

BAB II

TINJAUAN PUSTAKA

2.1. Kajian Relevan

Kajian relevan ini membahas penelitian-penelitian terdahulu yang berkaitan dengan prediksi curah hujan sebagai dasar perbandingan dan penguatan terhadap penelitian yang dilakukan.

Analisis Prediksi Curah Hujan di Kota Tangerang Menggunakan Metode LSTM dan GRU bertujuan mengatasi tantangan ketidakpastian curah hujan yang berdampak pada sektor pertanian, energi, dan infrastruktur, sekaligus keterbatasan metode tradisional dalam memodelkan pola hujan yang bersifat nonlinier. Studi ini membandingkan kinerja dua algoritma deep learning, yaitu Long Short-Term Memory (LSTM) dan Gated Recurrent Unit (GRU), dengan memanfaatkan data sekunder curah hujan dari BMKG Kota Tangerang periode 2014–2025. Hasil penelitian menunjukkan bahwa LSTM memberikan performa yang lebih unggul dibandingkan GRU, ditunjukkan oleh nilai evaluasi terbaik dengan MAE sebesar 0,084, RMSE 0,098, dan koefisien determinasi (R^2) sebesar 0,656, serta kemampuan yang lebih baik dalam menangkap pola musiman yang kompleks dan stabilitas konvergensi selama proses pelatihan (Supriatna et al., 2025).

Rainfall Prediction: A Comparative Analysis of Modern Machine Learning Algorithms for Time-Series Forecasting membahas tantangan dalam prediksi curah hujan yang bersifat nonlinier serta keterbatasan model statistik tradisional yang memerlukan biaya komputasi dan anggaran tinggi. Studi ini melakukan analisis komparatif antara algoritma machine learning konvensional, seperti XGBoost dan model ansambel, dengan berbagai arsitektur deep learning, yaitu LSTM, Stacked-LSTM, dan Bidirectional-LSTM, menggunakan data deret waktu cuaca per jam dari tahun 2000 hingga 2020 di lima kota besar di Inggris untuk memprediksi volume curah hujan. Hasil penelitian menunjukkan bahwa Stacked-LSTM dengan dua lapisan tersembunyi serta Bidirectional-LSTM menghasilkan performa terbaik dengan nilai kesalahan RMSE, MAE, dan RMSLE paling rendah dibandingkan model lainnya, sehingga menegaskan

bahwa arsitektur LSTM dengan jumlah lapisan yang relatif sederhana lebih efektif dan efisien untuk diterapkan dalam sistem prediksi curah hujan yang berorientasi pada efisiensi anggaran (Barrera-Animas et al., 2022)

Prediksi Curah Hujan Harian Menggunakan Model ARIMA dan LSTM di Stasiun Meteorologi Radin Inten II Lampung dilakukan untuk menjawab tantangan fluktuasi pola hujan yang semakin sulit diprediksi dan berdampak langsung pada sektor pertanian serta mitigasi bencana hidrometeorologi. Studi ini membandingkan kinerja model deret waktu konvensional ARIMA dengan model deep learning Long Short-Term Memory (LSTM) menggunakan data klimatologi periode 2022–2024, dengan tahapan prapemrosesan berupa normalisasi Min-Max serta evaluasi menggunakan metrik RMSE, MAE, MAPE, dan koefisien determinasi (R^2). Hasil penelitian menunjukkan bahwa LSTM memberikan performa yang lebih unggul dibandingkan ARIMA, ditunjukkan oleh nilai RMSE sebesar 19,57 dan R^2 mencapai 0,94, sehingga model ini dinilai lebih efektif dalam menangkap pola nonlinier dan variasi musiman yang kompleks pada karakteristik iklim tropis wilayah Lampung (Meteorologi & Inten, 2025)

Multidimensional Precipitation Index Prediction Based on CNN-LSTM Hybrid Framework mengangkat permasalahan tingginya kompleksitas dan ketidakpastian data curah hujan yang membuat pendekatan statistik dan model fisik konvensional semakin sulit memenuhi kebutuhan akurasi prediksi. Studi ini mengusulkan kerangka kerja hibrida CNN-LSTM yang menggabungkan kemampuan Convolutional Neural Network (CNN) dalam mengekstraksi fitur lokal serta pola fluktuasi musiman dengan keunggulan Long Short-Term Memory (LSTM) dalam memodelkan ketergantungan temporal jangka panjang pada data runtun waktu. Pengujian dilakukan menggunakan data curah hujan bulanan selama 31 tahun (1972–2002) di wilayah Pune, India. Hasil eksperimen menunjukkan bahwa model CNN-LSTM mampu mencapai nilai RMSE sebesar 6,752, yang menegaskan peningkatan akurasi dan kemampuan generalisasi yang lebih baik dibandingkan metode deret waktu tradisional. Model ini efektif dalam menangkap tren musiman dan fluktuasi jangka panjang, meskipun masih menghadapi keterbatasan dalam memprediksi kejadian cuaca ekstrem yang bersifat sangat variatif. (Wang et al., 2025)

Integrasi Sistem IoT untuk Pemantauan Cuaca Real-Time dan Prediksi Curah Hujan Berbasis LSTM dilatarbelakangi oleh keterbatasan prakiraan cuaca

regional yang belum mampu memberikan informasi real-time dan spesifik lokasi, sehingga menyulitkan pelaku UMKM sektor pertanian dalam mengambil keputusan berbasis kondisi cuaca aktual. Studi ini mengembangkan sistem pemantauan cuaca mikrosipial berbasis Internet of Things (IoT) menggunakan mikrokontroler ESP32 dan sensor DHT22 serta Hall-effect 3144, yang diintegrasikan dengan model prediksi Long Short-Term Memory (LSTM) yang dilatih menggunakan data historis BMKG dan data observasi lapangan. Hasil pengujian menunjukkan bahwa sistem mampu melakukan pemantauan cuaca secara real-time sekaligus menghasilkan prediksi curah hujan yang stabil, dengan nilai RMSE sebesar 17,05 mm dan MAE 12,40 mm. Temuan ini menegaskan bahwa integrasi IoT dan LSTM berpotensi menjadi solusi prakiraan cuaca yang terjangkau, mudah direplikasi, dan adaptif untuk mendukung sistem peringatan dini serta pengelolaan irigasi cerdas di sektor pertanian (No et al., 2025)

Application of RNN-LSTM in Predicting Drought Patterns in Pakistan: A Pathway to Sustainable Water Resource Management berangkat dari meningkatnya krisis kelangkaan air global akibat perubahan iklim, di mana Pakistan menghadapi transisi serius dari kondisi stres air menuju kelangkaan air yang parah. Kondisi ini mendorong kebutuhan mendesak akan model peramalan sumber daya air yang lebih akurat dan andal untuk mendukung mitigasi kekeringan di wilayah rentan. Studi ini menerapkan pendekatan Recurrent Neural Network–Long Short-Term Memory (RNN-LSTM) dengan memanfaatkan data sekunder jangka panjang berupa curah hujan, suhu, serta ketersediaan air tanah dan permukaan selama periode 1901–2020. Arsitektur model dirancang menggunakan lapisan LSTM dengan 50 neuron dan fungsi aktivasi ReLU guna menangkap ketergantungan temporal jangka panjang pada data deret waktu. Hasil penelitian menunjukkan bahwa model LSTM mencapai tingkat akurasi yang sangat tinggi, yakni 99%, dengan nilai RMSE masing-masing sebesar 13 untuk curah hujan, 15 untuk suhu, dan 20 untuk sumber daya air. Selain itu, model ini mampu memproyeksikan tren penurunan pasokan air di masa depan secara konsisten, sehingga dinilai sangat potensial sebagai alat pendukung perencanaan strategis dan pengelolaan sumber daya air yang berkelanjutan (Shah et al., 2024).

