

Tipologi Curah Hujan Tahunan Berbasis Autoencoder Dan Prediksi Hujan Lebat Harian Menggunakan 1D-CNN di Stasiun Meteorologi Maritim Serang

Dian Herdianingsih^{a*}, Sajarwo Anggai^b, Sudarno Wiharjo^c
dianherdianingsih72@gmail.com^a

Program Studi Magister Teknik Informatika, Universitas Pamulang, Tangerang Selatan^{abc}

Intisari

Kata kunci:
Autoencoder, 1D-CNN,
Curah Hujan, Curah
Hujan Dasarian.

Kawasan pesisir memerlukan produk klimatologis operasional untuk perencanaan infrastruktur dan kesiapsiagaan cuaca. Penelitian ini bertujuan membangun tipologi pola curah hujan tahunan dan mengembangkan sinyal deteksi hujan lebat harian di Stasiun Meteorologi Maritim Serang melalui integrasi *Autoencoder* (AE) dan *Convolutional Neural Network 1D* (1D-CNN). Periode data curah hujan harian pada periode 1995–2024 diagregasi menjadi 36 dasarian per tahun untuk tipologi, dan diturunkan menjadi fitur akumulasi untuk deteksi harian. *Autoencoder* merubah vektor dasarian menjadi dua dimensi, lalu diklasterisasi dengan *K-Means* yang menghasilkan dua pola iklim stabil. Tipologi 1 yaitu data curah hujan 21 tahun dengan rata-rata 1.502 mm dan mencirikan pola monsun basah sedang, sedangkan Tipologi 2 yaitu curah hujan 9 tahun dengan rata-rata 1.927 mm mencerminkan pola yang lebih basah. Selanjutnya, 1D-CNN dilatih menggunakan sekuens 14 hari dengan fitur akumulasi hujan lebat (≥ 50 mm/hari) yang berhasil mencapai AUROC 0,667 pada data uji. Keunggulan model ini terbukti melampaui akurasi metode *baseline* konvensional dengan skor CSI 0,029 berbanding 0,021. Peningkatan ini mengonfirmasi besarnya nilai tambah dari proses ekstraksi pola temporal. Secara keseluruhan, penelitian ini menghasilkan kerangka klimatologis stasiun tunggal yang efisien dan siap direproduksi untuk mendukung layanan peringatan dini lokal.

Tentang naskah:
-diterima, 23 Juni 2026
-diterbitkan, 1 Juli 2026

Abstract

Coastal areas require operational climatological products for infrastructure planning and weather preparedness. This study aims to build an annual rainfall pattern typology and develop a daily heavy rainfall detection signal at the Serang Maritime Meteorological Station by integrating an Autoencoder (AE) and a 1D Convolutional Neural Network (1D-CNN). Daily rainfall data from 1995 to 2024 were aggregated into 36 decadal units per year for the typology, and then derived into accumulation features for daily detection. The Autoencoder transformed the decadal vectors into two dimensions, which were then clustered using K-Means, resulting in two stable climate patterns. Typology 1 consists of 21 years of rainfall data with an average of 1,502 mm, characterizing a moderate wet monsoon pattern, while Typology 2 covers 9 years with an average of 1,927 mm, reflecting a wetter pattern. Furthermore, the 1D-CNN was trained using a 14-day sequence with heavy rainfall accumulation features (≥ 50 mm/day), successfully achieving an AUROC of 0.667 on the test data. The model's superiority surpassed the conventional threshold baseline accuracy with a CSI of 0.029 compared to 0.021. This improvement confirms the added value of temporal pattern extraction. Overall, this research produces an efficient and reproducible single-station climatological framework to support local early warning services.

Keywords:
Autoencoder, 1D-CNN,
Decadal Rainfall

1. Pendahuluan

Variabilitas curah hujan memiliki dampak yang luas terhadap pengelolaan sumber daya air, perencanaan infrastruktur, serta mitigasi risiko bencana di kawasan pesisir. Di Indonesia, pemahaman mengenai karakteristik pola

curah hujan lokal sangat penting untuk kebutuhan operasional. Kota Serang, sebagai ibu kota Provinsi Banten, berada pada zona iklim monsun di mana fluktuasi curah hujannya sering kali dipengaruhi oleh fenomena iklim global seperti ENSO (*El Niño-Southern Oscillation*) dan IOD (*Indian Ocean Dipole*) (Chang et

al., 2005; Mulsandi et al., 2023). Walaupun Stasiun Meteorologi Maritim Serang memiliki rekam jejak data historis yang sangat lengkap, hingga saat ini masih belum tersedia produk ringkasan iklim atau tipologi yang praktis di tingkat stasiun tunggal. Padahal, pemangku kepentingan seperti Badan Penanggulangan Bencana Daerah (BPBD) dan Dinas Pertanian sangat membutuhkan informasi cuaca yang spesifik dan mudah dipahami untuk pengambilan keputusan.

