

BAB II

KAJIAN LITERATUR

2.1 Curah Hujan dan Karakteristiknya

Curah hujan merupakan jumlah air yang jatuh ke permukaan tanah dalam bentuk air atau partikel es dalam periode waktu tertentu, yang umumnya dinyatakan dalam satuan milimeter (mm) [21]. Sementara itu, BMKG mendefinisikan bahwa curah hujan merupakan ketinggian air yang jatuh pada alat pengukur curah hujan yang diasumsikan bahwa air tersebut tidak menguap, meresap dan mengalir [4]. Penelitian oleh [4], menjelaskan bahwa Indonesia termasuk ke dalam negara dengan curah hujan yang cukup tinggi karena wilayah Indonesia terletak dalam wilayah tropis. Penelitian oleh [5] menambahkan, bahwa wilayah tropis seperti Indonesia memiliki curah hujan dengan pola yang sangat fluktuatif akibat pergantian musim dan pengaruh kondisi atmosfer. Menurut klasifikasi BMKG, intensitas curah hujan dapat dibagi menjadi 4 (empat) kategori utama yang ditunjukkan pada Tabel 2.1 [5].

Tabel 2. 1 Klasifikasi Curah Hujan Menurut BMKG

Kategori	Hujan Harian (mm/24jam)	Hujan Bulanan
Rendah	< 20	< 100
Menengah	21 – 50	100 – 300
Tinggi	51 – 100	300 – 500
Sangat Tinggi	> 100	> 500

Berdasarkan proses pembentukannya, hujan dapat dibedakan menjadi 3 (tiga) jenis utama, yaitu hujan Konvektif, hujan Orografis, dan hujan Frontal [22], [23], [24]:

1. Hujan Konvektif

Hujan Konvektif terbentuk akibat pemanasan di permukaan bumi yang menyebabkan udara lembab naik dan membentuk awan konvektif, seperti awan cumulonimbus. Proses ini menghasilkan hujan dengan intensitas tinggi dalam waktu yang relatif singkat dan hujan ini biasanya terjadi pada wilayah tropis.

2. Hujan Orografis

Hujan Orografis terbentuk ketika massa udara lembab bergerak menuju daerah pegunungan dan dipaksa naik mengikuti topografi wilayah tersebut. Saat udara naik,

suhu udara menurun sehingga terjadi hujan di wilayah pegunungan yang mengikuti arah datangnya angin.

3. Hujan Frontal

Hujan Frontal terbentuk dari pertemuan dua massa udara dengan suhu yang berbeda, yaitu massa udara hangat dengan massa udara dingin. Pada kondisi ini, udara hangat yang lebih ringan akan terdorong naik di atas udara dingin sehingga terjadi proses pembentukan awan yang menghasilkan hujan.

Penelitian oleh [5], menjelaskan bahwa pola curah hujan di Indonesia sangat dipengaruhi oleh kondisi geografis, pergerakan angin, kondisi laut, dan atmosfer. Penelitian tersebut juga menambahkan bahwa curah hujan di wilayah Medan menunjukkan pola musiman monsun, di mana puncak hujan terjadi antara Oktober hingga Januari, dan saat musim kemarau berlangsung pada Juni hingga September yang menunjukkan pola musiman yang signifikan namun tidak teratur. Fluktuasi curah hujan ini berpengaruh terhadap meningkatnya potensi genangan dan banjir perkotaan. [5]

Variabilitas curah hujan menggambarkan perubahan jumlah dan distribusi curah hujan di berbagai wilayah dan periode waktu yang dipengaruhi oleh kondisi atmosfer, suhu laut, serta faktor topografis. Variabilitas ini dapat terjadi dalam skala harian, bulanan, hingga tahunan, dan menjadi salah satu karakteristik iklim di wilayah tropis seperti Indonesia [5]. Selain itu, penelitian oleh [25], menjelaskan bahwa variabilitas penyebaran curah hujan di Indonesia sangat dipengaruhi oleh suhu permukaan laut di sekitar perairan nusantara. Peningkatan suhu laut memperbesar penguapan dan memperkuat pembentukan awan konvektif yang memicu hujan ekstrem di wilayah pesisir barat dan utara Sumatera. Penelitian lain oleh [2], juga menjelaskan bahwa variabilitas hujan di Sumatera Utara menunjukkan kecenderungan meningkat akibat perubahan suhu dan kelembapan atmosfer, sehingga menimbulkan pola hujan yang lebih acak dan sulit diprediksi.

Fluktuasi curah hujan merupakan hasil dari variabilitas iklim yang terjadi akibat interaksi kompleks antara faktor global, regional, dan lokal. Faktor-faktor utama yang memengaruhi fluktuasi curah hujan di Indonesia antara lain sebagai berikut [2], [5], [6], [25], [26], [27]:

1. *El Niño–Southern Oscillation (ENSO)*

ENSO merupakan fenomena iklim global yang berperan besar dalam memengaruhi curah hujan di wilayah tropis. Penelitian oleh [26] menjelaskan bahwa, Saat *El Niño*, suhu permukaan laut di Samudra Pasifik bagian tengah meningkat sehingga mengurangi konveksi awan di Indonesia, menyebabkan penurunan curah hujan dan potensi kemarau.

Sebaliknya, saat *La Niña* kelembapan atmosfer meningkat dan curah hujan ekstrem lebih sering terjadi di wilayah barat Indonesia. Penelitian tersebut juga menjelaskan bahwa variabel curah hujan di Asia Tenggara memiliki respon non-linear terhadap indeks ENSO, yang menyebabkan ketidakpastian dalam pola prediksi musiman. [26]

2. **Indian Ocean Dipole (IOD)**

IOD adalah fenomena anomali suhu laut di Samudra Hindia bagian barat dan timur. Pada fase IOD positif mengakibatkan pengurangan curah hujan di wilayah barat Indonesia, sedangkan saat fase IOD negatif memperkuat pergerakan massa udara lembab dari barat ke timur, sehingga meningkatkan curah hujan di Sumatera dan Jawa. [25]

3. **Suhu Permukaan Laut**

Suhu permukaan laut memiliki pengaruh langsung terhadap proses penguapan dan pembentukan awan. Semakin tinggi suhu laut, semakin besar jumlah uap air yang tersedia untuk pembentukan awan [2], [6]. Penelitian oleh [6], menjelaskan bahwa peningkatan suhu udara dan kelembapan berperan signifikan dalam memicu hujan ekstrem di wilayah pesisir timur Sumatera Utara. Penelitian oleh [2], juga menjelaskan bahwa peningkatan kelembapan udara yang terjadi di lapisan bawah atmosfer memperbesar potensi curah hujan dengan intensitas tinggi.

4. **Urbanisasi**

Perubahan penggunaan lahan dan pertumbuhan wilayah perkotaan berpengaruh terhadap peningkatan suhu permukaan dan sirkulasi udara. Fenomena *Urban Heat Island* (UHI) menyebabkan udara di perkotaan menjadi lebih panas dibandingkan daerah sekitarnya, yang kemudian meningkatkan potensi hujan lokal berintensitas tinggi [27]. Penelitian oleh [28], menjelaskan bahwa urbanisasi yang tidak diimbangi dengan sistem drainase yang baik, dapat memperbesar risiko banjir di musim hujan.

