

Deteksi Anomali Hasil Pengukuran Penakar Hujan Otomatis Menggunakan Metode Long Short Term Memory

Niko Wahyudi¹, Setyawan Widyarto²

nikowahyudi@gmail.com¹, setyawan.widyarto@budiluhur.ac.id²

^{1,2} Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur, Indonesia

Informasi Artikel

Diterima : 6 Feb 2025
Direvisi : 24 Feb 2025
Disetujui : 28 Feb 2025

Kata Kunci

Deteksi Anomali, Long Short Term Memory, Penakar Hujan Otomatis

Abstrak

Perkembangan teknologi memungkinkan Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) untuk melakukan pengamatan curah hujan secara otomatis menggunakan peralatan penakar hujan otomatis. Namun, peralatan ini berpotensi menghasilkan data curah hujan yang tidak valid akibat kerusakan sensor atau gangguan lingkungan. Penelitian ini bertujuan untuk mendeteksi anomali hasil pengukuran penakar hujan otomatis dengan metode Long Short Term Memory (LSTM) untuk memastikan validitas data dan mempercepat perbaikan peralatan yang mengalami malfungsi. Deteksi anomali dilakukan melalui metode quality control (QC) berbasis range dan step check, spatial check, serta error check yang menghasilkan label Total Anomali QC. Label ini kemudian ditransformasikan menggunakan one-hot encoding dan digunakan sebagai input dalam model klasifikasi berbasis LSTM. Hasil pengujian menunjukkan bahwa model mampu mengklasifikasikan anomali data dengan akurasi lebih dari 90% pada kluster barat, timur, dan pesisir, sehingga memungkinkan deteksi anomali yang lebih akurat dan efisien. Hasil penelitian ini berkontribusi dalam meningkatkan keandalan pengukuran curah hujan otomatis dan mendukung upaya BMKG dalam menjaga kualitas data cuaca dan iklim.

Keywords

Anomaly Detection, Automatic Rain Gauge, Long Short Term Memory

Abstract

Technological developments allow the Meteorology, Climatology and Geophysics Agency (BMKG) to carry out rainfall observations automatically using automatic rain measuring equipment. However, this equipment has the potential to produce invalid rainfall data due to sensor damage or environmental disturbances. This research aims to detect anomalies in the results of automatic rain gauge measurements using the Long Short Term Memory (LSTM) method to ensure data validity and speed up repairs to equipment that is malfunctioning. Anomaly detection is carried out using quality control (QC) methods based on range and step check, spatial check, and error check which produces a Total Anomaly QC label. These labels are then transformed using one-hot encoding and used as input in an LSTM-based classification model. Test results show that the model is able to classify data anomalies with an accuracy of more than 90% in the western, eastern and coastal clusters, thereby enabling more accurate and efficient anomaly detection. The results of this research contribute to improving the reliability of automatic rainfall measurements and support BMKG's efforts to maintain the quality of weather and climate data.

A. Pendahuluan

Badan Meteorologi Klimatologi dan Geofisika (BMKG) memiliki tugas melakukan pengamatan parameter cuaca dan iklim. Satu diantara parameter cuaca dan iklim tersebut adalah curah hujan. Seiring dengan perkembangan teknologi, pengamatan curah hujan tidak hanya dilakukan menggunakan peralatan konvensional yang mengukur curah hujan secara manual, namun secara bertahap juga menggunakan peralatan otomatis yang mana datanya tercatat otomatis [1].

Peralatan penakar hujan otomatis beroperasi secara mandiri yang ditempatkan pada suatu *remote area* tanpa membutuhkan pengawasan khusus. Data akumulasi curah hujan disimpan di dalam memori datalogger dan juga dikirim secara otomatis ke server menggunakan jaringan komunikasi internet, sehingga perlu dipastikan validitasnya melalui pemeliharaan dan kalibrasi peralatan yang lebih ketat agar dapat memberikan hasil pengukuran yang lebih akurat. Pemeliharaan rutin dan berkala dibutuhkan untuk memastikan peralatan penakar hujan beroperasi dengan baik. Sedangkan proses kalibrasi berguna untuk memastikan peralatan memberikan hasil pengukuran yang akurat dengan cara membandingkan hasil pengukuran peralatan dengan peralatan standar sebagai referensi yang sudah diketahui akurasi. Jika hasil pengukuran peralatan penakar hujan otomatis tidak sesuai dengan peralatan standar, maka perlu dilakukan penyetingan ulang atau perbaikan agar dapat memberikan hasil pengukuran yang lebih akurat. Namun demikian, pemeliharaan dan kalibrasi yang telah dilakukan secara berkala, tidak dapat menjamin sepenuhnya data yang dihasilkan adalah data yang valid. Dengan kondisi peralatan yang berada di *remote area*, petugas monitoring dan teknisi BMKG mengalami kesulitan dalam mendeteksi jika terjadi anomali dari hasil pengukuran peralatan penakar hujan otomatis, dimana jika terjadi ketidaksesuaian atau kesalahan dalam hasil pengukuran yang dihasilkan oleh peralatan dengan curah hujan yang sebenarnya untuk lokasi dan periode tertentu.