Kondisi cuaca ekstrem, khususnya curah hujan lebat yang memicu banjir pesisir dan kerusakan infrastruktur, terbukti semakin sering terjadi seiring dengan perubahan dinamika iklim regional (Hermawan et al., 2025). Sayangnya, ketersediaan sistem peringatan dini yang sensitif terhadap anomali cuaca ekstrem langka di tingkat lokal masih sangat terbatas. Sering kali, pemodelan prakiraan cuaca mengandalkan resolusi regional kasar. Hal ini menyebabkan pemodelan tersebut kesulitan menangkap karakteristik dinamika awan konvektif yang sangat spesifik jika diterapkan pada satu titik stasiun observasi (Mouassom et al., 2025). Di sisi lain, meskipun teknik regionalisasi iklim menggunakan data historis telah banyak dieksplorasi di berbagai belahan dunia (Darand & Mansouri Daneshvar, 2014), penerapannya jarang sekali bermuara pada produk operasional stasiun tunggal yang siap pakai.

Berdasarkan kajian literatur, penelitian terdahulu mengenai analisis cuaca umumnya masih berjalan secara terpisah. Sebagian penelitian berfokus pada pemetaan spasial antar-stasiun dalam skala wilayah yang luas (Ahmad Basri et al., 2021), sementara kelompok lainnya berfokus murni pada prediksi curah hujan harian tanpa mempertimbangkan latar belakang pola iklim tahunannya (Liu et al., 2023; Sari et al., 2020). Terdapat celah analisis (*gap analysis*) berupa minimnya pendekatan terintegrasi yang menggabungkan pengelompokan pola curah hujan tahunan dengan prediksi kejadian cuaca ekstrem harian pada skala stasiun tunggal. Untuk menjembatani kesenjangan tersebut, penelitian ini bertujuan mengkarakterisasi distribusi curah hujan tahunan, membangun tipologi pola iklim yang stabil menggunakan kombinasi *Autoencoder* dan

deteksi probabilistik untuk hujan lebat harian menggunakan arsitektur *1D-CNN*. Kebaruan (*novelty*) dari penelitian ini terletak pada integrasi hierarkis antara pengelompokan pola iklim tahunan dan prediksi cuaca ekstrem harian menggunakan *deep learning* pada skala stasiun tunggal, yang sebelumnya belum dieksplorasi secara utuh.

2. Kerangka Teori

2.1. Klasifikasi dan Klastering Pola Curah Hujan

Upaya untuk memetakan dan mengelompokkan pola curah hujan tahunan telah lama menjadi fokus dalam kajian klimatologi. Sa'adi et al. (2021) berhasil mengklasifikasikan wilayah Kalimantan menggunakan analisis kluster, membuktikan bahwa pemisahan pola curah hujan sangat bermanfaat secara operasional. Di Indonesia, Suryanto (2017) juga menunjukkan bahwa algoritma *K-Means* mampu memberikan hasil pengelompokan data curah hujan yang lebih valid dibandingkan metode lain. Terkait dengan tujuan penelitian ini, metode *K-Means* tidak hanya digunakan sebagai alat deskriptif, tetapi secara khusus diterapkan untuk mengelompokkan output ruang laten dari model saraf tiruan agar tipologi stasiun tunggal Serang dapat dipetakan secara presisi dan divalidasi dengan pengujian stabilitas *bootstrap*.

2.2. Autoencoder

Autoencoder (AE) adalah jaringan saraf tiruan yang dirancang untuk memampatkan data ke dalam dimensi yang lebih kecil (ruang laten) dan merekonstruksinya kembali (Hinton & Salakhutdinov, 2006). Kurihana et al. (2024) menemukan bahwa penggabungan klastering dengan *Autoencoder* mampu menangkap karakteristik iklim yang lebih spesifik dibandingkan statistik konvensional seperti *Principal Component Analysis* (PCA). Dalam konteks penelitian ini, *Autoencoder* dipilih untuk mengekstraksi fitur 36 dasarian yang sangat kompleks menjadi representasi reduksi yang lebih padat, sehingga batas pemisahan antar tipe sirkulasi iklim tahunan dapat diidentifikasi secara optimal (Ibebuchi, 2024).

