilai akurasi *Neural Network* maka dilakukan dengan menggunakan fitur *Selection* dengan metode *Evolutionary* dengan menggunakan *Cross validation* menghasilkan nilai akurasi 81,13%. Lalu di optimasi lagi dengan *Split Validation* dengan ratio 0,5 hingga 0,9, berdasarkan pengujian yang telah dilakukan nilai rata-rata akurasi yang di hasilkan adalah 83.61%.

DAFTAR PUSTAKA

- [1] J. Romm, *Extreme Weather and Climate Change*. 2018. doi: 10.1093/wentk/9780190866112.003.0002.
- [2] S. Marpaung *et al.*, "Analysis of water productivity in the Banda Sea based on remote sensing satellite data," *Int. J. Remote Sens. Earth Sci.*, vol. 17, no. 1, pp. 25–34, 2021.
- [3] S. Zainudin, D. S. Jasim, and A. A. Bakar, "Comparative analysis of data mining techniques for Malaysian rainfall prediction," *Int. J. Adv. Sci. Eng. Inf. Technol*, vol. 6, no. 6, pp. 1148–1153, 2016.
- [4] S. S. Baboo and I. K. Shereef, "ANN_AnEfficientWeather_ForecastingSystem," *Int. J. Environ. Sci. Dev.*, vol. 1, no. 4, pp. 321–326, 2010.
- [5] A. Supriyatna and W. P. Mustika, "Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kulit," *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 2, no. 2, pp. 152–161, 2018.
- [6] H. Annur, "Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering," *J. Inform. Upgris*, vol. 5, no. 1, 2019, doi: 10.26877/jiu.v5i1.3091.
- [7] Y. Firmansyah, "TA: ANALISIS DATA CURAH HUJAN YANG HILANG DENGAN MENGGUNAKAN METODE RASIONAL DAN METODE INVERSED SQUARE DISTANCE DI STASIUN CIPAKU, CIPEUSING, DAN CISAMPIH." Institut Teknologi Nasional, 2021.
- [8] Triatmodjo, "(Triatmodjo, 2008)," *Hidrol. Terap. Cetakan Pertama*, pp. 4–16, 2008.
- [9] J. Suntoro, *DATA MINING: Algoritma dan Implementasi dengan Pemrograman php*. Elex Media Komputindo, 2019.

- [10] H. Sulastri and A. I. Gufroni, "Penerapan data mining dalam pengelompokan penderita thalassaemia," *J. Nas. Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 299–305, 2017.
- [11] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation.," *Encycl. database Syst.*, vol. 5, pp. 532–538, 2009.
- [12] D. Normawati and D. P. Ismi, "K-Fold Cross Validation for Selection of Cardiovascular Disease Diagnosis Features by Applying Rule-Based Datamining," *Signal Image Process. Lett.*, vol. 1, no. 2, pp. 23–35, 2019, doi: 10.31763/simple.v1i2.3.
- [13] A. P. Windarto *et al.*, *Jaringan Saraf Tiruan: Algoritma Prediksi dan Implementasi*. Yayasan Kita Menulis, 2020.
- [14] M. Ahsan and W. Harianto, "KOMPARASI TINGKAT AKURASI INFORMATION GAIN DAN GAIN RATIO PADA METODE K-NEAREST NEIGHBOR," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 1, pp. 386–391, 2022.
- [15] I Gede Agus Widyadana and Lala Febriana, "PENERAPAN EVOLUTIONARY ALGORITHM PADA PENJADWALAN PRODUKSI (Studi Kasus di PT Brother Silver Product Indonesia)," *J. Tek. Ind.*, vol. 3, no. 2, pp. 43–47, 2001, [Online]. Available: <http://puslit2.petra.ac.id/ejournal/index.php/ind/article/view/15999>
- [16] X. Yu and M. Gen, *Introduction to evolutionary algorithms*. Springer Science & Business Media, 2010.
- [17] P. Ristoski, C. Bizer, and H. Paulheim, "Mining the Web of Linked Data with RapidMiner," *J. Web Semant.*, vol. 35, pp. 142–151, 2015, doi: 10.1016/j.websem.2015.06.004.
- [18] H. Sedarmayanti, "Syarifudin.(2002)," *Metodol. Penelitian. Bandung Mandar Maju*, 2022.
- [19] A. E. Hassanien and A. Darwish, *Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges*. Springer International Publishing, 2020. [Online]. Available: <https://books.google.co.id/books?id=qfsOEAAAQBAJ>
- [20] A. PSW, "Perbandingan Jaringan Syaraf Tiruan Backpropagation Dan Metode Deret Berkala Box-Jenkins (Arima) Sebagai Metode Peramalan Curah Hujan," *Fak. Mat. dan Ilmu Pengetah. Alam. Univ. Negeri Semarang. Semarang*, 2007.