

# Prediksi Penggunaan Energi Listrik pada Rumah Hunian Menggunakan Long Short-Term Memory

I Nyoman Kusuma Wardana<sup>1</sup>, Naser Jawas<sup>2</sup>, I Komang Agus Ady Aryanto<sup>3</sup>

<sup>1</sup>Jurusan Teknik Elektro, Politeknik Negeri Bali, Bali

<sup>2</sup>Program Studi Sistem Informasi, Institut Teknologi dan Bisnis STIKOM Bali, Bali

<sup>3</sup>Program Studi Teknologi Informasi, Institut Teknologi dan Bisnis STIKOM Bali, Bali

## ABSTRACT

The prediction of electricity consumption in residential homes is important to do. This is because household electrical equipment is one of the factors that has a major influence on the use of electricity nationally. This research implemented Long Short-Term Memory (LSTM) as a predictor model. Based on the available dataset, thirteen attributes were used as the inputs for the LSTM. The LSTM with eight neurons and seven *lookbacks* found to give the best performance with the error rates of 60.992 and 28.278 for RMSE and MAE, respectively.

**Keywords:** Home appliances, electricity, prediction, dataset, LSTM.

## ABSTRAK

Prediksi konsumsi listrik pada rumah hunian adalah penting untuk dilakukan. Ini dikarenakan peralatan listrik rumah tangga merupakan salah satu faktor yang berpengaruh besar terhadap penggunaan energi listrik secara nasional. Penelitian ini menggunakan Long Short-Term Memory (LSTM) sebagai model prediktor. Sebanyak 13 jenis atribut dari dataset yang tersedia digunakan sebagai input untuk LSTM. Berdasarkan penelitian, 8 neuron dalam LSTM dengan *lookback* sebanyak 7 memiliki kinerja paling baik. Besarnya nilai error terhadap data uji masing-masing sebesar 60,992 dan 28,278 untuk RMSE dan MAE.

**Kata kunci:** peralatan listrik, energi listrik, prediksi, dataset, LSTM

## Info Artikel

Diterima Redaksi : 24-06-2020

*This is an open access article under the [CC BY-SA](#) license.*

Selesai Revisi : 25-06-2020

Diterbitkan Online : 30-06-2020



## Penulis Korespondensi:

I Nyoman Kusuma Wardana

Jurusan Teknik Elektro

Politeknik Negeri Bali

Jalan Kampus Bukit Jimbaran, Kuta Selatan, Badung, Bali 80364

Email: kusumawardana@pnb.ac.id

## 1. PENDAHULUAN

Berdasarkan Laporan Tahunan PLN tahun 2018 mengenai daya tersambung per segmen pelanggan, rumah tangga menempati posisi tertinggi dengan nilai daya sebesar 63.577 MVA (48,8%) dari total daya tersambung sebesar 130,281 MVA. Pertumbuhan daya tersambung dari tahun 2016 sampai 2018 tercatat 7,3%, melebihi segmen industri dan bisnis dengan nilai masing-masing sebesar 5,6% dan 5,9% [1]. Berdasarkan data tersebut, dapat disimpulkan bahwa konsumsi listrik dari rumah tangga (atau rumah hunian) menjadi salah satu faktor penting yang mempengaruhi konsumsi listrik secara nasional.

Peralatan listrik (*appliances*) menjadi salah satu sumber pemakai energi listrik terbesar di suatu rumah hunian. Sebagai gambaran, penelitian yang dilakukan oleh Cetin, dkk., mencatat bahwa peralatan listrik dalam suatu rumah hunian di Amerika Serikat dapat mengkonsumsi energi listrik sampai sekitar 30%

[2]. Istilah *appliance* disini merujuk pada peralatan yang umumnya digunakan untuk membantu pekerjaan rumah tangga seperti kulkas, mesin cuci, *juicer*, *mixer*, *microwave*, oven, kompor listrik dan sebagainya. Istilah ini dibedakan dari berbagai peralatan lain yang membutuhkan energi listrik seperti alat penerangan, pemanas/pendingin ruangan, CCTV, dan sebagainya. Dalam penelitian ini, istilah *appliance* cukup ditulis sebagai peralatan rumah tangga.

Karena penggunaan peralatan rumah tangga sangat mempengaruhi total konsumsi energi listrik di suatu rumah hunian, maka prediksi mengenai penggunaan energi listrik untuk alat-alat rumah tangga ini menjadi penting untuk dilakukan [3]. Berbagai penelitian mengenai prediksi penggunaan peralatan listrik telah dilaksanakan, salah satunya oleh Candanedo, dkk [4]. Candanedo, dkk menggunakan empat prediktor yang berbeda dalam memprediksi konsumsi listrik dari suatu rumah hunian, yaitu menggunakan Linear Regression Model (LM), Support Vector Machine (SVM), Random Forest (RF) dan Gradient Boosting Machine (GBM). Penelitian dalam makalah ini merujuk pada penelitian sebelumnya yang telah dilakukan oleh Candanedo, dkk. Namun berbeda dengan penelitian sebelumnya, Penulis mencoba menerapkan salah satu model deep learning, yaitu Long Short-Term Memory untuk memprediksi penggunaan energi listrik dari peralatan rumah tangga.

Long Short-Term Memory (LSTM) [5] merupakan modifikasi struktur dari Recurrent Neural Network (RNN) dengan menambahkan sel memori di dalam lapisan tersembunyi sehingga dapat digunakan untuk mengontrol arus informasi dalam suatu data time-series [6]. Data yang diprediksi dalam penelitian ini tergolong time-series. Data time-series adalah rangkaian data yang diobservasi berdasarkan interval waktu tertentu. Data time-series dapat diterapkan untuk berbagai aplikasi, seperti regresi, klastering, dan klasifikasi [7]. LSTM memiliki kemampuan yang bagus dalam memprediksi kasus-kasus yang melibatkan time-series [8][9]. Selain digunakan dalam kasus time-series, contoh penerapan lainnya misalnya pengenalan tulisan tangan [10], klasifikasi teks [11], intrusi data dalam jaringan komputer [12], dan berbagai jenis penerapan lainnya.