2.3. 1D-CNN

Arsitektur *Convolutional Neural Network* satu dimensi (*1D-CNN*) sangat tangguh dalam melacak pola berurutan dari suatu data deret waktu (Sari et al., 2020). Dalam kajian iklim, Liu et al. (2023) telah memanfaatkan fitur konvolusi 1D untuk mendeteksi tingkat intensitas curah hujan berdasarkan pola ketergantungan jangka panjang (*long-term dependencies*). Pemilihan arsitektur ini sangat selaras dengan tujuan operasional penelitian, yaitu memproses runtunan data historis 14 hari sebelumnya guna mendeteksi probabilitas kejadian curah hujan ≥ 50 mm/hari di Stasiun Serang. Sebagai ringkasan rasionalisasi pemilihan metode pada penelitian ini, perbandingan teknis antara metode konvensional dan metode *deep learning* yang diusulkan disajikan pada Tabel 1.

Tabel 1 Perbandingan Metode Analisis Ekstraksi dan Prediksi

Pendekatan	Metode Konvensional	Metode Usulan
Ekstraksi Fitur (Reduksi)	<i>Principal Component Analysis</i> (PCA)	<i>Autoencoder</i> (AE)
Prediksi Hujan Lebat	Ambang Batas Akumulasi Harian Statis	<i>1D-CNN</i>

3. Metodologi

Penelitian ini menerapkan pendekatan kuantitatif komputasional yang mengintegrasikan dua metode *machine learning*. Pendekatan pertama menggunakan *unsupervised learning* (gabungan *Autoencoder* dan *K-Means*) untuk mengelompokkan pola iklim tahunan. Pendekatan kedua menggunakan *supervised learning* berbasis *1D-CNN* untuk memprediksi kejadian cuaca ekstrem harian. Sasaran utama dari penelitian ini adalah mengungkap distribusi tipe iklim tahunan sekaligus mendeteksi frekuensi hujan lebat dengan ambang batas ekstrem (≥ 50 mm/hari) di Stasiun Meteorologi Maritim Kelas I Serang, Banten.

Data utama yang digunakan adalah rekam observasi curah hujan harian selama 30 tahun (1 Januari 1995 hingga 31 Desember 2024), yang mencakup 10.958 baris data (kelengkapan 100%). Untuk pengelompokan pola tahunan, data diringkas menjadi matriks berisi 36 dasarian per tahun. Untuk prediksi harian, data diubah menjadi fitur jendela

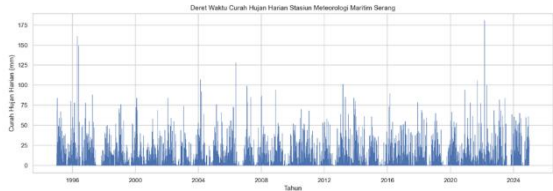
akumulasi bergulir (R1, R3, R5, R7, R10) dipadukan dengan komponen rotasi waktu (sinus-cosinus) guna merekam siklus harian.

Pemodelan ini dibangun menggunakan bahasa pemrograman Python 3 (TensorFlow, Keras, Scikit-Learn). Rincian parameter dan konfigurasi untuk pemodelan *Autoencoder* serta *1D-CNN* disajikan secara ringkas pada Tabel 2. Konfigurasi tersebut diatur agar jaringan mampu mempelajari kompleksitas pola curah hujan tanpa mengalami *overfitting*.

