Akreditasi Sinta 5 KEMRISTEKDIKBUD No. 225/E/KPT/2022; E-ISSN 2723-7958

Analisis Performa *Autoregressive Integrated Moving Average Model* dan *Deep Learning Long Short-Term Memory Model* untuk Peramalan Data Cuaca

Vithiaz Y. Montolalu[1], Cindy P. Munaiseche[2], Made Krisnanda[3]

Program Studi Teknik Informatika UNIMA
Email: vithiaz@gmail.com

***Abstract*— Weather is an aspect that cannot be separated from all activities carried out by humans, so information about the weather is very important. To meet the need for this information, it is necessary to do forecasting. Each data has its own characteristics, and choosing the right forecasting method is very important. The Autoregressive Integrated Moving Average (ARIMA) method is one of the popular statistical methods used in forecasting time-series data. Long Short-Term Memory (LSTM) is a modern deep learning algorithm model that is most suitable for forecasting time series data. In this study, an analysis was carried out to compare the traditional ARIMA method and the deep learning model, namely LSTM, in forecasting weather data in Manado city to see the best forecasting model that can be used. The results of this study indicate that in terms of the accuracy of the 18 tests performed, the LSTM forecasting model is superior to the ARIMA model as measured by Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). In terms of computational time in making forecasting models for 6 weather data attributes, the LSTM model is faster than the ARIMA model.**

***Keywords*— Weather Forecasting, ARIMA, Deep Learning, LSTM**

***Abstrak*— Cuaca merupakan suatu aspek yang tidak dapat dipisahkan dari segala aktivitas yang dilakukan manusia, sehingga informasi mengenai cuaca menjadi hal yang sangat penting. Untuk memenuhi kebutuhan akan informasi tersebut, maka dibutuhkan untuk dilakukannya peramalan. Setiap data memiliki karakteristik tersendiri, dan pemilihan metode peramalan yang tepat merupakan hal yang sangat penting. Metode *Autoregressive Integrated Moving Average* (ARIMA) merupakan salah satu metode statistik yang populer digunakan dalam melakukan peramalan data *time-series*. *Long Short-Term Memory* (LSTM) merupakan model algoritma modern *deep learning* yang paling cocok untuk peramalan data *time series*. Pada penelitian ini dilakukan analisis untuk membandingkan metode tradisional ARIMA dan model *deep learning* yaitu LSTM dalam melakukan peramalan terhadap data cuaca di daerah kota Manado untuk melihat model peramalan terbaik yang dapat digunakan. Hasil dari penelitian ini menunjukkan bahwa dari segi akurasi dari 18 pengujian yang dilakukan, model peramalan LSTM lebih unggul dari model ARIMA yang diukur dengan *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE). Dari segi waktu komputasi dalam pembuatan model peramalan untuk 6 atribut data cuaca, model LSTM lebih cepat dibandingkan model ARIMA.**

***Kata Kunci*— Peramalan Cuaca, ARIMA, Deep Learning, LSTM**

I. PENDAHULUAN

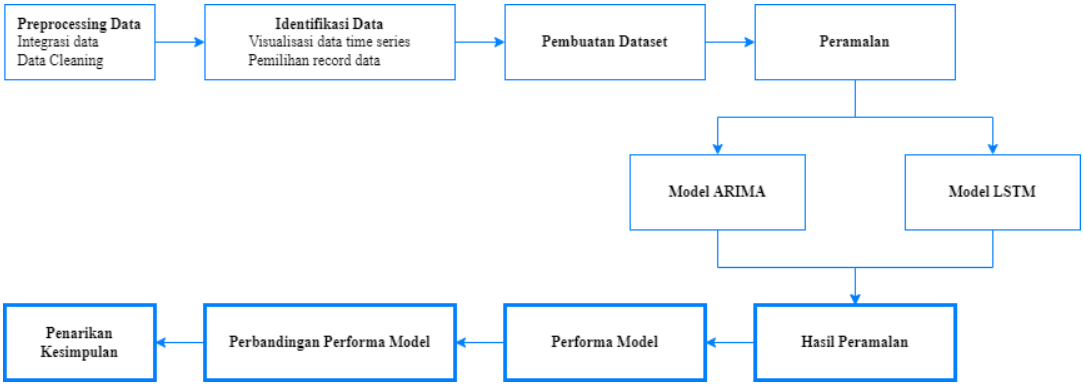
Cuaca merupakan suatu aspek yang tidak dapat dipisahkan dan berkaitan erat dengan segala aktivitas yang dilakukan manusia, sehingga informasi mengenai cuaca adalah merupakan hal yang sangat penting dalam menunjang aktivitas dan pekerjaan manusia. Oleh karena itu, banyak sekali pihak – pihak yang membutuhkan dan bahkan sangat bergantung pada informasi mengenai cuaca pada waktu tertentu.

Untuk memenuhi kebutuhan masyarakat mengenai informasi cuaca pada waktu tertentu, maka dibutuhkan untuk dilakukannya peramalan, yang merupakan suatu proses dengan menggunakan ilmu pengetahuan untuk memprediksi peristiwa yang akan terjadi pada masa mendatang.

Metode *Autoregressive Integrated Moving Average* (ARIMA) merupakan salah satu metode statistik yang populer digunakan dalam melakukan peramalan data *time-series*. Metode ini sering disebut sebagai metode peramalan tradisional atau *traditional forecasting method*. Kemudian dengan semakin berkembangnya kemampuan komputasi dari komputer serta kemajuan algoritma *machine learning* seperti *deep learning*, ditemukan algortima – algoritma baru untuk meramalkan data *time-series*. Algoritma deep learning yang paling cocok digunakan untuk peramalan data *time-series* yaitu *Long Short-Term Memory* (LSTM).

Oleh karena masih kurangnya informasi data perkiraan cuaca yang berhubungan dengan suhu udara, kelembapan, lama penyinaran matahari, dan curah hujan bagi masyarakat di kota Manado dan belum diketahui performa masing – masing dari metode ARIMA dan model LSTM menggunakan data cuaca di kota Manado, maka perlu untuk melakukan penelitian tentang bagaimana performa dari metode *Autoregressive Integrated Moving Average* (ARIMA) dan model *deep learning* yaitu *Long Short-Term Memory* (LSTM) dalam peramalan data cuaca

Akreditasi Sinta 5 KEMRISTEKDIKBUD No. 225/E/KPT/2022; E-ISSN 2723-7958



Gbr. 1 Tahapan Penelitian

di daerah kota Manado. Selain itu perlu juga untuk melakukan komparasi bagaimana performa dari model peramalan tradisional yaitu ARIMA, dan LSTM sebagai representasi dari algoritma *deep learning* dalam melakukan *training data* serta pembuatan model untuk mengetahui kondisi cuaca untuk waktu ke depan.

II. METODOLOGI

A. *Tahapan Penelitian*

Pada penelitian ini terdapat dua bagian tahapan penelitian, bagian pertama pertama adalah tahapan peramalan menggunakan metode ARIMA dan LSTM kemudian bagian kedua adalah analisa performa peramalan data cuaca dari kedua metode tersebut. Seperti terlihat pada Gbr. 1 bagian pertama penelitian ditunjukkan oleh tahapan dengan gambar border yang lebih tipis dan bagian kedua penelitian ditunjukkan oleh tahapan dengan gambar border yang lebih tebal.

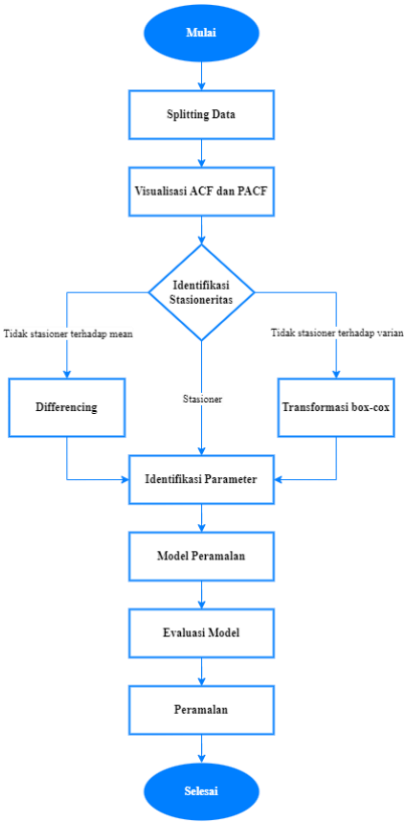
Penjelasan mengenai tahapan penelitian adalah sebagai berikut:

1) Preprocessing data: Dilakukan agar data awal yang diperoleh siap untuk digunakan pada tahap pembuatan model peramalan. Preprocessing data yang dilakukan antara lain: Integrasi data, yaitu penggabungan data dari beberapa file menjadi satu file agar memudahkan untuk diproses pada tahapan selanjutnya. Kedua adalah data cleaning, yaitu proses pembersihan yang dilakukan pada data dengan tujuan untuk menangani data yang kosong ataupun data yang tidak konsisten.

2) Identifikasi data: Dilakukan untuk mengetahui karakteristik data iklim daerah kota Manado dengan melakukan visualisasi time-series pada setiap atribut data iklim daerah kota Manado kemudian melakukan analisis deksriptif pada data. Pada tahapan ini juga dilakukan pemilihan record data untuk masing – masing atribut yang akan digunakan untuk pembuatan model peramalan

3) Pembuatan dataset: Dilakukan pembuatan dataset terpisah untuk setiap atribut data yaitu temperatur minimum, temperatur maksimum, temperatur rata – rata, kelembapan rata – rata, penyinaran matahari dan curah hujan. Setelah melalui tahapan preprocessing dan identifikasi, data sudah siap untuk digunakan dalam pembuatan model peramalan.

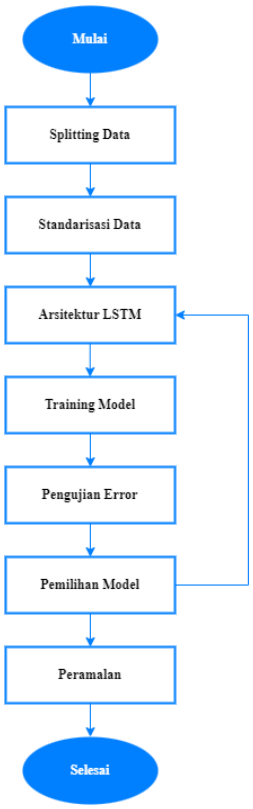
4) Peramalan dengan model ARIMA: Dilakukan peramalan dengan model ARIMA untuk setiap atribut dari data iklim, dengan langkah – langkah dapat dilihat pada *flowchart* di Gbr. 2.