Makalah ini disusun menjadi empat bagian. Bagian pertama mengulas latar belakang penelitian. Bagian kedua membahas tentang metode penelitian, yang mencakup deskripsi dari data yang digunakan dalam penelitian, penjelasan mengenai model prediktor yang digunakan, serta metode untuk evaluasi model yang telah diusulkan. Bagian ketiga membahas tentang pemilihan model yang paling optimal, evaluasi model serta perbandingan dengan penelitian yang telah dilakukan sebelumnya. Terakhir, bagian keempat merangkum hasil-hasil penelitian yang telah dilakukan.

## 2. METODE PENELITIAN

### 2.1. Deskripsi Dataset

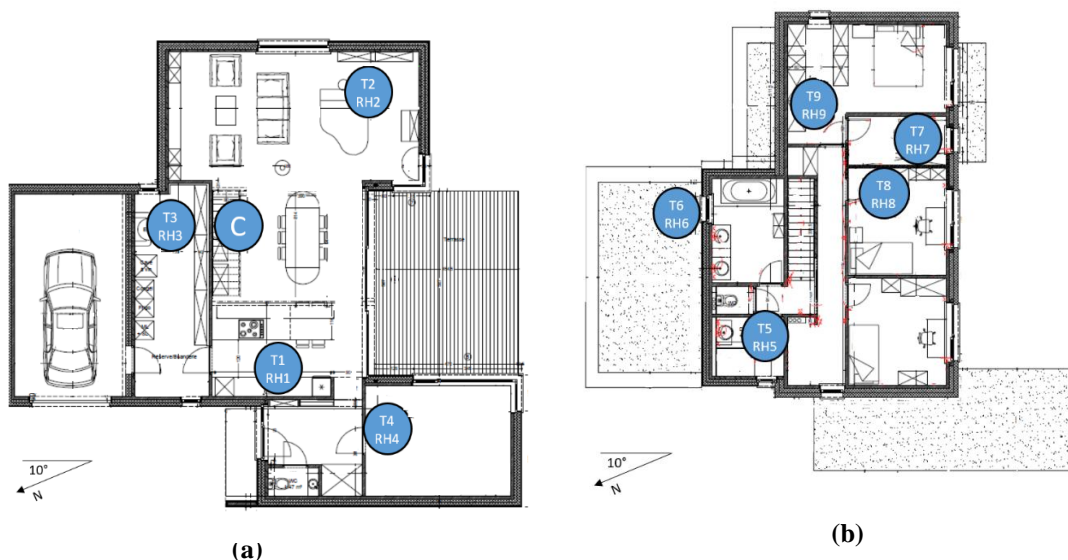
Penelitian ini tidak menggunakan data dari hasil pengukuran yang dilakukan oleh Penulis secara langsung, namun Penulis menggunakan dataset yang disediakan oleh Candanedo, dkk [13]. Dataset ini dapat diunduh pada laman UCI Machine Learning Repository. Atribut penyusun dataset adalah berupa data *indoor* dan *outdoor*. Data *indoor* berupa hasil pengukuran parameter udara menggunakan wireless sensor network (temperatur dan kelembaban), dan konsumsi energi listrik dari berbagai peralatan dan penerangan di suatu rumah hunian. Selain itu, dataset juga dilengkapi dengan data *outdoor* berupa parameter cuaca (tekanan, kelembaban, kecepatan angin, jarak pandang, dan titik embun) yang dikumpulkan dari stasiun bandara terdekat. Semua data tersebut dikumpulkan menjadi sebuah dataset dengan interval 10 menit untuk setiap baris data.

Untuk data *indoor*, beberapa sensor untuk mengukur temperatur dan kelembaban ruangan mengirimkan data kurang lebih setiap 3,3 menit menggunakan protocol ZigBee, sedangkan energi meter untuk mengukur konsumsi energi listrik mengumpulkan data setiap 10 menit. Data temperatur dan kelembaban tersebut selanjutnya dirata-ratakan untuk mendapatkan data setiap 10 menit. Selain energi meter utama, terdapat pula sub-energi meter yang khusus mengukur konsumsi energi dari alat-alat penerangan. Data dari alat penerangan dimaksudkan sebagai prediktor terhadap hunian ruangan ketika dikombinasikan kelembaban udara relatif [4]. Untuk data *outdoor*, berbagai parameter cuaca diperoleh dari stasiun cuaca di bandara terdekat. Karena pengukuran parameter cuaca ini dilakukan setiap jam, maka dilakukan interpolasi linier untuk mendapatkan interval setiap 10 menit. Atribut dataset yang diunduh dari laman UCI Machine Learning Repository diperlihatkan pada Tabel 1, sedangkan penempatan sensor di dalam ruangan diperlihatkan pada Gambar 1.

Tabel 1. Atribut dataset [13]

Atribut	Satuan	Deskripsi	Atribut	Satuan	Deskripsi	Atribut	Satuan	Deskripsi
Date	dd:mm:yy hh:mm:ss	Tanggal Waktu	T4	°C	Temperatur ruang kantor	RH_8	%	Kelembaban kamar anak
Appliances	Wh	Penggunaan energi total	RH_4	%	Kelembaban ruang kantor	T9	°C	Temperatur kamar orang tua
lights	Wh	Energi penerangan	T5	°C	Temperatur kamar mandi	RH_9	%	Kelembaban kamar orang tua
T1	°C	Temperatur dapur	RH_5	%	Kelembaban kamar mandi	T_out	°C	Temperatur outdoor
RH_1	%	Kelembaban dapur	T6	°C	Temperatur sisi luar utara	Press_mm_hg	mmHg	Tekanan outdoor
T2	°C	Temperatur ruang tamu	RH_6	%	Kelembaban sisi luar utara	Windspeed	m/s	Kecepatan angin
RH_2	%	Kelembaban ruang tamu	T7	°C	Temperatur ruang setrika	Visibility	km	Jarak pandang
T3	°C	Temperatur ruang cuci	RH_7	%	Kelembaban ruang setrika	Tdewpoint	°C	Titik embun
RH_3	%	Kelembaban ruang cuci	T8	°C	Temperatur kamar anak	rv1	-	Random variable 1
						rv2	-	Random variable 2

