

Prediction Of Bogor City Rainfall Parameters Using *Long Short Term Memory (LSTM)*

Sherly Amora Jofipasi, Admi Salma*, Dodi Vionanda, Dina Fitria

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: admisalma1@fmipa.unp.ac.id

Submitted : 06 Oktober 2023

Revised : 24 Oktober 2023

Accepted : 31 Oktober 2023

ABSTRACT

Bogor is a city that has high intensity of rainfall and has erratic rainfall. So it is necessary to predict Bogor's rainfall. Rainfall prediction can be done using the LSTM algorithm. In the LSTM algorithm, there are neuron hidden layer and epoch parameters. Neuron hidden layer and epoch greatly affect the resulting prediction results, therefore it is necessary to determine the best neuron hidden layer and epoch values to produce good prediction results in Bogor rainfall. The prediction parameters results obtained by LSTM have worked well using optimal neuron hidden values of 256, optimal epoch of 150, MAPE of 1,64%, and the comparison of actual data patterns and prediction data already has the same data patterns.

Keywords: *Long Short Term Memory, Rainfall, Prediction*



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Prediksi merupakan suatu kegiatan meramalkan atau memperkirakan kemungkinan kejadian untuk di masa depan dengan menggunakan metode tertentu (Diera, 2018). Untuk melakukan prediksi terdapat berbagai macam metode salah satunya yaitu metode *deep learning*. Menurut Wahyudi (2020) *deep Learning* merupakan bagian dari *machine learning*, dimana *deep learning* terdiri dari beberapa layer. Layer awal menghasilkan fitur yang sederhana sehingga menghasilkan layer akhir yang memiliki fitur yang kompleks. Pada *deep learning* terdapat algoritma yang baik digunakan untuk prediksi, yaitu algoritma LSTM (*Long Short Term Memory*). Algoritma LSTM merupakan algoritma lanjutan dari algoritma RNN, dimana algoritma LSTM ini memiliki tambahan *cell state* yang berguna untuk mengingat dan memberikan informasi dalam jangka waktu yang panjang. Pada algoritma LSTM terdapat parameter *neuron hidden layer* dan *epoch* yang digunakan untuk memperoleh hasil yang baik. Parameter *neuron hidden layer* dan *epoch* sangat berpengaruh terhadap hasil prediksi yang dihasilkan. Oleh karena itu, perlu dilakukan penentuan nilai *neuron hidden layer* dan *epoch* yang optimal untuk menghasilkan hasil prediksi yang baik.

Algoritma LSTM dapat diterapkan pada data musiman, salah satunya yaitu data curah hujan. Menurut BMKG (2020) curah hujan merupakan ketinggian air hujan dalam 1 mm alat penakar hujan yang mana berada di tempat yang datar, tidak meresap, dan tidak menyerap. Intensitas curah hujan di Indonesia berbeda-beda di setiap daerahnya, tergantung dengan faktor-faktor yang mempengaruhinya. Intensitas hujan di Indonesia ada yang memiliki intensitas hujan yang tinggi dan ada yang memiliki intensitas hujan yang rendah. Salah satu kota yang memiliki intensitas tinggi di Indonesia yaitu Kota Bogor. Menurut Diskominfo Kota Bogor (2021) Intensitas curah hujan Kota Bogor yaitu berada pada rentang 3.500 - 4.000 mm/ tahunnya dengan 300 - 500 mm/bulannya. Intensitas curah hujan tersebut termasuk ke dalam kategori tinggi. Oleh sebab itu perlu dilakukan prediksi parameter curah hujan Kota Bogor dengan algoritma LSTM upaya untuk mengurangi dampak dan menjadi acuan dan rancangan dalam menjalani kegiatan sehari-hari baik bagi pemerintah, instansi, maupun bagi masyarakat sekitarnya.