Deep Learning Rainfall–Runoff Predictions of Extreme Events dilakukan untuk menjawab keraguan mengenai keandalan model deep learning berbasis data dalam memprediksi kejadian hidrologi ekstrem, terutama jika dibandingkan dengan model berbasis proses fisik. Studi ini mengevaluasi kinerja Long Short-Term Memory (LSTM) dan variannya, Mass-Conserving LSTM (MC-LSTM), dengan membandingkannya terhadap model konseptual Sacramento Model dan model berbasis proses fisik US National Water Model menggunakan dataset CAMELS dari 498 daerah aliran sungai di Amerika Serikat. Hasil penelitian menunjukkan bahwa LSTM secara konsisten memberikan akurasi yang lebih tinggi dalam memprediksi kejadian ekstrem, bahkan ketika kejadian ekstrem tersebut tidak disertakan dalam data pelatihan. Menariknya, penerapan batasan fisik berupa konservasi massa pada MC-LSTM justru sedikit menurunkan kemampuan model dalam menangkap kejadian ekstrem dibandingkan LSTM standar, sehingga menegaskan keunggulan fleksibilitas LSTM murni dalam pemodelan dinamika hidrologi ekstrem. (Frame et al., 2022)

A Combined Monthly Precipitation Prediction Method Based on CEEMD and Improved LSTM dilaksanakan untuk mengatasi rendahnya akurasi prediksi curah hujan bulanan yang disebabkan oleh karakteristik data yang non-linear dan non-stasioner, terutama di tengah meningkatnya tekanan terhadap ketersediaan sumber daya air. Studi ini mengusulkan model hibrida CEEMD-PSO-LSTM, yang mengintegrasikan Complementary Ensemble Empirical Mode Decomposition (CEEMD) untuk memecah data curah hujan menjadi sub-sekuen yang lebih stabil, serta Long Short-Term Memory (LSTM) dengan parameter yang dioptimalkan menggunakan Particle Swarm Optimization (PSO). Pengujian menggunakan data curah hujan Kota Changde periode 1961–2020 menunjukkan bahwa model yang diusulkan menghasilkan akurasi dan stabilitas yang lebih tinggi dibandingkan model tunggal seperti ANN, SVM, dan BP, maupun model hibrida lain seperti EMD-LSTM dan EEMD-LSTM. Kinerja tersebut tercermin dari nilai RMSE dan MAE yang paling rendah, sehingga model ini dinilai efektif dalam menangkap tren perubahan curah hujan bulanan dan relevan untuk mendukung perencanaan manajemen sumber daya air yang lebih presisi. (Jiang, 2023).

Prediksi Curah Hujan di Kota Bandung Menggunakan Metode Long Short-Term Memory bertujuan menjawab tantangan tingginya variabilitas intensitas hujan di

Kota Bandung yang kerap berubah secara mendadak, sehingga dibutuhkan model prediksi yang lebih akurat untuk mendukung aktivitas masyarakat, khususnya kegiatan luar ruangan. Studi ini menerapkan pendekatan deep learning berbasis Long Short-Term Memory (LSTM) pada data curah hujan harian BMKG periode 2017–2021, dengan tahapan penelitian meliputi prapemrosesan data, normalisasi menggunakan MinMaxScaler, serta evaluasi kinerja model menggunakan metrik Root Mean Square Error (RMSE). Hasil pengujian menunjukkan bahwa model LSTM memberikan performa terbaik pada konfigurasi 50 epoch dan batch size 1, dengan nilai RMSE sebesar 12,24 pada data pelatihan dan 8,86 pada data pengujian, yang menandakan kemampuan LSTM dalam memodelkan pola temporal curah hujan harian secara cukup akurat .(Farikhul Firdaus & Paputungan, 2022)

Particle Swarm Optimized Deep Learning Models for Rainfall Prediction: A Case Study in Aizawl, Mizoram dilakukan untuk menjawab tingginya risiko tanah longsor di wilayah perbukitan Mizoram selama musim hujan, yang hingga kini masih minim didukung oleh studi prediksi curah hujan berbasis machine learning tingkat lanjut. Penelitian ini membandingkan dua belas model hibrida machine learning dan deep learning, termasuk berbagai varian ANN, LSTM, dan Bidirectional LSTM (BiLSTM), yang seluruh hyperparameternya dioptimalkan secara otomatis menggunakan algoritma Particle Swarm Optimization (PSO). Hasil evaluasi menunjukkan bahwa model PSO-BiLSTM-ANN II, yang mengombinasikan arsitektur BiLSTM, ANN, dan dropout layer, memberikan performa paling unggul dengan tingkat akurasi 6,4% lebih tinggi dibandingkan model PSO-SVR. Selain itu, model ini juga terbukti lebih efisien karena mampu mencapai konvergensi pada jumlah epoch yang rendah serta membutuhkan jumlah neuron tersembunyi yang lebih sedikit, sehingga berpotensi menjadi pendekatan standar untuk prediksi curah hujan di wilayah rawan bencana tersebut (Zoremsanga & Hussain, 2024)

Rainfall and runoff time-series trend analysis using LSTM recurrent neural network and wavelet neural network with satellite-based meteorological data: case study of Nzoia hydrologic basin dilakukan untuk mengatasi keterbatasan jaringan pemantauan meteorologi dan hidrologi di daerah aliran sungai yang menyebabkan tingginya ketidakpastian dalam prediksi respons hidrologis. Studi ini membandingkan

kinerja model Long Short-Term Memory (LSTM) dan Wavelet Neural Network (WNN) dalam memprediksi tren curah hujan dan limpasan menggunakan data meteorologi berbasis satelit selama periode 1980–2009. Hasil penelitian menunjukkan bahwa kedua model mampu menangkap pola deret waktu dengan baik, namun LSTM memberikan performa yang lebih unggul, dengan nilai akurasi sebesar 0,8967 untuk prediksi limpasan dan 0,8610 untuk curah hujan. Temuan ini menegaskan bahwa pemanfaatan data satelit yang dipadukan dengan pendekatan deep learning, khususnya LSTM, merupakan solusi yang andal untuk pemodelan hidrologi di wilayah dengan keterbatasan data observasi lapangan (Ouma et al., 2022)