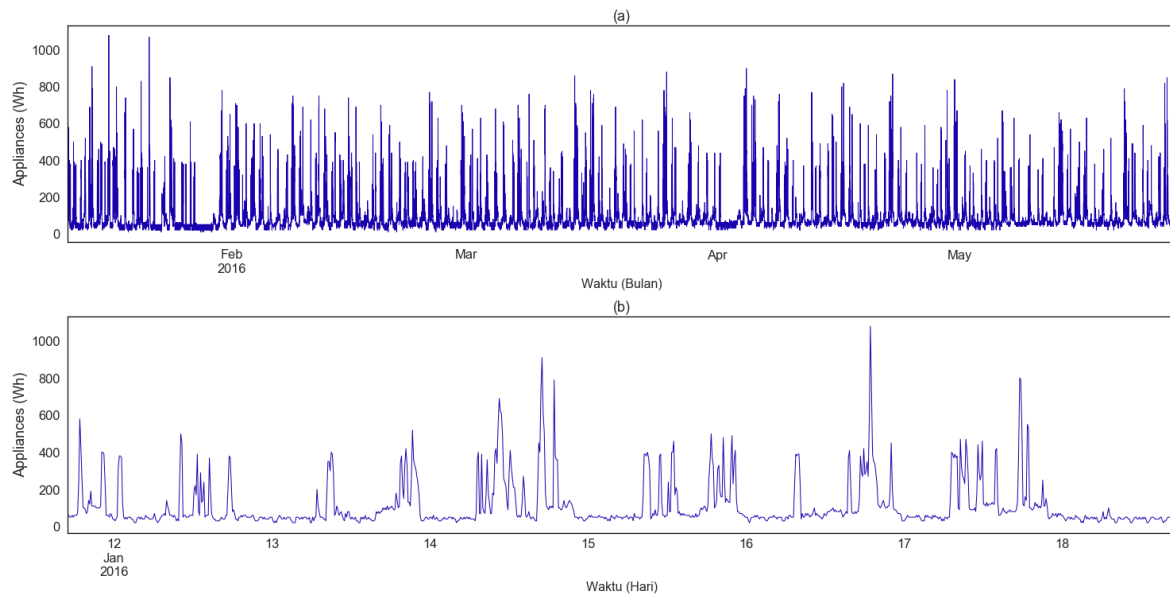
Dari atribut dataset seperti pada Tabel 1 tersebut, Candanedo, dkk menambahkan tiga atribut lagi berdasarkan atribut Date, yaitu jumlah detik yang dihitung dari tengah malam untuk setiap harinya (NSM), status hari (*weekend* atau *weekday*) dan nama-nama hari yang bersesuaian (*Monday - Sunday*). Penulis mengkodekan *weekend* sebagai 0 dan *weekday* sebagai 1. Demikian juga *Monday* sampai *Sunday* dikodekan sebagai 1 sampai 7. Penulis menambahkan satu atribut lagi, yaitu *hour*. Atribut ini digunakan untuk memperoleh informasi mengenai jam saat pengambilan data berlangsung, yang dapat diekstrak dari atribut Date. Pada penelitian ini, nilai *ev1* dan *rv2* tidak digunakan sebagai bagian dari metode penelitian. Atribut selengkapnya yang digunakan dalam penelitian ini ditunjukkan pada Apendiks A.



Gambar 1. (a) Penempatan sensor pada lantai satu. Terlihat node koordinator © untuk mengumpulkan semua data sensor diletakkan di tengah ruangan, (b) Penempatan sensor pada lantai dua. [13]

Berdasarkan atribut waktu yang disediakan dalam dataset, diperoleh keterangan bahwa pengumpulan data dimulai dari tanggal 11 Januari 2016 pada pukul 17.00 sampai dengan 27 Mei 2016 pada pukul 18.00. Dataset tersebut terdiri dari 19.735 baris (menyatakan banyaknya data) dan 32 kolom (menyatakan banyaknya atribut). Dari 19.735 data tersebut, Penulis membagi jumlah data sebanyak 60% (11.841 data) sebagai data latihan (*training*), 20% (3.947 data) sebagai data validasi (*validation*) dan sisanya sebanyak 20%

(3.947 data) sebagai data uji (*testing*). Pola konsumsi total energi listrik (*Appliances*) selama periode pengambilan data dan pola lebih detail untuk satu minggu pertama diperlihatkan pada Gambar 2.



Gambar 2. (a) Pola konsumsi energi listrik selama periode penuh, (b) Tinjauan lebih detail untuk pola konsumsi energi selama satu minggu pertama

## 2.2. Analisis Korelasi

Setelah mengetahui dataset yang digunakan, langkah selanjutnya adalah menghitung nilai korelasi. Untuk memperoleh model yang tepat, perhitungan koefisien korelasi antaratribut yang menyusun dataset menjadi penting untuk dilakukan, terutama ketika atribut tersebut tersedia dalam jumlah yang banyak. Ada kemungkinan bahwa tidak semua atribut tersebut relevan untuk memprediksi suatu target (dalam penelitian ini, target adalah *Appliances*). Dengan memilih atribut yang berkorelasi dengan target serta menyisihkan bagian yang tidak relevan, proses untuk membangun serta melatih model untuk mengenali target akan lebih efektif. Manfaat lain dari proses seleksi ini adalah mengurangi kompleksitas atau bahkan mengurangi ukuran file akhir dari model yang dibangun, terutama ketika model tersebut akan diterapkan pada alat yang memiliki sumber daya terbatas seperti mikrokontroler, *single-board computer*, FPGA, ASIC, dan sebagainya.