5. **Topografi**

Kondisi geografis dan topografi juga memegang peran penting dalam menentukan pola hujan suatu wilayah. Daerah pegunungan cenderung memiliki intensitas curah hujan yang lebih tinggi dibandingkan daerah dataran rendah karena adanya efek orografis, di mana udara lembab dipaksa naik kemudian terjadi hujan. Penelitian oleh [5], menjelaskan bahwa distribusi curah hujan di Kota Medan tidak merata, dengan intensitas tertinggi terjadi di wilayah barat yang berbukit dibandingkan dengan wilayah timur. [5]

2.2 Data Time Series

Data *Time Series* merupakan kumpulan data yang dicatat dalam jangka waktu yang tetap dan saling berhubungan dari waktu ke waktu. Analisis *Time Series* bertujuan untuk mengidentifikasi pola, tren, dan perilaku data dari waktu ke waktu agar dapat digunakan untuk peramalan di masa mendatang [1], [28]. Menurut penelitian oleh [29], data curah hujan adalah contoh unik dari *data deret waktu* karena nilainya dapat berubah terus menerus, hal ini dipengaruhi oleh dinamika atmosfer, suhu laut, dan geografis. Faktor-faktor ini menyebabkan pola curah hujan sulit diprediksi secara linier dan memerlukan pendekatan pemodelan yang mampu mengenali perubahan pola data. Data *Time Series* umumnya terdiri atas 4 (empat) komponen utama, yaitu tren, musiman, siklus, dan komponen acak [29], [30], [31].

1. **Tren (*Trend*)**

Komponen tren menggambarkan perubahan data dalam jangka panjang, bisa berupa kenaikan atau penurunan nilai rata-rata dari waktu ke waktu.

2. **Musiman (*Seasonality*)**

Komponen musiman menggambarkan fluktuasi yang berulang secara periodik dalam rentang waktu tertentu, biasanya tahunan.

3. **Siklus (*Cyclical*)**

Komponen siklus menggambarkan fluktuasi jangka panjang yang tidak berulang secara periodik.

4. **komponen acak (*irregular*)**

Komponen *irregular* merupakan variabel acak yang tidak dapat dijelaskan oleh tren, musiman, maupun siklus. Dalam model *Time Series*, komponen ini biasanya dianggap sebagai *noise* (fluktuasi acak) yang perlu diminimalkan agar hasil peramalan lebih akurat.

Data curah hujan memiliki karakteristik yang berbeda dibandingkan data *Time Series* lain seperti ekonomi atau demografi, beberapa ciri tersebut antara lain [11], [19], [32], [33], [34]:

1. **Non-linear**

Data non-linier adalah data yang hubungan antara variabelnya tidak membentuk garis lurus atau pola yang tetap, sehingga perubahan nilai pada suatu waktu tidak selalu diikuti perubahan yang sama pada waktu berikutnya [32]. Hubungan data dari waktu ke waktu pada data curah hujan bersifat non-linier karena dipengaruhi oleh berbagai kondisi cuaca yang saling berinteraksi dengan pola yang rumit [19]. Penelitian oleh [19],

menjelaskan bahwa model linier seperti ARIMA memiliki keterbatasan dalam menangkap pola ini. Oleh karena itu, model *deep learning* seperti *LSTM* sering digunakan karena mampu mengenali hubungan pola data non-linier yang panjang, contohnya hujan deras dapat terjadi setelah musim kemarau panjang tanpa pola linier yang jelas. Penelitian oleh [35], menjelaskan bahwa *LSTM* mampu menghasilkan prediksi curah hujan di Kabupaten Karo dengan MAPE di bawah 10%.

2. *Outlier*

Data curah hujan sering mengandung nilai ekstrem (*outlier*) akibat hujan lebat mendadak atau kesalahan pengukuran dari alat. Nilai-nilai ekstrem ini dapat menurunkan performa model jika tidak dilakukan prapemrosesan seperti *smoothing* atau *outlier removal*. Misalnya, curah hujan tiba-tiba melonjak dari 5 mm menjadi 150 mm dalam satu hari akibat badai lokal. Teknik *smoothing* atau deteksi *outlier* diperlukan sebelum dilakukan pemodelan [33]. Penelitian oleh [36], menjelaskan bahwa data dari *Automatic Weather Station* (AWS) sering mengandung *missing values*, *outlier*, dan data anomali yang disebabkan oleh gangguan sensor, kesalahan kalibrasi, maupun gangguan komunikasi data. Meskipun demikian, pada penelitian hidrometeorologi tidak semua nilai ekstrem dapat dianggap sebagai kesalahan data, karena itu merepresentasikan kejadian hujan lebat yang memang terjadi secara alami.

3. *Zero-inflated Data* (Banyak Nilai Nol)

Data curah hujan harian biasanya mengandung banyak nilai nol karena tidak setiap hari terjadi hujan, hal ini menyebabkan data memiliki proporsi nilai nol yang tinggi. Pada data curah hujan, nilai nol biasanya menggambarkan hari tanpa hujan (hari kering atau *dry days*) [34]. Penelitian oleh [37], menjelaskan bahwa tingginya proporsi nilai nol dapat mengurangi kemampuan model dalam menangkap pola jangka panjang, serta membuat model menjadi lebih sering memprediksi nilai nol sehingga menurunkan akurasi prediksi. Penelitian tersebut menyimpulkan bahwa model *LSTM* memiliki kemampuan dalam mempelajari pola pada data *Time Series* yang kompleks, termasuk data dengan banyak nilai nol.

4. *Non-stasioner*

Data non-stasioner adalah data yang memiliki pola statistik yang tidak stabil, seperti perubahan rata-rata, varians, atau tren dari waktu ke waktu. Data curah hujan sering kali tidak memenuhi syarat stasioneritas data, karena rata-rata dan variansnya berubah sepanjang waktu [11]. Penelitian oleh [38], menjelaskan bahwa data *Time Series* umumnya memiliki sifat non-linear dan non-stasioner, sehingga metode tradisional

sering mengalami keterbatasan dalam mengenali pola tersebut dalam data. Karakteristik ini menyebabkan proses pemodelan menjadi lebih sulit dilakukan, karena hubungan antar waktu dalam data bersifat tidak tetap dan mengandung pola yang kompleks. Oleh karena itu, diperlukan pendekatan pemodelan yang mampu menangkap perubahan pola dari waktu ke waktu dan hubungan antar data (hubungan temporal jangka panjang). Model *deep learning*, seperti LSTM banyak digunakan karena kemampuannya dalam memodelkan hubungan data dalam jangka waktu yang panjang dan pola non-linear pada data *Time Series*.