Tabel 2 Konfigurasi Pemodelan Autoencoder dan 1D-CNN

Komponen	Tebal
Input Autoencoder	36 Fitur Curah Hujan Dasarian Per Tahun StandarScaler
Normalisasi	
Klastering	K-Means, K=2 sampai 6
Evaluasi Klaster	Silhouette, Davies-Bouldin Index, Bootstrap ARI
Targen CNN 1D	Hujan lebat hari berikutnya (≥ 50 mm/hari)
Input sequence 1D-CNN	Curah Hujan 14 Hari Terakhir
Fitur tambahan	R1, R3, R5, R7, R10, sin dan cos day-of-year
Evaluasi Model	POD, FAR, CSI, AUROC

Dalam mengevaluasi kinerja prediksi *1D-CNN*, penelitian ini secara khusus memilih metrik AUROC (*Area Under the Receiver Operating Characteristic*), CSI (*Critical Success Index*), POD, dan FAR. Pemilihan metrik ini didasarkan pada karakteristik data cuaca ekstrem yang bersifat *imbalanced data* (kelas hujan lebat jauh lebih sedikit dibandingkan hari tidak hujan). Metrik CSI dan AUROC jauh lebih representatif dan objektif dalam mengukur keandalan model mendeteksi kejadian langka dibandingkan sekadar menggunakan akurasi keseluruhan (*accuracy*) yang sering kali menipu pada data yang tidak seimbang.



Gambar 1. Deret Curah Hujan Harian Stasiun Meteorologi Maritim Serang periode 1995-2024

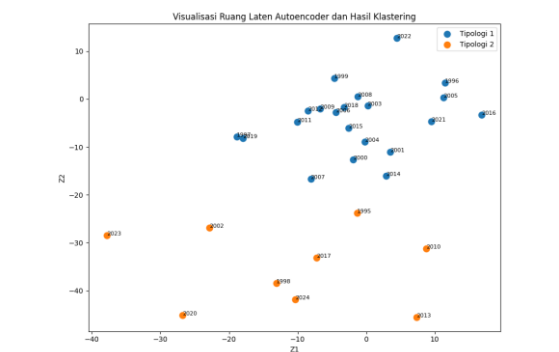
4. Hasil dan Pembahasan

Berdasarkan proses reduksi data curah hujan selama 30 tahun, terbukti bahwa penggunaan ruang laten berdimensi dua dari *Autoencoder* memberikan hasil pemisahan struktur pola monsun yang

paling optimal, meminimalkan fluktuasi cuaca lokal yang tidak beraturan. Proses klasterisasi *K-Means* pada representasi dua dimensi tersebut secara meyakinkan membagi data tahunan ke dalam dua tipologi utama. Kekokohan pengelompokan ini dibuktikan oleh perolehan skor *Silhouette* sebesar 0,518 dan stabilitas *Bootstrap ARI* 0,922 (lihat rincian evaluasi dimensi laten pada Tabel 3).

Tabel 3 Evaluasi Konfigurasi Kluster pada Berbagai Dimensi Laten

Dim. Laten	Klas ter	SC	DBI	B-ARI
2	2	0.518	0.759	0.922
2	3	0.490	0.740	0.698
3	3	0.420	0.790	0.764
3	4	0.410	0.660	0.742
4	3	0.490	0.770	0.770

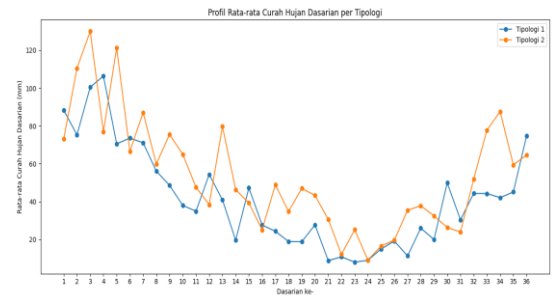


Gambar 2. Visualisasi ruang laten Autoencoder dan hasil klasterisasi K-Means

Tipologi 1 mendominasi 21 tahun pengamatan yang mencerminkan profil iklim "basah sedang", dengan puncak curah hujan di awal musim 106,44 mm dan akumulasi tahunan 1.502 mm. Sementara Tipologi 2 (9 tahun pengamatan) merepresentasikan pola "basah lebat" dengan intensitas tinggi, di mana puncak hujannya mencapai 130,01 mm dan volume tahunan melampaui 1.927 mm. Profil lengkap dari kedua tipologi ini dapat dilihat pada Tabel 4.