Gbr. 2 *Flowchart* model peramalan ARIMA

5) Peramalan dengan model LSTM: Dilakukan peramalan dengan model LSTM untuk setiap atribut dari data iklim, dengan langkah – langkah dapat dilihat pada *flowchart* di Gbr. 3

Akreditasi Sinta 5 KEMRISTEKDIKBUD No. 225/E/KPT/2022; E-ISSN 2723-7958



Gbr. 3 *Flowchart* model peramalan LSTM

6) Hasil Peramalan: Dilakukan analisis dan perbandingan terhadap hasil peramalan oleh model ARIMA dan model LSTM pada setiap atribut data iklim. Peramalan dilakukan dengan metode *rolling forecast* atau disebut juga dengan *walk-forward model validation* dimana model akan meramalkan setiap time step dari data test kemudian nilai observasi aktual pada time step tersebut di masukkan pada model untuk meramalkan nilai dari time step berikutnya [1].

7) Performa model: Setiap hasil peramalan, dihitung nilai RMSE dan MAE untuk mengetahui model mana yang menghasilkan hasil peramalan paling akurat.

8) Perbandingan performa model: Hasil pengukuran dari performa model ARIMA dan LSTM dibandingkan untuk melihat model yang lebih unggul.

9) Penarikan kesimpulan: Merupakan tahapan terakhir penelitian dengan menarik kesimpulan mengenai performa peramalan model ARIMA dan model LSTM berdasarkan hasil peramalan dan perbandingan kedua model tersebut.

III. HASIL DAN PEMBAHASAN

A. *Data yang digunakan*

Data yang digunakan dalam penelitian ini yaitu data laporan kondisi iklim harian di daerah kota Manado [2] pada rentang waktu Januari 2016 sampai Maret 2022. Data ini berupa file excel (.xlsx) sejumlah 75 file mewakili informasi iklim harian pada 75 bulan yang didalamnya memuat beberapa atribut data cuaca yang akan digunakan dalam penelitian ini yaitu temperatur minimum (ºC), temperatur maksimum (ºC), temperatur rata – rata (ºC), kelembapan rata – rata (%), curah hujan (mm), dan lamanya penyinaran matahari (jam).

B. *Preprocessing Data*

1) Integrasi Data: Sebagai persiapan terhadap data awal yang diperoleh dari website pusat database BMKG maka dibutuhkan proses integrasi untuk menyatukan data harian yang terbagi pada 75 file excel terpisah agar menjadi satu dataset yang memuat seluruh atribut data cuaca yang diperlukan serta semua record data dari atribut – atribut tersebut. Pada preprocessing awal ini, penulis membuat program python agar komputer dapat secara otomatis mengambil nilai yang diperlukan pada setiap file excel kemudian melakukan proses penggabungkan ke dalam file dengan tipe (*.csv*).

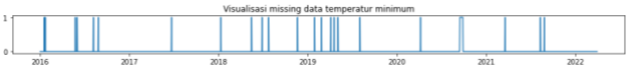
2) Cleaning Data: Untuk menghasilkan peramalan yang optimal maka data yang digunakan dalam pembuatan model peramalan harus memiliki kualitas yang baik. Untuk itu, diperlukan proses pembersihan pada dataset awal untuk menangani data yang hilang ataupun data yang tidak konsisten. Berdasarkan analisis terhadap dataset yang digunakan, record data untuk atribut temperatur minimum, temperatur maksimum, temperatur rata – rata, kelembapan rata – rata, dan lama penyinaran matahari memiliki nilai dan format yang sesuai dengan satuan dari atributnya sehingga dapat dikatakan nilai dari atribut – atribut tersebut sudah konsisten. Tetapi untuk atribut curah hujan, terdapat 190 record data yang memiliki nilai 8888 dan berdasarkan keterangan dari website pusat database BMKG, nilai tersebut merupakan tanda bahwa data tersebut tidak terukur. Oleh karena itu penulis menganggap bahwa record data yang memiliki nilai 8888 sebagai *missing data*.

Setelah setiap record data memiliki nilai yang konsisten, selanjutnya adalah menangani *missing data* yang terdapat pada masing – masing atribut pada dataset. Jumlah dari *missing data* untuk masing – masing atribut disajikan dalam tabel I.

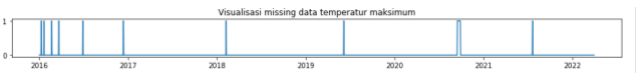
TABEL I. JUMLAH MISSSING DATA PER ATRIBUT

Atribut Data Iklim	Jumlah *Missing Data*
Temperatur minimum	35
Temperatur maksimum	22
Temperatur rata – rata	662
Kelembapan rata - rata	663
Penyinaran matahari	29
Curah hujan	567

Visualisasi *missing data* yang terdapat pada setiap atribut dapat dilihat pada Gbr. 4 sampai Gbr. 9 dimana sumbu x pada gambar masing – masing gambar merupakan rentang data dari atribut tersebut dan sumbu y menandakan ada atau tidaknya data pada rentang tersebut, angka 1 pada sumbu y menandakan bahwa pada rentang tersebut tidak terdapat adanya data.

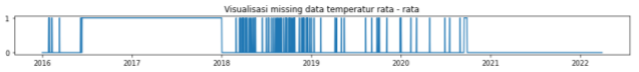


Gbr. 4 Visualisasi *missing data* temperatur minimum



Gbr. 5 Visualisasi *missing data* temperatur maksimum

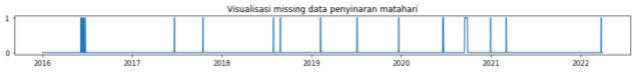
Akreditasi Sinta 5 KEMRISTEKDIKBUD No. 225/E/KPT/2022; E-ISSN 2723-7958



Gbr. 6 Visualisasi *missing data* temperatur rata-rata



Gbr. 7 Visualisasi *missing data* kelembapan rata-rata



Gbr. 8 Visualisasi *missing data* penyinaran matahari



Gbr. 9 Visualisasi *missing data* curah hujan

Berdasarkan Gbr. 6 dan Gbr. 7 yaitu visualisasi *missing data* untuk atribut temperatur rata – rata dan kelembapan rata – rata terlihat bahwa tidak terdapat record sama sekali pada pertengahan tahun 2016 sampai januari 2018. Penanganan *missing data* dengan cara mengganti nilai yang hilang dengan nilai rata – ratanya pada rentang yang panjang dapat menghilangkan informasi penting yang terdapat dalam dataset. Jika dilihat record data diatas bulan januari 2018, hanya terdapat 85 *missing data* untuk temperatur rata – rata dan 86 *missing data* untuk kelembapan rata - rata, sehingga penulis memutuskan untuk hanya mengisi *missing data* pada rentang waktu diatas bulan januari 2018 untuk kedua atribut tersebut.

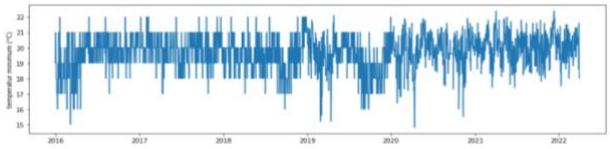
Untuk atribut curah hujan berdasarkan Gbr. 9 dan tabel I, terdapat 567 *missing data* yang tersebar di keseluruhan dataset. Dengan banyaknya jumlah data yang hilang dan posisi *missing data* yang menyebar maka tidak efektif untuk mengganti nilai yang hilang dengan nilai rata – ratanya, penulis juga tidak dapat mengganti nilai yang hilang dengan nilai pada observasi sebelumnya karena jumlah *missing data* yang besar dan tersebar.

Jika dilihat pada Gbr. 9 komposisi *missing data* terbesar berada pada rentang waktu dibawah tahun 2020, oleh karena itu penulis memutuskan untuk hanya mengisi *missing data* pada rentang waktu diatas januari 2020.

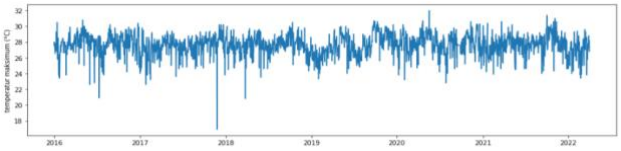
C. Identifikasi Data

Sebelum melakukan identifikasi pada data dengan analisis deskriptif, penulis menentukan rentang *record data* yang akan digunakan untuk setiap atribut berdasarkan proses *cleaning data* yang sudah dilakukan sebelumnya.

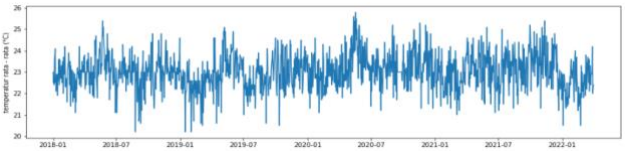
Visualisasi dari dataset yang akan digunakan untuk proses pembuatan model peramalan dapat dilihat pada Gbr. 10 sampai Gbr. 15. Terlihat bahwa untuk atribut temperatur rata – rata (Gbr. 12) dan kelembapan rata – rata (Gbr. 13) seperti sudah dijelaskan sebelumnya maka data yang akan digunakan adalah data mulai dari januari 2018, untuk atribut curah hujan (Gbr. 15) data yang akan digunakan adalah data mulai dari januari 2020, dan untuk atribut temperatur minimum (Gbr. 10), temperatur maksimum (Gbr. 11), dan penyinaran matahari (Gbr. 14) data yang digunakan adalah data mulai dari januari 2016, sama seperti rentang awal sebelum dilakukan proses *data cleaning*.