Beberapa penelitian terdahulu oleh M. Rizki (2020) menggunakan algoritma LSTM untuk memprediksi curah hujan Kota Malang, menyimpulkan bahwa algoritma LSTM berhasil memberikan hasil prediksi curah hujan di Kota Malang dengan tingkat keakurasian di atas 85%. Penelitian Jamilatul (2022) membandingkan algoritma LSTM dengan algoritma GRU dan algoritma RNN untuk prediksi curah hujan di Kota Surabaya. Menyimpulkan bahwa hasil pengukuran tingkat keakurasian untuk prediksi curah hujan di Kota Surabaya pada algoritma LSTM lebih kecil. Sehingga algoritma LSTM lebih baik digunakan untuk prediksi dibandingkan algoritma GRU dan RNN. Selain itu,

penelitian Satyo (2020) menggunakan algoritma LSTM untuk prediksi saham Bank BRI, menyimpulkan bahwa untuk mendapatkan tingkat keakuratan yang baik menggunakan LSTM, maka diperlukan nilai *epoch* yang optimal.

II. METODE PENELITIAN

A. Jenis Penelitian dan Sumber Data

Penelitian ini merupakan penelitian terapan yang digunakan untuk melakukan prediksi curah hujan di Kota Bogor dengan menggunakan metode *Long Short Term Memory* (LSTM). Penelitian ini menggunakan data sekunder yang diperoleh dari website BMKG yaitu www.bmkg.go.id. Data yang digunakan yaitu data *time series* harian curah hujan di Kota Bogor. Sampel yang digunakan sebanyak 1826 observasi dari 1 Januari 2018 hingga 31 Desember 2022 untuk variabel curah hujan.

B. Teknik Analisis Data

Teknik analisis data dilakukan dengan menggunakan algoritma LSTM dengan menggunakan bantuan *software Python*. Berikut tahapan analisis data yang dilakukan.

1. Analisis statistika deskriptif
Analisis statistika deskriptif dilakukan untuk melihat ringkasan data yang digunakan secara umum. Dimana statistika deskriptif berisi rata-rata, nilai maksimum, standar deviasi, nilai minimum, dan grafik curah hujan harian Kota Bogor pada tahun 2018-2022.
2. Pembagian data *training* dan data *testing*
Pembagian data *training* dan data *testing* dilakukan untuk melihat kinerja metode yang digunakan. Menurut Rogel (2018) nilai yang diperoleh digunakan untuk melihat apakah model telah bekerja dengan baik atau tidak pada data. Adapun rasio pembagian data *training* dan data *testing* yang digunakan yaitu 80% : 20%.
3. Melakukan normalisasi data
Menurut Francesca (2021) normalisasi data dilakukan untuk mentransformasikan data sehingga data memiliki rentang yang sama dalam rentang [0,1]. Adapun normalisasi yang dilakukan yaitu dengan rumus *MinMaxScaler* yang disajikan pada persamaan (1)

$$y_i = \frac{(x_i - x_{min})}{(x_{max} - x_{min})} \quad (1)$$

dimana y_i = nilai data baru ke-i, x_i = nilai data lama ke-i, x_{min} = nilai minimum pada data asli, dan x_{max} = nilai maksimum pada data asli.

4. Penentuan parameter *neuron hidden layer*
Neuron hidden layer merupakan lapisan tersembunyi yang terletak di antara *input layer* dan *output layer*. *Hidden layer* berfungsi untuk memproses informasi yang diterima pada *input layer* dan meneruskan ke *output layer* untuk menghasilkan *output*.
5. Penentuan parameter *epoch*
Epoch merupakan hyperparameter yang menentukan berapa kali algoritma *Deep Learning* bekerja melewati dataset, baik *forward* maupun *backward*. Penentuan *epoch* dilakukan untuk melihat hasil dari pelatihan model dengan menggunakan nilai *loss*. Apabila nilai *loss* mendekati 0 maka model telah memprediksi data dengan baik, begitu juga sebaliknya apabila nilai *loss* tidak mendekati 0 maka model tidak memprediksi data dengan baik (Kahlil, 2023).
6. *Long Short Term Memory*
Algoritma LSTM digunakan untuk memprediksi curah hujan harian Kota Bogor untuk masa yang akan datang. Menurut M.Rizki (2020) Untuk melakukan prediksi pada algoritma LSTM terdapat 3 gate, yaitu:
 - a. *Forget Gate*, dimana nilai pada setiap data dimasukkan dan di seleksi dengan menggunakan fungsi aktivasi *sigmoid* (σ). Dengan persamaan (2).

