

## Prediksi Curah Hujan di Kota Bandung Menggunakan Metode *Long Short Term Memory*

Riza Farikhul Firdaus<sup>\*1</sup>, Irving Vitra Papatungan<sup>2</sup>

<sup>1,2</sup>Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia, Indonesia  
Email: <sup>1</sup>rizaff16@gmail.com, <sup>2</sup>irving@uii.ac.id

### Abstrak

Perkiraan curah hujan adalah salah satu hal yang sangat penting untuk menunjang kegiatan diberbagai bidang yang bisa dijadikan acuan agar bisa menjalankan aktivitas tanpa terganggu keadaan cuaca yang tidak bersahabat. Hal tersebut disebabkan karena adanya faktor yang mempengaruhi peningkatan dan penurunan curah hujan. Oleh karena itu, memprediksi curah hujan sangat penting untuk berbagai kalangan terutama yang beraktivitas di luar ruangan. Penelitian ini menggunakan data curah hujan Kota Bandung yang diambil dari website Badan Meteorologi, Klimatologi dan Geofisika (BMKG) dengan metode *Long Short Term Memory* (LSTM). *Epoch* dan *batch size* merupakan parameter yang dapat mempengaruhi hasil dari metode LSTM. Alur dari metode LSTM yaitu pengumpulan data, preprocessing data, normalisasi data, *Long Short Term Memory* network, denormalisasi, evaluasi. Akurasi tertinggi diperoleh menggunakan *epoch* 50 serta *batch size* 1 dan nilai RMSE terbaik yaitu *Train Score* 12.24 dan *Test Score* 8.86.

**Kata kunci:** *Curah Hujan, Long Short Term Memory, Prediksi*

### Abstract

*Estimated rainfall is one of the most important things to support activities in various fields that can be used as a reference so that they can carry out activities without being disturbed by unfriendly weather conditions. This is due to the factors that influence the increase and decrease in rainfall. Therefore, predicting rainfall is very important for various groups, especially those who have outdoor activities. This study uses Bandung City rainfall data taken from the website of the Meteorology, Climatology and Geophysics Agency (BMKG) using the Long Short Term Memory (LSTM) method. Epoch and batch size are parameters that can affect the results of the LSTM method. The flow of the LSTM method is data collection, data preprocessing, data normalization, Long Short Term Memory network, denormalization, evaluation. The highest accuracy was obtained using epoch 100 and batch size 1 and the best RMSE values were Train Score 12.24 and Test Score 8.86.*

**Keywords:** *Batch size, Epoch, Long Short Term Memory*

## 1. PENDAHULUAN

Cuaca merupakan kondisi udara yang diamati dalam periode yang relatif singkat atau pendek pada teritori yang terbatas. *World Climate Conference* menyebutkan bahwa cuaca merupakan keadaan atmosfer yang diukur dengan kompleks meliputi perubahan, perkembangan, dan muncul atau hilangnya suatu fenomena udara (Luthfiarta et al., 2020). Unsur-unsur pembentuk cuaca yaitu suhu, udara, tekanan udara, kelembaban udara, laju uap air, awan, curah hujan dan angin. Curah hujan itu sendiri merupakan salah satu komponen iklim yang penting dari ekosistem alam, Indonesia merupakan negara dengan iklim tropis karena terletak di garis khatulistiwa (Rachmawati, 2021).

Ada dua musim di iklim tropis di antaranya musim kemarau dan musim hujan. Variasi regional ada dalam ketajaman iklim. Misalnya, pada musim hujan, daerah yang berbeda mengalami tingkat curah hujan yang berbeda berdasarkan garis lintang, ketinggian, kedekatan dengan sumber air, arah angin, suhu tanah, dan luas lahan. Cuaca di Indonesia tidak selalu mengikuti musim atau berjalan dengan baik karena beberapa faktor yang dapat mengubah curah hujan. Sebaliknya, sering dan tiba-tiba berubah. Sama halnya dengan kota Bandung, intensitas curah hujan tidak dapat diprediksi karena hujan turun setiap hari dalam seminggu dengan derajat yang bervariasi dan terkadang tidak ada hujan sama sekali

(Soekendro, 2021). Oleh karena itu, berbagai kalangan dapat memanfaatkan prakiraan cuaca sebagai acuan dalam menjalankan operasionalnya sehari-hari. Prakiraan cuaca sangat dibutuhkan dan sangat penting oleh banyak pihak (Rizki et al., 2020).