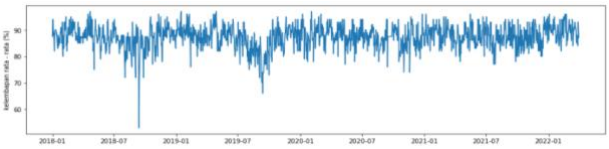
Gbr. 10 Grafik *time-series* temperatur minimum



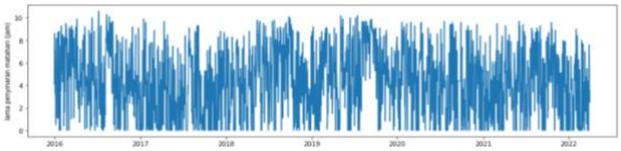
Gbr. 11 Grafik *time-series* temperatur maksimum



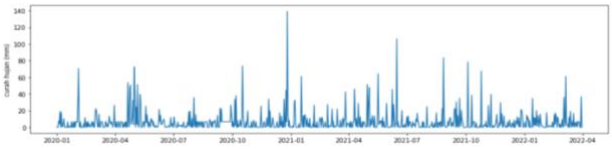
Gbr. 12 Grafik *time-series* temperatur rata-rata



Gbr. 13 Grafik *time-series* kelembapan rata-rata



Gbr. 14 Grafik *time-series* penyinaran matahari



Gbr. 15 Grafik *time-series* curah hujan

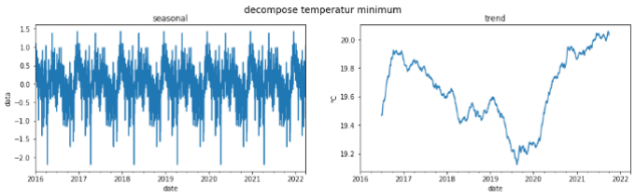
Statistik deskriptif untuk keseluruhan atribut pada dataset yang digunakan disajikan pada tabel II. Berdasarkan tabel tersebut, diketahui bahwa jumlah data yang digunakan untuk penelitian ini bervariasi yaitu 2282 record data untuk atribut temperatur minimum, temperatur maksimum, dan penyinaran matahari, 1551 record data untuk atribut temperatur rata – rata dan kelembapan rata – rata, dan 821 record data untuk atribut curah hujan. Nilai rata – rata untuk temperatur minimum adalah 19.6 ^{o}C dengan nilai diantara 14.8 ^{o}C hingga 22.4 ^{o}C. Nilai rata – rata untuk temperatur maksimum adalah 27.6 ^{o}C dengan nilai diantara 16.9 ^{o}C hingga 32 ^{o}C. Nilai rata – rata untuk temperatur rata – rata adalah 23 ^{o}C dengan nilai diantara 20.2 ^{o}C hingga 25.8 ^{o}C. Nilai rata – rata untuk kelembapan rata – rata adalah 87.3% dengan nilai diantara 53% hingga 97%. Nilai rata – rata untuk penyinaran matahari adalah 4.4 jam dengan nilai diantara 0 hingga 10.6 jam. Nilai rata – rata untuk curah hujan adalah 6.7 mm dengan nilai diantara 0 hingga 139.2 mm.

Akreditasi Sinta 5 KEMRISTEKDIKBUD No. 225/E/KPT/2022; E-ISSN 2723-7958

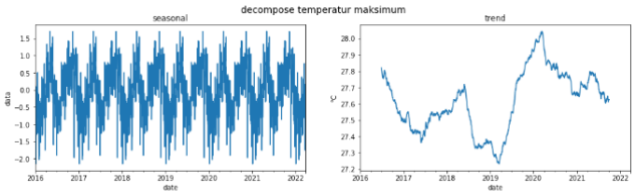
TABEL II. STATISTIK DESKRIPTIF ATRIBUT DATA IKLIM PADA DATASET YANG DIGUNAKAN

Atribut	Jumlah record data	Nilai rata - rata	Rentang Nilai	Standar Deviasi
temperatur minimum (°C)	2282	19.637	14.8 - 22.4	1.199
temperatur maksimum (°C)	2282	27.613	16.9 - 32	1.355
temperatur rata – rata (°C)	1551	23.027	20.2 - 25.8	0.846
kelembapan rata – rata (%)	1551	87.306	53 - 97	4.376
lamanya penyinaran matahari (jam)	2282	4.415	0 - 10.6	2.855
curah hujan (mm)	821	6.688	0 - 139.2	12.497

Berdasarkan visualisasi *time – series* untuk setiap atribut data iklim seperti yang terlihat pada Gbr. 10 sampai Gbr. 15, dilakukan dekomposisi terhadap dataset untuk melihat pola musiman dan trend yang terjadi pada setiap atributnya sehingga menghasilkan grafik seperti pada Gbr. 16 sampai Gbr. 21.

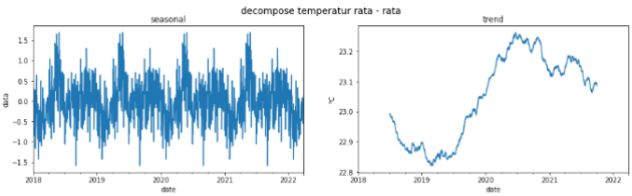


Gbr. 16 Dekomposisi *time-series* temperatur minimum



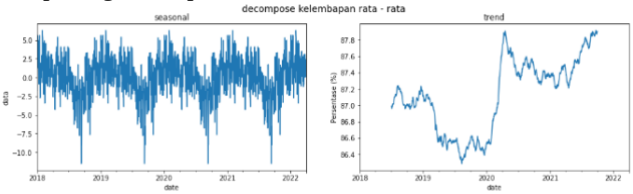
Gbr. 17 Dekomposisi *time-series* temperatur maksimum

Berdasarkan Gbr. 16 diketahui bahwa pola musiman dari temperatur minimum cenderung mengalami penurunan pada bulan pertama hingga bulan ke-3 dan di sekitar bulan ke-6 hingga ke-9. Sebaliknya pola musiman dari temperatur maksimum (Gbr. 17) cenderung mengalami kenaikan pada bulan pertama hingga bulan ke-3 dan bulan ke-6 hingga ke-9. Trend temperatur maksimum yang terjadi mengalami penurunan dari akhir 2016 hingga pertengahan tahun 2019 sekitar rata – rata 1°C, dan mengalami kenaikan kembali hingga tahun 2022. Trend temperatur maksimum (Gambar 1.14) mulai mengalami kenaikan dari awal tahun 2019 hingga 2020 sekitar rata – rata 1°C, kemudian mulai mengalami penurunan secara perlahan hingga tahun 2022.



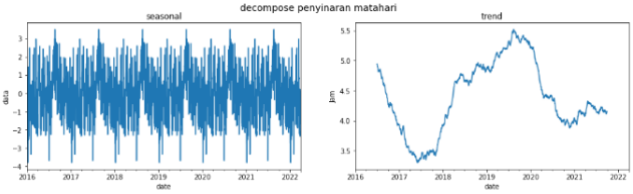
Gbr. 18 Dekomposisi *time-series* temperatur rata-rata

Untuk pola musiman dari temperatur rata – rata seperti terlihat pada Gbr. 18 lebih mirip dengan pola musiman temperatur maksimum (Gbr. 17), dimana grafik mengalami kenaikan pada awal tahun hingga sekitar bulan ke-4 dan turun secara fluktuatif hingga akhir tahun, begitu juga dengan garis trend temperatur rata – rata yang kurang lebih mirip dengan temperatur maksimum.



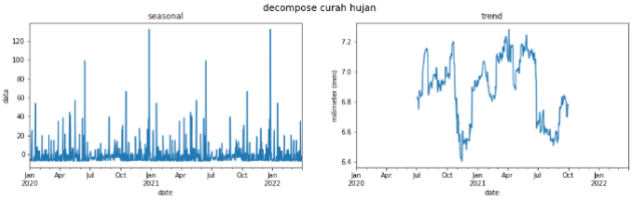
Gbr. 19 Dekomposisi *time-series* kelembapan rata-rata

Untuk pola musiman dari kelembapan rata – rata seperti terlihat pada Gbr. 19 cenderung mengalami penurunan pada sekitar bulan ke-8, mengalami kenaikan hingga awal tahun dan kembali turun secara perlahan hingga bulan ke-8 periode berikutnya. Trend yang terjadi yaitu pola menurun sebelum awal tahun 2020 dan cenderung naik hingga tahun 2022.



Gbr.20 Dekomposisi *time-series* penyinaran matahari

Pola musiman dari penyinaran matahari seperti terlihat pada Gbr. 20 membentuk pola yang sangat fluktuatif per tahunnya, dimana grafik cenderung rendah pada awal tahun dan nilai tertinggi berada pada sekitar bulan ke-8 setiap tahunnya. Dari trend juga terlihat fluktuatif dimana grafik mulai naik pada pertengahan tahun 2017 dan mulai turun kembali dari pertengahan tahun 2019.



Gbr. 21 Dekomposisi *time-series* curah hujan

Jika dilihat pola musiman dari atribut curah hujan pada Gbr. 21, nilai curah hujan terbesar itu terjadi setiap akhir tahun hingga menjelang awal tahun. Sedangkan sekitar bulan juli setiap tahunnya nilai curah hujan selalu rendah. Untuk trend yang terjadi, juga sangat mencerminkan pola musiman dari atribut curah hujan. Jika dilihat secara

Akreditasi Sinta 5 KEMRISTEKDIKBUD No. 225/E/KPT/2022; E-ISSN 2723-7958

kuantitatif, data record dari curah hujan tergolong sedikit dibandingkan atribut lainnya. Tidak hanya itu, pada proses sebelumnya juga terdapat cukup banyak data tidak konsisten yang kemudian digantikan dengan nilai rata – rata dari atribut ini.

Sebagai tahap akhir setelah proses identifikasi data, penulis melakukan pembuatan dataset terpisah untuk setiap atribut data iklim dengan format (*.csv*) untuk memudahkan tahapan analisis berikutnya.