2.3 Model Prediksi *Time Series*

2.3.1 *Prophet Facebook*

Prophet merupakan algoritma peramalan yang digunakan untuk menganalisis data deret waktu (*Time Series*). Model ini mampu menyesuaikan pola tren non-linear dengan komponen musiman tahunan, mingguan, maupun harian, serta mempertimbangkan pengaruh hari libur [39]. Menurut penelitian oleh [40] *Prophet* efektif untuk menangani data dengan pola tren non-linear, pola musiman yang kompleks, dan kemampuannya untuk menangani data yang hilang (*missing value*) atau tidak teratur. Penelitian oleh [39] menambahkan, bahwa *Prophet* dapat menangani nilai yang tidak wajar (*outlier*) dengan baik dan juga dapat menangani faktor eksternal dengan menambahkan regresor lainnya, sehingga dapat mengolah lebih banyak data untuk meningkatkan akurasi peramalan. Kemudian, penelitian oleh [29] menjelaskan bahwa model ini mengasumsikan bahwa suatu deret waktu terdiri atas komponen utama yang dapat dipisahkan, yaitu tren, musiman, pengaruh khusus (seperti hari libur atau kejadian tertentu), serta komponen *error*. Pada dasarnya *Prophet* memiliki 5 (lima) parameter utama yaitu model *type*, *changepoints*, *seasonality*, *holidays*, dan *error* yang dapat dilihat pada persamaan di bawah ini [39]:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \dots \dots \dots (1)$$

Dimana komponen dari persamaan (1), terdiri atas:

- t : waktu saat ini
- y(t) : nilai saat ini
- g(t) : menggambarkan variabel nonperiodik dari deret waktu
- s(t) : variabel siklus atau musiman dalam deret waktu
- h(t) : efek kejadian atau periode khusus
- ϵ_t : error acak yang tidak dijelaskan oleh model

Berdasarkan penelitian oleh [29] dan [39], kedua penelitian ini menjelaskan bahwa faktor $g(t)$ dalam *Prophet* dapat dimodelkan dalam 2 (dua) model yaitu model tren: *piecewise linear trend* dan *logistic growth trend*. Pemilihannya bergantung pada bentuk perubahan jangka panjang data.

1. *Logistic Growth Model*: model ini mewakili pertumbuhan dalam berbagai tahap, pada tahap pertama pertumbuhan terjadi dengan cepat, kemudian ketika mendekati batas maksimum, pertumbuhan akan melambat dan mengikuti pola linear. Bentuk persamaannya sebagai berikut [29], [39]:

$$f(x) = \frac{L}{1+e^{-k(x-x_0)}} \dots\dots\dots (2)$$

Dimana komponen dari persamaan (2), terdiri atas:

- L : nilai maksimum dari model kurva
- k : laju pertumbuhan model
- x_0 : nilai x di titik sigmoid

2. *Piecewise Linear Model*: merupakan versi modifikasi dari model linier di mana rentang x yang berbeda memiliki perbedaan hubungan linier. Bentuk persamaannya sebagai berikut [29], [39]:

$$y = \beta_0 + \beta_1 x + \beta_2 (x - c)^+ + \varepsilon \dots\dots\dots (3)$$

Istilah periodik (t), yang memodelkan periodisitas deret waktu menggunakan deret *Fourier*, melalui persamaan di bawah ini [29]:

$$s(t) = \sum_{n=1}^N \left[a_n \cos\left(\frac{2\pi nt}{p}\right) + b_n \sin\left(\frac{2\pi nt}{p}\right) \right] \dots\dots\dots (4)$$

Dimana komponen dari persamaan (4), terdiri atas:

- p : periode waktu, parameternya dapat dinyatakan sebagai $[a_1, b_1, \dots, a_N, b_N]^T$

untuk hari libur (*holiday-event*) model $h(t)$, dapat dilihat pada persamaan di bawah ini [29], [39]:

$$h(t) = [1(t \in D_1), \dots, 1(t \in D_L)]k \dots\dots\dots (5)$$

Dimana komponen dari persamaan (5), terdiri atas:

- D : tanggal libur masa lalu dan masa depan
- t : jumlah atau rentang hari libur
- k : pengaruh hari libur terhadap hasil prediksi

Dalam penelitian yang dilakukan oleh [33], menggunakan *Prophet* untuk memprediksi curah hujan bulanan di *catchment Brisbane River* Australia. Model *Prophet* dibandingkan dengan beberapa algoritma *machine learning* (MLP, SVR, LGB, XGB, RDF,

STC) menggunakan data observasi BMKG Australia dan data dekadal CMIP5. Penelitian tersebut menjelaskan bahwa *Prophet* menunjukkan kinerja terbaik untuk periode kering. Penelitian lain oleh [15], menggunakan *Prophet* untuk memprediksi curah hujan bulanan di Jakarta Pusat menunjukkan bahwa *Prophet* memberikan hasil prediksi dengan RMSE dan MSE terendah dibanding LSTM dan SARIMA, sehingga menjadi model paling akurat dalam memproyeksikan tren hujan dua tahun ke depan. Hasil penelitian tersebut menjadi dasar penerapan kebijakan mitigasi banjir oleh Jakarta *Smart City*. Sementara itu, penelitian oleh [17] melakukan perbandingan antara *Prophet* dan LSTM pada prediksi curah hujan dasarian (10 harian) di Kabupaten Banjarnegara. *Prophet* menghasilkan RMSE 69,55 dan MAE 53,05, lebih rendah dibandingkan LSTM (RMSE 73,03 dan MAE 55,72). Penelitian tersebut juga menjelaskan bahwa *Prophet* memiliki keunggulan dalam menangkap pola musiman dan tren jangka panjang di daerah dengan variabilitas curah hujan tinggi.

Penggunaan model *Prophet* juga efektif dalam sistem mitigasi banjir yang memanfaatkan IoT, seperti penelitian yang dilakukan oleh [16] yang mengembangkan sistem prediksi banjir di Sungai Melaka, Malaysia. Penelitian tersebut menggunakan model ARIMA dan *Prophet*, hasilnya *Prophet* menunjukkan MAPE 6% dan RMSE 5, lebih baik dibanding ARIMA dalam menangani data sensor waktu secara *real-time*. Selain dalam bidang hidrometeorologi, *Prophet* juga diterapkan pada bidang lain seperti prediksi polusi udara dan penjualan. Seperti penelitian oleh [40] menunjukkan bahwa metode hybrid model *LSTM-Prophet* memberikan akurasi hingga 99,74% dalam prediksi konsentrasi gas SO₂ dan NO₂ di Kota Medan, sedangkan penelitian oleh [39], menjelaskan bahwa *Prophet* efektif untuk peramalan penjualan mingguan dan bulanan, terutama pada data dengan pola musiman dan tren yang signifikan.