Anomali hasil pengukuran peralatan penakar hujan otomatis adalah penyimpangan data hasil pengukuran terhadap nilai data normal curah hujan menurut kriteria-kriteria tertentu. Contoh kasus anomali hasil pengukuran penakar hujan otomatis adalah ketika nilai curah hujan yang dihasilkan sangat tinggi atau sangat rendah dibandingkan dengan hasil pengukuran pada lokasi pengukuran curah hujan lain yang berada di lokasi yang berdekatan. Data pengukuran curah hujan peralatan penakar hujan otomatis berpotensi mengalami anomali akibat faktor teknis dan lingkungan. Faktor teknis disebabkan oleh kerusakan sensor tipping bucket, reed switch, kotornya corong penakar, kegagalan power supply dan gangguan jaringan komunikasi. Sementara itu, faktor lingkungan disebabkan oleh penguapan air pada corong, lokasi yang sudah tidak representatif akibat penghalang berupa bangunan dan pepohonan dan bencana alam [1]. Untuk menyelesaikan permasalahan tersebut, maka diperlukan suatu model untuk mendeteksi anomali hasil pengukuran curah hujan peralatan penakar hujan otomatis sehingga dapat segera dilakukan pemeliharaan pada peralatan yang terdeteksi mengalami anomali.

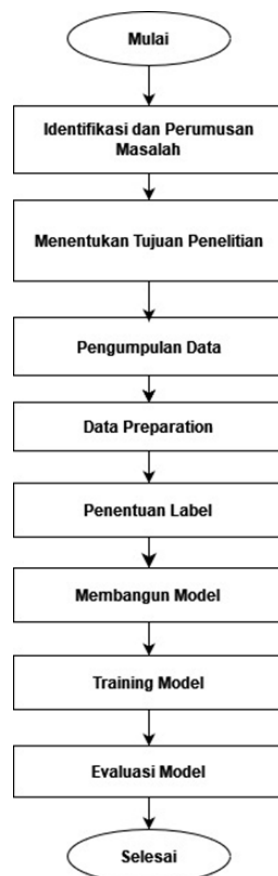
Tujuan penelitian ini adalah untuk mendeteksi dengan akurat anomali hasil pengukuran penakar hujan otomatis dengan menggunakan metode Artificial Neural Network (ANN) Long Short Term Memory (LSTM) sehingga dapat

membantu petugas monitoring dan teknisi BMKG dalam mengidentifikasi kesalahan pengukuran secara cepat sehingga peralatan dapat segera diperbaiki. Selain itu, penelitian ini juga bertujuan untuk mengetahui tingkat akurasi LSTM dalam mendeteksi anomali hasil pengukuran penakar hujan otomatis.

Penelitian ini memiliki relevansi dengan penelitian yang dilakukan oleh [2], yang mengembangkan sistem deteksi anomali jaringan pengukuran curah hujan secara online untuk meminimalisir dampak kesalahan pengukur curah hujan pada layanan peringatan banjir perkotaan menggunakan metode Support Vector Regression (SVR). Selain itu penelitian ini juga memiliki relevansi dengan penelitian [3] dalam hal pemilihan metode yang digunakan yaitu metode Long-Short Term Memory (LSTM), di smana penelitian tersebut membandingkan penggunaan LSTM dan model algoritma konvensional untuk memprediksi kondisi motor 10 kV pada PLTU Batubara.

B. Metode Penelitian

Metode pada penelitian ini terdiri dari langkah-langkah yang dilakukan dalam rangka mengembangkan model untuk mendeteksi anomali hasil pengukuran penakar hujan otomatis yang diilustrasikan pada Gambar 1.



Gambar 1. Metode Penelitian

Berdasarkan Gambar 1, penelitian ini diawali dengan melakukan identifikasi permasalahan- permasalahan yang dihadapi petugas monitoring dan teknisi

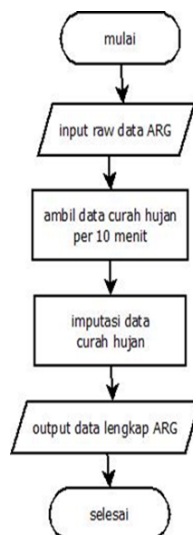
BMKG dalam melaksanakan tugas monitoring dan pemeliharaan peralatan penakar hujan otomatis. Dari permasalahan yang ada kemudian dilakukan perumusan masalah yang akan diselesaikan melalui penelitian ini yaitu bagaimana mendeteksi anomali hasil pengukuran penakar hujan otomatis menggunakan metode LSTM. Dari rumusan masalah yang diangkat, kemudian ditetapkan tujuan dilakukannya penelitian ini yaitu untuk mendeteksi dengan akurat anomali hasil pengukuran penakar hujan otomatis menggunakan metode LSTM.

1. Pengumpulan Data

Tahap pengumpulan data adalah tahap dimana dilakukannya pengumpulan data pengukuran curah hujan dari penakar hujan otomatis. Data yang digunakan adalah data sekunder dimana data diperoleh dari server AWS Center BMKG yang dikelola oleh Pusat Database BMKG. Data tersebut adalah data *time series* data data curah hujan per 10 menit yang dikirim ke server dari 27 titik pengukuran yang terdapat di wilayah DKI Jakarta, Tangerang, Tangerang Selatan, Bekasi dan Depok selama 2 tahun dari 1 Januari 2021 sampai dengan 31 Desember 2022. Sebaran sampel lokasi peralatan penakar hujan otomatis dipilih atas dasar area dengan jarak yang hampir sama dengan memperhatikan kondisi lingkungan yang relatif sama antar lokasi penakar hujan otomatis.

