



**UNIVERSITAS INDONESIA**

**EVALUASI METODE PREDIKSI DAN PENGARUH DATA  
CUACA TERHADAP PENGGUNAAN LISTRIK GEDUNG  
KAMPUS**

**SKRIPSI**

**Yogie Wisesa**

**1806148851**

**FAKULTAS TEKNIK  
PROGRAM STUDI TEKNIK KOMPUTER**

**DEPOK**

**JUNI 2022**



**UNIVERSITAS INDONESIA**

**EVALUASI METODE PREDIKSI DAN PENGARUH DATA  
CUACA TERHADAP PENGGUNAAN LISTRIK GEDUNG  
KAMPUS**

**SKRIPSI**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar  
Sarjana Teknik**

**Yogie Wisesa**

**1806148851**

**FAKULTAS TEKNIK  
PROGRAM STUDI TEKNIK KOMPUTER**

**DEPOK  
JUNI 2022**

## **HALAMAN PERNYATAAN ORISINALITAS**

**Skripsi ini adalah hasil karya saya sendiri, dan  
semua sumber baik yang dikutip maupun dirujuk  
telah saya nyatakan dengan benar.**

**Nama : Yogie Wisesa**

**NPM : 1806200066**

**Tanda Tangan :  \_\_\_\_\_**

**Tanggal : 24 Juni 2022**

## LEMBAR PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Yogie Wisesa

NPM : 1806148851

Program Studi : Teknik Komputer

Judul Skripsi : Evaluasi Metode Prediksi dan Pengaruh Data Cuaca Terhadap  
Penggunaan Listrik Gedung Kampus

**Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagaibagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana mata kuliah Skripsi pada Program Studi Teknik Komputer Fakultas Teknik Universitas Indonesia.**

## DEWAN PENGUJI

Pembimbing : I Gde Dharma Nugraha S.T., M.T., Ph.D.



Penguji : Dr. Prima Dewi Purnamasari S.T., M.Sc.



Penguji : Dr. Ruki Harwahyu S.T., M.T., M.Sc.



Ditetapkan di : Fakultas Teknik

Tanggal : 18 Juli 2022

## KATA PENGANTAR

Puji syukur saya panjatkan ke hadirat Tuhan Yang Maha Esa, karena berkat rahmat dan karunia-Nya saya dapat menyelesaikan buku skripsi ini. Penulisan skripsi ini dilakukan dengan tujuan untuk menuntaskan salah satu syarat kelulusan gelar Sarjana Teknik di Universitas Indonesia. Penulisan skripsi ini bertujuan untuk melakukan perancangan metode prediksi dan pengaruh cuaca terhadap penggunaan listrik gedung kampus. Penulis berterima kasih kepada:

1. Bapak I Gde Dharma Nugraha S.T., M.T., Ph.D. selaku dosen pembimbing mata kuliah seminar dan skripsi yang sudah membantu dan membimbing penulis dalam penelitian dan penulisan buku skripsi ini selama satu tahun ke belakang.
2. Ibu Dr. Eng. Mia Rizkinia, S.T., M.T selaku dosen pembimbing akademis penulis dari awal perkuliahan hingga akhir yang banyak membantu penulis dalam menyelesaikan studi.
3. Kedua orang tua penulis yang selalu mendukung dan memotivasi penulis untuk menyelesaikan studi
4. Seluruh pihak yang sudah membantu penulis untuk menyelesaikan skripsi ini.

Penulis memohon maaf yang sebesar-besarnya jika ada salah kata dan perbuatan sebelumnya.

Depok, Juni 2022



Yogie Wisesa

## HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Indonesia, saya yang tertandatangan di bawah ini:

Nama : Yogie Wisesa  
NPM : 1806148851  
Program Studi : Teknik Komputer  
Fakultas : Fakultas Teknik  
Jenis Karya : Skripsi

demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Indonesia **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul:

### EVALUASI METODE PREDIKSI DAN PENGARUH DATA CUACA TERHADAP PENGGUNAAN LISTRIK GEDUNG KAMPUS

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Indonesia berhak menyimpan, mengalihmedia/formatkan, mengelola, dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Depok  
Pada tanggal : 24 Juni 2022  
Yang menyatakan