2.3.2 Long Short-Term Memory

Long Short-Term Memory (LSTM) merupakan arsitektur jaringan saraf berulang (*Recurrent Neural Network/RNN*) yang dikembangkan oleh Hochreiter & Schmidhuber (1997), dirancang untuk memodelkan hubungan data dalam jangka waktu yang panjang pada deret waktu. LSTM digunakan untuk mengatasi keterbatasan RNN tradisional yang sulit mempertahankan informasi ketika jarak waktu antar data semakin panjang. Hal ini disebabkan oleh fenomena *vanishing gradient*, yaitu melemahnya gradien selama proses pembelajaran sehingga model terbatas dalam mengenali pola data dari waktu ke waktu. LSTM memperkenalkan unit memori khusus yang memiliki mekanisme pengaturan masuk dan keluarnya informasi melalui 3 (tiga) jenis gerbang (*gates*) yaitu *forgetting gate*, *input*

gate, dan *output gate*, sehingga model dapat mempertahankan, memperbarui, atau menghapus informasi dengan lebih teratur. [41]

Forget gate (f_t) yaitu menentukan informasi apa yang akan dihilangkan dari unit tersebut. *Forget gate* ini dapat dinyatakan dengan persamaan berikut [41]:

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f) \dots \dots \dots (6)$$

Input gate (i_t) atau disebut juga *update gate* untuk menentukan informasi baru apa yang harus disimpan di unit tersebut. *Input gate* ini dapat dinyatakan dengan persamaan berikut [41]:

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \dots \dots \dots (7)$$

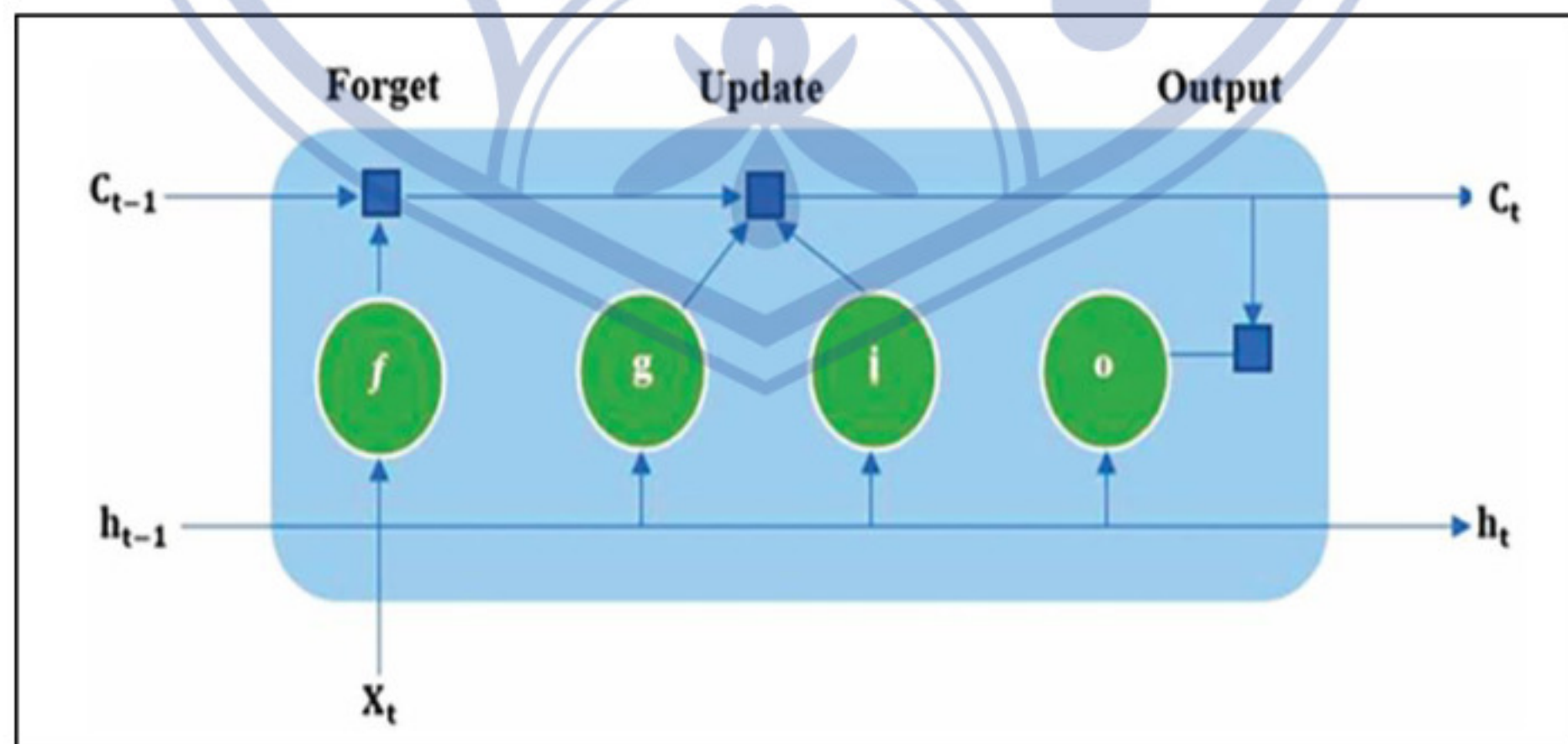
$$C_t = \tanh(W_c * [h_{t-1}, x_t] + b_c) \dots \dots \dots (8)$$

Output gate (o_t) yaitu suatu kondisi tersembunyi di mana kontrol akan dialihkan ke unit neuron berikutnya. *Output gate* ini dapat dinyatakan dengan persamaan berikut [41]:

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o) \dots \dots \dots (9)$$

$$h_t = o_t * \tanh(C_t) \dots \dots \dots (10)$$

Yang dimana, h_{t-1} adalah layer *output* tersembunyi dari unit sebelumnya. h_t adalah layer *output* tersembunyi dari unit saat ini. x_t adalah *input* dari unit saat ini. C_t adalah memori dari unit saat ini, yang akan diteruskan sepanjang jaringan *neuron*. W_f, W_i, W_c, W_o adalah bobot metrik. B_f, B_i, B_c, B_o adalah istilah bias, yang membantu model mempelajari jarak antar data yang panjang dalam urutan saat memproses mengurutkan data. *Neuron* LSTM standar digunakan untuk membangun jaringan *deep learning* untuk membuat prediksi dengan cara yang sepenuhnya terhubung, dan sebuah neuron LSTM standar dibentuk sesuai dengan Gambar 2.1 di bawah ini.



Gambar 2. 1 A Standard LSTM Neuron [41]

LSTM banyak digunakan dalam penelitian *rainfall forecasting* karena kemampuannya menangkap hubungan non-linear antar variabel cuaca. Seperti penelitian

oleh [42], menerapkan LSTM untuk prediksi curah hujan di Kabupaten Malang dengan data BMKG tahun 2010-2021. Hasil terbaik diperoleh pada konfigurasi 3 *layers* dengan 50 neuron per layer, menghasilkan nilai MAE 7.90 dan RMSE 10.16, menunjukkan akurasi yang baik dalam mendeteksi tren musiman dan anomali curah hujan ekstrem. Penelitian lain oleh [35], menunjukkan LSTM memprediksi curah hujan bulanan dengan MAPE < 10%. Penelitian tersebut juga menjelaskan bahwa model ini mampu memetakan pola data dari waktu ke waktu dan faktor musiman yang tidak dapat dijelaskan dengan metode statistik klasik seperti ARIMA.