2. Data Preparation

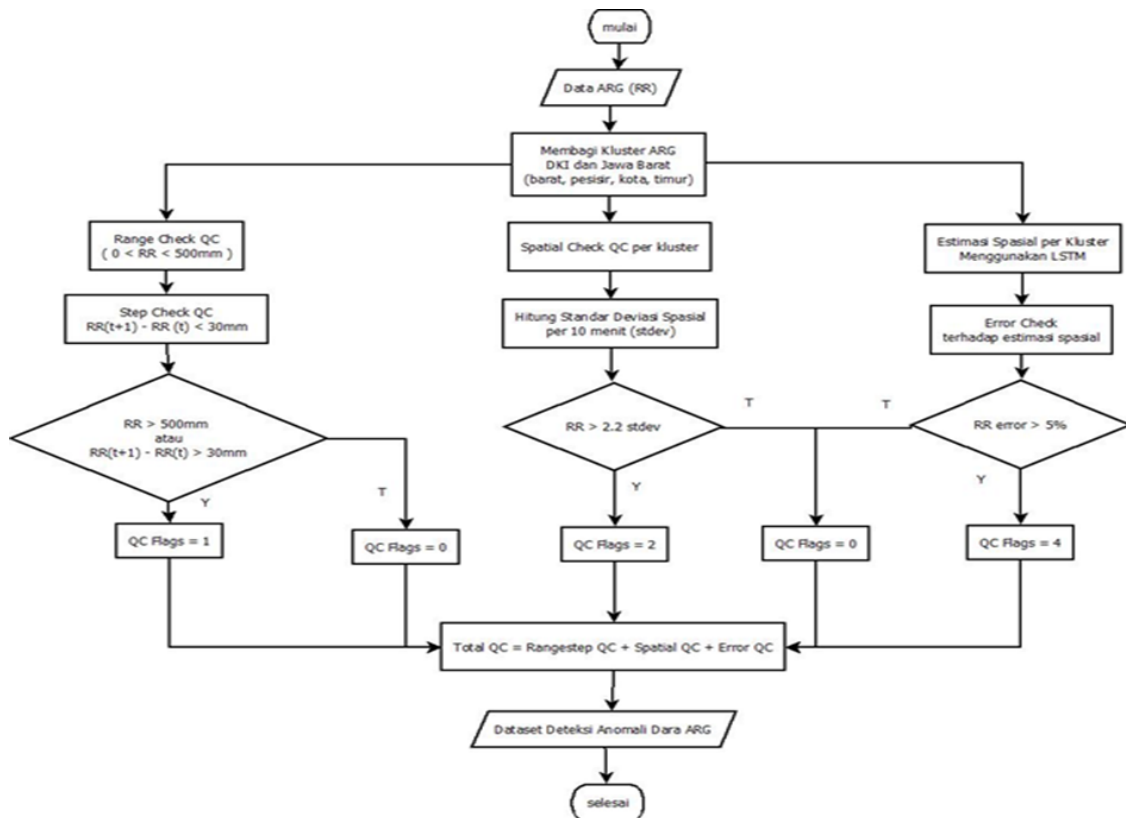
Pada tahap ini dilakukan pra-pemrosesan data untuk pengecekan ketersediaan data, kualitas dan kebersihan data. Data sampel pada setiap lokasi peralatan penakar hujan otomatis dilakukan penghitungan jumlah baris data serta menentukan gab data dari jumlah ideal ketersediaan data. Kemudian dilanjutkan dengan melakukan proses imputasi data kosong untuk akurasi saat proses permodelan deteksi anomali. Hasil dari tahap ini adalah output data lengkap dari setiap lokasi sampel. Diagram alir data *preparation* dapat dilihat pada Gambar 2.



Gambar 2. Diagram Alir Data Preparation

3. Penentuan Label

Pada tahap ini dilakukan pelabelan data dengan metode *quality control* (QC) berbasis *range* dan *step check*, *spatial check* dan *error check*. Hasil dari tahap ini adalah klasifikasi label anomali data yang didapat berdasarkan total nilai QC dari *range* dan *step check*, *spatial check* dan *error check* yang direpresentasikan sebagai *Label Total QC*. Diagram alir pelabelan data dapat dilihat pada Gambar 3.



Gambar 3. Diagram Alir Pelabelan Data

4. Membangun Model

Pada tahap ini dilakukan pengembangan model dengan menerapkan metode LSTM untuk mendeteksi anomali dalam data pengukuran penakar hujan otomatis. Pada model yang dibangun, disesuaikan arsitektur jaringan, termasuk jumlah lapisan LSTM, jumlah neuron, fungsi aktivasi, dan parameter lainnya berdasarkan karakteristik data dan tujuan penelitian ini. Dataset pada tahap permodelan ini adalah data curah hujan per kluster dan label total QC yang didapat pada tahap pelabelan deteksi anomali.

5. Training Model

Setelah model berhasil dibangun tahap selanjutnya adalah melakukan pelatihan terhadap model dengan menggunakan data pengukuran curah hujan yang telah dikumpulkan. Proses ini membagi data menjadi 80% data *training* dan 20% data *testing*. Model hasil *training* dipakai untuk memprediksi menggunakan data *testing*.

6. Evaluasi Model

Tahap selanjutnya adalah tahap evaluasi model untuk mengukur performa model yang dihasilkan dalam mendeteksi anomali hasil pengukuran penakar hujan otomatis. Evaluasi dilakukan menggunakan *confusion matrix* yang ditunjukkan dengan nilai akurasi, presisi, recall, dan F1-score.