Jika suatu data time series memiliki vektor  $X = (x_1, x_2, \dots, x_n)$  dan terdapat vektor lainnya  $Y = (y_1, y_2, \dots, y_n)$ , maka koefisien korelasi  $r$  dari kedua vektor tersebut dihitung dengan menggunakan persamaan berikut [14]:

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}} \quad (1)$$

Nilai  $r$  pada persamaan (1) juga dikenal sebagai koefisien korelasi Pearson. Ketika  $0 < r < 1$ , maka dikatakan kedua atribut memiliki korelasi positif, dan ketika  $-1 < r < 0$  maka dikatakan sebagai korelasi negatif. Nilai 0 menunjukkan tidak ada korelasi antaratribut. Ketika nilai mutlak dari  $r$  mendekati 1, maka kedua atribut memiliki korelasi yang semakin besar. Contoh korelasi dari beberapa atribut dalam dataset diperlihatkan pada Tabel 2, dan tabel koefisien korelasi selengkapnya dapat dilihat pada Apendiks B.

Berdasarkan data pada Tabel 2, terlihat korelasi positif antara konsumsi energi listrik oleh berbagai peralatan (*Appliances*) dengan penggunaan alat penerangan (*lights*). Demikian juga untuk sensor T1 dan RH\_1 memiliki korelasi positif terhadap *Appliances*, walaupun korelasinya rendah. Korelasi yang sama juga terlihat pada Temperatur udara luar T\_out dan kecepatan angin *Windspeed*. Sebaliknya, korelasi negatif untuk *Appliances* tampak pada RH\_9 dan *WeekStatus*. Hal ini wajar, sebab ketika hari libur, pemakaian peralatan listrik justru meningkat. Penjelasan lebih detail mengenai korelasi antaratribut dapat merujuk pada [13].

Tabel 2. Koefisien korelasi untuk beberapa atribut dalam dataset

	Appliances	lights	T1	RH_1	T9	RH_9	T_out	Windspeed	Visibility	WeekStatus
Appliances	1,00	0,20	0,06	0,09	0,01	-0,05	0,10	0,09	0,00	-0,02
lights	0,20	1,00	-0,02	0,11	-0,16	0,00	-0,07	0,06	0,02	0,05
T1	0,06	-0,02	1,00	0,16	0,84	0,07	0,68	-0,09	-0,08	-0,01
RH_1	0,09	0,11	0,16	1,00	0,12	0,76	0,34	0,20	-0,02	0,02
T9	0,01	-0,16	0,84	0,12	1,00	0,00	0,67	-0,18	-0,10	0,01
RH_9	-0,05	0,00	0,07	0,76	0,00	1,00	0,22	0,24	0,00	-0,03
T_out	0,10	-0,07	0,68	0,34	0,67	0,22	1,00	0,19	-0,08	-0,04
Windspeed	0,09	0,06	-0,09	0,20	-0,18	0,24	0,19	1,00	0,00	-0,09
Visibility	0,00	0,02	-0,08	-0,02	-0,10	0,00	-0,08	0,00	1,00	0,06
WeekStatus	-0,02	0,05	-0,01	0,02	0,01	-0,03	-0,04	-0,09	0,06	1,00

### 2.3. Pemilihan Input Fitur dan Model Prediktor

Candanedo, dkk., menggunakan metode *recursive feature elimination* (REF) untuk melakukan pemeringkatan terhadap berbagai atribut yang akan digunakan sebagai input model prediktor. Pada penelitian ini, Penulis menggunakan pendekatan yang berbeda. Berdasarkan nilai koefisien korelasi antara *Appliances* dengan semua atribut lainnya, maka dilakukan penyaringan besarnya nilai korelasi dengan ambang batas (*threshold*) yang diperbolehkan sebagai input untuk model. Tidak semua atribut akan diikuti sebagai input ke model. Jika diatur nilai ambang koefisien  $r > 0,01$ , maka terdapat sebanyak 26 input yang akan terpilih. Jika ditentukan  $r > 0,02$ , maka sebanyak 21 atribut terpilih. Demikian seterusnya sampai nilai  $r > 0,07$ . Tabel 3 menunjukkan atribut terpilih selengkapnya.

Tabel 3. Atribut yang terpilih berdasarkan ambang batas koefisien korelasi

No	$r > 0,01$	$r > 0,02$	$r > 0,03$	$r > 0,04$	$r > 0,05$	$r > 0,06$	$r > 0,07$
1	lights	lights	lights	lights	lights	lights	lights
2	T1	T1	T1	T1	T1	RH_1	RH_1
3	RH_1	RH_1	RH_1	RH_1	RH_1	T2	T2
4	T2	T2	T2	T2	T2	RH_2	T3
5	RH_2	RH_2	RH_2	RH_2	RH_2	T3	T6
6	T3	T3	T3	T3	T3	T6	RH_6
7	RH_3	RH_3	RH_3	T4	T6	RH_6	RH_8
8	T4	T4	T4	T6	RH_6	RH_8	T_out
9	RH_4	T6	T6	RH_6	RH_7	T_out	RH_out
10	T5	RH_6	RH_6	RH_7	RH_8	RH_out	Windspeed
11	T6	T7	RH_7	RH_8	RH_9	Windspeed	NSM
12	RH_6	RH_7	T8	RH_9	T_out	NSM	hour
13	T7	T8	RH_8	T_out	RH_out	hour	
14	RH_7	RH_8	RH_9	Press_mm_hg	Windspeed		
15	T8	RH_9	T_out	RH_out	NSM		
16	RH_8	T_out	Press_mm_hg	Windspeed	hour		
17	T9	Press_mm_hg	RH_out	NSM			
18	RH_9	RH_out	Windspeed	hour			
19	T_out	Windspeed	NSM				
20	Press_mm_hg	NSM	hour				
21	RH_out	hour					
22	Windspeed						
23	Tdewpoint						
24	NSM						
25	WeekStatus						
26	hour						

Sebagai seleksi awal, semua atribut yang terpilih untuk setiap nilai ambang akan dilatih dan dipilih yang terbaik. Sebelum dilatih, data input dan target terlebih dahulu dinormalisasi, sehingga semua data masuk dalam rentang antara 0 dan 1. Pada tahapan ini, jumlah neuron dari model akan ditetapkan sama. Setelah diketahui jenis input yang paling baik, maka langkah selanjutnya adalah mengotimalkan model dengan mengatur berbagai parameter dari model tersebut.