Short-Term Weather Forecasting Using Spatial Feature Attention Based LSTM Model mengkaji keterbatasan model prediksi cuaca konvensional yang kurang mampu menangkap ketergantungan jangka panjang dan interaksi antarvariabel meteorologi, sehingga performanya menurun saat terjadi perubahan pola cuaca secara tiba-tiba. Untuk mengatasi hal tersebut, penulis mengusulkan model Spatial Feature Attention Long Short-Term Memory (SFA-LSTM) berbasis arsitektur *encoder–decoder*, di mana lapisan LSTM pada tahap encoder digunakan untuk mempelajari dependensi temporal, sementara mekanisme *spatial feature attention* pada decoder berperan dalam menangkap hubungan timbal balik antar fitur cuaca seperti suhu, titik embun, dan tekanan. Hasil pengujian menunjukkan bahwa model SFA-LSTM mampu menjelaskan hingga 99% varians data dengan nilai MSE sebesar 0,0871 dan MAE sebesar 0,2317, serta secara konsisten mengungguli model baseline dalam memprediksi suhu, terutama pada kondisi fluktuasi cuaca yang signifikan, berkat kemampuannya menginterpretasikan kontribusi fitur spasial secara lebih akurat (Suleman & Shridevi, 2022).

Berdasarkan kajian terhadap berbagai penelitian terdahulu, dapat disimpulkan bahwa pendekatan *deep learning*, khususnya model Long Short-Term Memory (LSTM) dan variannya, telah menunjukkan kinerja yang baik dalam memodelkan dan memprediksi curah hujan serta variabel iklim yang bersifat nonlinier dan kompleks. Literatur yang ada menegaskan bahwa LSTM unggul dalam menangkap ketergantungan temporal jangka panjang dan pola musiman pada data runtun waktu, baik pada skala harian maupun bulanan, serta pada berbagai kondisi iklim dan wilayah. Selain itu, hasil penelitian terdahulu juga menunjukkan bahwa pemilihan arsitektur model, variabel input

meteorologis, dan strategi prapemrosesan data berperan penting dalam meningkatkan akurasi prediksi. Berangkat dari temuan-temuan tersebut, penelitian ini mengadopsi model LSTM untuk memprediksi curah hujan harian di Stasiun Meteorologi Pattimura, Kota Ambon, dengan memanfaatkan data historis dan beberapa variabel meteorologis pendukung, guna menghasilkan prediksi yang mampu merepresentasikan pola curah hujan lokal secara lebih akurat dan informatif.

2.2. Landasan Teori

2.2.1 Deep Learning

Deep Learning merupakan pendekatan pembelajaran mesin yang mempelajari representasi data secara hierarkis melalui arsitektur jaringan saraf tiruan dengan banyak lapisan tersembunyi. Setiap lapisan berperan dalam mengekstraksi fitur dari tingkat sederhana hingga kompleks, sehingga informasi yang diperoleh semakin abstrak pada lapisan berikutnya. Dengan mekanisme ini, Deep Learning mampu menangani data berdimensi tinggi dan berpola nonlinier secara lebih efektif dibandingkan metode pembelajaran mesin tradisional, terutama karena proses ekstraksi fitur dan pembelajaran dilakukan secara simultan dalam satu model (Shiri et al., 2024).

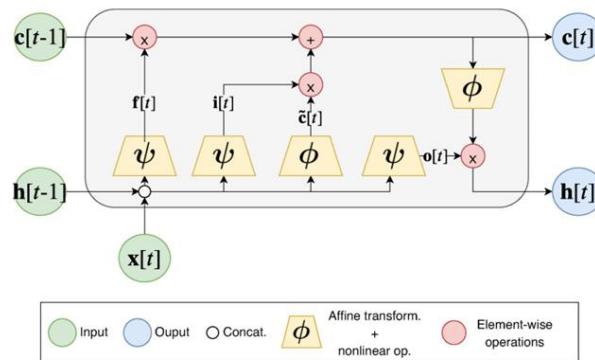
Perkembangan pesat Deep Learning tidak terlepas dari meningkatnya kapasitas komputasi, ketersediaan data dalam jumlah besar, serta kemajuan algoritma optimasi. (Mu & Zeng, 2019) dalam studi tinjauannya menjelaskan bahwa Deep Learning telah menjadi fondasi utama dalam berbagai bidang aplikasi, seperti pengolahan citra, pengenalan suara, pemrosesan bahasa alami, dan sistem prediksi berbasis data. Keunggulan utama pendekatan ini terletak pada kemampuannya melakukan pembelajaran fitur secara otomatis, sehingga mengurangi ketergantungan pada proses rekayasa fitur manual yang umumnya dibutuhkan pada metode pembelajaran mesin konvensional.

Deep Learning juga banyak diterapkan pada permasalahan data runtun waktu karena kemampuannya dalam memodelkan hubungan temporal yang kompleks. Dalam konteks peramalan, Deep Learning telah menunjukkan kinerja yang unggul pada berbagai aplikasi karena mampu memanfaatkan data historis untuk menangkap pola jangka pendek maupun jangka panjang. Berbagai arsitektur Deep Learning, seperti jaringan saraf berumpan maju dan jaringan saraf berulang, digunakan untuk meningkatkan akurasi prediksi pada data deret waktu yang bersifat nonlinier dan dinamis (Gasparin et al., 2022).

Hasil survei komprehensif oleh (Kong et al., 2025) menunjukkan bahwa pendekatan Deep Learning secara konsisten memberikan kinerja yang lebih unggul dibandingkan model statistik klasik dalam tugas peramalan time series, terutama pada domain meteorologi, energi, dan lingkungan. Keunggulan tersebut berasal dari kemampuan Deep Learning dalam mengekstraksi representasi temporal laten, memodelkan ketergantungan jangka panjang, serta menangani data multivariat dengan tingkat kompleksitas yang tinggi. Oleh karena itu, Deep Learning dipandang sebagai pendekatan yang relevan dan efektif untuk prediksi curah hujan, yang secara alami memiliki karakteristik nonlinier dan variabilitas temporal yang signifikan.

2.2.2 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) merupakan salah satu pengembangan dari Recurrent Neural Network (RNN) yang dirancang untuk memodelkan data runtun waktu dengan ketergantungan jangka panjang. RNN konvensional memiliki keterbatasan utama berupa *vanishing gradient problem*, yaitu kondisi ketika nilai gradien mengecil selama proses pelatihan sehingga informasi dari waktu sebelumnya sulit dipertahankan. Akibatnya, RNN kurang efektif dalam menangkap pola temporal yang panjang. Untuk mengatasi permasalahan tersebut, LSTM memperkenalkan mekanisme memori internal yang diatur melalui beberapa gerbang (gate), sehingga aliran informasi dapat dikendalikan secara lebih stabil dan selektif (Gasparin et al., 2022).



