

Prediksi Kejadian Banjir dengan Ensemble Machine Learning Menggunakan BP-NN dan SVM

Flood Prediction with Ensemble Machine Learning using BP-NN and SVM

Ike Fitriyaningsih^{*)}, Yuniarta Basani

Program Studi D3 Teknik Informatika, Fakultas Teknik Informatika dan Elektro, Institut Teknologi Del
Jalan Sisingamangaraja, Sitoluama, Laguboti, Toba Samosir, Indonesia 22381

Cara sitasi: I. Fitriyaningsih and Y. Basani, "Prediksi Kejadian Banjir dengan Ensemble Machine Learning menggunakan BP-NN dan SVM," *Jurnal Teknologi dan Sistem Komputer*, vol. 7, no. 3, pp. 93-97, 2019. doi: 10.14710/jtsiskom.7.3.2019.93-97, [Online].

Abstract - This study aims to examine the prediction of rainfall and river water debit using the Back Propagation Neural Network (BP-NN) method. Prediction results are classified using the Support Vector Machine (SVM) method to predict flooding. The parameters used to predict rainfall with BP-NN are minimum, maximum and average temperature, average relative humidity, sunshine duration, and average wind speed. The debit of Ular Pulau Tagor river is predicted by BP-NN. BPNN and SVM modeling using software R. Daily climate data from 2015-2017 were taken from three stations, namely Sampali climatology station, Kualanamu meteorological station, and Tuntung geophysics station. Prediction of river water debit is for 6 days and 30 days in the future. The best dataset is a 6-day prediction with a combination of 60% training and 40% testing. Flood prediction accuracy with SVM was 100% in predicting flood events for the next 6 days.

Keywords – ensemble machine learning; flood prediction, BP-NN and SVM; rainfall prediction; river water debit prediction

Abstrak - Penelitian ini bertujuan untuk mengkaji prediksi curah hujan dan debit air sungai menggunakan metode Back Propagation Neural Network (BP-NN). Hasil prediksi diklasifikasikan menggunakan metode Support Vector Machine (SVM) untuk memprediksi banjir. Parameter yang digunakan untuk memprediksi curah hujan dengan BP-NN adalah suhu minimum, maksimum dan rata-rata, kelembaban relatif rata-rata, lama penyinaran matahari, dan kecepatan angin rata-rata. Data debit air sungai Ular Pulau Tagor diprediksi dengan BP-NN. Pemodelan BP-NN dan SVM menggunakan software R. Data iklim harian 2015-2017 berasal dari tiga stasiun, yaitu stasiun klimatologi Sampali, stasiun meteorologi Kualanamu, dan stasiun geofisika Tuntungan. Prediksi debit air sungai adalah 6 hari dan 30 hari ke depan. Dataset terbaik adalah prediksi

6 hari dengan kombinasi training 60% dan testing 40%. Prediksi banjir dengan SVM secara akurat 100% dapat menyatakan kejadian banjir selama 6 hari kedepan.

Kata kunci – ensemble machine learning; prediksi banjir; BP-NN dan SVM; prediksi curah hujan; prediksi debit air sungai

I. PENDAHULUAN

Banjir merupakan bencana alam yang sangat mengganggu aktifitas masyarakat. Banjir juga menyebabkan kerusakan infrastruktur dan merugikan kegiatan perekonomian. Kejadian banjir sering menelan korban jiwa. Prediksi banjir dapat dilakukan dengan melihat curah hujan dan debit air.

Tinggi dan rendahnya curah hujan disebabkan oleh musim yang sedang berlangsung. Namun, dengan adanya perubahan iklim, prediksi curah hujan dengan metode statistika klasik memberikan hasil yang kurang bagus. Di sisi lain, beragam metode dapat digunakan untuk meramalkan curah hujan, yaitu antara lain *Seasonal Autoregressive Integrated Moving Average* (SARIMA) [1], *ensemble* (gabungan) antara *Adaptive Neuro Fuzzy Inference System* (ANFIS) dan *ARIMA* [2], dan *Artificial Neural Networks* (ANN) [3]-[5].

Beberapa sungai memiliki debit air rata-rata yang cukup besar yang dapat dimanfaatkan untuk sumber air irigasi dan bahan baku air bersih, meskipun di sisi lain dapat menimbulkan ancaman bahaya banjir di musim penghujan akibat terdegradasinya lahan pada daerah hulu [6]. Oleh karena itu, selain menggunakan variabel curah hujan untuk memprediksi kejadian banjir, besarnya debit air juga patut diperhitungkan. Banjir bandang dapat terjadi karena luapan air sungai yang tidak mampu lagi menampung air karena derasny aliran dari dataran yang lebih tinggi. Debit air juga dapat diprediksi dengan BP-NN [7], [8].