## 2.5. Long Short-Term Memory (LSTM)

Kombinasi input pada Tabel 3 akan dilatih menggunakan LSTM. Struktur dari LSTM diperlihatkan seperti pada Gambar 3. Input dan output jaringan pada struktur LSTM dideskripsikan sebagai berikut [6]:

$$F_t = \sigma(W_f \cdot [H_{t-1}, X_t] + b_f) \quad (2)$$

$$I_t = \sigma(W_i \cdot [H_{t-1}, X_t] + b_i) \quad (3)$$

$$\tilde{C}_t = \tanh(W_c \cdot [H_{t-1}, X_t] + b_c) \quad (4)$$

$$C_t = F_t * C_{t-1} + I_t * \tilde{C}_t \quad (5)$$

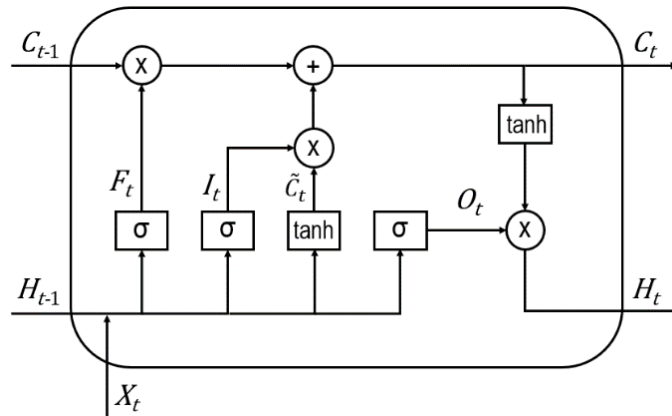
$$O_t = \sigma(W_o \cdot [H_{t-1}, X_t] + b_o) \quad (6)$$

$$H_t = O_t * \tanh(C_t) \quad (7)$$

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (8)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (9)$$

Dengan  $W_f$ ,  $W_i$ ,  $W_c$  dan  $W_o$  adalah bobot input,  $b_f$ ,  $b_i$ ,  $b_c$ , dan  $b_o$  adalah bias,  $t$  adalah waktu saat ini,  $t-1$  mewakili satu keadaan sebelumnya,  $X$  adalah input,  $H$  adalah output, dan  $C$  adalah status dari sel. Notasi  $\sigma$  adalah fungsi sigmoid, yang menghasilkan input antara 0 dan 1. Nilai 0 berarti tidak membiarkan nilai apapun lolos ke tahapan berikutnya, sedangkan nilai 1 berarti membiarkan sepenuhnya output memasuki tahapan berikutnya. Fungsi tangen hiperbolik ( $\tanh$ ) digunakan untuk mengatasi hilangnya gradien selama proses pelatihan, yang umumnya terjadi pada struktur RNN.



Gambar 3. Struktur LSTM [6]

Proses pemodelan dan pengujian dilakukan dengan menggunakan bahasa pemrograman Python. Penelitian ini menggunakan framework Keras dengan Tensorflow sebagai backend. Beberapa pustaka Python lainnya yang dipergunakan, yaitu Scikit-learn, Pandas, Matplotlib, Numpy, dan Seaborn. Model dilatih dengan metode *backpropagation*, dengan menggunakan algoritma optimasi *Adam*.

## 2.5. Evaluasi Kinerja Prediktor

LSTM yang digunakan sebagai model prediktor pada penelitian ini menggunakan *root mean squared error* (RMSE) dan *mean average error* (MAE), sebagai parameter evaluasi. RMSE dan MAE masing-masing dapat dihitung dengan menggunakan persamaan (10) dan (11):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (10)$$

$$MAE = \frac{\sum_{i=1}^n |Y_i - \hat{Y}_i|}{n} \quad (11)$$

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Hasil Pemilihan Input

Setiap kelompok input pada Tabel 3 akan dilatih, dan selanjutnya ditentukan kelompok input mana yang memberikan hasil paling baik. Sebagai seleksi awal, LSTM dengan dengan jumlah 8 neuron ditetapkan untuk semua kelompok input, dan *lookback* input data ke model ditetapkan sebanyak 1. Lookback menunjukkan panjang data yang akan dipakai untuk memprediksi kondisi selanjutnya. Lookback 1 berarti satu buah data saat ini digunakan untuk memprediksi satu buah data di masa depan. Contoh lain, jika *lookback* ditetapkan dengan nilai 4, maka sebanyak 4 data digunakan untuk memprediksi satu data kedepan. Karena setiap data diambil dengan interval 10 menit, maka skenario tersebut berarti 10 menit yang telah saat ini digunakan utk memprediksi 10 menit kedepan. Untuk mendapatkan wawasan yang lebih baik, maka percobaan dilakukan sebanyak tiga kali untuk tiap-tiap kategori, dan dipilih hasil terbaik dari tiap-tiap kategori tersebut. Berdasarkan seleksi awal, diperoleh hasil seperti pada Tabel 4.