Yogie Wisesa

## ABSTRAK

Nama : Yogie Wisesa  
Program Studi : Teknik Komputer  
Judul : Evaluasi Metode Prediksi dan Pengaruh Data Cuaca Terhadap Penggunaan Listrik Gedung Kampus

Indonesia merupakan negara tropis yang memiliki dua musim yaitu musim kemarau dan musim hujan. Akan tetapi, kondisi cuaca setiap harinya tidak dapat diprediksi karena pengaruh perubahan iklim yang terjadi. Hujan bisa terjadi kapan saja di musim kemarau dan begitu juga sebaliknya. Hal ini menyebabkan perbedaan suhu di setiap harinya. Penelitian ini bertujuan untuk menganalisa pengaruh perubahan suhu dan cuaca terhadap penggunaan listrik di gedung kampus pendidikan khususnya penggunaan pendingin ruangan. Untuk itu dilakukan pemilihan fitur data cuaca menggunakan *Pearson Correlation Coefficient* (PCC) dan *user behaviour modelling* dengan menggunakan dua metode yaitu Long Short Term Memory (LSTM) dan penambahan lapisan *convolutional neural network* (CNN) pada metode LSTM. Setelah dilakukan *modelling*, maka akan dilakukan prediksi dan akurasi akan untuk menentukan metode mana yang lebih cocok untuk kasus ini. Metode gabungan CNN-LSTM mendapatkan skor tertinggi dengan skor  $R^2$  sebesar 0,83 dan MAE sebesar 136,55. Penggunaan data cuaca terbukti memiliki pengaruh terhadap akurasi model. Meskipun cuaca di Indonesia tidak menentu dan penggunaan pendingin ruangan yang tidak teratur, faktor cuaca seperti rata-rata suhu masih memiliki pengaruh terhadap penggunaan pendingin ruangan di gedung institusi pendidikan.

*LSTM; Prediksi; Akurasi;*

## ABSTRACT

Name : Yogie Wisesa  
Study Program : Computer Engineering  
Title : Evaluation of the Prediction Method and the Effect of Weather  
Data on the Use of Electricity in Campus Buildings

Indonesia is a tropical country that has two seasons, namely the dry season and the rainy season. However, daily weather conditions cannot be predicted due to the influence of climate change. Rain can occur at any time in the dry season and vice versa. This causes temperature differences on a daily basis. This study aims to analyze the influence of temperature and weather changes on electricity use in educational campus buildings, especially the use of air conditioning. For this reason, weather data will be selected using *Pearson Correlation Coefficient* (PCC) and *user behavior modeling* using two methods, namely Long Short Term Memory (LSTM) and the addition of a *convolutional neural network* (CNN) layer to an LSTM method. After modeling, predictions will be made and the accuracy will be measured to determine which method is more suitable for this case. The combined CNN-LSTM method got the highest score with an  $R^2$  score of 0.83 and MAE of 136.55. The use of weather data has proven to have an influence on model accuracy. Although the weather in Indonesia is erratic and the use of air conditioning is unmonitored, weather factors such as average temperature and rain still have an influence on the use of air conditioning in educational institution buildings.

*LSTM; Forecast; Prediction; Accuracy;*



## DAFTAR ISI

<b>HALAMAN PERNYATAAN ORISINALITAS .....</b>	<b>ii</b>
<b>KATA PENGANTAR.....</b>	<b>iv</b>
<b>HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS.....</b>	<b>v</b>
<b>ABSTRAK .....</b>	<b>vi</b>
<b>ABSTRACT .....</b>	<b>vii</b>
<b>DAFTAR ISI.....</b>	<b>viii</b>
<b>BAB 1 PENDAHULUAN .....</b>	<b>1</b>
1.1. Latar Belakang .....	1
1.2. Tujuan Penelitian .....	2
1.3. Manfaat Penelitian .....	2
1.4. Metodologi Penelitian .....	3
1.5. Sistematika Penulisan .....	4
<b>BAB 2 LANDASAN TEORI .....</b>	<b>5</b>
2.1. Prediksi Penggunaan Listrik .....	5
2.2. <i>Time Series</i> .....	6
2.2.1. Komponen <i>Time Series</i> .....	6
2.2.2. <i>Lag Function</i> .....	7
2.3. <i>Deep Learning</i> .....	7
2.3.1. Recurrent Neural Network .....	8
2.3.2. Convolutional Neural Network .....	9
2.4. Long Short-term Memory .....	10
2.5. Structured Query Language .....	12
2.6. Fungsi Koefisien Korelasi.....	13
2.7. <i>Mean Absolute Error</i> .....	14
2.8. <i>Coefficient of Determination</i> .....	14
<b>BAB 3 PERANCANGAN SISTEM.....</b>	<b>16</b>
3.1. Diagram alur sistem .....	16
3.1.1. Model LSTM Tanpa Data Cuaca .....	16
3.1.2. Model LSTM dengan Data Cuaca.....	17
3.1.3. Model CNN-LSTM.....	18
3.2. Dataset.....	19
3.3. Preprocessing dan Pembuatan Fitur .....	20
3.4. Skenario Pengujian.....	24