Prediksi banjir dalam beberapa penelitian menggunakan data saat ini (*realtime*). Aplikasi pendeteksi banjir tersebut menggunakan sensor dan IoT (*Internet of Things*) dengan laporan berupa SMS (*Short Message Service*) dan berbasis android dengan laporan SMS gateway [9]. Di sisi lain, *machine learning* dapat digunakan untuk mengenali secara otomatis pola-pola

^{*)} Penulis korespondensi (Ike Fitriyaningsih)
Email: ike.fitriyaningsih@gmail.com

yang kompleks dan membuat keputusan cerdas berdasarkan data empiris sebelumnya, misalnya untuk meramal cuaca dan iklim, memprediksi iklim yang ekstrim, dan mengkonstruksi ulang iklim di masa lampau [10], [11]. Algoritma *machine learning* untuk memprediksi jangka panjang menggunakan data historis yang dapat diterapkan, di antaranya adalah *Back Propagation Neural Network* (BP-NN) dan *Support Vector Machine* (SVM). Penerapan SVM dalam memprediksi banjir telah dilakukan dalam [12], [13]. Selain itu, prediksi tinggi muka air (TMA) untuk deteksi dini bencana banjir telah dilakukan dalam [14] dengan menggunakan *Support Vector Regression* (SVR).

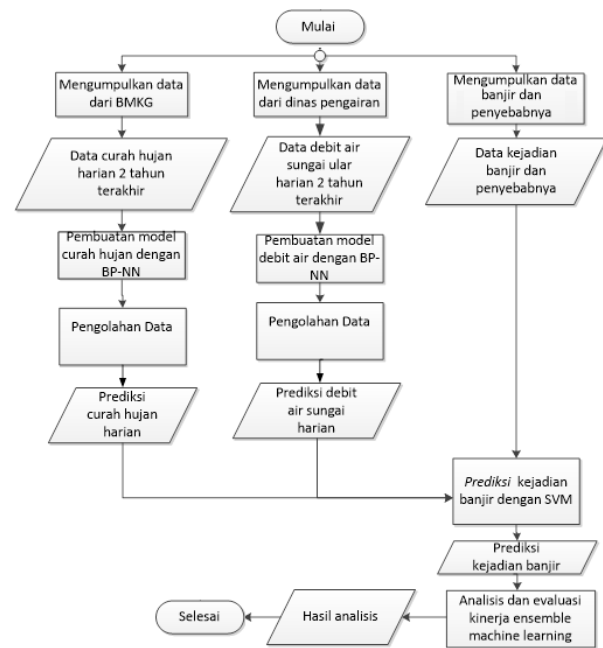
Belum banyak kajian tentang prediksi banjir ini berdasarkan data curah hujan dan debit air sungai menggunakan *machine learning*. Yunianto et al. [9] dan Soebroto et al. [14] mengkaji tentang deteksi dini banjir berdasarkan tinggi permukaan sungai, sedangkan pemodelan dan prakiraan curah hujan dilakukan dalam [15], [16]. Penelitian ini melakukan kajian untuk memprediksi banjir secara jangka panjang (6 dan 30 hari ke depan) menggunakan data historis curah hujan dan debit air selama dua tahun. Curah hujan dan debit air sungai diprediksi menggunakan BP-NN. Hasil prediksi tersebut digunakan untuk menentukan kapan terjadi banjir dengan metode SVM untuk klasifikasi *data mining* kejadian banjir beberapa hari ke depan. Hasil prediksi terbaik mengenai curah hujan dan debit air dilakukan dengan teknik *data mining* dengan kriteria RMSE [17].

II. METODE PENELITIAN

Penelitian ini melakukan prediksi banjir secara jangka panjang (6 dan 30 hari ke depan) menggunakan data historis selama dua tahun. Lokasi yang digunakan adalah Kabupaten Deli Serdang, Sumatera Utara, Indonesia. Kabupaten ini merupakan daerah yang rawan banjir di Sumatera Utara. Variabel yang digunakan untuk memprediksi banjir adalah curah hujan dan debit air. Data curah hujan diambil dari tiga stasiun cuaca di Kabupaten Deli Serdang Indonesia, yaitu Sampali, Kualanamu dan Tuntungan. Data debit air sungai berasal dari sungai Ular. Tahapan penelitian ditampilkan dalam Gambar 1.