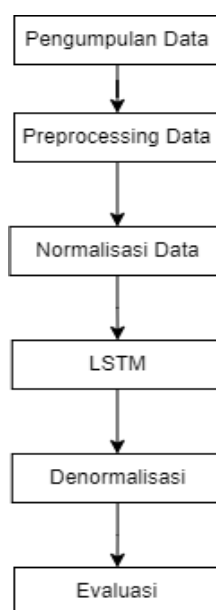
Informasi kondisi cuaca saat ini sudah dapat diketahui melalui BMKG (Badan Meteorologi Klimatologi dan Geofisika) dari setiap provinsi yang ada di Indonesia. BMKG Kota Bandung sendiri, sudah melakukan prakiraan cuaca dengan menggunakan banyak sumber data dan model analisis cuaca, baik yang dibuat oleh BMKG Pusat sendiri, atau melalui webnya BOM Australia (<http://www.bom.gov.au/>), Arpege Perancis dan KMA Korea (<http://www.kma.go.kr/eng/index.jsp>) yang dikombinasikan dengan kondisi beberapa parameter cuaca yang ada di BMKG Kota Bandung. Akan tetapi, akurasi dari ketepatan prediksi masih dirasa kurang oleh pakar analisa BMKG sendiri (Insani et al., 2020). Untuk itu, pemilihan metode yang tepat dalam penentuan cuaca terus dilakukan agar menemukan metode yang paling tepat dengan tingkat akurasi yang tinggi untuk menentukan prediksi cuaca (Insani et al., 2020).

Terdapat penelitian tentang prediksi di berbagai bidang akhir-akhir ini yang telah dilakukan menggunakan *deep learning* mampu menghasilkan akurasi di atas 85% seperti (*Klasifikasi Cuaca Menggunakan Metode Convolutional Neural Network (CNN)*). Jika dibandingkan dengan metode *machine learning* yang masih tradisional, *deep learning* tidak membutuhkan campur tangan manusia, tetapi masih membutuhkan data yang cukup besar. Pendekatan lain yang dilakukan yaitu mengubah kata menjadi vektor yang berisikan angka dengan *word embedding*. Kemudian algoritma dilatih dengan nilai dari suatu kata atau kalimat. Metode *deep learning* mengekstraksi dari *neural network* dan kemudian belajar dari *error*. *Neural network* biasanya terdiri dari beberapa lapisan yang dipetakan oleh fungsi aktivasi. Model *deep learning* yang biasa digunakan untuk menganalisis data berupa teks adalah CNN, RNN, LSTM, dan Gated Recurrent Unit (GRU) (Zhang et al., 2018).

*Long Short Term Memory (LSTM)* mempunyai akurasi yang baik. Oleh sebab itu, peneliti menggunakan teknik ini untuk menilai apakah prediksi curah hujan metode *Long Short Term Memory (LSTM)* lebih akurat daripada algoritma prediksi yang digunakan sebelumnya dalam penelitian. Penelitian ini dilakukan menggunakan data curah hujan Kota Bandung yang diambil dari website BMKG dengan menjadikan RMSE sebagai evaluasi dari prediksi yang dihasilkan.

## 2. METODE PENELITIAN

Metode yang digunakan penulis dalam penelitian ini adalah *Long Short Term Memory*. Di bawah ini merupakan tahapan penelitian dan penjelasannya.



Gambar 1. Tahapan Penelitian

## 2.1. Pengumpulan Data

Pada tahap ini, data akan dikumpulkan sebelum melanjutkan ke proses selanjutnya. Data yang dibutuhkan yaitu pola curah hujan dari tahun 2017-2021 di Kota Bandung yang didapatkan dari website BMKG ([http://dataonline.bmkg.go.id/dashboard\\_user](http://dataonline.bmkg.go.id/dashboard_user)).