3.5. Rencana Evaluasi .....	25
<b>BAB 4 HASIL PENELITIAN .....</b>	<b>26</b>
4.1. Implementasi Model.....	26
4.2. Hasil Pengujian .....	27
4.2.1. Metode Stacked LSTM dengan Data Penggunaan Listrik.....	28
4.2.2. Metode Stacked LSTM Dengan Data Cuaca .....	30
4.2.3. Metode CNN-LSTM .....	32
4.3. Percobaan Tambahan .....	34
4.3.1. Pengubahan Jumlah <i>Epoch</i> .....	35
4.3.2. Pengubahan Jumlah <i>Lag</i> dan <i>Epoch</i> .....	37
<b>BAB 5 PENUTUP.....</b>	<b>39</b>
5.1. Kesimpulan .....	39
5.2. Saran.....	40
<b>DAFTAR PUSTAKA .....</b>	<b>41</b>

## DAFTAR GAMBAR

Gambar 2.1 Struktur lapisan CNN [10] .....	9
Gambar 2.2 Struktur lapisan LSTM [13] .....	11
Gambar 3.1 Diagram alur metode LSTM tanpa data cuaca .....	16
Gambar 3.2 Diagram alur model LSTM .....	17
Gambar 3.3 Diagram alur model CNN-LSTM .....	18
Gambar 3.4 Grafik batang perbandingan nilai PCC setiap fitur .....	23
Gambar 4.1 Grafik <i>Validation Loss</i> dari model Stacked LSTM tanpa data cuaca	28
Gambar 4.2 Grafik penggunaan daya dan hasil prediksi .....	29
Gambar 4.3 Grafik <i>validation loss</i> model Stacked LSTM dengan data cuaca .....	30
Gambar 4.4 Grafik penggunaan daya sebenarnya serta hasil prediksi model.....	31
Gambar 4.5 Grafik <i>validation loss</i> model CNN-LSTM .....	32
Gambar 4.6 Grafik penggunaan daya sebenarnya dengan hasil prediksi model CNN-LSTM .....	33
Gambar 4.7 Grafik prediksi CNN-LSTM tanpa seleksi PCC .....	34
Gambar 4.8 Grafik <i>loss</i> CNN-LSTM dengan epoch 150.....	35
Gambar 4.9 Grafik hasil prediksi CNN-LSTM dengan epoch 150 .....	36
Gambar 4.10 Grafik <i>loss</i> CNN-LSTM dengan <i>lag</i> 14 dan <i>epoch</i> 100.....	37
Gambar 4.11 Grafik Hasil Prediksi CNN-LSTM dengan <i>lag</i> 14 dan <i>epoch</i> 100 .	38

## DAFTAR TABEL

Tabel 3.1 Daftar tanggal data kosong pada dataset.....	20
Tabel 3.2 Tabel korelasi fitur berdasarkan PCC .....	22
Tabel 4.1 Hasil Penilaian Model dan Error.....	27
Tabel 4.2 Tabel Skor Hasil Percobaan Tambahan .....	35

## BAB 1

### PENDAHULUAN

#### 1.1. Latar Belakang

Gedung lembaga pendidikan biasanya terdiri dari beberapa gedung yang dapat dipakai bersama antara beberapa jurusan atau departemen dan juga gedung milik masing-masing. Pada gedung milik sendiri, pemakai dapat lebih mudah untuk mengatur konsumsi listrik karena penggunaan ruangan dapat diatur. Sedangkan pada gedung kuliah bersama, *monitoring* dan penggunaan lebih sulit untuk diatur karena penggunanya selalu berbeda dan tentunya intensitas *penggunaan* ruangan yang lebih tinggi menyebabkan konsumsi listrik menjadi tidak termonitor. Selama ini penggunaan listrik di gedung kuliah bersama kampus belum terukur dan listrik masih digunakan tanpa kesadaran untuk melakukan penghematan. *Profiling* dapat dilakukan pada penggunaan listrik gedung kampus agar penggunaan listrik dapat lebih mudah termonitor dan prosedur penghematan dapat dilakukan dengan lebih mudah dengan mengacu pada hasil *profiling* tersebut.