Langkah pertama yang dilakukan adalah pengumpulan data dari BMKG dan Badan Pengawasan Sungai II (BWS II). Data yang digunakan adalah data harian sehingga ukuran data yang didapatkan adalah 731 data untuk 2 tahun terakhir (2016-2017). Data kejadian banjir di Deli Serdang diambil dari berita di internet.

Berdasarkan data yang didapatkan sebelumnya, dilakukan *preprocessing*, yaitu pengisian data yang hilang atau kosong dan normalisasi data. Pengisian data curah hujan yang kosong dari rata-rata stasiun yang lain. Jika seluruh data stasiun kosong pada hari tertentu, maka data curah hujan diisi dengan rata-rata curah hujan sebelum dan sesudah hari tersebut. Normalisasi dilakukan dengan Min-Max (minimum dan maksimum)



Gambar 1. Tahapan penelitian prediksi banjir

seperti dalam Persamaan 1 [18]. Parameter s' menyatakan nilai data normal, s nilai data aktual, dan k menunjukkan indeks data s .

$$s' = \frac{s - \min\{s_k\}}{\max\{s_k\} - \min\{s_k\}} \quad (1)$$

Tahap prediksi menggunakan BP-NN. Ukuran akurasi yang dihitung adalah yang mempunyai RMSE terkecil menggunakan Persamaan 2 [19]. Parameter e_t menunjukkan perbedaan antara nilai target dan prediksi, $e_t = X_t - F_t$, dan n menunjukkan jumlah data. Data asli dan hasil prediksi dilakukan pembelajaran SVM untuk menentukan kejadian banjir menggunakan data hasil prediksi curah hujan dan debit air sungai Ular dengan bahasa R. Skenario *data mining* yang dilakukan pada tahap ini yaitu membagi data menjadi *training:testing* antara lain 60:40, 70:30, 80:20 dan 90:10. Hal ini dilakukan sesuai dengan teknik data mining yang membagi data training dan testing menggunakan ukuran tersebut untuk mengetahui ukuran data training yang paling optimal untuk mendapatkan model. Tahap terakhir yang dilakukan adalah evaluasi performa *ensemble machine learning* untuk memprediksi curah hujan, debit air dan kejadian banjir di Deli Serdang.

$$RMSE = \sqrt{\frac{\sum (e_t)^2}{n}} \quad (2)$$

III. HASIL DAN PEMBAHASAN

Penelitian ini berfokus pada penggunaan *machine learning* untuk memprediksi banjir dalam jangka waktu

6 dan 30 hari ke depan dengan menggunakan teknik *data mining*. Variabel yang digunakan untuk memprediksi banjir adalah curah hujan dan debit air. Pembahasan dimulai dengan prediksi kedua variabel tersebut menggunakan BP-NN. Hasil prediksi terbaik digunakan untuk memperkirakan apakah terjadi banjir atau tidak selama jangka waktu prediksi.

A. Prediksi Curah Hujan dan Debit Air dengan BP-NN

Prediksi masing-masing data curah hujan dari tiga stasiun dengan BP-NN dilakukan sebanyak tiga kali untuk mengevaluasi apakah metode tersebut stabil dalam memprediksi data curah hujan. RMSE dicatat pada setiap *running* untuk menjadi ukuran *error* prediksi. RMSE data uji dari 48 kali *running* program dengan variasi persentase dataset dinyatakan dalam Tabel 1. Nilai yang diarsir menunjukkan RMSE terendah untuk tiap perbandingan dataset dan lokasi.

Prediksi 6 hari berarti data variabel independen dihapus sebanyak 6 hari di akhir, sedangkan variabel dependen atau target dihapus sebanyak 6 hari di awal. Hal ini dilakukan untuk membangun logika bahwa variabel dependen sekarang mempengaruhi target 6 hari setelahnya. Model terbaik curah hujan di stasiun Sampali dan Tuntungan adalah dengan kombinasi dataset 60:40 menghasilkan RMSE berturut-turut 13.651 dan 13.623 mm. Model terbaik curah hujan di stasiun Kualanamu adalah dengan kombinasi dataset 90:10 yang menghasilkan RMSE 16.616. Debit air menghasilkan prediksi terbaik pada dataset 70:30. Rata-rata RMSE dapat dilihat pada Gambar 2.