Tabel 4 Profil tipologi curah hujan tahunan hasil Autoencoder K-Means

Tipologi	Jumlah Tahun	SC	DBI	B-ARI
1	21	1502.24	106.44	8.04
2	9	1927.12	130.01	9.13

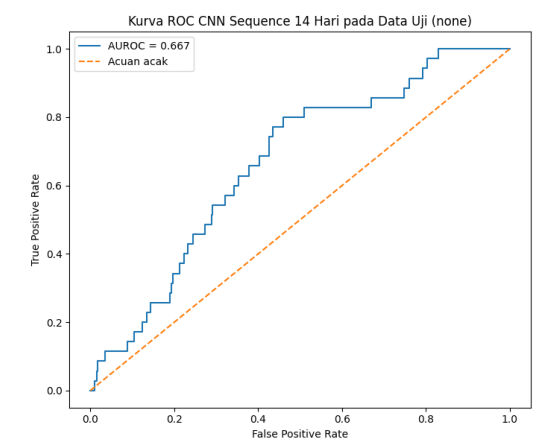


Gambar 3. Profil rata-rata curah hujan dasarian pada dua tipologi tahunan

Pada tahap prediksi kejadian ekstrem harian, model *1D-CNN* menunjukkan kinerja yang sangat unggul. Arsitektur yang mengandalkan riwayat 14 hari sebelumnya (*Seq-14*) mampu menyeleksi dan mereduksi kemunculan alarm palsu secara signifikan dibanding sistem konvensional. Seperti yang terangkum dalam Tabel 5, model tanpa pembobotan buatan mampu mencapai *CSI* 0,029 dan kurva *AUROC* 0,667, melampaui capaian parameter *baseline* R10 (akumulasi 10 hari) yang hanya menghasilkan *CSI* 0,021.

Tabel 5 Perbandingan varian 1D-CNN dan Model baseline pada data uji

Model	POD	FAR	CSI	AUROC
CNN Seq-14 (Tanpa Bobot)	0.086	0.958	0.029	0.667
CNN Seq-14 (1:2)	0.2	0.971	0.026	0.647
CNN Seq-14 (1:3)	0.029	0.987	0.009	0.647
CNN Seq-14 (balanced)	0.057	0.974	0.018	0.469
Baseline R10 ≥50 mm	0.514	0.979	0.021	-
Baseline R7 ≥50 mm	0.343	0.979	0.020	-
Baseline R1 ≥50 mm	0.000	1.000	0.000	-



Gambar 4. Kurva ROC CNN 1D-CNN pada data uji

Keberhasilan pemodelan terintegrasi ini memberikan implikasi praktis yang signifikan bagi para pemangku kepentingan. Badan Penanggulangan Bencana Daerah (BPBD) dapat memanfaatkan ambang batas probabilistik dari 1D-CNN untuk memicu sistem peringatan dini banjir pesisir yang minim *false alarm*. Di sisi lain, Dinas Pertanian setempat dapat menggunakan pemetaan Tipologi 1 (basah sedang) dan Tipologi 2 (basah lebat) untuk mengoptimalkan penyesuaian jadwal atau kalender tanam komoditas pangan setiap tahunnya.

Kendati demikian, penting untuk menggarisbawahi bahwa model 1D-CNN ini masih memiliki keterbatasan yang diakibatkan oleh fenomena ketidakseimbangan kelas (*data imbalance*). Mengingat kejadian ekstrem (≥ 50 mm/hari) sangat langka, nilai *False Alarm Ratio* (FAR) model secara keseluruhan masih cukup tinggi (0,958). Berbagai eksperimen rekayasa bobot (*class weight* 1:2, 1:3, hingga seimbang) telah diujikan, namun model *sequence* alami tanpa pembobotan terbukti tetap menyajikan nilai *CSI* operasional yang paling baik.

5. Simpulan

Integrasi arsitektur *deep learning* yang menggabungkan *Autoencoder* dan 1D-CNN terbukti efektif menangani kompleksitas data curah hujan di Stasiun Meteorologi Maritim Serang. Representasi dimensi ruang *laten* dari *Autoencoder* berhasil memetakan dua tipologi iklim utama secara andal (monsun basah sedang dan lebat). Kompresi fitur ini juga memfasilitasi algoritma 1D-CNN dalam menyaring alarm peringatan palsu secara teliti, sehingga mendongkrak performa deteksi harian (*AUROC* 0,667 dan *CSI* 0,029) melewati akurasi metode *baseline* konvensional. Kerangka kerja ganda ini sukses menghasilkan produk operasional yang lebih spesifik untuk kesiapsiagaan peringatan dini lokal.