## 2.2. Preprocessing Data

Menyiapkan data yang akan digunakan untuk pelatihan data (data *training*) dikenal sebagai data *preprocessing*. Ini melibatkan persiapan data terlebih dahulu untuk menangani data yang hilang atau kosong dalam berbagai metode, seperti menghitung rata-rata atribut untuk kelas yang sama. Data kemudian dinormalisasi menggunakan *MinMaxScaler* dengan rentang (0, 1).

## 2.3. Normalisasi Data

Normalisasi data adalah tahapan yang digunakan untuk menghilangkan data tidak terstruktur dan redundansi untuk memastikan penyimpanan data logis.

## 2.4. Long Short Term Memory Network

*Long Short Term Memory* (LSTM) adalah sistem penyimpanan data yang dapat memproses, memprediksi dan mengklasifikasikan informasi yang telah disimpan dalam jangka waktu yang lama.

## 2.5. Denormalisasi

Setelah mendapatkan hasil prediksi, maka sebelum menghitung akurasi hasil prediksi harus dilakukan denormalisasi yaitu data diubah menjadi nilai real. Dikarenakan data hasil prediksi masih berupa data berbentuk range interval yang dilakukan pada normalisasi data.

## 2.6. Evaluasi

Cara lain untuk menilai keakuratan keluaran prediksi model adalah dengan melihat *Root Mean Square Error* (RMSE) dari metodologi prediksi. Rerata kuadrat jumlah error pada model prediksi menghasilkan nilai RMSE. *Root Mean Square Error* (RMSE) adalah metode sederhana yang sering diterapkan dalam berbagai investigasi prediksi.

## 3. HASIL DAN PEMBAHASAN

Sebelum dilakukan pengujian untuk mengukur nilai kesalahan prediksi, diperlukan beberapa bahan untuk melakukan penelitian ini seperti kebutuhan *hardware* dan *software* yang digunakan dalam pengujian. Kebutuhan tersebut di antaranya:

### 3.1. Kebutuhan Hardware

- Laptop ASUS S430UN
- RAM 12 GB
- Processor Intel Core I5 Gen 8

### 3.2. Kebutuhan Software

- Sistem Operasi Windows 10
- Python
- Text Editor (Jupyter Notebook)

### 3.3. Transform Raw Dataset

Tahap ini merupakan tahap mentransformasikan raw dataset atau kumpulan dataset mentah untuk dilakukan pemrosesan data. Gambar 2 merupakan hasil dari transformasi dataset, dataset yang digunakan mulai dari tahun 2017 sampai dengan tahun 2021.

RR	
Tanggal	
2017-02-01	1.6
2017-02-02	6.0
2017-02-03	2.0
2017-02-04	1.7
2017-02-05	2.1
...	...
2021-01-27	10.6
2021-01-28	2.9
2021-01-29	3.6
2021-01-30	12.4
2021-01-31	0.3

1826 rows × 1 columns

Gambar 2. Data Raw Curah Hujan Harian

Pengujian dilakukan menggunakan data dari BMKG Kota Bandung yang diperoleh dari website BMKG ([http://dataonline.bmkg.go.id/dashboard\\_user](http://dataonline.bmkg.go.id/dashboard_user)) untuk melihat curah hujan Kota Bandung dan untuk melihat nilai kesalahan dari peramalan digunakan. Gambar 3 merupakan *source code* untuk mengimport *library* yang digunakan.

```
import pandas as pd
import numpy as np
from datetime import datetime
from matplotlib import pyplot
import matplotlib.pyplot as plt
from sklearn.preprocessing import MinMaxScaler
import tensorflow as tf

from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import mean_squared_error
```

Gambar 3. Source Code Import Library

Setelah *library* diimpor, tahap selanjutnya adalah normalisasi pada data curah hujan harian, `scaler = MinMaxScaler(feature_range=(0, 1))` mengubah fitur dengan menskalakan setiap fitur ke rentang tertentu pada set pelatihan, misalnya antara nol dan satu. Gambar 4 merupakan *source code* yang digunakan.

```
# normalize the dataset
scaler = MinMaxScaler(feature_range=(0, 1))
dataset = scaler.fit_transform(dataset)
```