C. Hasil dan Pembahasan

1. Hasil Data Preparation

Sebanyak 27 lokasi peralatan penakar hujan otomatis di wilayah DKI Jakarta, sebagian Jawa Barat dan Banten digunakan sebagai sampel pada penelitian ini. Dari 27 lokasi sampel tersebut, dilakukan pengelompokan sebaran lokasi menjadi 4 kluster untuk menemukan korelasi yang lebih erat antar lokasi yang dijadikan sampel sehingga model yang akan dikembangkan lebih optimal untuk setiap lokasi sampel. Pembagian kluster dilakukan berdasarkan grid pengamatan curah hujan (0.5° lintang dan 0.5° bujur), kondisi topografi dan elevasi lokasi alat [4]. Satu kluster pengukuran memiliki karakteristik pola data yang berkorelasi erat satu sama lain. Keempat kluster tersebut diuraikan sebagai berikut:

- a. Kluster Barat
Kluster ini terdiri dari 6 (enam) lokasi penakar hujan otomatis. Kluster ini terletak di wilayah Provinsi Banten dengan elevasi 6 - 18 meter.
- b. Kluster Pesisir
Kluster ini terdiri dari 7 (tujuh) lokasi penakar hujan otomatis. Kluster ini terletak di dekat sepanjang garis pesisir teluk Jakarta dengan elevasi 0 - 3 meter.
- c. Kluster Kota
Kluster ini terdiri dari 10 (sepuluh) lokasi penakar hujan otomatis. Kluster ini terletak di wilayah perkotaan padat pemukiman dengan elevasi 15 - 80 meter.
- d. Kluster Timur
Kluster ini terdiri dari 4 (empat) lokasi penakar hujan otomatis. Kluster ini terletak di wilayah Provinsi Jawa Barat bagian Bekasi dengan elevasi 2 - 20 meter.

Pengecekan ketersediaan data untuk lokasi sampel pada periode 1 Januari 2021 - 31 Desember 2022 dengan interval data per 10 menit dilakukan menggunakan program python. Rentang waktu dengan interval 10 menit periode waktu 1 Januari 2021 sampai dengan 31 Desember 2022 digunakan sebagai sumbu waktu sehingga diperoleh jumlah titik waktu sebanyak 105.120 yang merepresentasikan jumlah baris ideal data untuk setiap lokasi sampel. Dengan demikian, total jumlah raw data setiap lokasi dapat ditemukan dengan cara melakukan pengecekan titik waktu disetiap data setiap lokasi dengan sumbu waktu per 10 menit dari rentang periode 1 Januari 2021 sampai dengan 31 Desember 2022. Terhadap lokasi sampel yang memiliki *missing data*, dilakukan teknik imputasi untuk mengisi data kosong berdasarkan data yang tersedia menggunakan metode tertentu [5]. Metode imputasi yang digunakan adalah metode MICE (Multivariate Imputation Chained Equation) yang

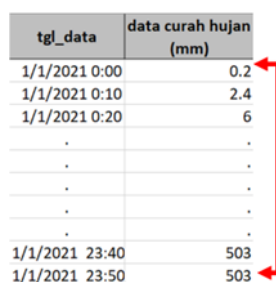
mengkombinasikan metode joint modelling (JM) dan fully conditional specification (FCS) untuk mengimputasi data multivariabel. Kemudian, FCS menentukan model imputasi variabel per variabel secara bergantian berdasarkan teknik Markov Chain Monte Carlo. MICE menghasilkan beberapa nilai imputasi terhadap satu titik missing data melalui proses iterasi [6].

2. Hasil Pelabelan Anomali

Anomali data curah hujan hasil pengukuran penakar hujan otomatis dideteksi melalui metode *quality control* (QC) berbasis *range* dan *step check*, *spatial check* dan *error check*. Setiap proses QC akan menghasilkan label untuk setiap baris data. Jadi untuk satu baris data curah hujan dari satu lokasi sampel akan mempunyai 3 label, yaitu label dari QC *range* dan *step check*, *spatial check* dan *error check*.

A. Pelabelan Anomali Range Check dan Step Check

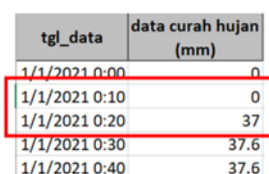
Range check adalah pengecekan data berdasarkan rentang histori per parameter. Jika data tidak berada dalam rentang normal, maka data menjadi *suspect* berdasarkan *range check*. Berdasarkan WMO, rentang curah hujan harian normal berkisar 0-500 mm [1]. Referensi lain menyebut rentang normal yaitu 0-508 mm [7]. *Step check* adalah pengecekan berdasarkan kaitan temporal data saat ini dengan data sebelumnya. Jika selisih data terkini dengan data sebelumnya melebihi ambang tertentu, maka data menjadi *suspect* berdasarkan *step check*. Selisih data maksimum pengukuran curah hujan setelah 10 menit tidak melebihi 30 mm [8]. Setiap baris data pada atribut lokasi yang dijadikan sampel di setiap kluster akan dilakukan pengecekan rentang data dan *step* data, yaitu tidak boleh lebih dari 500 mm untuk rentang data dan tidak boleh lebih dari 30 mm untuk *step* data (selisih antara data saat ini dengan data 10 menit sebelumnya). Gambar 4 menunjukkan ilustrasi anomali *range* data curah hujan.



