

ANALISIS *TIME SERIES* PADA DATA METEOROLOGI BMKG KOTA BANTEN MENGGUNAKAN METODE *GATED RECURRENT UNIT*

Radika Yudha Riyanto

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara
Jl. Letjen S. Parman No.1, Jakarta Barat, DKI Jakarta, Indonesia 11410
e-mail: radika.535190028@stu.untar.ac.id

ABSTRAK

Dalam kehidupan sehari-hari keadaan cuaca dan iklim sangat mempengaruhi segala aktivitas manusia. Prediksi cuaca yang lengkap dan akurat sangat dibutuhkan agar dapat meningkatkan kinerja dari berbagai bidang aktivitas manusia. Metode dalam melakukan prediksi telah diuji dalam proses berbagai penelitian, salah satunya adalah *Gated Recurrent Unit* (GRU). Penelitian dilakukan dengan data cuaca berupa temperatur rata-rata, kelembapan rata-rata, dan kecepatan angin rata-rata dengan metode *Gated Recurrent Unit* (GRU). Parameter yang mempengaruhi hasil prakira model GRU dalam penelitian ini yaitu epoch. Akurasi terbaik rata-rata diperoleh dengan epoch 100 dan nilai RMSE dan MAE terbaik diperoleh yaitu 1,31 dan 1,04%.

Kata kunci: *Peramalan, Time Series, Cuaca, Meteorologi, GRU.*

ABSTRACT

In everyday life, weather and climate conditions greatly affect all human activities. Complete and accurate weather forecasts are needed to improve performance in various fields of human activity. Methods for making forecasts have been tested in various research processes, one of which is the Gated Recurrent Unit (GRU). The research was conducted using weather data in the form of average temperature, average humidity, and average wind speed with the Gated Recurrent Unit (GRU) method. The parameter that influenced the results of the GRU model prediction in this study was the epoch. The best average accuracy was obtained with an epoch of 100, and the best RMSE and MAE values were 1.31 and 1.04%.

Keywords: *Forecasting, Time Series, Weather, Meteorology, GRU.*

1. PENDAHULUAN

Cuaca merupakan kondisi udara yang diamati dalam periode yang relatif singkat atau pendek pada teritori yang terbatas. *World Climate Conference* menyebutkan bahwa cuaca merupakan keadaan atmosfer yang diukur dengan kompleks meliputi perubahan, perkembangan, dan muncul atau hilangnya suatu fenomena udara [1]. Dalam kehidupan sehari-hari keadaan cuaca dan iklim sangat mempengaruhi segala aktivitas manusia [2]. Cuaca merupakan suatu kondisi udara di suatu tempat pada waktu yang relatif singkat, yang dinyatakan dengan nilai berbagai parameter seperti suhu, tekanan udara, kecepatan angin, kelembapan udara, dan berbagai fenomena atmosfer lainnya. Dengan tingginya pengaruh dalam kehidupan manusia dibutuhkan penelitian-penelitian terkait atmosfer cuaca dan juga mengembangkan metode-metode peramalan untuk menentukan kondisi cuaca sehingga mampu mendapatkan hasil yang baik dan akurat untuk di masa depan.

Peramalan adalah kegiatan memprediksi apa yang akan terjadi di masa depan. Sedangkan forecast adalah situasi atau kondisi yang diperkirakan akan terjadi di masa yang akan datang. Deret waktu atau deret waktu adalah urutan pengamatan yang diukur pada rentang waktu pada interval waktu yang sama. Analisis data deret waktu sangat aplikatif dalam dunia bisnis untuk memprediksi atau memperkirakan nilai akuisisi data di masa mendatang berdasarkan data masa lalu. Peramalan memiliki perkiraan nilai atau karakteristik masa depan, yaitu prediksi, proyeksi, dan tren. Ramalan tidak pasti dan permintaan tidak pasti karena beberapa faktor, yaitu persaingan, perilaku konsumen, siklus bisnis, upaya penjualan, siklus hidup produk, variasi acak, dan lain-lain. [3]

Tujuan penelitian ini adalah untuk mempelajari data iklim historis guna mengidentifikasi faktor penyebab perubahan sebagai landasan perencanaan masa depan. Selain itu, penelitian ini bertujuan untuk menjadi referensi bagi peneliti selanjutnya dalam penggunaan metode *Gated Recurrent Unit* (GRU) untuk peramalan data BMKG. Terakhir, penelitian ini akan mengkaji penerapan metode GRU dalam memprediksi data *time series* iklim di Kota Tangerang Selatan (Banten).