Gambar 4. Source Code Normalisasi

Setelah dilakukan normalisasi maka selanjutnya adalah membagi dataset menjadi data pelatihan dan data uji. Gambar 5 merupakan *source code* yang digunakan.

```
# split into train and test sets
train_size = int(len(dataset) * 0.9)
test_size = len(dataset) - train_size
train, test = dataset[0:train_size,:], dataset[train_size:len(dataset),:]
```

Gambar 5. Source Code Split Data

Gambar 6 merupakan *source code* yang berfungsi menyiapkan rangkaian data pelatihan dan data uji untuk pemodelan

```
# reshape into X=t and Y=t+1
look_back = 3
trainX, trainY = create_dataset(train, look_back)
testX, testY = create_dataset(test, look_back)
```

Gambar 6. Source Code Persiapan Data

Gambar 7 merupakan merupakan proses pengujian *epoch* yang dapat dilihat pada *source code* berikut.

```
model.fit(trainX, trainY, epochs=50, batch_size=batch_size,
shuffle=False)
```

Gambar 7. Source Code Pengujian Epoch

*Source code* diatas berfungsi untuk menguji *epoch*, dengan nilai trainX dan trainY dengan nilai *epoch* sebesar 50. Di bawah ini menampilkan hasil pengujian *epoch* yang dapat dilihat pada gambar dibawah. Proses ini menggambarkan proses pengujian *epoch* dari data yang telah dinormalisasi. Gambar 8 merupakan hasil dari pengujian *epoch*.

```
Epoch 1/50
1639/1639 [=====] - 6s 2ms/step - loss: 0.0063
Epoch 2/50
1639/1639 [=====] - 4s 2ms/step - loss: 0.0062
Epoch 3/50
1639/1639 [=====] - 4s 3ms/step - loss: 0.0062
Epoch 4/50
1639/1639 [=====] - 4s 3ms/step - loss: 0.0061
Epoch 5/50
1639/1639 [=====] - 4s 3ms/step - loss: 0.0061
Epoch 6/50
1639/1639 [=====] - 4s 2ms/step - loss: 0.0061
Epoch 7/50
1639/1639 [=====] - 4s 3ms/step - loss: 0.0061
Epoch 8/50
1639/1639 [=====] - 4s 2ms/step - loss: 0.0061
Epoch 9/50
1639/1639 [=====] - 4s 3ms/step - loss: 0.0061
Epoch 10/50
1639/1639 [=====] - 4s 3ms/step - loss: 0.0060
Epoch 11/50
1639/1639 [=====] - 5s 3ms/step - loss: 0.0060
Epoch 12/50
1639/1639 [=====] - 4s 3ms/step - loss: 0.0060
Epoch 13/50
1639/1639 [=====] - 4s 2ms/step - loss: 0.0060
Epoch 14/50
1639/1639 [=====] - 4s 2ms/step - loss: 0.0060
Epoch 15/50
1639/1639 [=====] - 4s 2ms/step - loss: 0.0060
Epoch 16/50
1639/1639 [=====] - 4s 2ms/step - loss: 0.0060
Epoch 17/50
1639/1639 [=====] - 4s 2ms/step - loss: 0.0060
```

Gambar 8. Pengujian *Epoch*

10. Kemudian dilakukan prediksi data train dan data test yang dapat dilihat pada Gambar 9 dan Gambar

```
# make predictions
trainPredict = model.predict(trainX, batch_size=batch_size)
model.reset_states()
testPredict = model.predict(testX, batch_size=batch_size)
```

Gambar 9. Source Code Melakukan Prediksi

```
1639/1639 [=====] - 3s 1ms/step
179/179 [=====] - 0s 1ms/step
```

Gambar 10. Prediksi Data Train dan Data Test

Kemudian dilakukan evaluasi yang bertujuan mengukur kerja dari model yang telah dibuat. Metode evaluasi yang digunakan adalah RMSE. Metode ini dilakukan pengujian terhadap dua model yaitu data train dan data test. Berikut merupakan kode program yang dapat dilihat pada gambar 11, sebagai berikut:

```
# calculate root mean squared error
trainScore = np.sqrt(mean_squared_error(trainY[0],
trainPredict[:,0]))
print('Train Score: %.2f RMSE' % (trainScore))
testScore = np.sqrt(mean_squared_error(testY[0], testPredict[:,0]))
print('Test Score: %.2f RMSE' % (testScore))
```