D. Peramalan dengan Model ARIMA

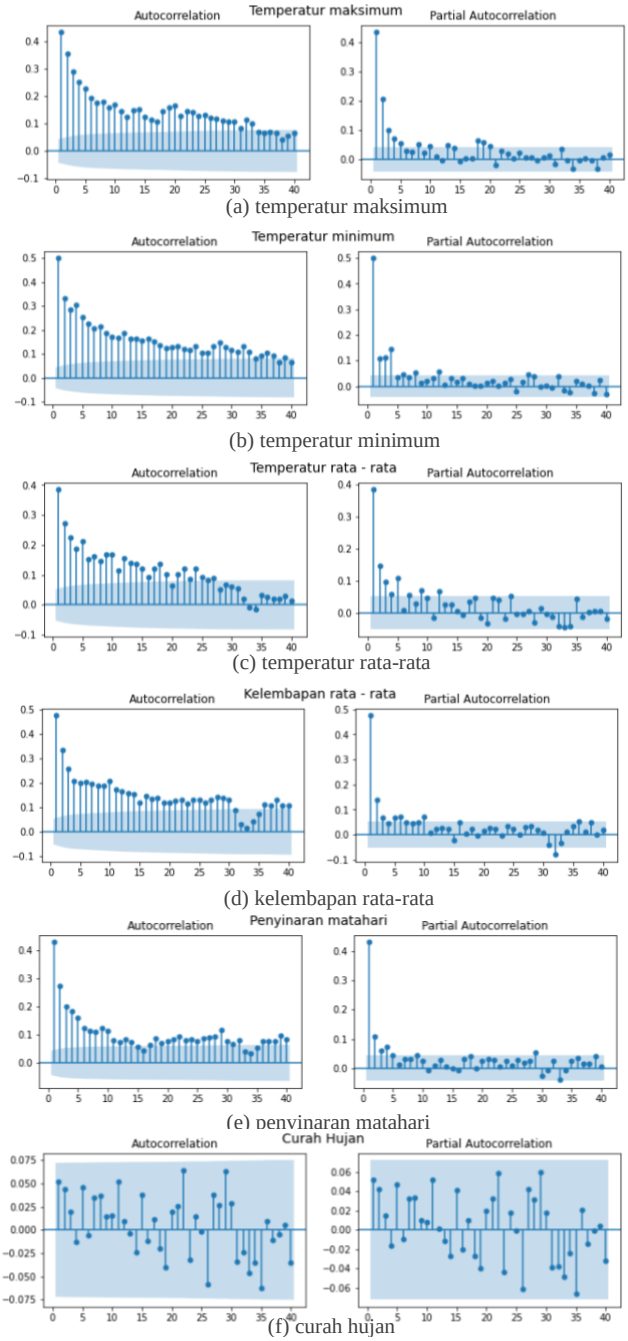
1) Data train dan Data Test: Sebelum melakukan peramalan dengan model ARIMA, dataset yang digunakan dibagi menjadi data train dan data test, dimana data train merupakan data yang digunakan untuk melakukan training pada model atau fitting pada model dan data test digunakan untuk memverifikasi hasil peramalan oleh model yang terbentuk terhadap nilai aktual dengan mengukur tingkat error menggunakan *Root Mean Squared Error* (RMSE) atau menggunakan *Mean Absolute Error* (MAE). Untuk peramalan dengan model ARIMA, data train dan data test dibagi dengan perbandingan 9:1. Tidak ada batasan khusus dalam penentuan rasio ini, namun nilai rasio pada data train harus lebih besar dari data test [3].

2) Autocorrelation Function (ACF) dan Partial Autocorrelation Function (PACF): Fungsi korelasi dapat digunakan untuk mencari tau sifat – sifat dari suatu data time-series. Visualisasi grafik ACF dan PACF untuk setiap atribut dari dataset disajikan pada Gbr. 22.

Jika dilihat pada setiap poin Gbr. 22, dapat dipastikan bahwa temperatur maksimum, temperatur minimum, temperatur rata – rata, kelembapan rata – rata, dan lama penyinaran matahari, memiliki korelasi antara nilai pada suatu observasi dengan nilai sebelumnya pada atribut tersebut. Hal ini dapat dilihat pada Gbr. 22 (a), Gbr. 22 (b), Gambar 4.19 (c), dan Gbr. 22 (d), dimana terdapat nilai pada setiap lag yang keluar dari batas signifikan atau area berwarna biru pada grafik. Berbeda dengan visualisasi ACF dan PACF atribut curah hujan yang dapat dilihat pada Gbr. 22 (e), dimana tidak terdapat nilai pada setiap lag yang keluar dari batas signifikan atau area berwarna biru pada grafik. Ini mengindikasikan bahwa tidak terdapat korelasi antara nilai pada suatu observasi dengan nilai sebelumnya.

Kondisi dimana tidak terdapat korelasi antara nilai pada suatu observasi dengan nilai sebelumnya pada suatu data *time – series* memungkinkan bahwa *time – series* tersebut adalah merupakan *white noise*. [4] menjelaskan bahwa *white noise* merupakan konsep penting dalam peramalan data *time-series* karena data yang bersifat *white-noise* menunjukkan data tersebut sulit untuk diprediksi.

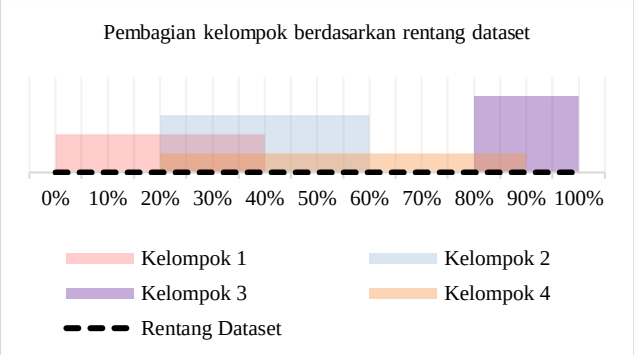
3) Identifikasi Stasioneritas: Dalam proses untuk membangun model peramalan ARIMA, data harus dalam keadaan stasioner karena model ARIMA tidak akan bekerja dengan optimal terhadap data yang tidak stasioner. Stasioneritas pada data *time-series* terbagi menjadi dua yaitu stasioner dalam *mean* dan stasioner dalam varian [5]. Data yang tidak stasioner terhadap varian dapat ditransformasikan menggunakan metode *box-cox*, dan data yang tidak stasioner terhadap *mean* atau rata – rata dapat ditransformasikan dengan metode *differencing*.



Gbr. 22 Visualisasi grafik ACF dan PACF

Untuk mengetahui atribut data iklim dengan data yang tidak stasioner dalam varian dapat diuji dengan menggunakan *levene test*. Data training pada setiap atribut dibagi menjadi empat kelompok, kemudian menggunakan levene test kelompok – kelompok tersebut diuji apakah memiliki nilai varian yang sama atau homogen. Kelompok pertama diambil dari 40% data pertama pada rentang dataset, kelompok kedua diambil dari 20% sampai 60% pada rentang dataset, kelompok ketiga diambil dari 80% sampai 100% pada rentang dataset, dan kelompok keempat diambil dari 20% sampai 90% dari rentang dataset. Secara visual, rentang masing – masing kelompok terhadap dataset-nya dapat dilihat pada Gbr. 23.

Akreditasi Sinta 5 KEMRISTEKDIKBUD No. 225/E/KPT/2022; E-ISSN 2723-7958



Gbr. 23 Rentang kelompok pengujian *Levene Test*

Hasil identifikasi stasioneritas dalam varian untuk setiap atribut pada dataset menggunakan *levene test* disajikan dalam tabel III, nilai signifikasi (*p-value*) yang lebih besar dari 0.05 menunjukkan bahwa kelompok data yang diuji memiliki nilai varian yang sama atau stasioner dan berdasarkan tabel tersebut, dataset dari atribut temperatur minimum, temperatur rata – rata, kelembapan rata – rata dinyatakan tidak stasioner dalam varian, dan untuk data dari atribut temperatur maksimum, dan penyinaran matahari dinyatakan sudah stasioner dalam varian. Untuk pengujian levene test dari data atribut curah hujan tidak dapat dihasilkan nilai statistik dan *p-value* dikarenakan atribut tersebut adalah bersifat *white noise*. Seperti dijelaskan sebelumnya bahwa untuk menangani data yang tidak stasioner dalam varian dapat ditransformasikan menggunakan metode *box-cox*.

TABEL III HASIL PENGUJIAN LEVENE TEST

Atribut Data Iklim	Levene Statistic	*p-value*	Stasioner Dalam Varian
Temperatur minimum	5.6555	0.0007	Tidak
Temperatur maksimum	1.1872	0.3130	Ya
Temperatur rata – rata	4.4933	0.0038	Tidak
Kelembapan rata - rata	4.0760	0.0067	Tidak
Penyinaran matahari	0.6582	0.5777	Ya
Curah hujan	-	-	*White noise*

Untuk mengetahui atribut data iklim dengan data yang tidak stasioner dalam *mean* dapat diuji dengan menggunakan tes ADF dan KPSS. Nilai signifikasi (*p-value*) dari tes ADF yang lebih kecil dari 0.05 menunjukkan bahwa data sudah stasioner, sebaliknya *p-value* dari tes KPSS yang lebih kecil dari 0.05 menunjukkan bahwa data belum stasioner. Dengan menggunakan python dapat diketahui jumlah *differencing* yang diperlukan untuk membuat data menjadi stasioner.

Hasil pengujian stasioneritas dalam mean dengan tes ADF dan KPSS disajikan pada tabel IV, berdasarkan tabel tersebut untuk ADF test setiap atribut kecuali atribut curah hujan yang merupakan *white noise*, didapatkan hasil *p-value* yang lebih kecil dari 0.05 artinya atribut – atribut tersebut dinyatakan sudah staioner dalam mean, tetapi menurut pengujian dari KPSS test, atribut temperatur minimum dan kelembapan rata – rata masih harus ditransformasikan untuk membuatnya menjadi stasioner.