2. METODE PENELITIAN

GRU merupakan metode pengembangan dari *Recurrent Neural Network* (RNN) yang bertujuan untuk membuat setiap recurrent unit bisa dapat menangkap hubungan (dependensi) dalam skala waktu yang berbeda-beda secara adaptif. Secara analogi, metode GRU menggunakan beberapa informasi pada masa lalu (sebelumnya) yang relevan untuk dapat membuat keputusan sekarang [4][5].

Teknik evaluasi yang digunakan dalam penyusunan laporan ini adalah dengan melakukan pengumpulan data, menentukan metode yang digunakan, melakukan pembersihan data (*cleaning data*) dari data yang terjadi *missing value*, melakukan uji coba serta membuat laporan hasil evaluasi. Pada prediksi ini digunakan *Root Mean Square Error* (RMSE) dan *Mean Absolute Error* (MAE) sebagai metode evaluasi. RMSE adalah rata-rata perbedaan yang dikuadratkan di antara nilai yang diramalkan dengan yang diamati. MAE adalah salah satu metode yang digunakan untuk mengukur tingkat keakuratan model peramalan [6][7].

3.1 Dataset

Data time series yang sudah di kumpulkan adalah data time series Kota Tangerang Selatan (Banten) yang bersumber dari data BMKG online yang dapat di akses melalui <http://dataonline.bmkg.go.id> adalah data-data iklim untuk Januari 2010 – Oktober 2022 yang bertujuan agar metode ini untuk menemukan pola-pola pada deret kondisi cuaca masa lalu dan mengekstrapolasi pola-pola tersebut di masa yang akan datang sehingga hasilnya dapat digunakan sebagai acuan untuk memprediksi nilai-nilai yang akan datang. Ada 10 sampel yang digunakan pada pengujian data time series Kota Tangerang Selatan. Berikut pada Gambar 1 ditampilkan contoh dari bentuk dataset yang sudah di unduh.

	A	B	C	D	E	F	G	H	I	J	K
	Tanggal	Ts	Tm	Tang	rh_hum	rh	u	u_s	dbt_h	rh_hum	dbt_car
1	2010-01-01	21	33.4	27.8	78	0.1	4.8	4	270	2 W	
2	2010-01-02	23	33.4	27.8	78	0	5.8	4	270	2 W	
3	2010-01-03	25	33.4	28.8	79	14.8	4.2	4	270	3 W	
4	2010-01-04	26	34.4	29.7	72	0	8	5	270	4 W	
5	2010-01-05	26	32.8	28.1	77	0	1.7	4	225	3 SW	
6	2010-01-06	26	33.4	28.1	74	2.94	5.4	5	270	3 W	
7	2010-01-07	24	34.8	28.4	76	0	1	3	260	2 N	
8	2010-01-08	24	28.4	28.7	86	61	0.3	3	260	0 N	
9	2010-01-09	24	31.4	27.8	84	14	0.7	3	260	2 N	
10	2010-01-10	24	31.8	27	84	21	1.6	3	260	2 N	
11	2010-01-11	24	32	26.8	87	0	1.8	2	270	2 W	
12	2010-01-12	26	28.8	25.7	91	1	1	4	270	3 W	
13	2010-01-13	24	25.8	26.1	87	1	1	3	270	2 W	
14	2010-01-14	24	28.4	25.8	86	6	2.8	3	260	2 N	
15	2010-01-15	23	31	25	88	42.7	1.7	2	260	1 N	
16	2010-01-16	22	28.8	24.8	88	11.858	0	2	270	0 N	
17	2010-01-17	24	28.4	26	91	7	0	4	270	2 W	
18	2010-01-18	23	32.8	27.2	85	8.8	2.8	3	260	2 N	
19	2010-01-19	23	30	26	88	20.3	1	3	260	2 N	
20	2010-01-20	24	30.8	25.4	90	1.2	0.2	1	270	2 W	
21	2010-01-21	24	31.2	27.7	80	10	6.8	3	260	2 N	
22	2010-01-22	24	31.2	25.5	89	0	2.2	3	270	2 W	
23	2010-01-23	24	33.4	28.1	87	8.4	1.8	3	270	2 W	
24	2010-01-24	25	34	27.8	84	0	7.2	3	260	2 N	
25	2010-01-25	24	33.4	27.8	84	4	1.8	3	260	2 N	
26	2010-01-26	24	31.2	27.2	88	5.2	1	3	260	2 N	
27	2010-01-27	24	34	27.4	84	2.2	6.8	3	270	2 W	
28	2010-01-28	24	33.8	28.1	78	1.2	5.8	3	260	2 N	