tgl_data	data curah hujan (mm)
1/1/2021 0:00	0.2
1/1/2021 0:10	2.4
1/1/2021 0:20	6
.	.
.	.
.	.
.	.
1/1/2021 23:40	503
1/1/2021 23:50	503

Gambar 4. Ilustrasi Anomali Range Data Curah Hujan

Jika data tidak memenuhi kriteria *range check* atau *step check*, maka data curah hujan diberi label 1. Jika data memenuhi kriteria maka data curah hujan diberi label 0. Gambar 5 menunjukkan ilustrasi anomali *step* data curah hujan.



tgl_data	data curah hujan (mm)
1/1/2021 0:00	0
1/1/2021 0:10	0
1/1/2021 0:20	37
1/1/2021 0:30	37.6
1/1/2021 0:40	37.6

Gambar 5. Ilustrasi Anomali Step Data Curah Hujan

B. Pelabelan Anomali *Spatial Check*

Spatial check adalah pengecekan data berdasarkan data spasial di sekitar sensor. Data spasial diperoleh dari data peralatan-peralatan terdekat di sekitar penakar hujan otomatis yang diamati. Anomali hasil pengukuran peralatan penakar hujan otomatis berdasarkan *spatial check* terjadi apabila data curah hujan menyimpang jauh dibanding data-data hasil pengukuran peralatan penakar hujan otomatis di sekitarnya. Parameter penyimpangan ini dapat dideteksi melalui nilai standar deviasi secara spasial [7]. Nilai *spatial check* data curah hujan normal jika nilai pengukuran satu alat tidak melebihi 2,2 kali standar deviasi pengukuran peralatan per kluster [9]. Gambar 6 menunjukkan ilustrasi anomali *spatial* data curah hujan.

tgl_data	data curah hujan di Lokasi 1 (mm)	data curah hujan di Lokasi 2 (mm)	data curah hujan di Lokasi 3 (mm)	data curah hujan di Lokasi 4 (mm)
1/1/2021 7:30	8	5.4	2.6	20

Standard Deviasi kluster = 7.65

Nilai curah hujan di lokasi 4 melebihi 2.2 kali standar deviasi (terjadi anomali spatial)

Gambar 6. Ilustrasi Anomali Spatial Data Curah Hujan

Untuk pelabelan anomali *spatial check*, maka data hujan dari lokasi-lokasi dalam satu kluster berperan untuk menentukan label, yaitu nilai standar deviasi dari nilai curah hujan kluster. Jika data tidak memenuhi kriteria *spatial check*, maka data curah hujan diberi label 2. Jika data memenuhi kriteria maka data curah hujan diberi label 0.

C. Pelabelan Anomali *Error Check*

Error check adalah pengecekan data anomali berdasarkan kinerja akurasi sensor peralatan penakar hujan otomatis. Persyaratan akurasi peralatan penakar hujan otomatis menurut WMO yaitu memiliki error maksimum 5% terhadap alat standar [1]. Dilakukan *error check* terhadap masing-masing lokasi sampel, dimana error maksimum tidak boleh melebihi dari 5% terhadap data referensi. Gambar 7 menunjukkan ilustrasi anomali *error* data curah hujan.

tgl_data	data curah hujan di Lokasi yg diuji (mm)	data curah hujan penakar hujan standar (mm)
1/1/2021 07:30	7.8	8.6

Error(%) = 9.302325581

Gambar 7. Ilustrasi Anomali Error Data Curah Hujan

Jika data tidak memenuhi kriteria *error check*, maka data curah hujan diberi label 4. Jika data memenuhi kriteria maka data curah hujan diberi label 0.

Hasil estimasi spasial dari lokasi lain dalam satu kluster berperan sebagai data referensi. Estimasi spasial menggunakan metode regresi berbasis algoritma LSTM. Tabel 1 menunjukkan struktur algoritma LSTM sebagai estimator data curah hujan referensi [10].

Tabel 1. Struktur Algoritma LSTM sebagai Estimator Curah Hujan Referensi

Input Model	Jumlah Neuron	Detail Fungsi Aktivasi	Batch Size	Jumlah Epoch
Data curah hujan per kluster	Layer 1 : 40	Layer 1 : sigmoid	20	50
	Layer 2 : 40	Layer 2 : sigmoid		
	Layer 3 : 1	Layer 3: linear		

Untuk model estimasi spasial LSTM data curah hujan digunakan data curah hujan sampel lokasi dari setiap kluster. Jumlah data yang digunakan dari setiap sampel lokasi dibagi menjadi 80% data *training* dan 20% data testing.

Pada pelatihan model estimasi spasial menggunakan algoritma LSTM dihasilkan performa sebagai terlihat pada Tabel 2.

Tabel 2. Performa Model Estimasi Spasial Menggunakan LSTM

Kluster	Lokasi Sampel	RMSE	MAPE
Barat	ARWS Stageof Tangerang	6.529	1.842
Pesisir	ARG Tanjung Priok	9.033	3.897
Kota	ARG Ciganjur	3.767	1.879
Timur	ARG PJT II Muara	6.768	2.075

Data curah hujan hasil estimasi spasial LSTM dikomparasi terhadap data curah hujan dari lokasi sampel untuk memperoleh persentase nilai error [11]. Nilai persentase error dinyatakan sebagai berikut [12] :

$$\% \text{ err} = \frac{\text{nilai aktual} - \text{nilai referensi}}{\text{nilai referensi}} \times 100\% \quad (1)$$

Jumlah input model bergantung pada total lokasi penakar hujan otomatis per kluster. Data curah hujan dari lokasi sampel diestimasi melalui curah hujan dari lokasi lain per kluster.