TABEL IV HASIL PENGUJIAN ADF TEST DAN KPSS TEST

Atribut Data Iklim	ADF Test (p-value)	KPSS Test (p-value)	Jumlah differencing yang dibutuhkan
Temperatur minimum	4.304152175816683e-13	0.01	1
Temperatur maksimum	1.213413142581551 3e-07	0.1	0
Temperatur rata – rata	2.013472013858 67e-09	0.1	0
Kelembapan rata - rata	2.963019868156 6646e-10	0.1	0
Penyinaran matahari	2.391166969074 6513e-19	0.01	1
Curah hujan	0	0.1	0

4) Model Peramalan: Setelah data dikatakan stasioner, selanjutnya adalah mengidentifikasi parameter model peramalan yang akan digunakan, kemudian melakukan evaluasi terhadap model dengan melihat normalitas distribusi residual/error. Model ARIMA terdiri dari proses *Autoregressive* (AR) yang dinotasikan dengan (p), dan proses *Moving Average* yang dinotasikan dengan (q) pada data yang sudah stasioner dengan notasi (d) yaitu jumlah proses *differencing* yang dibutuhkan untuk membuat data tersebut menjadi stasioner. Pada penelitian ini, untuk mengidentifikasi model dengan kombinasi *p* dan *q* yang paling optimal, maka digunakan fungsi *auto_arima* yang disediakan oleh modul pmdarima pada bahasa pemograman python untuk secara otomatis mencari model dengan kombinasi notasi (p, d, q) dengan nilai AIC paling kecil.

Untuk parameter musiman yang pada metode ARIMA dinotasikan dengan *P* untuk parameter *Autoregressive* musiman, *D* untuk *differencing* terhadap data musiman, dan *Q* untuk parameter *Moving Average* musiman, dalam melakukan penelitian, penulis tidak dapat melakukan proses training atau fitting pada model dikarenakan periode musiman yang besar sehingga proses tersebut membutuhkan performa yang tinggi dan memori yang besar dari perangkat keras yang digunakan. Sebagai alternatif untuk menangkap periode musiman pada data, penulis melakukan transformasi *fourier* terhadap data *time series* yang akan diramalkan, kemudian dijadikan sebagai *exogenous variable* pada model. Hal ini didukung karena transformasi *fourier* telah digunakan pada paruh kedua abad ke-sembilan belas untuk mencari komponen periodik terhadap data meteorologi [6]. Selain itu, penggunaan *exogeneous variable* pada ARIMA memungkinkan model untuk melakukan peramalan tidak hanya mengacu pada satu atribut time-series, tetapi lebih dari satu atribut time-series sehingga dapat meningkatkan akurasi dari model [7].

Model peramalan terbaik untuk setiap atribut data iklim dapat dinilai dari skor AIC (*Akaike's Information Criterion*) yang paling kecil. Model peramalan yang terpilih untuk meramalkan masing – masing atribut dari data cuaca dapat dilihat pada tabel V.

Akreditasi Sinta 5 KEMRISTEKDIKBUD No. 225/E/KPT/2022; E-ISSN 2723-7958

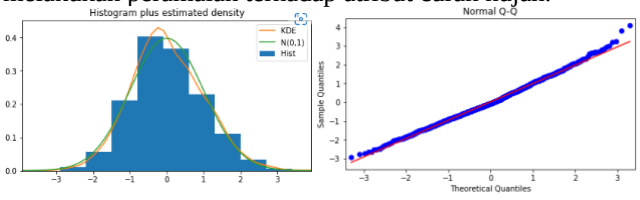
TABEL V PEMILIHAN MODEL PERAMALAN ARIMA

Atribut data iklim	Model Terpilih	AIC	Total waktu pencarian model
Temperatur minimum	ARIMAX(1,1,3)	44750.617	25min 43s
Temperatur maksimum	ARIMAX(1,0,4)	6748.568	28min 47s
Temperatur rata – rata	ARIMAX(0,0,5)	2692.512	22min 24s
Kelembapan rata - rata	ARIMAX(4,0,6)	58248.393	34min 40s
Penyinaran matahari	ARIMAX(1,1,2)	9960.197	37min 14s
Curah hujan	ARIMAX(0,0,0)	6055.965	1min 59s

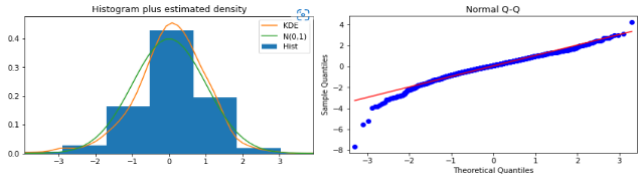
Berdasarkan tabel V diketahui model peramalan ARIMA untuk masing – masing atribut data cuaca dengan nilai AIC yang paling kecil. Tetapi untuk atribut curah hujan, terlihat bahwa program python tidak dapat melakukan estimasi parameter model peramalan sehingga memberikan model dengan order (0,0,0). Ini terjadi karena atribut curah hujan adalah merupakan *white noise* yang sebelumnya juga telah terlihat pada grafik ACF dan PACF di Gbr. 22 (f).

Dari model – model yang didapatkan pada tabel V, selanjutnya dilakukan analisis distribusi residual/error untuk melihat apakah residual/error berdistribusi normal sehingga model dapat dikatakan layak digunakan untuk peramalan. Untuk normalitas distribusi residual/error dapat dilihat dari grafik Q-Q Plot (*Quantile-Quantile Plot*) dan grafik Histogram.

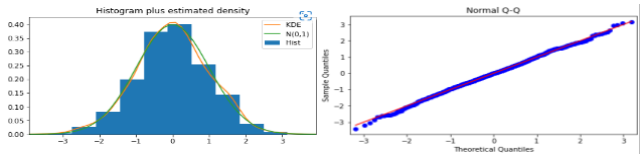
Q-Q Plot dan Histogram untuk masing – masing atribut dapat dilihat pada Gbr. 24 sampai Gbr. 29. Dari visualisasi tersebut terlihat bahwa residual/error dari model temperatur minimum (Gbr. 24), temperatur maksimum (Gbr. 25), temperatur rata-rata (Gbr. 26), kelembapan rata-rata (Gbr. 27), dan penyinaran matahari (Gbr. 28) berdistribusi normal. Hal ini terlihat dari Histogram dan Q-Q Plot, normalitas pada histogram ditandai dengan garis yang menyerupai kurva yang berbentuk seperti lonceng terbalik. Sedangkan normalitas melalui Q-Q Plot ditandai dengan sebaran residual/error yang mendekati garis diagonal QQ-Plot tersebut. Terlihat pada Gbr. 29 yaitu histogram dan Q-Q Plot model curah hujan, model tersebut memiliki residual/error yang tidak memenuhi normalitas, sehingga dapat dipastikan bahwa model ARIMA tidak dapat melakukan peramalan terhadap atribut curah hujan.



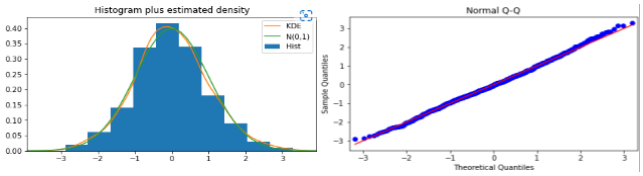
Gbr. 24 Histogram dan Q-Q Plot temperatur minimum



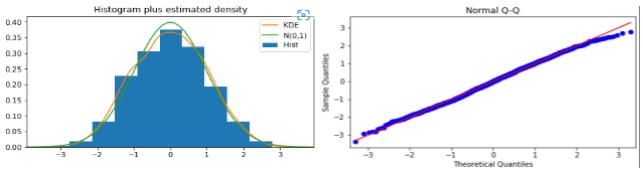
Gbr. 25 Histogram dan Q-Q Plot temperatur maksimum



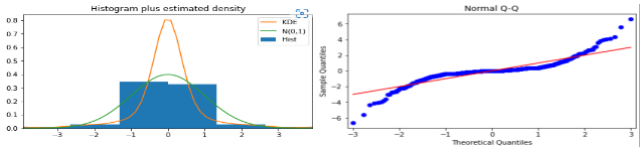
Gbr. 26 Histogram dan Q-Q Plot temperatur rata-rata



Gbr. 27 Histogram dan Q-Q Plot kelembapan rata-rata



Gbr. 28 Histogram dan Q-Q Plot penyinaran matahari

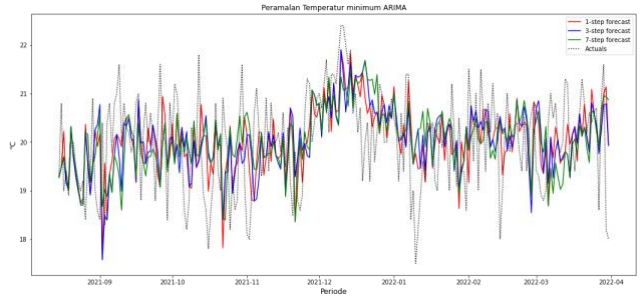


Gbr. 29 Histogram dan Q-Q Plot curah hujan

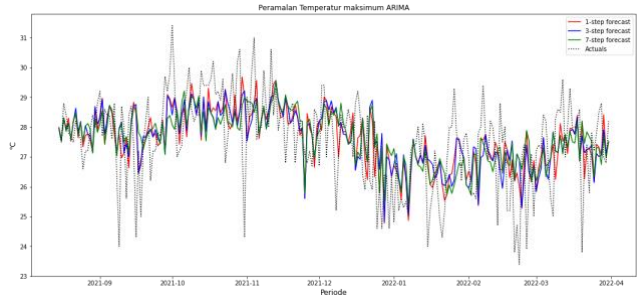
5) *Hasil Peramalan dengan Model ARIMA:* Peramalan dilakukan dengan skenario *rolling-forecast* menggunakan model yang telah terbentuk sebelumnya. Model akan meramalkan setiap *time step* dari data test kemudian nilai observasi aktual pada *time step* tersebut dimasukkan pada model untuk meramalkan nilai dari *time step* berikutnya, seperti pada penelitian [1], ini mencerminkan kondisi di lapangan mengingat informasi cuaca pada sumber yang digunakan pada penelitian ini diperbaharui perhari sehingga informasi tersebut dapat digunakan untuk peramalan hari berikutnya.

Pada penelitian ini *time step* yang digunakan yaitu 1, 3 dan 7. Pemilihan peramalan 1 hari, 3 hari dan 7 hari kedepan disesuaikan dengan produk prakiraan cuaca pada website BMKG dimana tersedia informasi prakiraan cuaca sampai 7 hari kedepan. Hasil peramalannya dapat dilihat pada Gbr. 30 sampai Gbr. 34.