Tabel 4. Hasil seleksi jenis input untuk model LSTM

Koef. korelasi	RMSE	MAE
$r > 0,01$	65,373	28,990
$r > 0,02$	66,137	40,965
$r > 0,03$	64,446	29,757
$r > 0,04$	72,525	55,316
$r > 0,05$	64,577	32,687
$r > 0,06$	62.196	29.997
$r > 0,07$	63.954	35.948

Berdasarkan hasil yang dirangkum pada Tabel 4, error paling rendah diperoleh untuk input model dengan koefisien  $r > 0.06$ , yang menghasilkan error terkecil yaitu masing-masing sebesar 62.196 dan 29.997 untuk RMSE dan MAE. Terdapat 13 jenis input untuk  $r > 0.06$ , yaitu *lights*, *RH\_1*, *T2*, *RH\_2*, *T3*, *T6*, *RH\_6*, *RH\_8*, *T\_out*, *RH\_out*, *Windspeed*, *NSM*, dan *hour*. Hasil dari tahapan ini menunjukkan bahwa belum tentu semua atribut yang direncanakan dalam tahapan awal akan memberikan pengaruh pada kinerja dari model prediktor. Atribut yang tidak terlalu signifikan berpengaruh kepada target tidak akan memberikan kontribusi secara optimal. Sebaliknya, jika terlalu banyak menghilangkan jumlah input, maka akan membuat model tidak dapat belajar secara baik karena informasi yang diterima terlalu sedikit. Selanjutnya, ketigabelas atribut ini yang akan dilanjutkan ke tahapan berikutnya.

#### 3.2. Pemilihan Jumlah Neuron

Salah satu hiperparameter yang paling banyak diatur dalam pemodelan menggunakan LSTM adalah mengatur jumlah neuron. Pada penelitian ini, neuron akan dikombinasikan dengan jumlah tertentu, dan diamati hasil akhir yang diperoleh. Setiap tahapan akan diulang sebanyak tiga kali percobaan, dan hasil terbaik ditunjukkan seperti pada Tabel 5.

Tabel 5. Hasil pemilihan jumlah neuron

Jumlah Neuron	RMSE	MAE
8	62.196	29.997
16	67.042	30.754
32	64.542	38.953
64	64.294	38.660
80	63,341	28,218
100	65.169	29.843
128	65.169	41.008

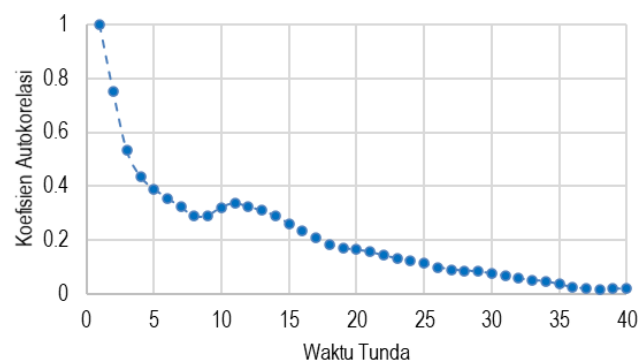
Berdasarkan hasil percobaan, diperoleh hasil bahwa meningkatkan jumlah neuron belum tentu akan meningkatkan kinerja dari model. Terlihat bahwa jumlah neuron sebanyak 8 masih menghasilkan nilai error yang lebih kecil, yaitu masing-masing sebesar 62.196 dan 29.997 untuk RMSE dan MAE. Jumlah neuron kecil memiliki manfaat tersendiri. Dari sisi ukuran file, semakin kecil jumlah neuron, maka semakin berkurang juga ukuran file dan kompleksitas perhitungan yang harus dilakukan oleh model. Hal ini tentu sangat menguntungkan terutama jika model tersebut harus dijalankan di mesin yang memiliki sumber daya terbatas, misalnya mikrokontroler, *single-board computer*, atau perangkat bergerak lainnya. Dengan demikian, arsitektur model dengan 8 neuron ini yang akan dipakai pada tahapan berikutnya.

### 3.3. Pemilihan Jumlah *Lookback*

Dalam pemodelan time-series, pemilihan yang tepat terhadap banyaknya data saat ini (atau yang telah berlalu) untuk memprediksi data di masa mendatang dapat meningkatkan kinerja dari model. Banyaknya data yang telah berlalu dikenal sebagai *lookback*. *Lookback* dalam penelitian ini diatur mulai dari 1 sampai dengan 10. Skenario ini menyatakan bahwa Penulis melakukan kombinasi dari 1 sampai 10 data saat ini (dan yang telah lampau) untuk memprediksi satu data di masa depan. Karena setiap data memiliki interval 10 menit, maka untuk *lookback* 1 mengindikasikan bahwa nilai 10 menit saat ini digunakan untuk memprediksi nilai 10 menit kedepan. *Lookback* 10 berarti Penulis menggunakan 10 data untuk memprediksi 1 data di depan. Hasil dari proses ini ditunjukkan seperti pada Tabel 5.

Tabel 6. Hasil pemilihan jumlah neuron

Lookback	RMSE	MAE
1	62.196	29.997
2	65,961	41,222
3	71,726	54,310
4	64.347	40.742
5	61,559	31,919
6	67,819	36,252
7	60.992	28.278
8	61,628	35,404
9	61.284	30.879
10	79,857	68,069



Gambar 4. Nilai fungsi autokorelasi terhadap waktu tunda

Sebagai gambaran dalam pemilihan nilai *lookback* ini, nilai fungsi autokorelasi dari suatu data time-series dapat diterapkan. Jika kondisi saat ini  $y_t$  disederhanakan sebagai  $A$ , dan kondisi dimasa depan  $y_{t+k}$  sebagai  $B$ , dengan  $k$  adalah waktu tunda, maka fungsi autokorelasi dihitung menggunakan persamaan (12):