Untuk pengembangan penelitian di masa depan, direkomendasikan agar pendekatan pemodelan ini diujikan menggunakan *masukan* data multi-stasiun guna memperluas dan memvalidasi cakupan spasial peringatan cuaca. Selain itu, eksperimen menggunakan arsitektur pemrosesan deret waktu yang lebih mutakhir, seperti model *Transformer*, sangat disarankan untuk dieksplorasi demi menekan laju

false alarm pada dataset cuaca yang sangat tidak seimbang.

Ucapan Terima Kasih

Penulis menghaturkan ucapan terima kasih yang sebesar-besarnya kepada Stasiun Meteorologi Maritim Kelas I Serang. Ketersediaan akses data curah hujan historis telah mendukung penyusunan riset berjalan dengan lancar dan komprehensi

Daftar Pustaka

- Ahmad Basri, M., Shahrudin, S. M., Kismiantini, Tan, M. L., Mohd Najib, S. A., Zainuddin, N. H., & Andayani, S. (2021). Regionalization of rainfall regimes using hybrid RF-Bs couple with multivariate approaches. *ISPRS International Journal of Geo-Information*, 10(10), 689.
- BMKG. (2022). *Pemutakhiran Zona Musim Indonesia Periode 1991–2020*. BMKG.
- Chang, C.-P., Wang, Z., McBride, J., & Liu, C.-H. (2005). Annual cycle of Southeast Asia–Maritime Continent rainfall and the asymmetric monsoon transition. *Journal of Climate*, 18(2), 287–301.
- Chen, P., Chen, A., Yin, S., Li, Y., & Liu, J. (2024). Clustering the diurnal cycle of precipitation using global satellite data. *Geophysical Research Letters*, 51(23), e2024GL111513.
- Darand, M., & Mansouri Daneshvar, M. R. (2014). Regionalization of precipitation regimes in Iran using principal component analysis and hierarchical clustering analysis. *Environmental Processes*, 1(4), 517–532.
- Hermawan, E., Risyanto, R., Purwaningsih, A., Ratri, D. N., Ridho, A., Harjana, T., Andarini, D. F., Satyawardhana, H., & Sujalu, A. P. (2025). Characteristics of mesoscale convective systems and their impact on heavy rainfall in Indonesia's new capital city. *Advances in Atmospheric Sciences*, 42(2), 342–356.
- Hinton, G., & Salakhutdinov, R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504–507.
- Ibebuchi, C. C. (2024). Fuzzy time series clustering using autoencoders neural network. *AIMS Geosciences*, 10(3), 524–539.
- Kurihana, T., Mastilovic, I., Wang, L., Meray, A., Praveen, S., Xu, Z., Memarzadeh, M., Lavin, A., & Wainwright, H. (2024). Identifying climate patterns using clustering autoencoder techniques. *Artificial Intelligence for the Earth Systems*, 3(3), e230035.
- Liu, Y., Liu, S., & Chen, J. (2023). RLFormer: A rainfall levels nowcasting model based on Conv1D-Transformer for the northern Xinjiang area of China. *Water*, 15(20), 3650.
- Mahmoud, W. H., Elagib, N. A., Gaese, H., & Heinrich, J. (2014). Rainfall conditions and rainwater

harvesting potential in the urban area of Khartoum. *Resources, Conservation and Recycling*, 91, 89–99.

Mouassom, F. L., Tamoffo, A. T., & Cardoso-Bihlo, E. (2025). Convolutional neural network-based insights into extreme precipitation regional dynamics over Central Africa. *Journal of Geophysical Research: Atmospheres*, 130(16), e2025JD044341.

Mulsandi, A., Koesmaryono, Y., Hidayat, R., Faqih, A., & Sopaheluwakan, A. (2023). On the interannual variability of Indonesian Monsoon Rainfall (IMR): A literature review. *Jurnal Meteorologi Dan Geofisika*, 24(2), 115–127.

Sa'adi, Z., Shahid, S., & Shiru, M. S. (2021). Defining climate zone of Borneo based on cluster analysis. *Theoretical and Applied Climatology*.

Sari, Y., Djamal, E., & Nugraha, F. (2020). Daily rainfall prediction using one dimensional convolutional neural networks. *Proceedings of CENIM 2020*, 391–396.

Suryanto, J. (2017). Analisa perbandingan pengelompokan curah hujan 15 harian Provinsi DIY menggunakan fuzzy clustering dan K-Means clustering. *Agrifor*, 16(2), 229–242.