Berdasarkan [18], semakin besar data *training*, maka model yang terbentuk akan semakin bagus. Namun, tidak semua data menghasilkan kesimpulan tersebut. Pada data curah hujan, RMSE terendah mayoritas terdapat pada pembagian data *training:testing* 60:40. Hal ini menunjukkan bahwa 60% atau 438 data paling baik digunakan untuk memprediksi curah hujan maupun debit air 6 hari ke depan.

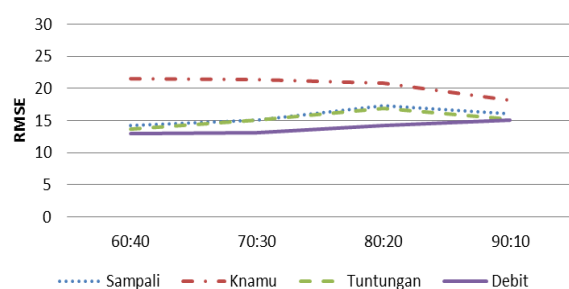
Prediksi 30 hari berarti data variabel independen dihapus sebanyak 30 hari di akhir, sedangkan variabel dependen atau target dihapus sebanyak 30 hari di awal. RMSE data testing dari 48 kali *running* program dengan variasi persentase dataset *training:testing* seperti ditunjukkan dalam Tabel 2. Nilai yang diarsir menunjukkan RMSE terendah untuk tiap perbandingan dataset dan lokasi.

Model terbaik curah hujan di stasiun Sampali dan Tuntungan adalah dengan kombinasi dataset 60:40 menghasilkan RMSE berturut-turut 14.749 dan 14.291 mm. Model terbaik curah hujan di stasiun Kualanamu adalah dengan kombinasi dataset 90:10 menghasilkan RMSE 17.027. Debit air menghasilkan prediksi terbaik pada dataset 70:30. Namun, RMSE yang dihasilkan prediksi 30 hari lebih besar daripada prediksi 6 hari. Rata-rata RMSE dapat dilihat pada Gambar 3.

Pola RMSE yang dihasilkan prediksi 30 hari sama dengan prediksi 6 hari. RMSE terendah pada prediksi curah hujan di Sampali dan Kualanamu adalah pada

Tabel 1. RMSE data testing hasil prediksi 6 hari

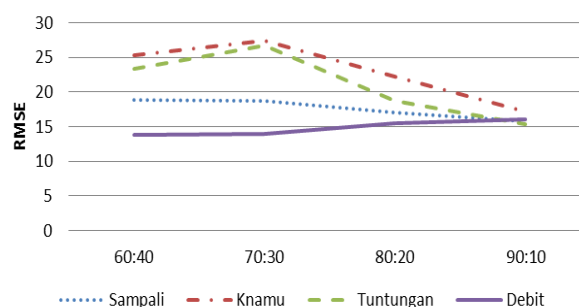
Training: Testing	RMSE			
	Sampali	Tuntungan	K.Namu	Debit
60:40	14.896	22.049	13.686	12.971
	13.651	20.827	13.623	12.973
	13.961	21.563	13.643	12.979
70:30	15.358	21.462	15.272	12.948
	14.881	19.802	14.934	13.221
	14.917	22.958	14.987	13.224
80:20	18.419	21.444	16.872	14.23
	16.79	20.927	16.961	14.233
	16.896	20.255	16.907	14.235
90:10	15.071	19.531	14.919	15.048
	17.492	16.616	15.751	15.049
	15.582	18.163	14.789	15.051



Gambar 2. Grafik rata-rata RMSE prediksi 6 hari dengan BP-NN

Tabel 2. RMSE data testing hasil prediksi 30 hari

Training: Testing	RMSE			
	Sampali	Tuntungan	K.Namu	Debit
60:40	21.705	33.101	13.666	13.737
	14.749	21.858	14.291	13.754
	20.305	20.759	42.022	13.758
70:30	20.192	43.160	20.151	13.901
	19.328	18.933	44.870	13.913
	16.486	20.207	15.261	13.920
80:20	16.344	21.151	17.409	15.430
	17.670	24.520	17.707	15.430
	17.090	21.186	20.870	15.431
90:10	15.463	17.265	15.616	16.089
	15.812	17.027	14.917	16.092
	15.933	17.411	15.694	16.127



Gambar 3. Rata-rata RMSE prediksi 30 Hari dengan BP-NN

pembagian data 60:40 sedangkan debit air pada 70:30 dan Tuntungan 90:10.