$$cor(A, B) = \frac{cov(A, B)}{std(A) std(B)} \quad (12)$$



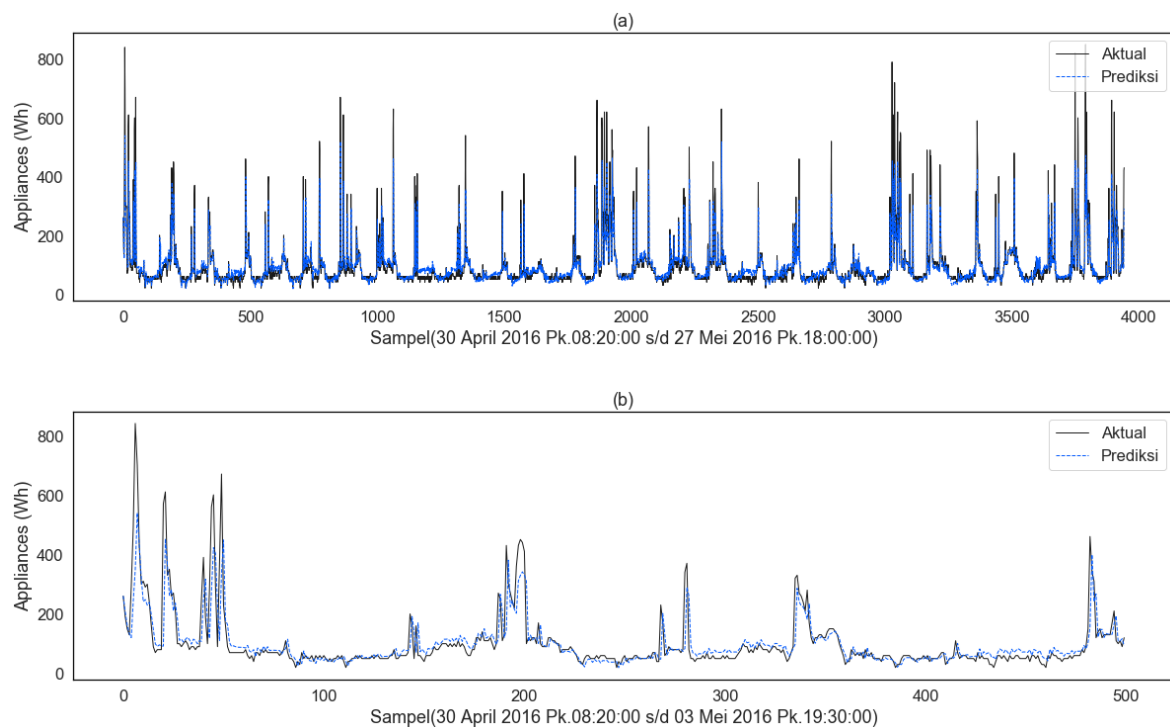
Dengan  $cov(A,B)$  adalah kovarian antara  $A$  dan  $B$ , sedangkan  $std(A)$  dan  $std(B)$  masing-masing adalah standar deviasi dari  $A$  dan dari  $B$ .

Gambar 4 menunjukkan koefisien autokorelasi dari atribut *Appliances* terhadap waktu tunda (*time lag*). Pada gambar tersebut, waktu tunda lebih dari 10 tidak memiliki korelasi yang signifikan. Pada penelitian ini, nilai *lookback* 7 dianggap sebagai yang paling optimal. Nilai *lookback* kecil tidak menjamin untuk cukup menyimpan memori jangka panjang yang menjadi keunggulan dari LSTM, sedangkan *lookback* terlalu besar membuat model terlalu banyak mengolah informasi berulang yang belum tentu optimal.

### 3.4 Gambaran Nilai Aktual dengan Nilai Prediksi

Seperti yang telah dibahas pada bagian 2.1, dataset memiliki 19.735 baris data. Sebanyak 60% data pertama digunakan sebagai data latih (11.841 baris), 20% data berikutnya sebagai data validasi (3.947 baris), dan 20% data terakhir sebagai data tes (3.947 baris). Jika merujuk pada waktu pengambilan data dalam dataset, maka data latih dimulai dari tanggal 11 Januari 2016 pukul 17.00 sampai dengan tanggal 02 April 2016 pukul 22.20. Data validasi dimulai pada tanggal 02 April 2016 pukul 22.30 sampai dengan 30 April 2016 pukul 08.10. Terakhir, data tes dimulai pada tanggal 30 April 2016 pukul 08.20 sampai dengan 27 Mei 2016 pukul 18.00.

Ploting antara hasil prediksi dengan nilai actual pada data tes dan ploting 500 data tes pertama diperlihatkan seperti pada Gambar 5. Garis kontinu menunjukkan nilai actual, sedangkan garis putus-putus menunjukkan hasil prediksi. Berdasarkan gambar tersebut, terlihat bahwa secara umum hasil prediksi telah mengikuti pola nilai actual. Fluktuasi untuk nilai Wh rendah dapat diikuti dengan baik. Namun demikian, model belum secara sempurna menangkap lonjakan nilai Wh yang tinggi.



Gambar 5. (a) Perbandingan nilai actual dan nilai prediksi untuk satu periode penuh, (a) perbandingan untuk 500 data tes pertama

### 3.5 Perbandingan dengan Penelitian Sebelumnya

Peneliti memiliki pendekatan yang berbeda dalam membagi data dengan apa yang dilakukan oleh Candanedo, dkk. Peneliti membagi dataset dengan skenario 60% data pertama sebagai data latih, 20% data berikutnya sebagai data validasi, dan 20% data terakhir sebagai data tes atau data uji. Dengan demikian, seluruh segmen data merupakan data yang terurut berdasarkan time-series tertentu. Berbeda dengan Penulis, Candanedo, dkk. membagi dataset sebanyak 75% sebagai data latih, dan sisanya 25% sebagai data uji. Namun demikian, pembagian data ini dilakukan secara acak, tidak menurut rangkaian time-series tertentu. Dengan demikian, data latih dan data uji yang untuk model prediktor yang dilakukan oleh Penulis dengan

peneliti sebelumnya adalah berbeda. Sebagai perbandingan, Penulis memperoleh hasil akhir masing-masing sebesar 60,992 RMSE dan 28, 278 MAE untuk data yang diujikan. Disisi lain, Candanedo, dkk menggunakan empat prediktor yaitu LM (93,18 RMSE dan 51,91 MAE), SVM (70,74 RMSE dan 31), GBM (66,65 RMSE dan 35,22 MAE), dan RF (68,48 RMSE dan 31,85 MAE).