Cuaca tropis di Indonesia yang tidak menentu menyebabkan perbedaan suhu dan kelembaban di setiap harinya. Secara umum, seharusnya penggunaan listrik saat musim hujan lebih rendah dibandingkan dengan musim kemarau karena adanya perbedaan suhu. Akan tetapi, akibat dari cuaca yang tidak menentu menyebabkan penggunaan listrik yang tidak menentu juga. Seharunya penggunaan pendinginan ruangan dapat dikurangi di saat suhu ruangan tidak terlalu panas dan penggunaan di saat suhu tinggi akan meningkat. Berdasarkan pengalaman penulis, ruangan kelas untuk kuliah di Fakultas Teknik Universitas Indonesia selalu menggunakan pendinginan ruangan sepanjang hari. Untuk itu perlu dibuktikan apakah data penggunaannya mencerminkan dan apakah ada korelasinya dengan data cuaca.

Modeling akan dilakukan dengan dua metode yaitu Stacked LSTM dan CNN-LSTM. Algoritma LSTM adalah algoritma *machine learning* berbasis Recurrent Neural Network (RNN) yang memiliki feedback connection. Tidak seperti algoritma yang lain, LSTM dapat memproses seluruh data secara

sekaligus. Untuk memperoleh hasil yang akurat, LSTM memerlukan data yang banyak karena LSTM akan mempelajari pola yang terbentuk dari data tersebut. Namun pada penelitian kali ini terdapat keterbatasan jumlah data yang tersedia dari kampus sehingga jumlah data hanya setahun lebih 11 hari atau 376 hari.

### **1.2. Tujuan Penelitian**

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mempelajari metode prediksi penggunaan listrik yang tepat untuk gedung institusi pendidikan
2. Membandingkan model prediksi yang dibuat dengan data penggunaan listrik dan data listrik ditambah data cuaca di gedung institusi pendidikan.
3. Menganalisis pengaruh perbedaan cuaca terhadap penggunaan pendinginan ruangan di gedung institusi pendidikan.

### **1.3. Manfaat Penelitian**

Manfaat dari penelitian ini adalah sebagai berikut:

1. Mengetahui metode prediksi yang tepat untuk memprediksi penggunaan listrik di gedung institusi pendidikan.
2. Mengetahui apakah cuaca memiliki pengaruh terhadap penggunaan pendinginan ruangan di gedung institusi pendidikan.

#### 1.4. Metodologi Penelitian

Metode yang digunakan selama penelitian ini adalah sebagai berikut:

1. Studi literatur

Studi literatur dilakukan dengan mempelajari metode prediksi penggunaan listrik yang sudah dilakukan sebelumnya dari berbagai macam sumber serta mempelajari hasil prediksi tersebut.

2. Konsultasi dengan dosen pembimbing

Konsultasi sudah dilakukan dengan dosen pembimbing yang meliputi pertemuan langsung di kampus dan pertemuan secara daring via Zoom atau pesan WhatsApp.

3. Desain Sistem

Desain sistem dilakukan dengan membuat rancangan kode *machine learning* atau model statistika yang akan digunakan untuk *preprocessing* serta pengolahan data.

4. Analisis

Hasil *output* dari sistem dan akan dianalisis untuk mencapai kesimpulan

5. Kesimpulan

Kesimpulan dari penelitian yaitu algoritma mana yang paling efektif untuk prediksi penggunaan listrik dan apakah ada korelasi antara cuaca dan penggunaan pendingin ruangan di gedung institusi pendidikan.

### 1.5. Sistematika Penulisan

Sistematika Penulisan pada penelitian ini dibagi ke dalam 5 bab, yaitu:

#### BAB I        PENDAHULUAN

Bab ini akan membahas latar belakang, tujuan, manfaat, metodologi, dan sistematika penelitian

#### BAB II       LANDASAN TEORI

Bab ini akan membahas *state of the art* hasil studi literatur dan dasar teori dari perancangan sistem yang akan digunakan pada penelitian.