Hasil pelabelan deteksi anomali dari ketiga QC akan dijumlahkan menjadi label Total QC. Secara keseluruhan, hasil deteksi anomali data pengukuran penakar hujan otomatis dapat diklasifikasikan berdasarkan total nilai QC per prosedur pengecekan. Tabel 3 menunjukkan daftar klasifikasi per pelabelan anomali data hasil pengukuran penakar hujan otomatis.

Tabel 3. Klasifikasi Pelabelan Anomali Data Hasil Pengukuran Penakar Hujan Otomatis

Label Total QC	Jenis Anomali Data Pengukuran
0	Tidak ada anomali data
1	Data pengukuran di luar rentang normal
2	Data pengukuran anomali secara spasial
3	Data pengukuran di luar rentang normal dan anomali secara spasial
4	Data pengukuran terindikasi error dan kurang akurat
5	Data pengukuran di luar rentang normal dan terindikasi error
6	Data pengukuran anomali secara spasial dan terindikasi error
7	Data pengukuran di luar rentang normal, anomali secara spasial dan terindikasi error

3. Hasil Membangun dan Training Model Klasifikasi Deteksi Anomali

Klasifikasi deteksi anomali data pengukuran penakar hujan otomatis menggunakan algoritma LSTM. Dataset pemodelan klasifikasi meliputi data hasil pengukuran curah hujan per kluster dan label total QC sesuai Tabel 3. Output deteksi anomali dihasilkan per 10 menit sesuai interval pengukuran penakar hujan otomatis.

Pemodelan deteksi anomali diawali dengan mentransformasi data total QC. Total QC tergolong data kategori. Data ini ditransformasi melalui metode *one-hot encoding*. Transformasi ini mengkonversi data kategori total QC menjadi nilai 0 dan 1. Posisi nilai 1 menjadi pembeda antara data kategori satu dengan data lainnya [13]. Hasil konversi data total QC melalui *one-hot encoding* ditunjukkan pada Tabel 4.

Tabel 4. Hasil Konversi Data Total QC Melalui One-Hot Encoding

Total QC	Hasil Transformasi One-Hot Encoding							
0	1	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0
2	0	0	1	0	0	0	0	0
3	0	0	0	1	0	0	0	0
4	0	0	0	0	1	0	0	0
5	0	0	0	0	0	1	0	0
6	0	0	0	0	0	0	1	0
7	0	0	0	0	0	0	0	1

Dataset hasil transformasi selanjutnya dilakukan segmentasi. Proses ini membagi menjadi 80% data *training* dan 20% data *testing*. Model hasil *training* dipakai untuk memprediksi data *testing*. Struktur algoritma LSTM dinyatakan pada Tabel 5.

Tabel 5. Struktur Algoritma LSTM Sebagai Classifier

Dataset Model	Jumlah Neuron	Detail Fungsi Aktivasi	Batch Size	Jml Epoch
Data Curah Hujan per kluster (input)	Layer 1 : 40	Layer 1 : relu	20	50
	Layer 2 : 40	Layer 2 : relu		
dan total QC (output)	Layer 3 : 8	Layer 3: softmax		

4. Hasil Evaluasi Model Klasifikasi Deteksi Anomali

Pengujian pemodelan deteksi anomali data hasil pengukuran penakar hujan otomatis diukur melalui uji performa klasifikasi. Output klasifikasi dikonversi menjadi *confusion matrix*. Matriks ini mengukur performa klasifikasi melalui pemetaan total data prediksi terhadap data aktual. Tabel 6 menunjukkan *confusion matrix* klasifikasi anomali data hasil pengukuran penakar hujan otomatis per kluster.

Tabel 6. Confusion Matrix Klasifikasi Anomali Data Pengukuran Curah Hujan

Kluster	Sampel	Confusion Matrix									
Barat	ARWS Stageof Tangerang		PREDIKSI								
			0	1	2	3	4	5	6	7	
		A	0	17727	0	195	0	145	0	0	0
		1	571	0	1	0	6	0	0	0	
		K	2	5	0	690	0	0	0	97	0
		T	3	0	0	0	0	0	0	0	0
		U	4	111	0	0	0	250	0	133	0
		A	5	1	0	0	0	5	0	0	0
Pesisir	ARG Tanjung Priok		PREDIKSI								
			0	1	2	3	4	5	6	7	
		A	0	15505	132	1	0	0	0	0	0
		1	165	11	0	0	0	0	0	0	
		K	2	300	6	2655	543	0	0	1	0
		T	3	6	12	18	63	0	0	0	0
		U	4	1	0	0	0	20	0	77	0
		A	5	0	0	0	0	0	0	0	0
Kota	ARG Ciganjur		PREDIKSI								
			0	1	2	3	4	5	6	7	
		A	0	10197	250	1145	19	213	0	0	0
		1	1839	76	188	2	23	0	0	0	
		K	2	406	156	5099	10	0	0	0	0
		T	3	39	16	732	3	0	0	0	0
		U	4	412	0	0	0	120	0	0	0
		A	5	65	0	0	0	14	0	0	0
Timur	ARG PJT II Muara		PREDIKSI								
			0	1	2	3	4	5	6	7	
		A	0	6495	0	10	0	128	0	0	0
		1	21	0	0	0	0	0	0	0	
		K	2	13	0	11306	0	0	0	58	0
		T	3	0	0	38	0	0	0	0	0
		U	4	204	0	0	0	2019	0	0	0
		A	5	1	0	0	0	8	0	0	0