B. Prediksi Banjir

RMSE yang dihasilkan prediksi 30 hari lebih besar daripada prediksi 6 hari. Hal ini menjadikan prediksi banjir dalam jangka panjang tidak dapat dilakukan. Agar menghasilkan prediksi banjir dalam jangka panjang beberapa hari ke depan, hasil prediksi debit air dan curah hujan digunakan untuk memprediksi kejadian banjir. Hasil prediksi kejadian banjir selama 6 hari ke depan dengan model SVM menggunakan pembelajaran data tahun 2016-2017 sesuai output software R dapat dilihat pada Gambar 4.

Tabel 3 memperlihatkan hasil prediksi SVM mengenai kejadian banjir di Deli Serdang. Berdasarkan prediksi 6 hari sejak tanggal 26 sampai 31 Desember 2017 tidak akan terjadi banjir. Model yang terbentuk terdiri dari 43 *support vector* menggunakan kernel radial dengan $\gamma=0,25$ dan $cost=1$. Selama 6 hari hasil prediksi banjir di Deli Serdang dengan SVM adalah akurat.

Hasil ini melengkapi sistem EWS dalam [9] yang lebih fokus pada penggunaan perangkat keras untuk menampilkan data ketinggian air harian, bulanan, dan tahunan serta pemberian peringatan dini banjir air sungai. Demikian juga dengan prediksi tinggi muka air untuk deteksi dini bencana banjir dalam [14] yang menggunakan TVIWPSO dan menghasilkan nilai MAE (*Mean Absolute Error*) dari hasil prediksi paling kecil dengan nilai sebesar 0.142659943. Dengan hasil penelitian ini, maka penggabungan metode BP-NN dan SVM telah berhasil dilakukan untuk memprediksi apakah terjadi banjir atau tidak selama 6 hari ke depan. RMSE prediksi curah hujan dari penelitian ini sebesar 12.948 mm lebih baik daripada [15] yang menggunakan ANFIS dengan RMSE 149,798 mm. Kelebihan ANN ini dari ANFIS ini juga selaras dengan [16].

IV. KESIMPULAN

Prediksi curah hujan dan debit air terbaik di Deli Serdang menggunakan BP-NN adalah prediksi 6 hari dengan kombinasi *training:testing* yaitu 60:40 dengan RMSE terendah dari Stasiun Sampali dan Tuntungan. Prediksi banjir ini menggunakan SVM dengan variabel hasil prediksi curah hujan dan debit air ini telah menghasilkan kesimpulan yang benar bahwa sejak tanggal 26 desember 2018 tidak akan terjadi banjir selama 6 hari ke depan.

UCAPAN TERIMA KASIH

Terima kasih kepada Kementrian Riset, Teknologi, dan Pendidikan Tinggi Indonesia yang mendanai penelitian ini.

```
> x=read.table("C:/Users/IT Del/Documents/runsvm pdp/asli.txt",header=TRUE)
> xData=cbind(x$Sampali,x$Knamu, x$Tuntungan,x$debit)
> pred=read.table("C:/Users/IT Del/Documents/runsvm pdp/prediksi.txt",header=TRUE)
> library(e1071)
> modelSVM <- svm(x$as.matrix(xData),y=x$Banjir,kernel="radial")
> resultSVM <- predict(modelSVM,pred)
> modelSVM

Call:
svm.default(x = as.matrix(xData), y = x$Banjir, kernel = "radial")

Parameters:
SVM-Type: C-classification
SVM-Kernel: radial
cost: 1
gamma: 0.25

Number of Support Vectors: 43

> resultSVM
 1 2 3 4 5 6
No No No No No No
Levels: No Yes
```

Gambar 4. Output software R untuk Model SVM Banjir

Tabel 3. Hasil prediksi banjir dan validasinya

No.	Tanggal	Prediksi Banjir	Kenyataan
1.	26/12/2017	No	Tidak terjadi banjir
2.	27/12/2017	No	Tidak terjadi banjir
3.	28/12/2017	No	Tidak terjadi banjir
4.	29/12/2017	No	Tidak terjadi banjir
5.	30/12/2017	No	Tidak terjadi banjir
6.	31/12/2017	No	Tidak terjadi banjir