Gambar 2. 1 Arsitektur LSTM

Sumber : (Gasparin et al., 2022)

Mekanisme kerja arsitektur LSTM pada Gambar 2.1 dapat dijelaskan secara matematis melalui persamaan-persamaan berikut (Gasparin et al., 2022):

1) Forget Gate

Forget gate berperan dalam menentukan informasi dari cell state sebelumnya yang perlu dipertahankan atau dihapus. Gate ini menggunakan fungsi aktivasi sigmoid sehingga menghasilkan nilai antara 0 dan 1. Nilai yang mendekati 0 menunjukkan informasi yang dihilangkan, sedangkan nilai yang mendekati 1 menunjukkan informasi yang tetap disimpan.

Secara matematis, forget gate dirumuskan sebagai berikut:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (2.1)$$

Keterangan:

f_t : nilai *forget gate* pada waktu ke- t

h_{t-1} : *hidden state* pada waktu sebelumnya

x_t : input pada waktu ke- t

W_f : bobot *forget gate*

b_f : bias *forget gate*

σ : fungsi aktivasi sigmoid

2) *Input Gate dan Candidate Cell State*

Input gate berfungsi mengontrol seberapa besar informasi baru yang akan disimpan ke dalam *cell state*. Proses ini dilakukan melalui dua tahapan, yaitu perhitungan input gate dan pembentukan *candidate cell state*. Input gate menentukan tingkat kontribusi informasi baru, sedangkan *candidate cell state* merepresentasikan informasi baru yang telah dinormalisasi menggunakan fungsi aktivasi *tanh*.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (2.2)$$

Keterangan:

i_t : nilai *input gate*

h_{t-1} : *hidden state* pada waktu sebelumnya

x_t : input pada waktu ke- t

W_i : matriks bobot *input gate*

b_i : bias *input gate*

$\sigma(\cdot)$: fungsi aktivasi sigmoid

$$\tilde{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (2.3)$$

Keterangan:

$\tilde{c}_t$: kandidat *cell state*

h_{t-1} : *hidden state* pada waktu sebelumnya

x_t : input pada waktu ke- t

W_c : matriks bobot kandidat *cell state*

b_c : bias kandidat *cell state*

$\tanh(\cdot)$: fungsi aktivasi hiperbolik tangen

3) Pembaruan *Cell State*

Setelah nilai forget gate dan input gate diperoleh, *cell state* diperbarui dengan mengombinasikan informasi lama dan informasi baru. Proses ini memungkinkan LSTM untuk mempertahankan informasi yang relevan dalam jangka panjang.

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (2.5)$$

Keterangan:

c_t : *cell state* pada waktu ke- t

c_{t-1} : *cell state* pada waktu sebelumnya

f_t : nilai *forget gate*

i_t : nilai *input gate*

$\tilde{c}_t$: kandidat *cell state*

$\odot$: operasi perkalian elemen-per-elemen

4) *Output Gate* dan *Hidden State*

Output gate berfungsi menentukan nilai keluaran (*hidden state*) yang akan diteruskan ke langkah waktu berikutnya. Gate ini mengatur bagian dari *cell state* yang ditampilkan sebagai keluaran jaringan.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (2.6)$$

Keterangan:

o_t : nilai *output gate*

h_{t-1} : *hidden state* pada waktu sebelumnya

x_t : input pada waktu ke- t

W_o : matriks bobot *output gate*

b_o : bias *output gate*

$\sigma(\cdot)$: fungsi aktivasi sigmoid

$$h_t = o_t \odot \tanh(c_t) \quad (2.7)$$

Keterangan:

h_t : *hidden state* pada waktu ke- t

o_t : nilai *output gate*

c_t : *cell state* pada waktu ke- t

$\tanh(\cdot)$: fungsi aktivasi hiperbolik tangen

$\odot$: operasi perkalian elemen-per-elemen

2.2.2.1 Parameter Bobot dan Bias pada Inisialisasi LSTM

Penentuan parameter bobot W dan bias b dalam algoritma LSTM memiliki landasan ilmiah yang sangat krusial dalam menentukan efektivitas model saat mempelajari pola data. Dalam simulasi perhitungan manual pada penelitian ini, nilai bobot seperti $W_f = 0,5$ dan $W_i = 0,4$ ditetapkan sebagai nilai inisialisasi awal yang merujuk pada prinsip *symmetry breaking* sebagaimana diuraikan oleh (Heaton, 2018). Berdasarkan teori tersebut, inisialisasi bobot tidak boleh bernilai nol agar setiap neuron dalam jaringan saraf dapat mempelajari fitur yang berbeda secara independen. Pemilihan nilai skalar di rentang menengah ini bertujuan untuk menjaga agar hasil transformasi linear tetap berada pada area sensitif fungsi aktivasi, sehingga mencegah terjadinya masalah *vanishing gradient* atau hilangnya informasi saat proses perambatan balik dilakukan. Dengan demikian, penggunaan bobot ini berfungsi sebagai parameter awal yang memastikan stabilitas alur matematis pada setiap gerbang LSTM sebelum model mencapai kondisi optimal melalui pelatihan.

Di sisi lain, penetapan nilai bias sebesar 0,01 didasarkan pada temuan empiris (Jozefowicz et al., 2015). dalam studinya yang berjudul "*An Empirical Exploration of Recurrent Network Architectures*". Jozefowicz menyoroti bahwa pemberian bias

positif dalam skala kecil pada gerbang-gerbang LSTM—khususnya pada *forget gate*—sangat penting untuk menjaga stabilitas memori jangka panjang sejak awal iterasi. Secara teknis, bias berperan menggeser fungsi aktivasi guna memastikan gerbang memori tidak tertutup sepenuhnya saat model mulai bekerja. Kehadiran bias positif ini menjamin *forget gate* tetap mampu meneruskan informasi relevan dari masa lalu ke tahapan berikutnya, sehingga mekanisme *cell state* dapat beroperasi maksimal tanpa kehilangan ketergantungan data jangka panjang (*long-term dependencies*).

Penting untuk dipahami bahwa meskipun nilai bobot dan bias dalam hitungan manual ini bersifat statis demi kemudahan demonstrasi matematis, pada praktiknya nilai-nilai ini bersifat dinamis. Saat model dilatih menggunakan perangkat lunak, parameter tersebut akan dibangkitkan secara acak (*stochastic*) dan diperbarui secara berkala melalui mekanisme *Backpropagation Through Time* (BPTT). Melalui proses iteratif ini, bobot akan menyesuaikan diri secara otomatis demi meminimalkan tingkat kesalahan (*loss function*) hingga mencapai titik konvergensi. Oleh sebab itu, penggunaan nilai tetap dalam simulasi manual ini berperan sebagai titik mula (*starting point*) untuk menunjukkan bagaimana interaksi antar parameter dan data input menghasilkan prediksi yang akurat sebelum model mencapai performa optimalnya setelah pelatihan.