Gambar 1. Alur Penelitian

3.2 Pra-pemrosesan Data

Data Time Series yang sudah di kumpulkan adalah Data Time Series Kota Tangerang Selatan (Banten) yang bersumber dari data BMKG yang dapat di akses melalui <http://dataonline.bmkg.go.id> adalah data-data cuaca untuk beberapa tahun ke belakang (2010 - 2022) yang bertujuan untuk mengekstrapolasi pola-pola tersebut di masa yang akan datang sehingga hasilnya dapat digunakan sebagai acuan untuk memprediksi nilai-nilai yang akan datang.

Preprocessing data adalah sebuah teknik yang digunakan untuk mengubah data mentah ke dalam format yang lebih berguna serta efisien. Inisiatif tersebut diperlukan karena data mentah seringkali tidak lengkap dan mempunyai format yang tidak konsisten. Kualitas dari data sendiri mempunyai korelasi langsung dengan keberhasilan pada setiap proyek yang melibatkan analisis data. *Preprocessing* melibatkan validasi serta imputasi data. Tujuan dari validasi sendiri adalah untuk menilai tingkat kelengkapan serta akurasi data yang tersaring. Salah satu masalah yang paling umum pada tahap *preprocessing* data adalah *missing value* yaitu data yang tidak akurat karena adanya informasi yang hilang sehingga menyebabkan informasi yang ada di dalamnya tidak relevan. *Missing value* sering terjadi pada tahap pengumpulan data. Gambar sebelumnya merupakan dataset yang belum dilakukan *preprocessing* dimana pada beberapa field dari data set tersebut terdapat *missing value* yang ditandai dengan *value* 8888 dan 9999. Kemudian dapat dilihat pada Gambar 2 adalah dataset yang telah dilakukan *preprocessing* data dimana *missing value* yang sebelumnya bernilai 8888 sudah digantikan oleh *value* yang telah dihitung sesuai dengan persentase *missing value* dari setiap variabel yang dapat dilihat pada Tabel 1.