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (2)$$

Dimana, f_t = *forget gate* ke-t, σ = fungsi aktivasi *sigmoid*, W_f = nilai *weight forget gate*, h_{t-1} = nilai output sebelum t, x_t = nilai input baru ke-t, dan b_f = nilai bias *forget gate*.

- b. *Input Gate*, dimana pada *input gate* terdapat 2 tahapan, yang pertama yaitu *input gate layer* mengambil keputusan nilai yang akan diperbarui dan menghasilkan nilai i_t . Dengan persamaan (3)

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (3)$$

Dimana i_t = *input gate* ke-t, W_i = nilai *weight input gate*, b_i = nilai bias untuk *input gate*. Yang kedua yaitu menggunakan fungsi aktivasi *tanh* yang akan membuat nilai baru (\bar{C}_t) kemudian ditambahkan pada *cell state*. Dengan persamaan (4)

$$\bar{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (4)$$

Dimana C_t = nilai baru yang akan ditambahkan pada *cell state* ke-t, \tanh = fungsi aktivasi *tanh*, W_c = nilai *weight* pada *cell state*, b_c = nilai bias pada *cell state*.

- c. *Output Gate*, dimana fungsi aktivasi *sigmoid* dan fungsi aktivasi *tanh* bekerja untuk menghasilkan output yang sama pada *cell state*. Dengan persamaan (5)

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (5)$$

Dimana o_t = *output gate* ke-t, W_o = nilai *weight* pada *output gate*, b_o = nilai bias pada *output gate*.

7. Evaluasi Model

Dalam melakukan suatu analisis, untuk memperoleh model yang akurat (baik) maka diperlukan pengecekan kinerja metode yang digunakan. Menurut Makridakis, et al (1999) untuk mengetahui tingkat akurasi hasil prediksi, terdapat beberapa estimator yang umum digunakan, salah satunya yaitu MAPE (*Mean Absolute Percentage Error*). MAPE merupakan salah satu metode yang digunakan untuk mengukur akurasi metode yang digunakan. MAPE dihitung menggunakan kesalahan mutlak setiap periode dibagi dengan nilai aktual periode tersebut. Berikut rumus yang digunakan untuk MAPE pada persamaan (6).

$$MAPE = \frac{\sum \frac{|\hat{y}_t - y_t|}{\hat{y}_t}}{m} \times 100\% \quad (6)$$

Dimana \hat{y}_t = data aktual ke-t, y_t = data hasil prediksi ke-t, dan m = jumlah periode data.

III. HASIL DAN PEMBAHASAN

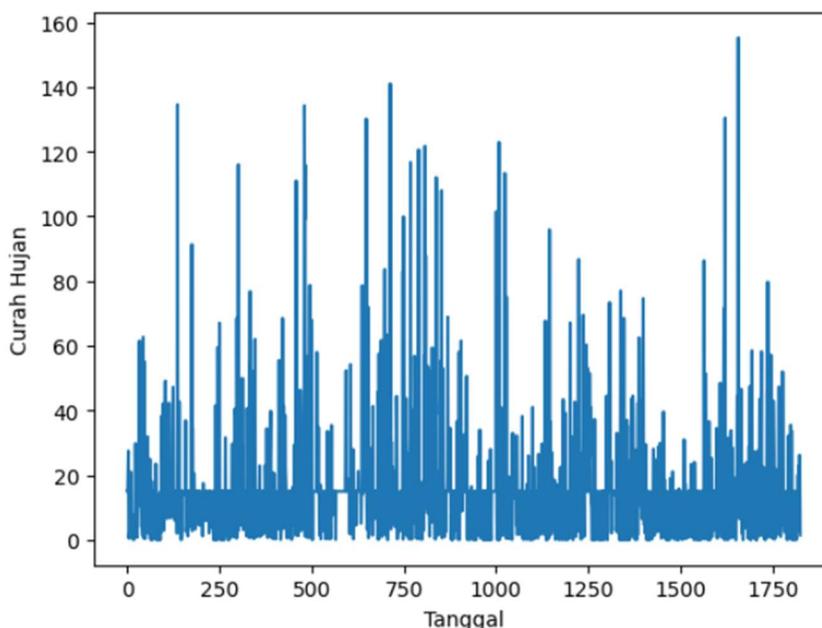
A. Statistika Deskriptif

Sebelum melakukan prediksi perlu dilakukan deskripsi data untuk melihat ringkasan data. Tabel 1 menyajikan deskripsi variabel penelitian