2.2.3 Evaluasi Model (MAE & RMSE)

Evaluasi model dilakukan untuk menilai tingkat akurasi hasil prediksi dengan cara membandingkan nilai keluaran model terhadap data observasi. Dalam penelitian prediksi curah hujan harian berbasis *Long Short-Term Memory* (LSTM), proses evaluasi menjadi tahap penting untuk memastikan bahwa model mampu merepresentasikan pola data secara kuantitatif. Metrik evaluasi yang digunakan pada penelitian ini adalah *Mean Absolute Error* (MAE) dan *Root Mean Square Error* (RMSE), karena keduanya umum digunakan dalam penelitian prediksi berbasis jaringan saraf dan data runtun waktu untuk mengukur besarnya kesalahan prediksi secara objektif (Haris et al., 2025).

2.2.3.1 *Mean Absolute Error (MAE)*

Mean Absolute Error (MAE) merupakan metrik evaluasi yang digunakan untuk mengukur rata-rata selisih absolut antara nilai prediksi dan nilai aktual. MAE memberikan gambaran langsung mengenai besarnya kesalahan prediksi tanpa

mempertimbangkan arah kesalahan. Karena tidak memberikan penalti berlebih terhadap kesalahan dengan nilai ekstrem, MAE dinilai stabil dan mudah diinterpretasikan, sehingga banyak digunakan dalam evaluasi model prediksi hidrometeorologi, termasuk curah hujan harian (Haris et al., 2025).

$$MAE = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{y}_i] \quad (2.8)$$

Keterangan:

n : jumlah data observasi

y_i : nilai aktual curah hujan

$\hat{y}_i$: nilai hasil prediksi model

2.2.3.2 Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) merupakan metrik evaluasi yang menghitung akar dari rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual. Berbeda dengan MAE, RMSE bersifat lebih sensitif terhadap kesalahan dengan nilai besar karena adanya proses pemangkatan, sehingga metrik ini efektif untuk mengidentifikasi kesalahan ekstrem pada data dengan fluktuasi tinggi, seperti data curah hujan harian (Haris et al., 2025).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.9)$$

keterangan:

n : jumlah data observasi

y_i : nilai aktual curah hujan

$\hat{y}_i$: nilai hasil prediksi model

Tabel 2. 1Klasifikasi Tingkat Akurasi Model Berdasarkan Nilai Kesalahan

Sumber : (Loukas et al., 2021)

Rentang	Nilai	MAE / RMSE	Kategori Akurasi	Interpretasi
< 5 mm			Sangat Baik	Model memiliki tingkat presisi yang sangat tinggi dalam memprediksi curah hujan harian.

5 mm – 10 mm	Baik	Model mampu memprediksi curah hujan dengan akurasi yang baik dan kesalahan relatif kecil.
10 mm -20 mm	Cukup/Sedang	Model cukup mampu menangkap pola umum curah hujan, namun masih memiliki selisih pada kondisi fluktuasi curah hujan yang besar.
> 20 mm	Kurang/Rendah	Tingkat kesalahan model tergolong tinggi dan menunjukkan ketidaksesuaian prediksi yang signifikan, setara dengan perbedaan satu kategori hujan.

Tabel 2.1 menyajikan klasifikasi tingkat akurasi model prediksi berdasarkan nilai kesalahan MAE dan RMSE. Klasifikasi ini digunakan sebagai acuan untuk menilai kinerja model dalam memprediksi curah hujan harian secara kuantitatif. Semakin kecil nilai MAE dan RMSE yang dihasilkan, maka semakin tinggi tingkat akurasi model, sedangkan nilai kesalahan yang besar menunjukkan kemampuan model yang rendah dalam merepresentasikan variabilitas curah hujan.

2.2.4 Curah Hujan

Curah hujan merupakan salah satu unsur meteorologi yang memiliki peranan penting dalam menggambarkan kondisi iklim dan cuaca suatu wilayah. Curah hujan merujuk pada jumlah air hujan yang jatuh ke permukaan bumi dalam periode waktu tertentu dan umumnya dinyatakan dalam satuan milimeter (mm). Intensitas dan pola curah hujan yang tinggi dapat berdampak langsung terhadap lingkungan, khususnya dalam meningkatkan risiko bencana hidrometeorologi seperti banjir (Wijayanto et al., 2024).

Penggunaan satuan milimeter (mm) pada curah hujan merujuk pada tinggi atau ketebalan lapisan air hujan yang menutupi suatu permukaan datar dan kedap air dalam periode waktu tertentu. Secara sederhana, apabila air hujan yang jatuh pada suatu

wilayah dikumpulkan dan tidak mengalami peresapan maupun penguapan, maka curah hujan sebesar 1 mm berarti air tersebut memiliki tinggi setebal 1 milimeter di atas permukaan tersebut.

Satuan ini digunakan karena mampu merepresentasikan volume air hujan secara proporsional terhadap luas permukaan. Secara matematis, 1 mm curah hujan setara dengan 1 liter air per meter persegi (1 L/m^2). Dengan demikian, satuan milimeter tidak hanya menunjukkan tinggi air, tetapi juga mencerminkan jumlah volume air hujan yang jatuh pada suatu area tertentu. Penggunaan satuan ini telah menjadi standar internasional dalam pengukuran curah hujan karena mudah dipahami, konsisten, dan relevan dalam analisis hidrologi serta perencanaan sumber daya air.

Perubahan iklim dan ketidakstabilan pola cuaca menyebabkan distribusi curah hujan menjadi semakin tidak teratur, termasuk meningkatnya kejadian hujan ekstrem. Kondisi tersebut menyulitkan proses pengendalian banjir serta perencanaan infrastruktur sumber daya air. Oleh karena itu, pemahaman terhadap karakteristik curah hujan menjadi aspek penting dalam manajemen sumber daya air dan perancangan infrastruktur yang tangguh terhadap kejadian cuaca ekstrem (Wijayanto et al., 2024)

Curah hujan memiliki karakteristik yang bersifat dinamis dan bervariasi baik secara temporal maupun spasial. Variasi temporal menunjukkan perubahan curah hujan dari waktu ke waktu, sedangkan variasi spasial menggambarkan perbedaan distribusi curah hujan antarwilayah. Kompleksitas ini menjadikan curah hujan sebagai parameter yang menantang untuk dianalisis dan diprediksi, terutama dalam konteks peramalan jangka panjang (Wijayanto et al., 2024).

Selain dianalisis berdasarkan pola dan distribusinya, curah hujan juga dapat diklasifikasikan berdasarkan tingkat intensitasnya. Klasifikasi intensitas curah hujan digunakan untuk mengidentifikasi kondisi hujan yang berpotensi menimbulkan dampak tertentu, seperti genangan atau banjir. Pengelompokan ini menjadi acuan penting dalam analisis klimatologi serta dalam mendukung proses pengambilan keputusan terkait mitigasi risiko bencana (Wijayanto et al., 2024).