Gambar 11. Proses Menghitung RMSE

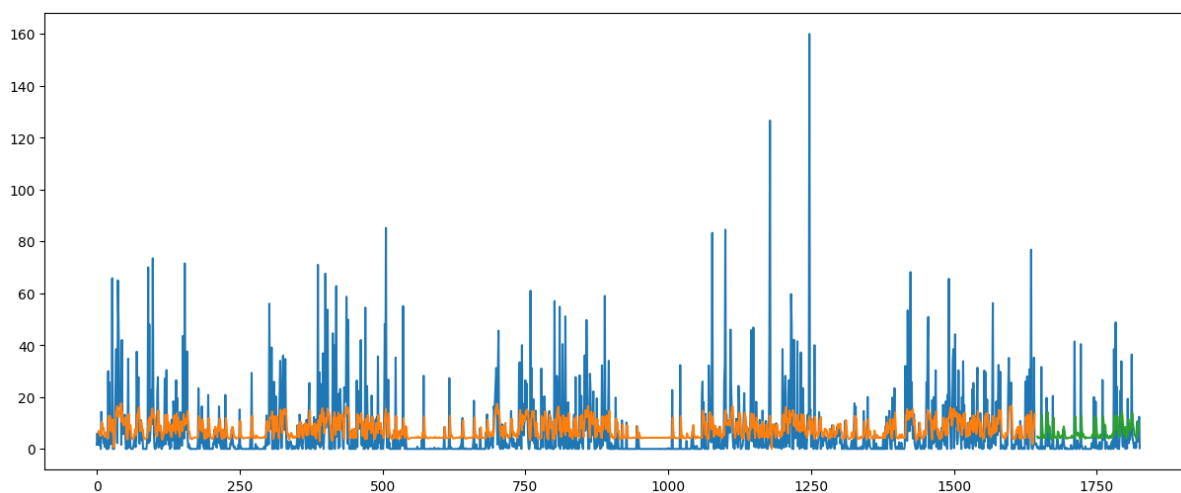
Hasil dari *source code* Gambar 11 berupa persentase kesalahan, di mana skor data training yaitu 12.24 sedangkan testing skor yaitu 8.86, dapat dilihat pada Gambar 12, sebagai berikut:

```
Train Score: 12.24 RMSE
Test Score: 8.86 RMSE
```

Gambar 12. Hasil Menghitung Kesalahan

Dari hasil di atas menunjukkan skor dari data testing yaitu 8.86. Semakin kecil skor data testing, maka semakin baik akurasi dari prediksi yang dihasilkan. Apabila skor data testing besar, maka hasil tersebut masih tergolong kurang baik.

Berikut merupakan visualisasi grafik prediksi yang dapat dilihat pada Gambar 13. Set data asli ditunjukkan dengan warna biru, prediksi untuk set data training berwarna orange, dan prediksi pada set data testing dengan warna hijau.



Gambar 13. Visualisasi Prediksi

#### 4. KESIMPULAN

Setelah melakukan analisis, perancangan dan implementasi serta hasil pengujian dari implementasi deep learning menggunakan arsitektur *Long Short Term Memory* (LSTM) untuk prediksi curah hujan harian di Kota Bandung, menunjukkan hasil prediksi yang diperoleh cukup baik. Dalam melakukan prediksi curah hujan harian di Kota Bandung dilakukan pengumpulan data, preprocessing, normalisasi, dan evaluasi dengan menggunakan RMSE. Penelitian ini menghasilkan nilai akurasi tertinggi yaitu dengan *Train Score* RMSE sebesar 12.24 dan *Test Score* RMSE sebesar 8.86 pada prediksi curah hujan

harian Kota Bandung. Dengan hasil *Test Score* 8.86 maka bisa dikatakan hasil prediksi cukup baik karena semakin kecil nilai *Test Score* maka akan semakin baik akurasi dari prediksi.