Tabel 1. Statistik Deskriptif

Variabel	N	Mean	Median	Standar Deviasi	Min	Max
X ₁	1826	15,43	15,43	17,55	0,00	155,20

Pada Tabel 1, terlihat bahwa terdapat sebanyak 1826 data yang digunakan. Selain itu intensitas curah hujan tertinggi sebesar 155,2 mm/hari, terendah sebesar 0,00 mm/hari, dan rata-rata perharinya yaitu sebesar 15,43 mm. Berikut disajikan grafik curah hujan harian kota Bogor 1 Januari 2018 hingga 31 Desember 2022.



Gambar 1. Curah Hujan Harian Kota Bogor tahun 2018-2022

Berdasarkan Gambar 1, terdapat grafik pengamatan data curah hujan harian Kota Bogor dari Januari 2018 sampai Desember 2022. Dapat dilihat bahwa terjadi kenaikan dan penurunan intensitas curah hujan yang ekstrem. Berikut disajikan tabel intensitas curah hujan bulanan Kota Bogor tahun 2018-2022.

Tabel 2. Intensitas Curah Hujan Bulanan Kota Bogor

Bulan	2018	2019	2020	2021	2022
Januari	293,7	423,6	519,9	337,1	203,2
Februari	445,3	363,1	585,4	531,2	205,8
Maret	225	381	811,3	338	277,1
April	530,4	820,8	629,5	336,8	406,6
Mei	547,6	582,2	575,5	677,6	437,5
Juni	500,7	422,5	474,2	431,7	527
Juli	418,3	427,8	367,3	355,6	583,9
Agustus	388,8	545,3	341,1	577,2	474,9
September	442,9	496,6	403,3	392,4	375,1
Oktober	564,2	587,1	611	671,4	619,6
November	551,9	570,1	369,5	252,5	428
Desember	467,1	658	301,8	414,1	331
Total	5.5375,9	6.278,1	5.989,9	5.315,6	4.869,7

Berdasarkan Tabel 2, intensitas curah hujan bulanan Kota Bogor tahun 2018-2022 berada pada rentang 200 – 850mm/bulannya. Sedangkan intensitas curah hujan tahunan berada pada rentang 4.500 - 6.500 mm/tahunnya. Dan berdasarkan Tabel 2 tersebut, Kota Bogor dapat dikategorikan sebagai kota yang memiliki intensitas curah hujan yang tinggi.

B. Data Training dan Data Testing

Pembagian data *training* dan data *testing* dilakukan untuk melihat kinerja suatu model pada data, apakah model telah bekerja dengan baik terhadap data atau sebaliknya. Data *training* dan data *testing* yang digunakan yaitu 80% : 20%. Tabel 3 menyajikan pembagian data training dan data testing.

Tabel 3. Data Training dan Data Testing

Data Training	Data Testing
1460	366

Berdasarkan Tabel 3, diketahui bahwa jumlah sampel untuk data *training* sebesar 1460 sampel, sedangkan untuk data *testing* sebesar 366 sampel.

C. Normalisasi Data

Normalisasi dilakukan untuk membuat data curah hujan lebih konsisten, sehingga data curah hujan berada dalam rentang yang sama yaitu 0 sampai 1. Berikut hasil nilai normalisasi data ditampilkan pada Tabel 4.

Tabel 4. Nilai Normalisasi

No	Normalisasi
1	0,02
2	0,02
3	0,14
⋮	⋮
1826	0,00

Berdasarkan Tabel 4, normalisasi data dilakukan untuk semua data yaitu sebanyak 1826 data. Berdasarkan Tabel 3 terlihat bahwa nilai normalisasi data telah berada pada rentang 0 sampai 1.

D. Penentuan Parameter *Neuron Hidden*

Sebelum melakukan prediksi dengan LSTM, maka perlu dilakukan pemilihan *neuron hidden* optimal dengan menggunakan data *training* agar menghasilkan hasil yang baik. Berikut ditampilkan pada Tabel 5 hasil evaluasi model berdasarkan kombinasi parameter *neuron hidden*.