Tabel 2. 2 Kategori Intensitas Curah hujan

Sumber : (Wijayanto et al., 2024)

Keadaan curah hujan	Intensitas curah hujan (mm)	Potensi banjir
Hujan sangat ringan	< 5 (24jam)	Tidak
Hujan ringan	5 – 20 (24 jam)	Tidak
Hujan normal	20 – 50 (24 jam)	Tidak
Hujan lebat	50 – 100 (24 jam)	Ya
Hujan sangat lebat	> 100 (24 jam)	Ya

Dalam pengembangan arsitektur prediksi ini, **RR (*Rainfall*)** ditetapkan sebagai variabel target atau luaran (*output*) utama. Variabel RR ini merepresentasikan akumulasi total curah hujan harian yang terukur dalam rentang waktu 24 jam. Merujuk pada studi yang dilakukan oleh (Supriatna et al., 2025), pemilihan RR sebagai target prediksi bertujuan untuk memperoleh estimasi kuantitatif yang presisi mengenai intensitas hujan di waktu mendatang dengan mengeksplorasi pola-pola tersembunyi dari data historis.

Penting untuk dicatat bahwa dalam pemodelan *time-series*, nilai RR tidak dipandang sebagai variabel statistik yang berdiri sendiri. Sebaliknya, RR merupakan hasil akhir dari interaksi atmosferik yang kompleks, yang dipengaruhi oleh berbagai faktor meteorologi pendukung seperti suhu dan tingkat kelembapan udara. Dengan memfokuskan RR sebagai target utama, model LSTM diharapkan memiliki sensitivitas yang tinggi dalam mengenali ambang batas intensitas hujan—mulai dari kategori ringan hingga ekstrem (sangat lebat). Ketajaman prediksi ini menjadi sangat krusial agar hasilnya dapat dijadikan rujukan tepercaya dalam sistem peringatan dini serta perumusan kebijakan mitigasi bencana, khususnya di wilayah kerja Stasiun Meteorologi Pattimura.

2.2.4.1 Variabel Meteorologi Pendukung

Akurasi dalam memprediksi curah hujan sangat bergantung pada keterlibatan berbagai variabel meteorologi sebagai fitur masukan (*input features*). Hal ini dikarenakan curah hujan merupakan hasil dari interaksi dinamis di atmosfer. Penelitian oleh (Farisi et al., 2024), menegaskan bahwa mengintegrasikan beberapa variabel cuaca sekaligus ke dalam model *deep learning* mampu menghasilkan prediksi yang lebih stabil dan presisi jika dibandingkan dengan penggunaan variabel tunggal.

Dalam penelitian ini, variabel meteorologi yang dianalisis merujuk pada standar parameter pengamatan BMKG di Stasiun Meteorologi Pattimura, yang meliputi:

1. Suhu Udara (TAVG, TN, TX): Mencakup temperatur rata-rata, minimum, dan maksimum harian. Suhu menjadi faktor krusial yang mengatur kapasitas atmosfer dalam menyimpan uap air. Selain itu, suhu juga memicu proses konveksi yang menjadi penggerak utama terbentuknya awan hujan (Supriatna et al., 2025).
2. Kelembapan Udara (RH_AVG): Parameter ini menggambarkan tingkat kejenuhan uap air di udara. Angka kelembapan yang tinggi menjadi sinyal kuat bahwa massa udara sudah cukup jenuh untuk memulai proses kondensasi.
3. Penyinaran Matahari (SS): Durasi matahari bersinar memengaruhi seberapa cepat air di permukaan menguap (*evaporasi*). Bagi wilayah kepulauan seperti Ambon, intensitas penguapan ini sangat vital karena menjadi penyedia utama "bahan baku" uap air di langit.
4. Kecepatan Angin (FF_AVG): Angin berperan dalam mendistribusikan massa udara lembap. Menurut (Chortimah & Norman, 2026), karakteristik wilayah kepulauan seperti Maluku memiliki variabilitas curah hujan yang sangat tinggi, di mana dinamika atmosfer lokal dan pergerakan massa udara sangat menentukan intensitas hujan harian.

2.2.5 Prediksi (*Forecasting*)

Forecasting atau peramalan merupakan proses untuk memperkirakan kondisi atau kebutuhan di masa mendatang dengan memanfaatkan pola data pada periode sebelumnya. Peramalan berperan sebagai dasar dalam perencanaan dan pengambilan keputusan karena mampu memberikan gambaran proyeksi secara sistematis dan terukur. Dengan menggunakan data historis yang relevan, forecasting membantu dalam mengantisipasi perubahan yang mungkin terjadi sehingga proses perencanaan dapat dilakukan secara lebih efektif dan efisien (Ningrum et al., 2014)

Forecasting juga dapat dipahami sebagai pendekatan analitis yang digunakan untuk memperkirakan nilai di masa depan, baik secara kuantitatif maupun kualitatif, dengan menjadikan data masa lalu sebagai dasar perhitungan. Dalam analisis data, forecasting sering diterapkan untuk mengidentifikasi kecenderungan (*trend*) dan pola perubahan data dari waktu ke waktu. Pendekatan ini memungkinkan pemahaman yang

lebih baik terhadap dinamika data serta mendukung pengambilan keputusan berbasis data pada berbagai bidang, termasuk sistem informasi dan analisis data historis (Titik et al., n.d.).

2.2.6 Data Runtun Waktu (*Time Series*)

Data runtun waktu (*time series*) merupakan sekumpulan data yang dicatat secara berurutan berdasarkan interval waktu tertentu, seperti harian, bulanan, atau tahunan. Karakteristik utama data ini adalah adanya ketergantungan temporal, yaitu kondisi ketika nilai pada suatu waktu dipengaruhi oleh nilai pada periode sebelumnya. Ketergantungan tersebut membentuk pola tertentu, seperti tren, musiman, dan fluktuasi jangka panjang, sehingga data time series tidak bersifat acak maupun independen antarobservasi (Shah et al., 2024).

Dalam kajian analisis dan peramalan, data time series memiliki struktur internal yang terdiri atas beberapa komponen utama, yaitu tren, musiman, dan residual. Komponen-komponen tersebut merepresentasikan dinamika data dari waktu ke waktu dan menjadi dasar dalam memahami perilaku data historis. Pemodelan yang tidak mempertimbangkan struktur temporal ini berpotensi mengabaikan informasi penting yang berpengaruh terhadap akurasi prediksi (Mendis et al., 2024).

Curah hujan termasuk dalam kategori data time series karena pengamatannya dilakukan secara periodik dan menunjukkan keterkaitan antarwaktu yang kuat. Nilai curah hujan pada suatu periode dipengaruhi oleh kondisi atmosfer sebelumnya, seperti suhu, kelembapan, tekanan udara, dan pola angin. Selain itu, data curah hujan umumnya bersifat nonlinier dan menunjukkan variasi temporal yang kompleks, sehingga membutuhkan pendekatan pemodelan yang mampu menangkap dinamika tersebut secara efektif (Barrera-Animas et al., 2022).

Perlakuan data curah hujan sebagai data potong lintang (*cross-sectional data*) tidak tepat karena mengabaikan hubungan temporal antarobservasi. Pendekatan tersebut dapat menghilangkan informasi historis yang penting dalam memahami perubahan curah hujan dari waktu ke waktu. Oleh karena itu, diperlukan model yang dirancang khusus untuk data sekuensial, seperti Recurrent Neural Network (RNN) dan Long Short-Term Memory (LSTM), yang mampu menangkap ketergantungan jangka pendek maupun jangka panjang dalam data time series meteorologi (Shah et al., 2024).