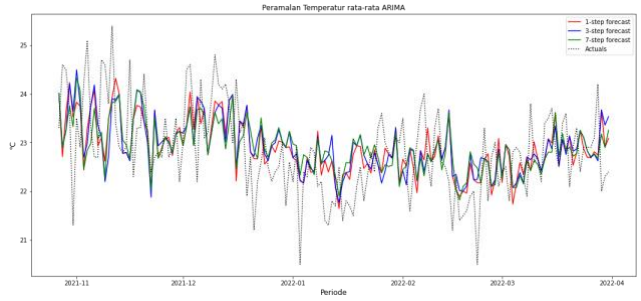
Akreditasi Sinta 5 KEMRISTEKDIKBUD No. 225/E/KPT/2022; E-ISSN 2723-7958



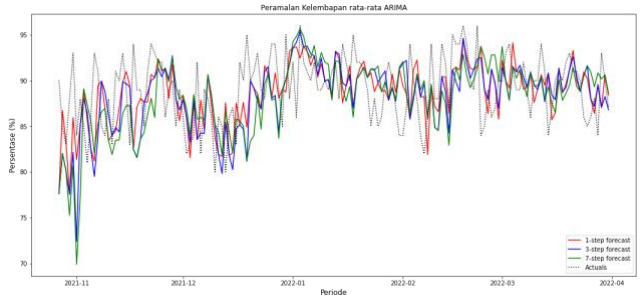
Gbr. 30 Hasil Peramalan temperatur minimum ARIMA



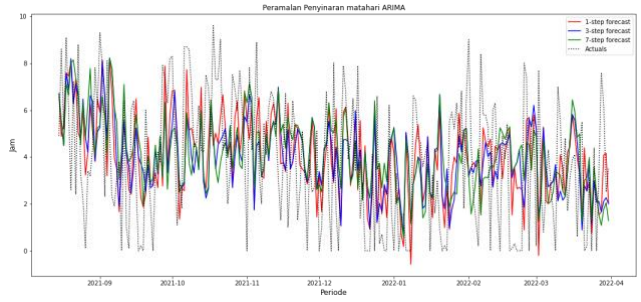
Gbr. 31 Hasil Peramalan temperatur maksimum ARIMA



Gbr. 32 Hasil Peramalan temperatur rata-rata ARIMA



Gbr. 33 Hasil Peramalan kelembapan rata-rata ARIMA



Gbr. 34 Hasil Peramalan penyinaran matahari ARIMA

Untuk mengetahui secara lebih jelas sejauh apa nilai hasil prediksi dan nilai aktual yang terjadi di lapangan dilakukan pengujian error menggunakan *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE). Nilai MAE dan RMSE dari peramalan setiap 1, 3 dan 7 hari untuk setiap atribut data cuaca menggunakan model ARIMA dapat dilihat pada tabel VI, berdasarkan tabel tersebut terlihat bahwa semakin kecil *time-step* yang digunakan untuk peramalan maka akan menghasilkan nilai error yang lebih kecil juga.

TABEL VI PENGUJIAN RMSE DAN MAE HASIL PERAMALAN MODEL ARIMA

Atribut Data Cuaca	Model ARIMA	Error	Rolling Forecast Period		
			1 time-step	3 time-step	7 time-step
Temperatur minimum	ARIMAX (1,1,3)	MAE	0.8127	0.8563	0.8891
		RMSE	1.0266	1.0716	1.0919
Temperatur maksimum	ARIMAX (1,0,4)	MAE	1.0898	1.1329	1.1505
		RMSE	1.4047	1.4432	1.4605
Temperatur rata – rata	ARIMAX (0,0,5)	MAE	0.6884	0.7352	0.7330
		RMSE	0.8586	0.9040	0.8955
Kelembapan rata – rata	ARIMAX (4,0,6)	MAE	3.3843	3.6883	3.9943
		RMSE	4.1828	4.6167	4.9710
Penyinaran matahari	ARIMAX (1,1,2)	MAE	2.4334	2.5337	2.5968
		RMSE	2.9177	2.9874	3.0441

E. Peramalan dengan Model LSTM

1) Data Train, Data Validasi dan Data Test: Sebelum melakukan peramalan dengan model LSTM, dataset yang digunakan dibagi menjadi data train, data validasi dan data test. Sama seperti model ARIMA sebelumnya, data train merupakan data yang digunakan untuk melakukan training pada model dan data test digunakan untuk memverifikasi hasil peramalan oleh model yang terbentuk terhadap nilai aktual dengan mengukur tingkat error menggunakan *Root Mean Squared Error* (RMSE) atau menggunakan *Mean Absolute Error* (MAE). Pada model LSTM, data validasi digunakan dalam proses *training* atau *fitting data* pada model untuk melihat nilai RMSE model dalam setiap iterasi yang dilakukan untuk mempelajari data train. Untuk peramalan dengan model LSTM, penulis membagi data train dengan rasio 80% data validasi 10% dan data test sebesar 10% yang masing – masing diambil dari dataset setelah *preprocessing*.

2) Standarisasi Data: Untuk menghasilkan model dengan kemampuan peramalan yang optimal serta mencegah penyimpangan dari pengolahan data train, maka dilakukan standarisasi agar data train memiliki *mean* dan varian nol [3]. Persamaan standarisasi yang digunakan dapat dilihat pada (1). Data validasi dan data test juga distandarisasi menggunakan persamaan tersebut.

$\sigma_{std} = data - \mu_{train} / \sigma_{train}$ (1)
$y_{pred} = y_{pred} * \sigma_{train} + \mu_{train}$ (2)
Dengan keterangan yaitu :
σ_{std} = data yang sudah distandarisasi
y_{pred} = nilai hasil prediksi
μ_{train} = rata – rata data train
σ_{train} = standar deviasi data train

Perlu diperhatikan bahwa hasil prediksi oleh model yang di-training dengan data yang sudah distandarisasi juga akan menghasilkan nilai standarisasi, untuk itu perlu dilakukan standarisasi balik dengan (2).

3) *Model Peramalan:* Setelah dilakukan standarisasi pada data, selanjutnya dilakukan proses training pada model LSTM. Data yang digunakan untuk melatih model LSTM dibagi menjadi dua variabel yaitu X dan Y dengan bentuk array. Variabel X adalah merupakan prediktor dan variabel Y sebagai variabel respon, variabel X adalah nilai – nilai pada setiap lag variabel Y sampai panjang yang ditentukan atau disebut *time step*. Pada penelitian ini panjang *time step* untuk variabel X adalah 20, ini berarti bahwa digunakan data dari 20 observasi sebelumnya untuk memprediksi nilai pada hari berikutnya. Jika dilihat pada grafik ACF sebelumnya yang ditampilkan pada Gbr. 22, setiap atribut kecuali curah hujan yang terindikasi sebagai *white noise*, terlihat pada area sekitar lag ke-20 masih terdapat garis yang keluar dari batas signifikan yang menandakan bahwa masih terdapat korelasi antara data pada satu observasi dengan 20 observasi sebelumnya.

Model LSTM pada penelitian ini dibuat menggunakan bahasa pemograman python dengan keras library, dengan arsitektur optimizer Adam dan learning rate 0.0001. Tidak ada aturan pasti mengenai jumlah Epoch dan neuron dalam memprediksi *time series*, oleh karena itu dibutuhkan pengujian untuk mengetahui model dengan peramalan paling optimal [8].

Jumlah iterasi atau Epoch setiap model pada penelitian ini dibatasi maksimal 250 iterasi. Pada setiap iterasi dilakukan pengujian nilai RMSE terhadap data validasi dan program hanya akan menyimpan model dengan nilai RMSE paling rendah untuk menghindari kondisi *over-fitting* pada model.

Penulis melakukan pengujian pembuatan model LSTM dengan arsitektur neuron layer berjumlah 128 dan 256 neuron, kemudian diuji nilai RMSE terhadap data validasi untuk melihat model dengan jumlah neuron yang paling optimal untuk masing masing atribut data cuaca. Hasil pengujian tersebut dapat dilihat pada tabel VII.

TABEL VII HASIL PENGUJIAN NEURON LAYER MODEL LSTM

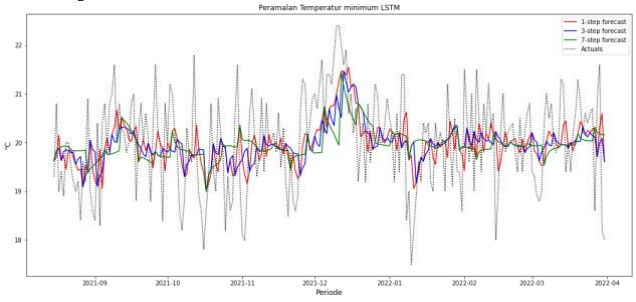
Atribut Data Cuaca	Neuron	RMSE Data Validasi	Durasi Training
Temperatur minimum	256	0.9131	14min
	128	0.9156	5min 32s
Temperatur maksimum	256	1.1853	17min 33s
	128	1.1844	4min 53s
Temperatur rata – rata	256	0.8231	10min 13s
	128	0.8295	4min 56s
Kelembapan rata – rata	256	3.7302	9min 10s
	128	3.7258	2min 40s
Penyinaran matahari	256	2.5160	15min 9s
	128	2.5189	5min 44s
Curah hujan	256	10.2165	4min 40s
	128	10.1846	1min 37s

Berdasarkan tabel VII terlihat bahwa tidak ada penurunan nilai RMSE yang signifikan untuk model dengan arsitektur 128 neuron layer ataupun 256 neuron layer, tetapi model dengan 256 neuron membutuhkan durasi training lebih dari dua kali lipat model dengan 128 neuron. Meskipun begitu, penulis memilih model dengan nilai error paling kecil sebagai model terbaik untuk melakukan peramalan. Pemilihan model peramalan LSTM untuk setiap atribut data cuaca dapat dilihat pada tabel VIII.