Tabel 5. Kombinasi Parameter Jumlah *Neuron Hidden*

<i>Neuron Hidden</i>	<i>Epoch</i>	MAPE
32	50	4,52
64	50	4,21
128	50	4,09
256	50	4,03

Berdasarkan Tabel 5, terlihat bahwa kombinasi parameter *neuron hidden* yang menghasilkan evaluasi model terbaik berdasarkan MAPE sebesar 4,03% yaitu *neuron hidden* dengan jumlah *layer* sebanyak 256 *layer*. *Neuron hidden* dilakukan untuk menyelesaikan tugas nilai *input* dan menyampaikannya ke nilai *output*.

E. Penentuan Parameter *Epoch*

Selain pemilihan parameter *neuron hidden*, juga dilakukan pemilihan *epoch* optimal dengan menggunakan data *training* agar menghasilkan hasil yang baik. Tabel 6 menyajikan *epoch* yang optimal dengan melihat nilai MAPE.

Tabel 6. Kombinasi Parameter Jumlah *Epoch*

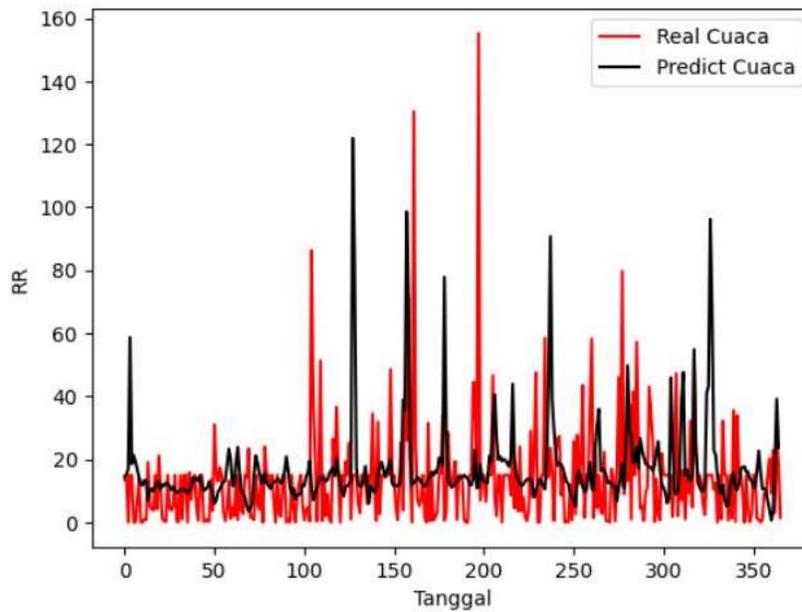
<i>Neuron Hidden</i>	<i>Epoch</i>	MAPE
256	50	4,73
256	100	4,44
256	150	4,12
256	200	4,22

Berdasarkan Tabel 6, menunjukkan bahwa kombinasi nilai *epoch* optimal yang menghasilkan evaluasi model terbaik berdasarkan MAPE sebesar 4,12% yaitu *epoch* dengan jumlah 150 *epoch*. Pengecekan nilai optimal *epoch* digunakan untuk meminimalkan terjadinya kesalahan hasil prediksi. Nilai *epoch* optimal ini diperoleh dari nilai *loss* yang dihasilkan model. Pada hasil *training* model menunjukkan nilai *loss* terendah pada data *train* yaitu sebesar

0,0042 pada *epoch* bagian 149. Sedangkan nilai *loss* tertinggi pada data *train* yaitu sebesar 0,0180 pada *epoch* bagian 1.

F. Long Short Term Memory

Berikut ditampilkan pada Gambar 2 perbandingan hasil data prediksi dan data aktual pada data curah hujan Kota Bogor dengan menggunakan data *testing*.