2.2.6.1 Representasi Fitur Temporal Siklik

Data runtun waktu (time series) dalam bidang meteorologi, seperti curah hujan, umumnya menunjukkan pola musiman yang bersifat siklik, yaitu pola yang berulang secara periodik pada interval waktu tertentu, misalnya harian, bulanan, atau tahunan. Pola ini terbentuk akibat dinamika atmosfer dan sistem iklim yang memiliki kecenderungan berulang dari tahun ke tahun. Oleh karena itu, cara merepresentasikan informasi waktu menjadi komponen penting dalam pemodelan data runtun waktu agar pola musiman tersebut dapat dipelajari secara efektif oleh model prediksi.

Dalam praktik analisis data, fitur waktu seperti *day of year* (DOY), bulan, atau jam sering kali direpresentasikan dalam bentuk nilai numerik linier. Pendekatan ini relatif sederhana, namun memiliki keterbatasan karena tidak mencerminkan sifat periodik waktu. Sebagai contoh, hari ke-365 dan hari ke-1 secara numerik memiliki jarak yang jauh, padahal secara temporal keduanya berada pada posisi yang berdekatan dalam satu siklus tahunan. Kondisi ini menimbulkan *artificial discontinuity* pada batas periode, yang berpotensi menghambat model dalam mempelajari hubungan temporal secara konsisten.

Untuk mengatasi permasalahan tersebut, digunakan pendekatan representasi fitur temporal secara siklik (*cyclical representation*). Pendekatan ini memetakan fitur waktu ke dalam ruang kontinu dengan tetap mempertahankan hubungan periodik antarwaktu. Salah satu teknik yang paling umum digunakan adalah transformasi sinus dan cosinus, yang dikenal sebagai *sinusoidal encoding*, khususnya untuk fitur waktu yang memiliki sifat periodik.

Secara matematis, apabila suatu fitur waktu dilambangkan sebagai t dengan periode P , maka representasi sikliknya dapat dinyatakan sebagai berikut:

$$x_{\sin}(t) = \sin\left(2\pi \frac{t}{P}\right) \quad (2.10)$$

Dengan

keterangan:

t : adalah nilai fitur waktu (misalnya hari ke- t dalam satu tahun),
 P : adalah panjang periode (misalnya 365 hari untuk satu siklus tahunan),
 $x_{\sin}(t)$ dan $x_{\cos}(t)$: merupakan representasi siklik fitur waktu dalam bentuk fungsi sinus dan cosinus.

Transformasi sinus dan cosinus tersebut memetakan nilai waktu ke dalam ruang dua dimensi yang bersifat melingkar. Dengan representasi ini, nilai waktu pada awal dan akhir periode berada pada posisi yang berdekatan secara geometris. Akibatnya, transisi antarperiode terjadi secara halus dan kontinu, tanpa lonjakan nilai yang tidak mencerminkan hubungan temporal yang sebenarnya.

Secara konseptual, kedekatan antara dua titik waktu t_1 dan t_2 dalam representasi sinusoidal dapat dipahami melalui jarak geometris pada ruang fitur dua dimensi. Representasi ini memastikan bahwa waktu-waktu yang berdekatan secara temporal juga memiliki kedekatan dalam ruang fitur, meskipun nilai numeriknya tampak berjauhan. Keunggulan ini menjadikan representasi siklik lebih efektif dibandingkan representasi waktu linier atau ordinal.

Penelitian yang dilakukan oleh (Bansal et al., 2025) menunjukkan bahwa penerapan *sinusoidal encoding* mampu menangkap pola temporal siklik secara lebih baik dibandingkan metode representasi waktu konvensional. Hasil penelitian tersebut memperlihatkan peningkatan akurasi prediksi serta penurunan kesalahan yang disebabkan oleh diskontinuitas pada batas periode waktu (Bansal et al., 2025).

2.2.7 Python



Gambar 2. 2 Python

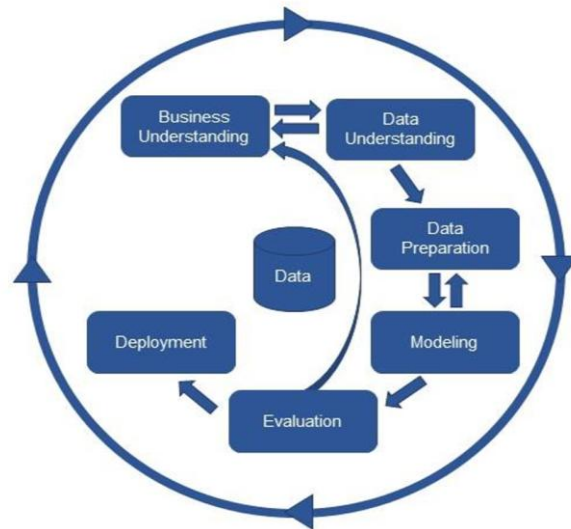
Sumber : <https://sl.bing.net/g4YRLTbqGWW>

Python adalah bahasa pemrograman tingkat tinggi yang banyak dipakai dalam pengembangan data science dan machine learning. Bahasa ini mendukung berbagai struktur data, menerapkan sistem tipe dinamis, memiliki dukungan manajemen memori, serta tersedia di banyak sistem operasi melalui interpreter, sehingga memudahkan pengembangan lintas platform (Arora, 2024)

Dalam konteks data science, Python sering dipilih karena mampu mengakomodasi alur kerja lengkap mulai dari pengolahan data, analisis, visualisasi, hingga pembuatan model melalui ekosistem library yang sangat luas. Pada praktiknya, pustaka seperti Scikit-learn, Keras, dan TensorFlow umum digunakan baik untuk pembelajaran maupun implementasi machine learning berbasis data (Arora, 2024).

Pada penelitian ini, beberapa library utama digunakan untuk mendukung proses komputasi, pengelolaan data, pemodelan, hingga evaluasi. NumPy digunakan untuk komputasi numerik berbasis array/matriks dan memanfaatkan vectorization agar perhitungan lebih efisien dibandingkan perulangan eksplisit. Pandas dimanfaatkan untuk menangani data tabular serta mendukung proses manipulasi, transformasi, dan pembersihan data. TensorFlow/Keras digunakan untuk membangun dan melatih model deep learning, termasuk penyusunan arsitektur seperti LSTM. Scikit-learn digunakan pada tahap preprocessing dan evaluasi, misalnya normalisasi (MinMaxScaler) serta perhitungan metrik kesalahan seperti MAE dan MSE (Arora, 2024).