TABEL VIII PEMILIHAN MODEL PERAMALAN LSTM

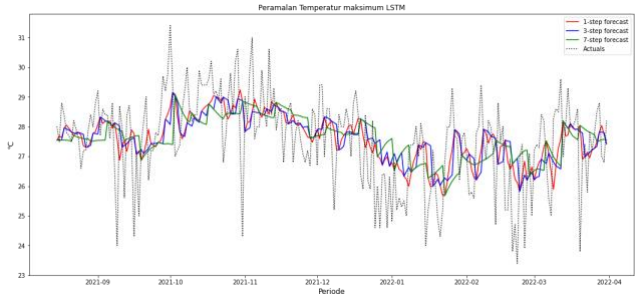
Atribut Data Cuaca	Model LSTM		RMSE Data Validasi	Durasi Training
	Jumlah Neuron	Max Epoch		
temperatur minimum	256	250	0.9131	14min
temperatur maksimum	128	250	1.1844	4min 53s
temperatur rata – rata	256	250	0.8231	10min 13s
kelembapan rata - rata	128	250	3.7258	2min 40s
penyinaran matahari	256	250	2.5160	15min 9s
Curah hujan	256	250	10.1846	1min 37s

4) Hasil Peramalan dengan Model LSTM: Setelah diketahui model peramalan terbaik untuk masing – masing atribut, selanjutnya adalah melakukan peramalan dan melakukan pengujian tingkat error dengan *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE) terhadap nilai aktualnya yaitu data test. Sama seperti peramalan dengan model ARIMA sebelumnya, dilakukan *rolling forecast* setiap 1, 3, dan 7 periode kedepan, kemudian nilai observasi aktual pada periode atau *time step* tersebut di masukkan pada model untuk meramalkan nilai dari *time step* berikutnya. Hasilnya dapat dilihat pada Gbr. 35 sampai Gbr. 40.

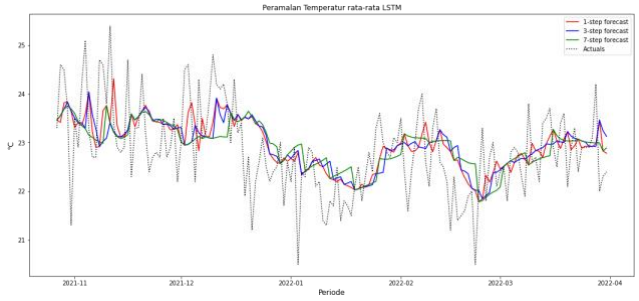


Gbr. 35 Hasil Peramalan temperatur minimum LSTM

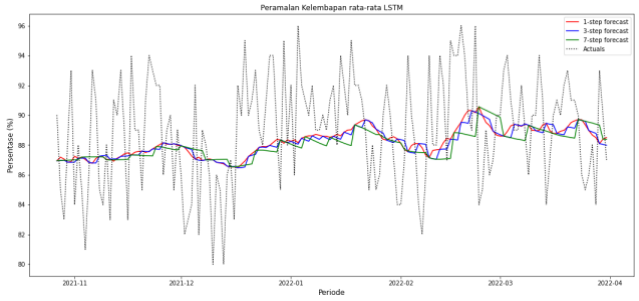
Akreditasi Sinta 5 KEMRISTEKDIKBUD No. 225/E/KPT/2022; E-ISSN 2723-7958



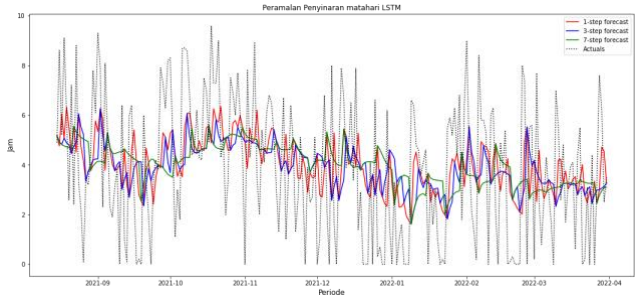
Gbr. 36 Hasil Peramalan temperatur maksimum LSTM



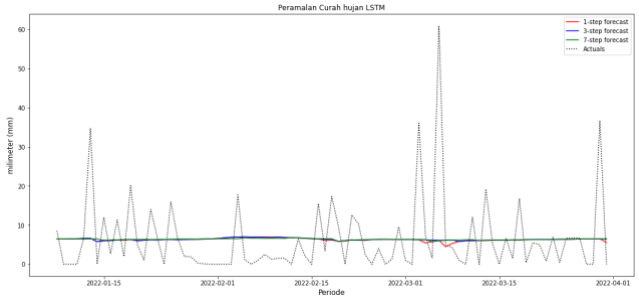
Gbr. 37 Hasil Peramalan temperatur rata-rata LSTM



Gbr. 38 Hasil Peramalan kelembapan rata-rata LSTM



Gbr. 39 Hasil Peramalan penyinaran matahari LSTM



Gbr. 40 Hasil Peramalan curah hujan LSTM

Untuk mengetahui secara lebih jelas sejauh apa nilai hasil prediksi dan nilai aktual yang terjadi di lapangan dilakukan pengujian error menggunakan *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE). Nilai MAE dan RMSE dari peramalan setiap 1, 3 dan 7 hari untuk setiap atribut data cuaca menggunakan model LSTM dapat dilihat pada Tabel 4.9, berdasarkan tabel tersebut terlihat bahwa semakin kecil *time-step* yang digunakan untuk peramalan maka akan menghasilkan nilai error yang lebih kecil juga.

TABEL IX PENGUJIAN RMSE DAN MAE HASIL PERAMALAN MODEL LSTM

Atribut Data Cuaca	Neuron / Max Epoch	Error	Rolling forecast period		
			1 time-step	3 time-step	7 time-step
temperatur minimum	256 / 250	MAE	0.7415	0.7758	0.7935
		RMSE	0.9006	0.9388	0.9614
temperatur maksimum	128 / 250	MAE	1.0297	1.0964	1.1348
		RMSE	1.3098	1.3960	1.4403
temperatur rata – rata	256 / 250	MAE	0.6205	0.6515	0.6728
		RMSE	0.7943	0.8321	0.8573
kelembapan rata – rata	128 / 250	MAE	3.0398	3.1252	3.1956
		RMSE	3.6490	3.7224	3.8158
penyinaran matahari	256 / 250	MAE	2.2182	2.2952	2.3810
		RMSE	2.6211	2.7057	2.7662
Curah hujan	256 / 250	MAE	6.4733	6.5069	6.4929
		RMSE	9.9741	9.9865	9.9909

F. Analisis dan Perbandingan Hasil Peramalan Model ARIMA dan LSTM

Berdasarkan kondisi saat ini, keadaan cuaca berada pada kondisi tidak stabil dan berubah dengan cepat. Jika model peramalan tidak dapat merespon dengan cepat terhadap perubahan, maka model tersebut tidak memenuhi syarat [1]. Oleh karena itu, model peramalan yang fleksibel sangat dibutuhkan untuk menghadapi tantangan ini sehingga pada penelitian ini digunakan skenario peramalan yang disebut sebagai *rolling forecast.*

Algoritma dari LSTM dan model ARIMA dikembangkan untuk meramalkan data time series berdasarkan "*Rolling Forecasting Origin*" [9]. Rolling forecasting origin berfokus untuk meramalkan data point selanjutnya untuk setiap kumpulan data. Terdapat beberapa variasi *rolling forecast* [10] dan pada penelitian ini diterapkan variasi yaitu *multi-step forecast with re-estimation*.

Hasil peramalan dengan skenario *rolling forecast* diuji tingkat error dengan melihat nilai MAE dan RMSE untuk masing – masing model peramalan kemudian dibandingkan untuk melihat model peramalan yang lebih optimal. Perbandingan error hasil peramalan oleh model ARIMA dan model LSTM disajikan dalam tabel X.

Akreditasi Sinta 5 KEMRISTEKDIKBUD No. 225/E/KPT/2022; E-ISSN 2723-7958

Berdasarkan hasil pengujian tingkat error pada tabel tersebut, terlihat bahwa model LSTM memiliki performa peramalan yang lebih baik dibandingkan model ARIMA karena model LSTM memiliki nilai error yang lebih rendah pada setiap peramalan.

TABEL X PERBANDINGAN ERROR HASIL PERAMALAN MODEL ARIMA DAN LSTM

Atribut Data Cuaca	Rolling Forecast Period	Model ARIMA		Model LSTM	
		MAE	RMSE	MAE	RMSE
temperatur minimum	1 time-step	0.8127	1.0266	0.7415	0.9006
	3 time-step	0.8563	1.0716	0.7758	0.9388
	7 time-step	0.8891	1.0919	0.7935	0.9614
temperatur maksimum	1 time-step	1.0898	1.4047	1.0297	1.3098
	3 time-step	1.1329	1.4432	1.0964	1.396
	7 time-step	1.1505	1.4605	1.1348	1.4403
temperatur rata – rata	1 time-step	0.6884	0.8586	0.6205	0.7943
	3 time-step	0.7352	0.9040	0.6515	0.8321
	7 time-step	0.7330	0.8955	0.6728	0.8573
kelembapan rata – rata	1 time-step	3.3843	4.1828	3.0398	3.649
	3 time-step	3.6883	4.6167	3.1252	3.7224
	7 time-step	3.9943	4.9710	3.1956	3.8158
penyinaran matahari	1 time-step	2.4334	2.9177	2.2182	2.6211
	3 time-step	2.5337	2.9874	2.2952	2.7057
	7 time-step	2.5968	3.0441	2.381	2.7662
curah hujan	1 time-step	-	-	6.4733	9.9741
	3 time-step	-	-	6.5069	9.9865
	7 time-step	-	-	6.4929	9.9909

Dalam peramalan atribut curah hujan menggunakan metode ARIMA, pada penelitian ini penulis tidak dapat menemukan model untuk peramalan. Hal ini disebabkan karena pada visualisasi grafik ACF Gbr. 22 (f), data curah hujan diidentifikasi sebagai *white noise* sehingga metode ARIMA tidak dapat memberikan model yang layak untuk peramalan atribut tersebut. Jika dilihat pada tabel X, model LSTM mampu membuat model peramalan untuk atribut curah hujan, tetapi model tersebut memberikan hasil dengan error yang tinggi. Sehingga seperti sifat white noise yang tidak dapat diprediksi, begitu juga pada penelitian ini model LSTM tidak dapat memprediksi nilai curah hujan.