Berdasarkan Tabel 6, sebagian besar data pengukuran curah hujan dari lokasi sampel merupakan data normal. Anomali data curah hujan terbanyak disebabkan oleh penyimpangan secara spasial. Data pengukuran curah hujan di luar rentang normal dan menyimpang secara temporal (label total QC=1) cukup sulit diprediksi secara benar oleh algoritma LSTM. Anomali tipe ini sebaiknya tetap dideteksi melalui fungsi logika “if-then” sederhana. Sementara itu, anomali data pengukuran curah hujan secara spasial cukup mudah dideteksi melalui algoritma LSTM, karena bersifat cukup kompleks.

Confusion matrix menghasilkan 4 atribut hasil prediksi [14] :

- True positive (TP), jumlah data positif yang benar diprediksi
- True negative (TN), jumlah data negatif yang benar diprediksi
- False positive (FP), jumlah data positif yang salah diprediksi
- False negative (FN), jumlah data negatif yang salah diprediksi

Parameter uji performa klasifikasi meliputi nilai akurasi, presisi, *recall* dan *F-1 score*. Parameter-parameter ini mengacu pada atribut *confusion matrix*. Akurasi menyatakan rasio total prediksi klasifikasi yang benar terhadap total seluruh hasil prediksi. Akurasi secara matematis dinyatakan sebagai berikut [14]:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Presisi menyatakan rasio antara total prediksi data positif terhadap total data yang diprediksi positif. Presisi berusaha memperkecil terjadinya *false positive*. Presisi secara matematis dinyatakan sebagai berikut [14]:

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (3)$$

Recall menyatakan rasio antara total prediksi data positif terhadap total data prediksi positif yang benar dengan total data prediksi negatif yang salah. *Recall* berusaha memperkecil terjadinya *false negative*. *Recall* secara matematis dinyatakan sebagai berikut [14]:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

F1 score menyatakan rata-rata harmonik antara *presisi* dan *recall* [15]. *F1 score* secara matematis dinyatakan sebagai berikut [14]:

$$\text{F1 score} = \frac{2 \times \text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (5)$$

Tabel 7 menunjukkan performa klasifikasi deteksi anomali data pengukuran curah hujan oleh penakar hujan otomatis.

Tabel 7. Performa Klasifikasi Deteksi Anomali Data Pengukuran Curah Hujan

Kluster	ARG Sampel	Akurasi	Presisi	Recall	F1-score
Barat	ARWS Stageof Tangerang	0.939	0.913	0.939	0.925
Pesisir	ARG Tanjung Priok	0.939	0.961	0.939	0.946
Kota	ARG Ciganjur	0.737	0.662	0.737	0.690
Timur	ARG PJT II Muara	0.975	0.972	0.975	0.974

Akurasi menunjukkan kemampuan algoritma dalam mengklasifikasikan data secara benar berdasarkan keseluruhan data. Presisi menunjukkan kemampuan algoritma dalam mengklasifikasikan data secara benar berdasarkan total data positif. Sementara itu, *recall* menunjukkan kemampuan algoritma dalam mengklasifikasikan data secara benar berdasarkan total data positif yang benar dan data negatif yang salah. Kedua parameter ini saling mengkompensasi satu sama lain terhadap ketidakseimbangan komposisi data klasifikasi. Peningkatan presisi diiringi dengan penurunan *recall* dan sebaliknya. Parameter *F1-score* menjadi penengah dalam menyatakan kemampuan presisi dan *recall* secara

bersamaan. Nilai presisi, recall dan F1-score deteksi anomali di kluster barat, pesisir dan timur cukup baik yaitu lebih dari 0.9.

D. Simpulan

Penelitian sebelumnya yang dilakukan oleh [2] yang mengembangkan metodologi deteksi anomali jaringan pengukuran curah hujan secara online untuk meminimalisir dampak kesalahan pengukur curah hujan pada layanan peringatan banjir perkotaan menggunakan metode Support Vector Regression (SVR), berhasil memudahkan dalam mengevaluasi pengukur hujan tertentu terkait ketidaksesuaian antara hasil pengukuran pengukur hujan sebenarnya dan estimasi SVR dalam periode waktu tertentu, serta kesalahan pengukur curah hujan yang dapat terdeteksi secara efektif selama kejadian curah hujan.

Pada penelitian lain yang dilakukan oleh [16] yang membuat suatu sistem yang dapat memudahkan pengontrolan kondisi sensor khususnya sensor suhu, tekanan udara, kelembaban udara, dan curah hujan menggunakan korelasi antar parameter cuaca yang digunakan untuk mengontrol kondisi sensor yang dilatih dan diuji menggunakan metode artificial neural network (ANN) backpropagation. Hasil akurasi antara output model dengan target saat training dan testing menunjukkan nilai yang sangat tinggi dan estimasi tersebut dapat dijadikan peringatan untuk dapat dilakukan pengecekan sensor secepatnya untuk menjaga kualitas data.