2.2.8 Cross Industry Standard Process for Data Mining



Gambar 2. 3 Disain penelitian CRISP-DM
Sumber : (Rifai et al., 2019)

CRISP-DM (*Cross-Industry Standard Process for Data Mining*) merupakan metodologi standar yang dikembangkan untuk memberikan kerangka kerja sistematis dalam pelaksanaan proyek data mining. Metodologi ini dirancang agar dapat diterapkan pada berbagai bidang, baik industri maupun penelitian, sehingga proses pengolahan dan

analisis data dapat dilakukan secara terstruktur. CRISP-DM menekankan pentingnya pemahaman tujuan sejak tahap awal serta keterkaitan antar tahapan dalam suatu siklus data mining yang bersifat iteratif dan adaptif terhadap kebutuhan proyek(Rifai et al., 2019)

CRISP-DM memandang proses data mining sebagai suatu siklus yang terdiri atas beberapa fase yang saling berkaitan. Setiap fase tidak berdiri sendiri, melainkan memungkinkan adanya umpan balik dan pengulangan tahap apabila hasil yang diperoleh belum sesuai dengan tujuan yang ditetapkan. Pendekatan siklus ini memberikan fleksibilitas bagi peneliti untuk melakukan penyesuaian terhadap data, metode, maupun model yang dikembangkan berdasarkan hasil evaluasi yang dilakukan pada setiap tahap.(Karmila & Rezki, 2020)

Dalam praktiknya, CRISP-DM banyak digunakan sebagai pendekatan sistematis untuk meningkatkan kualitas hasil analisis data. Dengan mengikuti tahapan yang telah ditetapkan, proses data mining menjadi lebih terkontrol dan terarah, mulai dari pemahaman permasalahan hingga penerapan hasil analisis. Pendekatan ini membantu memastikan bahwa hasil data mining tidak hanya akurat secara teknis, tetapi juga relevan dan dapat dimanfaatkan sesuai dengan tujuan awal penelitian atau kebutuhan organisasi(Siswipraptini et al., 2023).

CRISP-DM terdiri atas enam tahapan utama yang membentuk satu siklus proses data mining, yaitu sebagai berikut:

1) Business Understanding

Tahap awal yang berfokus pada pemahaman tujuan bisnis atau tujuan penelitian. Pada tahap ini dilakukan identifikasi permasalahan, penentuan sasaran, serta perumusan kebutuhan yang ingin dicapai melalui proses data mining.

2) Data Understanding

Tahap ini bertujuan untuk memahami data yang digunakan, meliputi proses pengumpulan data awal, deskripsi data, eksplorasi data, serta identifikasi kualitas data guna mengetahui potensi permasalahan yang mungkin muncul.

3) Data Preparation

Pada tahap ini, data yang telah dipahami diproses agar siap digunakan dalam pemodelan. Proses yang dilakukan mencakup seleksi data, pembersihan data, integrasi data, serta transformasi data sesuai kebutuhan analisis.

4) Modeling

Tahap modeling melibatkan pemilihan dan penerapan teknik atau algoritma data mining yang sesuai dengan karakteristik data dan tujuan penelitian. Proses ini dapat dilakukan secara berulang untuk memperoleh model dengan kinerja terbaik.

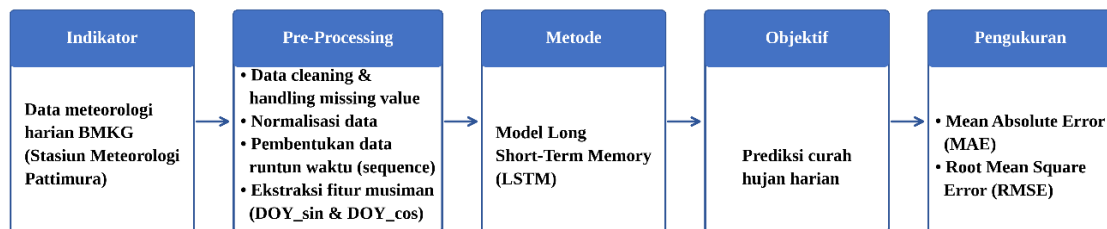
5) Evaluation

Tahap evaluasi bertujuan untuk menilai kesesuaian model yang dihasilkan dengan tujuan yang telah ditetapkan pada tahap awal. Evaluasi dilakukan tidak hanya berdasarkan performa teknis model, tetapi juga dari sisi relevansi hasil terhadap kebutuhan penelitian atau bisnis.

6) Deployment

Tahap akhir berupa penerapan hasil data mining. Pada tahap ini, pengetahuan atau model yang dihasilkan dapat diimplementasikan dalam bentuk sistem, laporan, atau rekomendasi yang dapat digunakan oleh pihak terkait.

2.2. Kerangka Pemikiran



Gambar 2. 4 kerangka pemikiran

proses kerangka penelitian yang akan di lakukan penulis dalam penelitian ini digambarkan di atas ,dan berikut penejalsnaya :

1. Indikator

Penelitian ini diawali dengan pengumpulan data meteorologi harian yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Stasiun

Meteorologi Pattimura. Data yang digunakan meliputi curah hujan harian sebagai variabel utama, serta beberapa variabel meteorologi pendukung yang berkontribusi terhadap pembentukan pola curah hujan. Seluruh data tersebut menjadi dasar dalam proses pemodelan dan peramalan curah hujan harian.

2. Pre-Processing

Tahap *pre-processing* dilakukan untuk memastikan data berada dalam kondisi yang layak digunakan sebagai input model. Proses ini mencakup pembersihan data dari nilai hilang (*missing value*), normalisasi untuk menyeragamkan skala antarvariabel, serta pembentukan data runtun waktu agar sesuai dengan kebutuhan model LSTM. Selain itu, ditambahkan fitur musiman berbasis *day of year* yang direpresentasikan melalui fungsi sinus dan cosinus (DOY_sin dan DOY_cos) guna membantu model mengenali pola periodik tahunan.

3. Metode

Pada tahap pemodelan, digunakan metode Long Short-Term Memory (LSTM). LSTM dipilih karena kemampuannya dalam menangkap ketergantungan jangka pendek dan jangka panjang pada data runtun waktu. Model ini dilatih menggunakan data historis curah hujan dan variabel pendukung untuk mempelajari pola temporal, sehingga mampu menghasilkan prediksi curah hujan harian secara berurutan.

4. Objektif

Tujuan utama penelitian ini adalah menghasilkan prediksi curah hujan harian berdasarkan data meteorologi historis. Prediksi yang dihasilkan diharapkan dapat merepresentasikan fluktuasi curah hujan secara realistis dan menjadi dasar dalam analisis pola curah hujan pada periode waktu berikutnya.

5. Pengukuran

Tahap akhir penelitian adalah evaluasi kinerja model. Pengukuran dilakukan menggunakan dua metrik evaluasi, yaitu Mean Absolute Error (MAE) dan Root Mean

Square Error (RMSE). MAE digunakan untuk menggambarkan rata-rata kesalahan absolut antara nilai aktual dan prediksi, sedangkan RMSE memberikan penalti yang lebih besar terhadap kesalahan berukuran besar. Kedua metrik ini digunakan untuk menilai tingkat akurasi dan keandalan model LSTM dalam memprediksi curah hujan harian.