Untuk hasil peramalan model ARIMA atribut kelembapan rata-rata seperti terlihat pada Gbr. 33, terlihat bahwa untuk hasil peramalan pada rentang waktu 1 bulan pertama, model memberikan hasil yang sedikit jauh dari nilai aktualnya, kemudian pada rentang waktu setelah 1 bulan peramalan, model mulai memberikan hasil yang mendekati nilai aktualnya. Kondisi ini menjelaskan bahwa seiring model mempelajari data baru setiap dilakukannya peramalan, maka model peramalan dapat mengevaluasi hasil peramalan pada periode – periode sebelumnya sehingga pada periode peramalan selanjutnya model dapat memberikan hasil yang lebih mendekati nilai aktualnya.

Meskipun model LSTM terbukti memiliki performa yang lebih baik dibandingkan model ARIMA, akurasi dari model ARIMA juga tidak dapat dibilang buruk. Jika dilihat selisih error hasil peramalan oleh model ARIMA dan LSTM pada tabel XI, model peramalan ARIMA dan LSTM untuk atribut temperatur minimum, temperatur maksimum, temperatur rata – rata, kelembapan rata – rata, dan penyinaran matahari hanya memiliki selisih RMSE dibawah 1.2 dan selisih MAE dibawah 0.8, diluar peramalan untuk atribut curah hujan.

TABEL XI SELISIH ERROR HASIL PERAMALAN MODEL ARIMA DAN LSTM

Atribut data cuaca	Rolling Forecast Period	Selisih Error model ARIMA dan LSTM	
		MAE	RMSE
temperatur minimum	1 time-step	0.0712	0.126
	3 time-step	0.0805	0.1328
	7 time-step	0.0956	0.1305
temperatur maksimum	1 time-step	0.0601	0.0949
	3 time-step	0.0365	0.0472
	7 time-step	0.0157	0.0202
temperatur rata – rata	1 time-step	0.0679	0.0643
	3 time-step	0.0837	0.0719
	7 time-step	0.0602	0.0382
kelembapan rata – rata	1 time-step	0.3445	0.5338
	3 time-step	0.5631	0.8943
	7 time-step	0.7987	1.1552
penyinaran matahari	1 time-step	0.2152	0.2966
	3 time-step	0.2385	0.2817
	7 time-step	0.2158	0.2779
Curah hujan	1 time-step	6.4733	9.9741
	3 time-step	6.5069	9.9865
	7 time-step	6.4929	9.9909

Jika dilihat dari segi waktu komputasi seperti pada tabel XII, terlihat bahwa model ARIMA dalam proses untuk pencarian model terbaik dengan melakukan estimasi parameter pada model dan training model terpilih untuk peramalan suatu atribut data cuaca memakan waktu paling lama 37 menit dan paling cepat 22 menit, sedangkan model LSTM dalam melakukan training pada model, dengan data yang sama dan dengan selisih error yang kecil hanya memakan waktu paling lama 10 menit dan paling cepat yaitu sekitar 2 menit.

TABEL XII WAKTU KOMPUTASI MODEL ARIMA DAN LSTM

Atribut data cuaca	Waktu Komputasi	
	ARIMA	LSTM
temperatur minimum	25min 43s	14min
temperatur maksimum	28min 47s	4min 53s
temperatur rata-rata	22min 24s	10min 13s
kelembapan rata-rata	34min 40s	2min 40s
penyinaran matahari	37min 14s	15min 9s
curah hujan	-	1min 37s

Pada model ARIMA, yang paling menentukan durasi waktu komputasi adalah jumlah data training yang digunakan, program python akan mencari model dengan melakukan estimasi parameter untuk setiap kombinasi order *p* dan *q*. Pada setiap kombinasi akan dilakukan *fitting* atau training dataset pada model sehingga dapat memberikan informasi nilai AIC model tersebut kemudian dipilih model dengan nilai AIC paling kecil.

Pada model LSTM, yang paling mempengaruhi durasi waktu komputasi selain jumlah data training, durasi waktu komputasi juga dipengaruhi oleh epoch dan jumlah neuron yang ditentukan. Epoch merupakan jumlah iterasi yang akan dilakukan model dalam mempelajari suatu dataset sehingga semakin banyak iterasi yang dilakukan maka semakin banyak waktu yang diperlukan. Pada penelitian ini, model LSTM menggunakan data 20 hari kebelakang sebagai input untuk dapat melakukan peramalan, setiap input akan diproses oleh neuron yang memiliki 3 gate didalamnya, sehingga semakin banyak neuron yang ditentukan maka semakin banyak waktu yang diperlukan.

IV. KESIMPULAN

Berdasarkan hasil peramalan dan analisis performa model menggunakan metode ARIMA dan model *deep learing* dengan algoritma LSTM dalam peramalan data cuaca di daerah kota Manado maka dapat disimpulkan hal – hal sebagai berikut:

1. Model ARIMA memberikan performa peramalan terbaik pada atribut temperatur minimum, temperatur maksimum, dan temperatur rata-rata dengan nilai RMSE dibawah 1.47 dan nilai MAE dibawah 1.16. Selanjutnya untuk penyinaran matahari model bekerja dengan cukup baik dengan nilai RMSE pada rentang 2.91 sampai 3.05. Peramalan kelembapan rata-rata memiliki nilai error tertinggi dengan RMSE mencapai 4.97 dan MAE mencapai 3.99. Pada penelitian ini, tidak dapat dihasilkan model ARIMA yang layak untuk peramalan data curah hujan dan berarti model ARIMA tidak dapat meramalkan atribut tersebut.
   Waktu komputasi model ARIMA dipengaruhi oleh jumlah data training yang digunakan. Semakin banyak data training maka semakin lama model memproses data untuk menghasilkan peramalan.
2. Model LSTM memberikan performa peramalan terbaik pada atribut temperatur minimum, temperatur maksimum, dan temperatur rata-rata dengan nilai RMSE dibawah 1.45 dan nilai MAE dibawah 1.14. Selanjutnya untuk kelembapan rata-rata dan penyinaran matahari, model peramalan bekerja dengan cukup baik dengan nilai RMSE pada rentang 2.62 sampai 3.82 serta MAE pada rentang 2.21 sampai 3.2. Peramalan curah hujan memiliki nilai error tertinggi dengan RMSE mencapai 9.99 dan MAE mencapai 6.5. Jika dibandingkan dengan tingkat error hasil peramalan atribut lainnya, model LSTM kurang optimal dalam memprediksi data curah hujan.
   Waktu komputasi model LSTM dipengaruhi oleh jumlah data training, epoch dan neuron pada model, semakin banyak jumlah yang digunakan maka semakin lama model memproses data untuk menghasilkan peramalan. Akan tetapi, jumlah epoch dan neuron yang besar belum tentu menghasilkan peramalan dengan error yang lebih rendah. Jumlah epoch dan neuron yang terlalu banyak ataupun terlalu sedikit dapat mempengaruhi performa model dalam melakukan peramalan.
3. Model LSTM lebih unggul dibandingkan model ARIMA dari segi hasil peramalan dan dari segi waktu komputasi. Dari setiap peramalan yang dilakukan, model LSTM menghasilkan nilai error yang lebih kecil dibandingkan model ARIMA. Meskipun demikian, nilai error hasil peramalan dari kedua model hanya memiliki selisih RMSE dibawah 1.2 dan selisih MAE dibawah 0.8. Selisih error yang besar hanya pada peramalan curah hujan, dimana model ARIMA sama sekali tidak dapat melakukan peramalan.
   Jika dilihat dari waktu komputasi untuk menghasilkan peramalan dan dengan tingkat error yang lebih rendah, model LSTM dapat menghasilkan model peramalan dengan lebih cepat dibandingkan dengan model ARIMA, dihitung dari proses fitting data pada model dan termasuk estimasi paramater peramalan pada model ARIMA.

UCAPAN TERIMA KASIH

Penelitian ini tidak dapat terselesaikan dengan baik tanpa ada bantuan dari berbagai pihak, maka dari itu pada kesempatan ini peneliti mengucapkan banyak terima kasih kepada Rektor Universitas Negeri Manado, Dekan Fakultas Teknik, seluruh Dosen Program Studi Teknik Informatika, Dosen – dosen Pembimbing dan rekan – rekan mahasiswa serta Orang Tua.

DAFTAR PUSTAKA

[1] R. Zhang *et al.*, “Comparison of arima and lstm in forecasting the incidence of hfmd combined and uncombined with exogenous meteorological variables in Ningbo, China,” *Int J Environ Res Public Health*, vol. 18, no. 11, 2021, doi: 10.3390/ijerph18116174.

[2] K. dan G. (BMKG) Badan Meteorologi, “Pusat Database - BMKG,” 2022.

[3] E. Supriyadi, “PREDIKSI PARAMETER CUACA MENGGUNAKAN DEEP LEARNING LONG-SHORT TERM MEMORY (LSTM),” *Jurnal Meteorologi dan Geofisika*, vol. 21, no. 2, 2021, doi: 10.31172/jmg.v21i2.619.

[4] Clive W. J. Granger, *Essays in Econometrics vol II: Collected Papers of Clive W. J. Granger*. 2001. doi: 10.1017/ccol052179207x.

[5] W. W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods Second Edition*, vol. SFB 373, no. Chapter 5. 2006.

[6] D. S. Stoffer and P. Bloomfield, “Fourier Analysis of Time Series: An Introduction,” *J Am Stat Assoc*, vol. 95, no. 452, 2000, doi: 10.2307/2669794.

[7] C. Kongcharoen and T. Kruangpradit, “Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) Model for Thailand Export,” *33rd International Symposium on Forecasting, South Korea*, no. June, 2013.

[8] M. Rizki, S. Basuki, and Y. Azhar, “Implementasi Deep Learning Menggunakan Arsitektur Long Short Term Memory(LSTM) Untuk Prediksi Curah Hujan Kota Malang,” *Jurnal Repositor*, vol. 2, no. 3, 2020, doi: 10.22219/repositor.v2i3.470.

[9] Rob J Hyndman and A. George, "Forecasting: Principles and Practice," *Principles of Optimal Design*, no. September, 2014.
[10] S. Siami-Namini and A. S. Namin, "Forecasting Economics and Financial Time Series: ARIMA vs. LSTM," arXiv.org, 2018. doi: https://doi.org/10.48550/arXiv.1803.06386.