Sementara itu penelitian oleh [43], yang menggunakan LSTM dan Vanilla RNN untuk prediksi curah hujan *dasarian* di Bali. Hasil menunjukkan bahwa LSTM lebih baik dibanding RNN dengan nilai R-squared = 0.61 dan RMSE = 28.43. Penelitian tersebut menyimpulkan bahwa model ini dapat dinilai lebih baik karena dapat menyesuaikan hubungan data jangka panjang dan variabel antarperiode dalam menentukan awal musim hujan dan kemarau. Penelitian oleh [18], menggunakan pendekatan LSTM-*Isolation Forest* untuk deteksi curah hujan ekstrem yang berpotensi banjir. Hasil evaluasi menunjukkan RMSE = 4.37 dan MSE = 19.11, hasilnya model tersebut dapat mengidentifikasi anomali hujan yang berpotensi menyebabkan genangan. Penelitian tersebut menjelaskan bahwa, dengan pendekatan *hybrid* ini model LSTM tidak hanya dapat digunakan dalam prediksi data curah hujan saja, tetapi juga sebagai *tools* (alat) untuk peringatan dini banjir.

LSTM juga digunakan dalam penelitian mengenai sistem prediksi curah hujan dengan basis data besar. Seperti penelitian oleh [25], yang menggunakan multivariate LSTM untuk memprediksi curah hujan setiap 10 menit menggunakan citra satelit *Himawari-8* dan data GPM IMERG. Model ini mendapatkan RMSE 1.54 mm/3 jam. Penelitian lain oleh [44], menjelaskan bahwa LSTM memberikan *loss value* hanya 0.09, jauh lebih rendah dibanding model regresi polinomial atau *random forest*, dalam peramalan banjir yang menggunakan data iklim historis. Penelitian tersebut juga menjelaskan bahwa dengan hasil ini menunjukkan LSTM mampu mengenali hubungan non-linear kompleks antara curah hujan, kelembapan, dan suhu udara.

2.4 Evaluasi Model Prediksi Curah Hujan

Evaluasi model merupakan tahapan yang mendasar dalam penelitian peramalan deret waktu (*Time Series Forecasting*) karena berfungsi untuk mengukur tingkat kesesuaian antara nilai prediksi dan data aktual. Evaluasi yang sistematis diperlukan agar kinerja antar model dapat dibandingkan secara objektif dan hasil prediksi dapat dipertanggungjawabkan secara

ilmiah [45]. Dalam penelitian prediksi curah hujan, evaluasi model menjadi proses penting karena data bersifat fluktuatif, non-linear, dan sering mengandung nilai ekstrem yang berkaitan langsung dengan risiko banjir dan bencana hidrometeorologi lainnya [46].

Secara umum, evaluasi dilakukan dengan membandingkan *output* model terhadap data yang dianalisis menggunakan berbagai metrik kesalahan (*error metrics*) serta ukuran tingkat kecocokan (*goodness-of-fit*) [46]. Penelitian oleh [45], menjelaskan bahwa penggunaan lebih dari satu metrik evaluasi diperlukan untuk memberikan gambaran kinerja model yang lebih menyeluruh, karena setiap metrik memiliki sensitivitas yang berbeda terhadap pola kesalahan tertentu. Untuk mengukur kinerja model prediksi penelitian ini akan mengevaluasi perbandingan antara hasil prediksi dan hasil aktualnya digunakan beberapa metrik evaluasi model yaitu:

1. *Mean Absolute Error (MAE)*

Mean Absolute Error (MAE) mengukur rata-rata kesalahan absolut antara nilai yang diprediksi dan nilai aktual, memberikan gambaran yang jelas tentang seberapa besar hasil prediksi model meleset dalam satuan yang sama dengan data [47]. Penelitian oleh [48], melakukan prediksi curah hujan di Kota Semarang menggunakan pendekatan *machine learning*, MAE digunakan untuk membandingkan performa beberapa model karena sifatnya yang stabil terhadap keberadaan *outlier*. Hasil penelitian tersebut menunjukkan bahwa model dengan nilai MAE lebih kecil cenderung memberikan prediksi yang lebih konsisten terhadap data yang dianalisis. Penelitian oleh [45] menambahkan, bahwa MAE banyak digunakan dalam penelitian hidrometeorologi karena tidak memberikan penalti (bobot kesalahan) berlebihan terhadap kesalahan ekstrem, sehingga cocok untuk mengevaluasi performa model dalam prediksi curah hujan. Bentuk persamaan dari MAE yaitu [47]:

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \dots\dots\dots (11)$$

Dimana komponen dari persamaan (11), terdiri atas:

- n : jumlah data
- y : nilai aktual
- $\hat{y}$: nilai prediksi

2. *Mean Square Error (MSE)*

Mean Square Error (MSE) merupakan rata-rata dari selisih kuadrat antara nilai yang diprediksi dan nilai yang diamati. MSE adalah metrik yang digunakan untuk menilai akurasi suatu prediksi. Semakin tepat prediksinya, semakin kecil nilai MSE yang

seharusnya [3]. Dalam penelitian peramalan curah hujan menggunakan model statistik dan *machine learning*, MSE dapat digunakan untuk menilai akurasi model. Seperti penelitian oleh [47], melakukan evaluasi terhadap model ARIMA-Kalman untuk prediksi curah hujan di Kota Medan, penelitian tersebut menjelaskan bahwa MSE efektif dalam membedakan model yang mampu menangkap fluktuasi curah hujan signifikan dengan model yang hanya mengikuti pola rata-rata. Penelitian oleh [29], menjelaskan bahwa MSE umumnya digunakan bersamaan dengan metrik lain karena nilai MSE yang besar dapat dipengaruhi oleh beberapa kesalahan ekstrem saja. Bentuk persamaan dari MSE yaitu [3]:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \dots\dots\dots (12)$$

Dimana komponen dari persamaan (12), terdiri atas:

- n : jumlah data
- y : nilai aktual
- $\hat{y}$: nilai prediksi

3. **Root Mean Square Error (RMSE)**

Root Mean Square Error (RMSE) menghitung akar kuadrat dari rata-rata kesalahan kuadrat antara nilai aktual dan nilai prediksi [47]. Penelitian oleh [48] menambahkan bahwa, metrik RMSE ini adalah akar kuadrat dari rata-rata selisih kuadrat antara prediksi. RMSE memberikan penalti yang lebih besar untuk kesalahan yang lebih besar, sehingga lebih sensitif terhadap *outlier* [47]. Dalam penelitian prediksi banjir dan curah hujan, RMSE digunakan untuk mengevaluasi kemampuan model dalam menangkap nilai puncak hujan yang berpotensi menyebabkan banjir. Seperti penelitian oleh [44], menjelaskan bahwa model LSTM menghasilkan nilai RMSE yang lebih rendah dibandingkan model regresi dan *random forest*. Penelitian oleh [45] menambahkan, bahwa RMSE juga digunakan dalam evaluasi model *deep learning* dan *hybrid models* karena mampu menggambarkan kualitas prediksi pada kondisi data yang ekstrem dalam sistem peringatan dini banjir. Bentuk persamaan dari RMSE yaitu [48]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \dots\dots\dots (13)$$

Dimana komponen dari persamaan (13), terdiri atas:

- n : jumlah data
- y : nilai aktual
- $\hat{y}$: nilai prediksi

2.5 Hubungan Curah Hujan dan Mitigasi Banjir

2.5.1 Hujan sebagai Pemicu Banjir

Curah hujan ekstrem merupakan pemicu utama banjir, terutama terjadinya banjir pluvial yang terjadi akibat intensitas hujan tinggi dalam durasi singkat. Banjir pluvial merupakan jenis banjir yang terjadi akibat curah hujan intensitas tinggi dalam durasi relatif singkat yang melebihi kapasitas kemampuan tanah dalam menyerap air dan sistem drainase perkotaan, tanpa keterkaitan langsung dengan meluapnya air sungai. Fenomena ini umumnya terjadi di wilayah perkotaan dengan permukaan tanah yang sulit menyerap air, serta sistem drainase yang tidak mampu menampung banyak air. Pada kondisi tersebut, air hujan langsung meluap ke permukaan seperti jalan, kawasan permukiman, dan infrastruktur perkotaan, hal ini menyebabkan genangan atau banjir secara tiba-tiba. Karakteristik utama banjir pluvial adalah aliran air yang cepat setelah hujan yang terjadi pada area tertentu, serta dipengaruhi oleh intensitas dan penyebaran hujan. [49]

Seperti penelitian oleh [26] di Daerah Aliran Sungai (DAS) Bekasi, menjelaskan bahwa kejadian banjir besar tahun 2016, 2020, 2022, dan 2025 itu disebabkan oleh hujan ekstrem dengan intensitas tinggi, bukan hanya disebabkan oleh luapan sungai. Hasil simulasi *Rainfall-Runoff-Inundation (RRI)* menunjukkan bahwa curah hujan ekstrem menjadi faktor dominan yang memicu banjir perkotaan. Hal yang sama juga dilaporkan pada penelitian oleh [48] di Kota Semarang, di mana peningkatan curah hujan harian berkorelasi langsung dengan kejadian banjir dan genangan air. Penelitian tersebut juga menjelaskan bahwa hujan dengan intensitas tinggi menjadi pemicu utama bencana hidrometeorologi seperti banjir, tanah longsor, dan kerusakan infrastruktur perkotaan. Penelitian oleh [44], juga menjelaskan bahwa sebagian besar kejadian banjir di Bangladesh dipicu oleh hujan intensitas tinggi yang berlangsung secara berurutan, terutama pada wilayah dengan sistem drainase yang tidak memadai.

2.5.2 Ambang Curah Hujan Ekstrem

Ambang curah hujan ekstrem (*extreme rainfall threshold*) merupakan nilai batas curah hujan tertentu, ketika nilai tersebut melewati batas tertentu dapat meningkatkan probabilitas terjadinya banjir. Penentuan ambang curah hujan ekstrem menjadi komponen penting dalam sistem peringatan dini banjir karena memungkinkan identifikasi potensi banjir sebelum banjir itu terjadi [26], [50], [51], [52]. Penelitian oleh [51] di Chennai, menjelaskan bahwa kejadian banjir umumnya terjadi ketika curah hujan harian berada pada kisaran 100-150 mm/hari, dengan faktor lain yang dipengaruhi oleh struktur tanah dan kapasitas drainase.

Penelitian tersebut juga menjelaskan bahwa, dengan hasil ini bahwa ambang hujan ekstrem dapat berbeda pada setiap wilayah dan bergantung pada kondisi wilayah tersebut.

Penelitian oleh [26], menjelaskan bahwa banjir besar di DAS (daerah aliran sungai) Bekasi disebabkan oleh curah hujan ekstrem yang melampaui 180 mm/hari, yang menyebabkan genangan hingga lebih dari 3 meter pada wilayah hilir. Nilai ambang ini dipengaruhi oleh perubahan tutupan lahan dan urbanisasi yang mempercepat air mengalir ke permukaan tanah. Sementara itu, penelitian oleh [48] di Kota Semarang menjelaskan bahwa risiko banjir meningkat secara signifikan ketika curah hujan harian melebihi 90 mm/hari, terutama pada kawasan dengan kepadatan penduduk tinggi dengan sistem drainase yang terbatas.

2.5.3 Time Lag antara Hujan dan Banjir

Time lag merupakan jarak waktu antara terjadinya hujan dan munculnya dampak banjir, seperti genangan atau luapan sungai. *Time lag* menjadi faktor penting dalam mitigasi banjir karena menentukan waktu respons yang tersedia bagi sistem peringatan dini dan pengambilan tindakan pencegahan sebelum kejadian banjir [50], [52]. Penelitian oleh [26], menjelaskan bahwa *time lag* antara puncak hujan dan awal kejadian banjir di DAS Bekasi berkisar 7 jam atau kurang, karena proses aliran air yang cepat pada wilayah perkotaan dataran rendah. Penelitian oleh [52], menjelaskan bahwa pada wilayah perkotaan *time lag* antara hujan dan banjir umumnya berkisar 1-6 jam, terutama pada kawasan dengan sistem drainase terbatas dan minim daerah resapan. Penelitian lain oleh [44], menjelaskan bahwa *time lag* antara puncak hujan dan puncak banjir di Bangladesh berkisar 6-12 jam, karena adanya proses penumpukan air di aliran air pada wilayah hulu sungai.

2.6 Penelitian Terkait dan *Research Gap*

Dalam penelitian ini, beberapa penelitian terdahulu digunakan sebagai acuan untuk memperkuat dasar teori dan metodologi dalam prediksi curah hujan. Rangkuman lima penelitian yang telah dipilih dapat dilihat pada Tabel 2.3.

Tabel 2. 2 Penelitian Terdahulu

No	Referensi	Model	Kontribusi Penelitian	Area Aplikasi	Evaluasi Model	Hasil	Analisis Gap Penelitian
1	[19]	LSTM	Mengembangkan model LSTM untuk memprediksi curah hujan dan debit sungai sebagai dasar penentuan status siaga banjir	Hidrologi & Mitigasi Banjir	MAE, RMSE	Model LSTM menghasilkan MAE hujan berkisar 0,041 - 0,054 dan RMSE 0,091 - 0,099	Tidak membandingkan LSTM dengan model <i>Time Series</i> lain seperti <i>Prophet</i> serta belum dianalisis khusus pada wilayah perkotaan
2	[42]	LSTM	Menerapkan LSTM untuk memprediksi curah hujan harian dengan dua skenario arsitektur jaringan	Klimatologi	MAE, RMSE, MSE	Nilai diperoleh dengan MAE 7,90, RMSE 10,16, dan MSE 103,37	Penelitian hanya menggunakan LSTM tanpa perbandingan dengan model <i>machine learning</i> lain seperti <i>Prophet</i>