Tanggal	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	ddd_car
2000-01-01	25	33,4	27,9	78	0,1	6,8	1	270	2 W	
2000-01-02	25	34,4	27,8	78	0	5,9	4	270	2 W	
2000-01-03	25	33,4	28,8	79	14,6	4,2	4	270	3 W	
2000-01-04	26	34,4	29,7	72	0	8	3	270	4 W	
2000-01-05	26	32,8	28,1	77	0	1,7	4	225	2 SW	
2000-01-06	26	33,4	29,1	74	2,94	0,4	3	270	3 W	
2000-01-07	24	34,8	28,4	76	0	5	3	350	2 N	
2000-01-08	24	29,4	26,7	86	61	0,5	3	350	0 N	
2000-01-09	24	31,4	27,8	84	14	0,7	3	360	2 N	
2000-01-10	24	31,8	27	84	22	1,8	3	350	2 N	
2000-01-11	24	32	26,9	87	0	1,9	2	270	2 W	
2000-01-12	24	29,8	25,7	91	5	1	4	270	3 W	
2000-01-13	24	29,8	26,1	87	1	3	3	270	2 W	
2000-01-14	23	29,4	25,3	96	8	2,5	3	350	2 N	
2000-01-15	21	31	25	88	42,7	1,7	2	360	1 N	
2000-01-16	22	28,6	24,9	86	31,356	0	2	270	0 N	
2000-01-17	24	29,4	26	91	7	0	4	270	2 W	
2000-01-18	23	32,8	27,2	85	6,8	2,6	1	360	2 N	
2000-01-19	23	30	26	88	20,1	1	1	360	2 N	
2000-01-20	24	30,8	25,4	90	3,2	0,3	5	270	2 W	
2000-01-21	24	33,2	27,7	80	10	4,3	1	360	2 N	
2000-01-22	24	31,2	25,5	89	0	2,2	1	270	2 W	
2000-01-23	24	33,4	28,1	81	9,4	5,9	1	270	2 W	
2000-01-24	25	34	27,9	84	0	7,1	1	350	2 N	
2000-01-25	24	31,4	27,3	84	4	1,8	1	360	2 N	
2000-01-26	24	31,2	27,1	86	5,2	1	1	360	2 N	
2000-01-27	24	34	27,4	84	2,2	6,6	1	270	2 W	
2000-01-28	24	33,8	28,3	79	3,1	5,8	1	360	2 N	
2000-01-29	25	33,2	27,9	83	0	6,2	4	360	2 N	
2000-01-30	25	33	28,4	80	9,6	6,4	1	360	2 N	
2000-01-31	24	31	26,6	88	13,5	0	1	360	0 N	

Gambar 2. Bentuk Dataset setelah *Preprocessing*

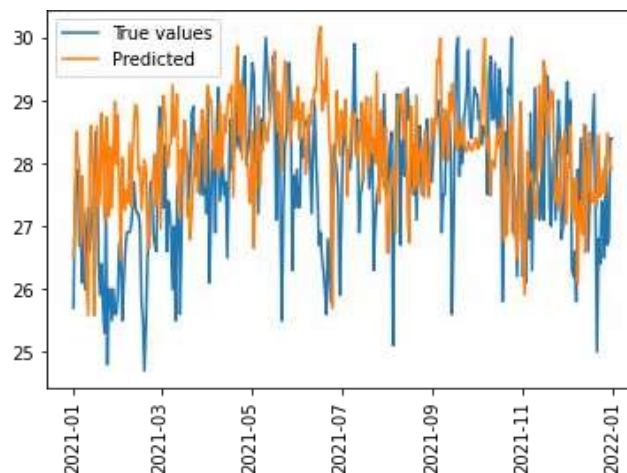
Tabel 1. Persentase *Missing Value*

Variabel	Persentase <i>Missing Values</i>
Tn: Temperatur minimum (°C)	5.397909
Tx: Temperatur maksimum (°C)	5.696608
Tavg: Temperatur rata-rata (°C)	5.568594
RH_avg: Kelembapan rata-rata (%)	5.568594
RR: Curah hujan (mm)	32.024749
ss: Lamanya penyinaran matahari (jam)	4.864519
ff_x: Kecepatan angin maksimum (m/s)	3.883081
ddd_x: Arah angin saat kecepatan maksimum (°)	3.883081
ff_avg: Kecepatan angin rata-rata (m/s)	7.126093
ddd_car: Arah angin terbanyak (°)	7.147429