Gambar 2. Perbandingan Data Prediksi dan Data Aktual Pada Data *Testing*

Berdasarkan Gambar 2, terdapat perbandingan antara data aktual dengan data prediksi curah hujan Kota Bogor dengan menggunakan data *test*. Pada perbandingan di atas, digunakan data *testing* dengan jumlah *neuron hidden* optimal sebesar 256 layer dan nilai *epoch* optimal sebesar 150 *epoch*. Selain itu, terlihat bahwa terdapat kenaikan intensitas curah hujan pada awal periode dan terjadi penurunan intensitas curah hujan pada akhir periode, dan hasil dari perbandingan di atas telah memiliki pola data yang sama.

G. Evaluasi Model

Berikut ditampilkan evaluasi model Gambar 2 pada persamaan (7).

$$MAPE = \frac{0,28+0,82+\dots+5,73}{365} \times 100\% = 1,64 \tag{7}$$

Berdasarkan persamaan 7, diperoleh nilai MAPE sebesar 1,64%.

IV. KESIMPULAN

Pada penelitian ini dilakukan prediksi parameter curah hujan Kota Bogor dengan menggunakan *algoritma Long Short Term Memory*. Berdasarkan hasil penelitian ini, diperoleh parameter *neuron hidden* terbaik sebesar 256 layer dan *epoch* terbaik sebesar 150 *epoch* dengan menggunakan data *training*. Dan berdasarkan Gambar 2, dilakukan perbandingan data *predict* dan data *actual* menggunakan data *testing* dengan *neuron hidden* 256 layer dan *epoch* 150 *epoch*, maka diperoleh MAPE sebesar 1,64%. Sehingga dapat disimpulkan bahwa model telah bekerja dengan baik terhadap data. Dimana model mampu menghasilkan nilai MAPE < 5%. Saran untuk peneliti selanjutnya yaitu menambahkan banyak data, membandingkan dengan metode prediksi lainnya, dan dapat mengaplikasikan metode LSTM lainnya seperti Bi-LSTM pada data curah hujan.

DAFTAR PUSTAKA

- Badan Meteorologi, Klimatologi, dan Geofisika. (2020). Prakiraan: tentang cuaca. <https://web.meteo.bmkg.go.id/id/prakiraan/cuaca-kota>. Diakses pada 8 Maret 2023.
- Diera, D., Tursina., & M.Azhar. I. (2018), “Prediksi Besaran Curah Hujan Menggunakan Metode Fuzzy Time Series”, *Jurnal Sistem dan Teknologi Informasi*, Vol.6, No.4, e-ISSN: 2620-8989.
- Dinas Komunikasi dan Informatika Kota Bogor. 2021. Letak Geografis. <https://www.kotabogor.go.id/index.php/page/detail/9/letak-geografis>. Diakses pada 13 Maret 2023.
- Francesca, L. (2021), “*Machine Learning for Time Series Forecasting With Python*”. Published Simultaneously in Canada, Manufactured in the United States of America.
- Jamilatul, B., Arna, F., & Tri, H., (2022). Prediksi Curah Hujan Menggunakan Long Short Term Memory. *Jurnal Media Informatika Budidarma*, vol.6, no.3, 1297-1303.
- Kahlil., Rizky, M., Laksono K., Adighuna, M., Nona, Z., Fina, N., & Alifya, F. (2023). *Computer Vision Berbasis Deep Learning Untuk Aplikasi Pertanian : Teori & Praktik*. Banda Aceh: Universitas Syah Kuala.
- M. Rizki., Setio, B., and Yufis., A, “Implementasi Deep Learning Menggunakan Arsitektur Long Short Term Memory Untuk Prediksi Curah Hujan Kota Malang”, *REPOSITOR*, vol.2, no. 3, pp.331-338, 2020.
- Makridakis, S. G., Wheelwright, S. C., and Hyndman, R.J. 1999. *Forecasting: Methods and Applications*. John Wiley and Sons Inc., USA.
- Rogel, S. J. (2018). *Data Science and Analytics with Python*. Chapman and Hall/CRC.
- Satyo, Adhitio. (2020). Prediksi Data Time Series Saham Bank BRI dengan Mesin Belajar LSTM (Long Short Term Memory). *Jurnal of Information and Information Security (JIFORTY)*, vol.1, no.2, 1-8 2020, e-ISSN: 2722-4058.
- Wahyudi, S. (2020). *Deep Learning Menggunakan Convolutional Neural Network : Teori & Aplikasi Cetakan I*. Malang: Media Nusa Crative.