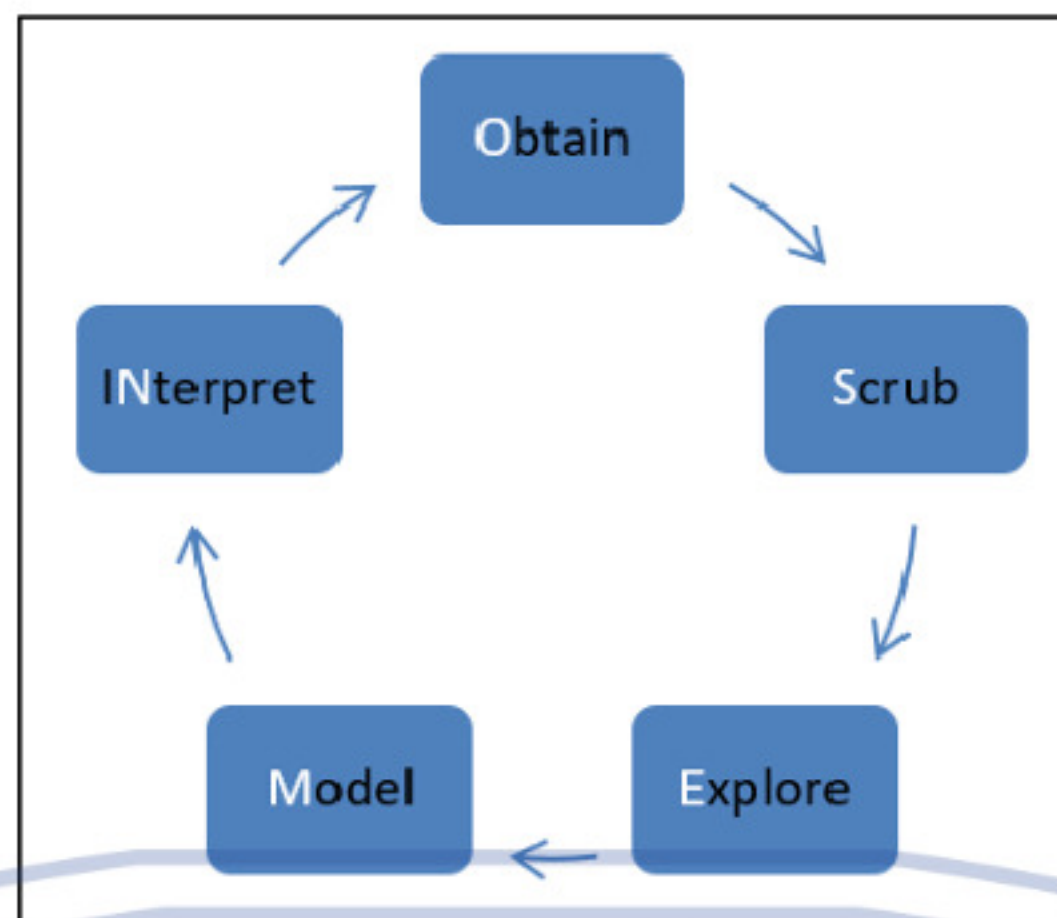
3	[16]	ARIMA, Prophet	Membandingkan ARIMA dan <i>Prophet</i> dalam sistem peringatan banjir dengan memanfaatkan IoT	<i>Flood Forecasting Real-Time</i>	MAPE, RMSE	<i>Prophet</i> menunjukkan performa lebih baik dengan MAPE = 6% dan RMSE = 5 dibandingkan ARIMA	Tidak menggunakan data curah hujan sebagai <i>input</i> utama dan belum membandingkan dengan model <i>deep learning</i> seperti LSTM
4	[17]	<i>Prophet</i> , LSTM	Membandingkan <i>Prophet</i> dan LSTM untuk prediksi curah hujan dasarian multi-stasiun	Klimatologi & Mitigasi Bencana	RMSE, MAE	<i>Prophet</i> lebih baik dengan RMSE rata-rata 69,55 dan MAE 53,05, lebih rendah dibandingkan LSTM dengan RMSE = 73,03 dan MAE = 55,72	Skala waktu masih dasarian, belum diterapkan pada data harian pada wilayah perkotaan
5	[15]	SARIMA, <i>Prophet</i> , LSTM	Membandingkan <i>Prophet</i> , SARIMA, dan LSTM untuk prediksi curah hujan bulanan sebagai dasar mitigasi banjir	Mitigasi Banjir Perkotaan	MSE, RMSE	<i>Prophet</i> menghasilkan nilai MSE = 23821,93 dan RMSE = 154,34. Terendah dibandingkan SARIMA (25006,03 dan 158,13) dan LSTM (46206,51 dan 214,96)	Hanya fokus pada prediksi skala bulanan, belum menganalisis data harian

Berdasarkan kajian penelitian terdahulu dan analisis *research gap* yang telah dilakukan, dapat disimpulkan bahwa meskipun berbagai metode peramalan curah hujan telah dikembangkan, masih terdapat beberapa keterbatasan pada pendekatan dan area aplikasinya. Sebagian besar penelitian berfokus pada penggunaan satu jenis model, khususnya *deep learning* seperti LSTM, tanpa melakukan perbandingan dengan model *Time Series* lain seperti *Prophet* dalam satu kerangka evaluasi yang sama. Selain itu, mayoritas penelitian masih menggunakan skala waktu bulanan atau dasarian, sementara kajian pada data curah hujan harian di wilayah perkotaan rawan banjir masih relatif terbatas. Kondisi ini menunjukkan perlunya penelitian yang secara khusus membandingkan kinerja *Prophet* dan LSTM pada data curah hujan di Kota Medan.

Sejalan dengan *research gap* tersebut, penelitian ini bertujuan untuk menerapkan dan membandingkan model *Time Series Prophet* dan LSTM dalam memprediksi curah hujan di Kota Medan, serta menentukan model dengan tingkat akurasi terbaik melalui evaluasi menggunakan beberapa metrik kesalahan. Selain itu, penelitian ini juga akan menghubungkan hasil prediksi curah hujan dengan upaya mitigasi banjir, sehingga informasi prediksi dapat dimanfaatkan sebagai dasar pendukung perencanaan dan pengambilan keputusan dalam pengurangan risiko banjir di Kota Medan. Dengan demikian, penelitian ini diharapkan mampu mengisi celah penelitian yang ada dengan melakukan perbandingan antara model *Prophet* dan LSTM pada data curah hujan Kota Medan, serta memberikan hasil prediksi tersebut sebagai informasi pendukung dalam upaya mitigasi banjir di Kota Medan.

2.7 Metode OSEMN

Metode OSEMN (*Obtain, Scrub, Explore, Model, iNterpret*) merupakan kerangka kerja dalam *data science* yang digunakan untuk mengelola proses analisis data secara sistematis, mulai dari pengumpulan data hingga interpretasi hasil analisis. OSEMN dapat digambarkan sebagai sebuah pendekatan yang praktis dan terstruktur karena setiap tahap dirancang saling berkaitan dan membentuk satu siklus analisis data yang utuh [39], [40]. OSEMN ini digunakan dalam penelitian yang menggunakan model *Time Series* karena memberikan kerangka kerja yang konsistens selama proses analisis dari awal hingga akhir. Seperti penelitian oleh [39] dan [40], OSEMN digunakan sebagai kerangka kerja penelitian dalam proses analisis untuk mengelola data deret waktu dan mengembangkan model prediksi, sehingga setiap keputusan analitis didasarkan pada tahapan yang jelas dan terdokumentasi. Tahapan metode OSEMN terdiri atas 5 (lima) tahap seperti yang ditunjukkan pada gambar 2.2 berikut.