Penulis berharap agar penelitian ini dapat dikembangkan oleh peneliti lain dimasa yang akan datang. Saran yang dapat diberikan oleh penulis yaitu peneliti perlu menambah beberapa parameter/metode lain yang dapat membuat penelitian semakin lebih baik dan dapat membandingkan mana metode yang lebih baik. Selanjutnya dapat menambah data yang digunakan untuk mendapatkan nilai prediksi yang semakin baik lagi.

## DAFTAR PUSTAKA

- Abrahamsen, E., Brastein, O. M., & Lie, B. (2018). Machine Learning in Python for Weather Forecast based on Freely Available Weather Data. *Proceedings of The 59th Conference on Imulation and Modelling (SIMS 59)*, 26-28 September 2018, Oslo Metropolitan University, Norway, 153(September), 169–176. <https://doi.org/10.3384/ecp18153169>
- Habibi, M. Y., & Riksakomara, E. (2017). Peramalan Harga Garam Konsumsi Menggunakan Artificial Neural Network Feedforward-Backpropagation (Studi Kasus : PT. Garam Mas, Rembang, Jawa Tengah). *Jurnal Teknik ITS*, 6(2). <https://doi.org/10.12962/j23373539.v6i2.23200>
- Insani, F., Fadilah, S., & Sanjaya, S. (2020). Prediksi Cuaca Pekanbaru Menggunakan Fuzzy Tsukamoto dan Algoritma Genetika. *Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI)*, 255–262.
- Intan, I., Rismayani, Ghani, S. A. D., Nurdin, & Koswara, A. T. C. (2021). Performance Analysis of Weather Forecasting using Machine Learning Algorithms. *Jurnal Pekommas*, 6(2), 1–8. <https://doi.org/10.30818/jpkm.2021.2060221>
- Luthfiarta, A., Febriyanto, A., Lestiawan, H., & Wicaksono, W. (2020). Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda. *JOINS (Journal of Information System)*, 5(1), 10–17. <https://doi.org/10.33633/joins.v5i1.2760>
- Nurhamiddin, F., & Sulisa, F. M. (2019). Peramalan Cuaca Menggunakan Metode Rantai Markov ( Studi Kasus : Rekaman Cuaca Harian Di Kantor BMKG Kota Ternate ). *Jurnal BIOSAINTEK*, 2(1), 16–22. <https://doi.org/10.24843/JMAT.2021.v11i01.p131>
- Rachmawati, R. N. (2021). Estimation of Extreme Rainfall Patterns Using Generalized Linear Mixed Model for Spatio-temporal data in West Java, Indonesia. *Procedia Computer Science*, 179(2020), 330–336. <https://doi.org/10.1016/j.procs.2021.01.013>
- Rizki, M., Basuki, S., & Azhar, Y. (2020). Implementasi Deep Learning Menggunakan Arsitektur Long Short Term Memory(LSTM) Untuk Prediksi Curah Hujan Kota Malang. *Jurnal Repositor*, 2(3), 331. <https://doi.org/10.22219/repositor.v2i3.470>
- Sari, F. R., & Anifah, L. (2019). *Decision Support Systems Prakiraan Cuaca Harian Berbasis Semi-Supervised Learning Menggunakan Recursive K-Means Di Bandar Udara Juanda Surabaya*. 1(2).
- Soekendro, C. A. (2021). *Prediksi Curah Hujan di Kab. Bandung Dengan Analisis Time Series, Menggunakan Model SARIMA (Seasonal Autoregressive Integrated Moving Average)*. 8(2), 2865–2875.
- Wiranda, L., & Sadikin, M. (2019). Penerapan Long Short Term Memory Pada Data Time Series Untuk Memprediksi Penjualan Produk Pt. Metiska Farma. *Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI*, 8(3), 184–196. <https://ejournal.undiksha.ac.id/index.php/janapati/article/view/19139>
- Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *WIRES Data Mining and Knowledge Discovery*, 8(4), e1253. <https://doi.org/https://doi.org/10.1002/widm.1253>