3. HASIL DAN PEMBAHASAN

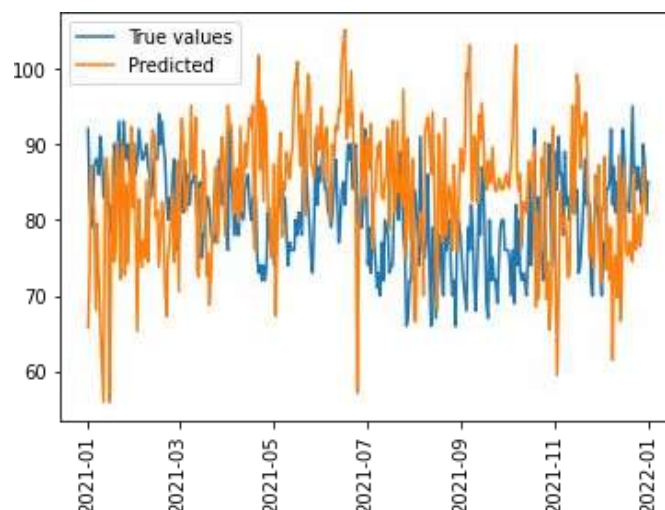
Dataset yang digunakan adalah data set BMKG kota Tangerang Selatan (Banten) yang memiliki 10 variabel, dari ke sepuluh variabel tersebut hanya 3 variabel yang akan digunakan antara lain, Tavg (Temperatur rata-rata (°C)), RH_avg (Kelembapan rata-rata (%)), ff_avg (Kecepatan angin rata-rata (m/s)). Dimana keseluruhan data dibagi menjadi data latih dan data uji. Untuk skenario yang digunakan adalah data latih yang digunakan adalah data dari tahun 2010 sampai tahun 2020 dan data uji yang digunakan adalah data pada tahun 2021.

Variabel yang digunakan adalah 6 dari 10 variabel yang telah ditentukan akan diuji dengan rentang waktu dari tahun 2010 sampai dengan tahun 2018 sebagai data latih untuk kemudian melakukan prediksi Temperatur minimum, Temperatur maksimum, Temperatur rata-rata, Kelembapan rata-rata, Curah hujan dan Lamanya penyinaran matahari pada tahun 2019 dengan menggunakan metode *Gated Recurrent Unit* atau bisa disebut sebagai GRU. Pelatihan *Gated Recurrent Unit* (GRU) dilakukan sebanyak 100 kali epoch. Hasil visualisasi data yang di prediksi dibandingkan dengan *data real* pada data Temperatur rata-rata (Tavg) untuk skenario data latih dari tahun 2010 sampai dengan 2020 dapat dilihat pada Gambar 3.



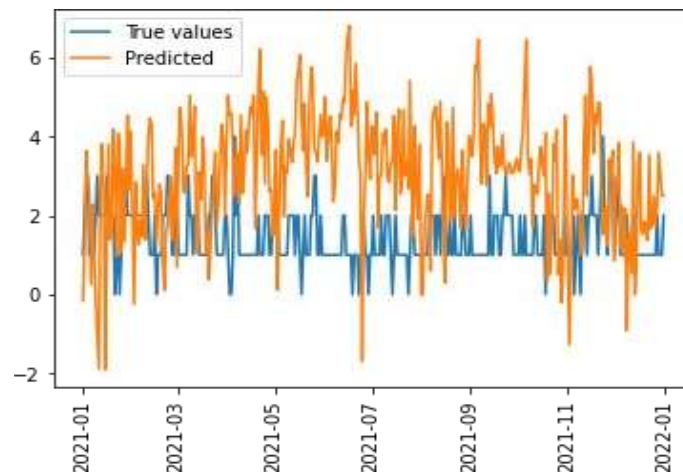
Gambar 3. Hasil Visualisasi Data Prediksi Dan Data Real Temperatur Rata-rata

Hasil visualisasi data yang di prediksi dibandingkan dengan data real pada data Temperatur rata-rata (RH_avg) untuk skenario data latih dari tahun 2010 sampai dengan 2020 dapat dilihat pada Gambar 4..



Gambar 4. Hasil visualisasi data prediksi dan data real pada data Kelembapan rata-rata

Hasil visualisasi data yang di prediksi dibandingkan dengan data real pada data Temperatur rata-rata (ff_avg) untuk skenario data latih dari tahun 2010 sampai dengan 2020 dapat dilihat pada Gambar 5.