Gambar 2. 2 Tahapan OSEM N [40]

1. *Obtain*

Tahap awal dalam metode OSEM N yaitu *Obtain*, yang berfokus pada proses pengumpulan data dari sumber yang relevan dengan tujuan penelitian [39]. Tahap *Obtain* juga mencakup proses penentuan atribut atau variabel yang relevan dengan tujuan penelitian [53]. Pada tahap ini data yang dikumpulkan harus memiliki data historis yang tersusun berdasarkan urutan waktu digunakan untuk mengidentifikasi pola tren, musiman, serta fluktuasi yang terjadi secara berulang [39]. Hal ini didukung oleh penelitian [33], yang menjelaskan bahwa penggunaan data historis dengan cakupan waktu yang panjang sangat penting dalam prediksi curah hujan karena model prediksi membutuhkan representasi pola musiman dan perubahan iklim jangka panjang untuk menghasilkan prediksi yang lebih akurat. Oleh karena itu, data yang diperoleh pada tahap *Obtain* harus memiliki interval waktu yang konsisten agar model dapat mempelajari pola data dengan lebih baik [33], [39]. Penelitian oleh [54], menjelaskan bahwa dataset yang digunakan pada proses *data science* harus dari sumber terbuka, terpercaya, mudah diakses, serta memiliki korelasi yang cukup kuat terhadap fenomena yang dianalisis. Kemudian penelitian oleh [39], menambahkan bahwa tahap *Obtain* dilakukan dengan mengumpulkan data deret waktu harian dari sumber resmi untuk memastikan kontinuitas dan validitas data.

2. *Scrub*

Tahap *scrub* merupakan proses membersihkan dan persiapan data sebelum dilakukan analisis dan pemodelan. Tahap ini bertujuan untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik, konsisten, dan siap diproses oleh model analisis maupun prediksi. Data yang tidak lengkap atau tidak konsisten dapat menyebabkan bias dan menurunkan performa model prediksi. Proses *Scrub* mencakup penanganan *missing*

value, penghapusan data duplikat, penyesuaian format dan tipe data, serta penyelarasan indeks waktu pada data deret waktu [39], [40]. Penelitian oleh [14], menjelaskan bahwa data curah hujan memiliki karakteristik yang kompleks serta non-linear sehingga membutuhkan proses *preprocessing* yang baik agar model *prediksi* dapat menghasilkan prediksi yang lebih akurat. Penelitian oleh [40], menambahkan bahwa tahap scrub dilakukan dengan memastikan keseragaman format waktu dan mengeliminasi data yang tidak valid agar model prediksi dapat mempelajari pola data yang sebenarnya. Penelitian lain oleh [55], menjelaskan bahwa proses pembersihan data yang sistematis berpengaruh langsung terhadap kualitas hasil pemodelan dan interpretasi.

3. *Explore*

Tahap *explore* merupakan proses eksplorasi dan visualisasi data untuk memahami karakteristik dasar data yang akan digunakan. Pada tahap ini dilakukan analisis terhadap distribusi data, pola tren, musiman, serta keberadaan anomali atau nilai ekstrem [39]. Tahap mengeksplorasi data dilakukan untuk membantu memahami pola data, mengidentifikasi pola periodik dan fluktuasi yang menjadi dasar dalam pemilihan model prediksi yang sesuai [56]. Hal ini didukung oleh penelitian [57], yang menjelaskan bahwa eksplorasi pola curah hujan membantu dalam menentukan model *Time Series* yang mampu menangani pola musiman dan fluktuasi data. Penelitian oleh [40], menjelaskan bahwa eksplorasi data diperlukan untuk identifikasi pola, musiman, ketidakteraturan, dan pengulangan data berdasarkan urutan waktu. Sehingga pola perubahan data dari waktu ke waktu dapat dianalisis. Kemudian penelitian oleh [57], Menambahkan bahwa data curah hujan memiliki karakteristik pola musiman, tren jangka panjang, serta non-linear sehingga dibutuhkan proses eksplorasi data sebelum dilakukan pemodelan. Penelitian lain oleh [58], menjelaskan bahwa eksplorasi data membantu mengidentifikasi pola curah hujan ekstrem yang penting dalam mendukung mitigasi bencana dan pengambilan keputusan.

4. *Model*

Tahap *model* merupakan proses pengembangan dan penerapan model analisis atau prediksi berdasarkan data yang telah melalui tahap pembersihan dan eksplorasi. Pada tahap ini dilakukan pemilihan model prediksi yang sesuai, pembagian data menjadi data *Training* dan data *Testing*, pelatihan model, serta evaluasi kinerja model. Pada analisis deret waktu, proses pembagian data *Training* dan *Testing* dilakukan berdasarkan urutan waktu agar pola perubahan data dari waktu ke waktu dapat dipelajari dengan baik oleh model prediksi [39]. Penelitian oleh [40], menambahkan bahwa tahap pemodelan

dilakukan dengan mempertimbangkan karakteristik data, seperti tren, musiman, dan non-linear agar model mampu menghasilkan prediksi yang lebih akurat. Penelitian oleh [57], menjelaskan bahwa pada tahap pemodelan perlu dilakukan perbandingan model untuk mengetahui model dengan performa terbaik berdasarkan data yang digunakan.

5. *INterpret*

Tahap *interpret* merupakan tahap akhir dalam metode OSEMN yang berfokus pada menginterpretasikan hasil analisis dan prediksi [39]. Tahap *iNterpret* ini untuk memahami kemampuan model dalam menangani pola data, sehingga dapat mengidentifikasi karakteristik dan keunggulan masing-masing model terhadap data yang digunakan [59]. Selain itu, pada tahap ini hasil model tidak hanya dinilai berdasarkan akurasi model, tetapi juga dianalisis manfaatnya atau kegunaannya terhadap permasalahan nyata yang menjadi fokus dan tujuan penelitian. Interpretasi hasil bertujuan agar hasil penelitian dapat digunakan sebagai dasar pengambilan keputusan, perencanaan, serta penyampaian informasi berdasarkan hasil analisis data. Tahap interpretasi memiliki peran penting karena hasil prediksi harus mampu diterjemahkan menjadi informasi yang mudah dipahami dan dapat digunakan terutama dalam proses pengambilan keputusan. [39], [56]. Penelitian oleh [40], menambahkan bahwa tahap *iNterpret* digunakan untuk mengaitkan hasil prediksi dengan permasalahan yang diteliti, tahap ini juga diperlukan agar hasil prediksi dapat divisualisasikan sehingga lebih mudah dipahami oleh pengguna akhir. Hal ini didukung oleh penelitian [60], yang menjelaskan bahwa proses interpretasi diperlukan agar hasil prediksi dapat digunakan dalam mendukung *pengambilan keputusan* dan perencanaan mitigasi dengan lebih efektif. Penelitian lain oleh [56], menjelaskan bahwa tanpa tahap interpretasi yang baik, hasil analisis data hanya bersifat numerik dan tidak dapat memberikan informasi yang relevan bagi pengguna akhir.