DAFTAR PUSTAKA

- [1] D. Eni and F. J. Adeyeye, "Seasonal ARIMA Modeling and Forecasting of Rainfall in Warri Town, Nigeria," *Journal of Geoscience and Environment Protection*, vol. 3, no. 6, pp. 91-98, 2015.
- [2] S. Suhartono et al., "Ensemble Method Based On ANFIS-ARIMA For Rainfall Prediction," in *2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE)*, Langkawi, Malaysia, Sept. 2012, pp. 1-4.
- [3] N. A. Charaniya and S. V. Dudul, "Design of Neural Network Models for Daily Rainfall Prediction," *International Journal of Computer Applications*, vol. 61, no. 14, pp. 23-27, 2013.
- [4] M. P. Darji, V. K Dabhi, and H. B. Prajapati, "Rainfall Forecasting using Neural Network: a Survey," in *2015 International Conference on Advances in Computer Engineering and Applications*, Ghaziabad, India, 2015, pp. 706-713.
- [5] M. Mislana, H. Havaluddin, S. Hardwinarto, S. Sumaryono, and M. Aipassa, "Rainfall Monthly Prediction Based on Artificial Neural Network: A Case Study in Tenggarong Station, East Kalimantan – Indonesia," *Procedia Computer Science*, vol. 59, pp. 142-151, 2015.
- [6] BAPPEDA, "Aspek Geografi dan Demografi Daerah Sumatera Utara," Bappeda Propinsi Sumatera Utara, 2012. [Online]. Available: <http://bappeda.sumutprov.go.id/index.php/potensi-daerah/141-aspek-geografi-dan-demografi>.
- [7] M. Rezaei, A. Motlaq, A. Mahmouei, and S. Mousavi, "River Flow Forecasting using Artificial

- Neural Network (Shoor Ghaen),” *Ciência e Natura*, vol. 37, pp. 207–215, 2015.
- [8] C. O. Maxwell, “Prediction of River Discharge Using Neural Networks,” Thesis, School of Computing and Informatics, University of Nairobi, Kenya, 2014. [Online]
- [9] M. Yunianto et al., “Smart EWS: Sebelas Maret Early Warning Sistem Aplikasi Deteksi Dini Bencana Banjir Sungai Bengawan Solo Berbasis Android,” in *Seminar Nasional Geografi UMS VII*, Solo, Indonesia, 2016, pp. 588-596.
- [10] T. M. Mitchell, *Machine Learning*. Mac Graw Hill, 1997.
- [11] C. Monteleoni et al., “*Climate Informatics*,” in *Computational Intelligent Data Analysis for Sustainable Development*, T. Yu, N. Chawla, and S. Simoff, Eds. Chapman and Hall/CRC, 2013, pp. 81-126.
- [12] D. Han, L. Chan, and N. Zhu, “Flood Forecasting using Support Vector Machines,” *Journal of Hydroinformatics*, vol. 9, no. 4, pp. 267-276, 2007.
- [13] N. S. Raghavendra and P. C. Deka, “Support Vector Machine Applications in the Field of Hydrology: A Review,” *Applied Soft Computing*, vol. 19, pp. 372–386, 2014.
- [14] A. A. Soebroto, I. Cholissodin, R. C. Wihandika, M. T. Frestantiya, and Z. El Arief, “Prediksi Tinggi Muka Air (TMA) Untuk Deteksi Dini Bencana Banjir Menggunakan SVR-TVIWPSO,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 2, pp. 79-86, 2015
- [15] G. A. F. Alfarisy and W. Mahmudy, “Rainfall Forecasting in Banyuwangi using Adaptive Neuro Fuzzy Inference System,” *Journal of Information, Technology and Computer Science*, vol. 1, no. 2. pp 65-71, 2016.
- [16] S. K. T. Kuni and C. Mohandas, “Rainfall Runoff Modelling Using ANN and ANFIS,” in *International Symposium on Integrated Water Resources Management (IWRM–2014)*, Kerala, India, Feb. 2014, pp. 9-16.
- [17] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*. Springer International Publishing, 2013.
- [18] N. Chamidah, W. Wiharto, and U. Salamah, “Pengaruh Normalisasi Data pada Jaringan Syaraf Tiruan Backpropagasi Gradient Descent Adaptive Gain (BPGDAG) untuk Klasifikasi,” *ITSMART: Jurnal Teknologi dan Informasi*, vol 1., no 1, pp. 28-33, 2012.
- [19] S. N. K. Kamarudin and A. A. Bakar, “Neural Network Algorithm Variants for Malaysian Weather. Second International Multi-Conference on Artificial Intelligence Technology,” in *Soft Computing Applications and Intelligent System*, S. A. Noah et al., Eds. Springer, 2013, pp. 121-134.