#### 4. KESIMPULAN

Penelitian ini menerapkan Long Short-Term memory (LSTM) untuk memprediksi besarnya penggunaan energi listrik dari suatu rumah hunian. Berdasarkan penelitian yang telah dilakukan, diperoleh hasil mengenai jenis atribut yang akan dipakai dan arsitektur dari model prediktor. Koefisien korelasi antara target dengan atribut lainnya dipilih lebih dari 0,07. Untuk skenario ini, sebanyak 13 jenis atribut yang digunakan sebagai input dari LSTM. Berdasarkan penelitian, 8 neuron dalam LSTM dengan *lookback* sebanyak 7 memiliki kinerja paling baik. Besarnya nilai error terhadap data uji masing-masing sebesar 60.992 dan 28.278 untuk RMSE dan MAE.

#### DAFTAR RUJUKAN

- [1] PLN, "Laporan Tahunan PT Perusahaan Listrik Negara (Persero) 2018." PT.PLN (Persero), 2018, [Online]. Available: <https://web.pln.co.id/stakeholder/laporan-tahunan>.
- [2] K. S. Cetin, P. C. Tabares-Velasco, and A. Novoselac, "Appliance daily energy use in new residential buildings: Use profiles and variation in time-of-use," *Energy Build.*, vol. 84, pp. 716–726, Dec. 2014, doi: 10.1016/j.enbuild.2014.07.045.
- [3] N. Arghira, L. Hawarah, S. Ploix, and M. Jacomino, "Prediction of appliances energy use in smart homes," *Energy*, vol. 48, no. 1, pp. 128–134, Dec. 2012, doi: 10.1016/j.energy.2012.04.010.
- [4] L. M. Candanedo and V. Feldheim, "Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models," *Energy Build.*, vol. 112, pp. 28–39, Jan. 2016, doi: 10.1016/j.enbuild.2015.11.071.
- [5] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [6] T. Li, M. Hua, and X. Wu, "A Hybrid CNN-LSTM Model for Forecasting Particulate Matter (PM2.5)," *IEEE Access*, vol. 8, pp. 26933–26940, 2020, doi: 10.1109/ACCESS.2020.2971348.
- [7] Á. Carmona-Poyato, N. L. Fernández-García, F. J. Madrid-Cuevas, and A. M. Durán-Rosal, "A new approach for optimal time-series segmentation," *Pattern Recognit. Lett.*, Apr. 2020, doi: 10.1016/j.patrec.2020.04.006.
- [8] H. Abbasimehr, M. Shabani, and M. Yousefi, "An optimized model using LSTM network for demand forecasting," *Comput. Ind. Eng.*, vol. 143, p. 106435, May 2020, doi: 10.1016/j.cie.2020.106435.
- [9] Y. Tian, K. Zhang, J. Li, X. Lin, and B. Yang, "LSTM-based traffic flow prediction with missing data," *Neurocomputing*, vol. 318, pp. 297–305, Nov. 2018, doi: 10.1016/j.neucom.2018.08.067.
- [10] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A Search Space Odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017, doi: 10.1109/TNNLS.2016.2582924.
- [11] X. Bai, "Text classification based on LSTM and attention," in *2018 Thirteenth International Conference on Digital Information Management (ICDIM)*, Sep. 2018, pp. 29–32, doi: 10.1109/ICDIM.2018.8847061.
- [12] A. H. Mirza and S. Cosan, "Computer network intrusion detection using sequential LSTM Neural Networks autoencoders," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, May 2018, pp. 1–4, doi: 10.1109/SIU.2018.8404689.
- [13] L. M. Candanedo, V. Feldheim, and D. Deramaix, "Data driven prediction models of energy use of appliances in a low-energy house," *Energy Build.*, vol. 140, pp. 81–97, Apr. 2017, doi: 10.1016/j.enbuild.2017.01.083.
- [14] Q. Tao, F. Liu, Y. Li, and D. Sidorov, "Air Pollution Forecasting Using a Deep Learning Model Based on 1D Convnets and Bidirectional GRU," *IEEE Access*, vol. 7, pp. 76690–76698, 2019, doi: 10.1109/ACCESS.2019.2921578.

**Apendiks A.**

Atribut dataset yang digunakan dalam penelitian

Atribut	Satuan	Deskripsi	Atribut	Satuan	Deskripsi	Atribut	Satuan	Deskripsi
Date	dd:mm:yy hh:mm:ss	Tanggal Waktu	RH_4	%	Kelembaban ruang kantor	RH_9	%	Kelembaban kamar orang tua
Appliances	Wh	Penggunaan energi total	T5	°C	Temperatur kamar mandi	T_out	°C	Temperatur outdoor
lights	Wh	Energi penerangan	RH_5	%	Kelembaban kamar mandi	Press_mm_hg	mmHg	Tekanan outdoor
T1	°C	Temperatur dapur	T6	°C	Temperatur sisi luar utara	Windspeed	m/s	Kecepatan angin
RH_1	%	Kelembaban dapur	RH_6	%	Kelembaban sisi luar utara	Visibility	km	Jarak pandang
T2	°C	Temperatur ruang tamu	T7	°C	Temperatur ruang setrika	Tdewpoint	°C	Titik embun
RH_2	%	Kelembaban ruang tamu	RH_7	%	Kelembaban ruang setrika	NSM	detik	Jumlah detik dari tengah malam
T3	°C	Temperatur ruang cuci	T8	°C	Temperatur kamar anak	WeekStatus	-	Hari kerja atau hari libur
RH_3	%	Kelembaban ruang cuci	RH_8	%	Kelembaban kamar anak	Day_of_week	-	Hari Senin s/d Minggu
T4	°C	Temperatur ruang kantor	T9	°C	Temperatur kamar orang tua	hour	jam	Jam saat pengambilan data

**Apendiks B.**

Koefisien korelasi untuk setiap atribut pada pada dataset. Nilai korelasi yang kurang dari 0,001 akan dibulatkan menjadi nol.