Penelitian yang dilakukan oleh [3], membandingkan pendekatan model LSTM dan dua model algoritma konvensional dalam memprediksi kondisi motor 10 kV. Metode LSTM menunjukkan prediksi sinyal temperatur winding motor 10 kV yang lebih akurat dan mendekati kondisi aktual dibandingkan dengan prediksi sinyal menggunakan metode algoritma konvensional ARMA dan RNN biasa. Nilai mean absolute percentage error (MAPE) model LSTM yang diperoleh yaitu 3,8%.

Pada penelitian, ini LSTM mampu mengklasifikasikan anomali data hasil pengukuran curah hujan pada kluster barat, timur dan pesisir dengan cukup baik. Nilai akurasi kluster-kluster tersebut lebih dari 0.9, sehingga >90% anomali diklasifikasikan secara benar. Namun demikian, performa klasifikasi deteksi anomali berbasis LSTM belum cukup baik diimplementasikan terhadap data pengukuran penakar hujan di kluster kota. Hal ini disebabkan adanya variabilitas data yang lebih tinggi di kluster kota dan jumlah penakar hujan yang cukup banyak dalam satu kluster. Lingkungan instalasi penakar hujan otomatis di daerah perkotaan juga mempengaruhi deteksi anomali data hasil pengukuran secara spasial di perkotaan, sehingga pembagian kluster perlu dipertimbangkan ulang.

E. Referensi

- [1] World Meteorological Organization, *Volume I – Measurement of meteorological variables*, vol. I, no. 8. 2018. [Online]. Available: https://library.wmo.int/index.php?lvl=notice_display&id=12407%0Ahttps://library.wmo.int/index.php?lvl=notice_display&id=12407#YkdSz3XMLio
- [2] G. Fattoruso *et al.*, "Online anomaly detection on rain gauge networks for robust alerting services to citizens at risk from flooding," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10406 LNCS, no. December, pp. 427–442, 2017, doi: 10.1007/978-3-319-

- 62398-6_30.
- [3] M. K. Wisyaldin, G. M. Luciana, and H. Pariaman, "Pendekatan Long Short-Term Memory untuk Memprediksi Kondisi Motor 10 kV pada PLTU Batubara," *J. Kilat*, vol. 9, no. 2, pp. 311–318, 2020.
 - [4] C. Funk *et al.*, "The climate hazards infrared precipitation with stations - A new environmental record for monitoring extremes," *Sci. Data*, vol. 2, pp. 1–21, 2015, doi: 10.1038/sdata.2015.66.
 - [5] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, and O. Tabona, *A survey on missing data in machine learning*, vol. 8, no. 1. Springer International Publishing, 2021. doi: 10.1186/s40537-021-00516-9.
 - [6] S. Van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in R," *J. Stat. Softw.*, vol. 45, no. 3, pp. 1–67, 2011, doi: 10.18637/jss.v045.i03.
 - [7] C. A. Fiebrich, Y. R. Morgan, A. G. Mccombs, P. K. Hall, and R. A. Mcpherson, "Quality assurance procedures for mesoscale meteorological data," *J. Atmos. Ocean. Technol.*, vol. 27, no. 10, pp. 1565–1582, 2010, doi: 10.1175/2010JTECHA1433.1.
 - [8] K. Kurnianto, "PENGEMBANGAN METODA VALIDASI SENSOR DENGAN MENGGUNAKAN AUTOASSOCIATIVE NEURAL NETWORK," *Pus. Pengemb. Teknol. Reakt. Ris. - BATAN*, pp. 247–257, 2005.
 - [9] D. Kim, B. Nelson, and L. Cedrone, "Reprocessing of historic Hydrometeorological Automated Data System (HADS) precipitation data," *86th AMS Annu. Meet.*, pp. 1–6, 2006.
 - [10] C. Hu, Q. Wu, H. Li, S. Jian, N. Li, and Z. Lou, "Deep learning with a long short-term memory networks approach for rainfall-runoff simulation," *Water (Switzerland)*, vol. 10, no. 11, pp. 1–16, 2018, doi: 10.3390/w10111543.
 - [11] S. Huang, F. Lin, and B. M. Chen, "Model-based fault detection, isolation and fault-tolerant control for a class of nonlinear uncertain systems using information estimation," *Chinese Control Conf. CCC*, vol. 2015-Septe, pp. 6164–6169, 2015, doi: 10.1109/ChiCC.2015.7260606.
 - [12] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.
 - [13] C. Seger, "An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing," *Degree Proj. Technol.*, p. 41, 2018.
 - [14] M. Kubat, *An Introduction to Machine Learning*. 2017. doi: 10.1007/978-3-319-63913-0.
 - [15] T. Iskandar Zulkarnain Maulana Putra, A. Farhan Bukhori, dan Ilmu Pengetahuan Alam, and U. Gadjah Mada, "Model Klasifikasi Berbasis Multiclass Classification dengan Kombinasi Indobert Embedding dan Long Short-Term Memory untuk Tweet Berbahasa Indonesia (Classification Model Based on Multiclass Classification with a Combination of Indobert Embedding and Long ," *J. Ilmu Siber dan Teknol. Digit.*, vol. 1, no. 1, pp. 1–28, 2022, [Online]. Available: <https://doi.org/10.35912/jisted.v1i1.1509>
 - [16] P. Wellyantama and S. Soekirno, "Temperature, pressure, relative humidity

and rainfall sensors early error detection system for automatic weather station (AWS) with artificial neural network (ANN) backpropagation," *J. Phys. Conf. Ser.*, vol. 1816, no. 1, 2021, doi: 10.1088/1742-6596/1816/1/012056.