Gambar 5. Hasil visualisasi data prediksi dan data real pada data Kecepatan angin rata-rata.

Melakukan evaluasi dengan menggunakan nilai *Mean Absolute Error* (MAE) dan *Root Mean square Error* (RMSE) seperti pada Tabel 2.

Tabel 2. Nilai Evaluasi MAE dan RMSE Prediksi Tahun 2021

Variabel	MAE	RMSE
Tavg: Temperatur rata-rata (°C)	1.041478593680722	1.310623477125447
RH_avg: Kelembapan rata-rata (%)	9.532072406925566	11.562294414954602
ff_avg: Kecepatan rata-rata(m/s)	1.9727252863781297	2.3481210831674337

4. KESIMPULAN

Dari hasil yang di dapatkan dari skenario pertama dengan melakukan pengujian pada data BMKG kota Banten dari tahun 2010 sampai tahun 2020 yang digunakan untuk memprediksi data pada tahun 2021 mendapatkan nilai MSE dan RMSE paling kecil sehingga memiliki akurasi yang paling bagus adalah Temperatur rata-rata (Tavg) yaitu sebesar 1,31 untuk nilai RMSE dan 1,04 untuk nilai MAE sedangkan untuk nilai RMSE dan MAE yang memiliki hasil paling tinggi sehingga memiliki akurasi yang paling buruk adalah Curah hujan (RH_avg) yaitu sebesar 9,53 untuk MAE dan 11,56 untuk nilai RMSE.

Kesimpulan yang dapat di ambil dari hasil analisis data time series BMKG Provinsi Banten Kota Tangerang Selatan adalah nilai MAE dan RMSE pada prediksi 2021 yang memiliki hasil yang baik adalah Tavg karena nilainya mendekati angka 0, Sedangkan yang memiliki hasil yang buruk adalah RH_avg karena nilainya jauh dari angka 0. Hasil yang buruk didapatkan karena adanya outlier yang memiliki nilai selisih yang tinggi dan juga tingginya missing value pada data original sebelum tahap preprocessing. Saran untuk penelitian selanjutnya dapat dilakukan perbandingan error pada variabel yang pada dengan metode RNN yang berbeda seperti LSTM agar didapatkan hasil akurasi yang lebih baik.

DAFTAR PUSAKA

- [1] Luthfiarta, Ardytha, Aris Febriyanto, Heru Lestiawan, and Wibowo Wicaksono. 2020. "Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda." *Journal of Information System*(JOINS) 10-17.
- [2] S. dan A. Rohmah, "Penerapan Jaringan Syaraf Tiruan Metode Hopfield Untuk Prediksicuaca Di Kota Palangka Raya," *Jurnal Sains Komputer dan Teknologi Informasi*, vol. 1, no. 3, pp. 52-59, 2020.
- [3] Miftahuddin. 2016. "Analisis Unsur-unsur Cuaca dan Iklim Melalui Uji ." *Jurnal Matematika, Statistika, & Komputasi* 26-38.
- [4] Qori, Putri Anis, Dhama Sekar Oktafani, and Iqbal Kharisudin. 2022. "Analisis Peramalan dengan Long Short Term Memory pada ." *PRISMA, Prosiding Seminar Nasional Matematika* 752-758.
- [5] Saputra, Rifky Ahmad, Yufis Azhar, and Vinna Rahmayati. 2020. "Prediksi Permintaan Kargo pada Cargo Service Center Tangerang City Menggunakan Metode Gated Recurrent Unit." *Jurnal Repositor* 1123-1132.
- [6] F. N. Adnan dan A. Nahrul, "Analisis Penentuan Data Latih pada PeramalanCurah Hujan Menggunakan Metode Simple Moving Average," *Journal of Information System*, vol. 1, no. 3, pp. 49-58, 2018.
- [7] Wiranda, Laras, and Mujiono Sadikin. 2019. "Penerapan Long Short Term Memory Pada DataTime Series Untuk Memprediksi Penjualan Produk Pt. Metiska Farma." *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI* 184-196.