#### BAB III      PERANCANGAN SISTEM

Bab ini akan membahas rancang bangun sistem yang akan digunakan untuk penelitian dan rencana evaluasi.

#### BAB IV      HASIL PENELITIAN

Pada bab terakhir ini akan membahas proses penelitian dan temuan yang ditemukan pada saat penelitian.

#### BAB V       PENUTUP

Pada bab terakhir ini akan berisi kesimpulan dari penelitian ini.



## BAB 2

### LANDASAN TEORI

#### 2.1. Prediksi Penggunaan Listrik

Prediksi adalah sebuah kegiatan yang dilakukan pada masa lampau untuk memperkirakan sesuatu yang akan terjadi di masa depan. Istilah lain dari prediksi adalah *educated guess* atau penebakan secara teredukasi atau ilmiah. Prediksi dilakukan untuk mengasumsikan apa yang kira-kira akan terjadi di masa depan sehingga seseorang bisa melakukan persiapan untuk menghadapi suatu kejadian. Semakin akurat sebuah prediksi maka manfaat dari prediksi tersebut akan semakin terasa dan seseorang akan lebih siap untuk menghadapi masa depan. Dalam melakukan prediksi dibutuhkan dua faktor utama, yaitu basis data masa lampau yang kuat serta metode prediksi yang tepat.

Prediksi penggunaan listrik dilakukan untuk mengetahui bagaimana listrik dikonsumsi. Prediksi dapat dilakukan dalam berbagai macam skala, mulai dari skala kecil seperti rumahan atau bangunan (*microgrid*) atau dalam skala besar seperti perkotaan dan pembangkit listrik. Jangka waktu prediksi penggunaan listrik juga dapat dibagi menjadi tiga macam yaitu jangka pendek, menengah, dan panjang. Prediksi jangka pendek dilakukan dalam satuan hari dan dilakukan untuk mengetahui konsumsi operasional masing-masing perangkat listrik serta biaya yang diperlukan. Data yang diperlukan untuk prediksi jangka pendek juga tidak terlalu besar yaitu cukup satu tahun saja. Prediksi jangka menengah biasanya dilakukan untuk mengetahui kapan waktu yang tepat untuk melakukan pemeliharaan pada sistem kelistrikan. Prediksi jangka panjang dilakukan untuk menentukan strategi penggunaan listrik jangka panjang seperti menentukan apakah membeli listrik dari perusahaan listrik dinilai lebih ekonomis dibandingkan membuat sistem pembangkit sendiri.

Prediksi penggunaan listrik sendiri sudah beberapa kali dilakukan oleh beberapa penelitian dan sudah diterapkan oleh beberapa perusahaan listrik. Metode seperti regresi linear dan algoritma Support Vector Machine sudah digunakan untuk melakukan load forecasting [1]. Jika data yang dimiliki bersifat *time series*,

beberapa penelitian juga menggunakan metodologi ARIMA [2], algoritma XGBoost [3], dan algoritma LSTM [4], [5]. Meskipun beberapa metode ini lebih sering diidentikkan dengan penggunaannya di bidang ekonomi seperti prediksi harga saham dan nilai tukar uang, tetapi metode ini terbukti memiliki akurasi yang cukup baik untuk memprediksi penggunaan listrik.

## 2.2. *Time Series*

*Time series* adalah sebuah jenis data set yang berisikan poin-poin data sekuensial yang diambil datanya secara berkala [6]. Poin data tersebut dapat digambarkan ke dalam grafik  $(x(t), t)$  di mana  $t$  merupakan variabel yang merepresentasikan waktu. Variabel  $x(t)$  dianggap sebagai variabel bebas yang *valuenya* akan didapatkan ketika pengambilan data pada waktu  $t$ . Data *time series* dapat dikatakan sebagai data *univariate* jika hanya terdapat satu variabel yang diambil. Jika variabel yang diambil melebihi satu, maka data tersebut dikatakan sebagai *multivariate*.

### 2.2.1. *Komponen Time Series*

Data time series secara umum memiliki empat komponen utama yang dapat dipisahkan dari data utama yaitu *Trend*, *Seasonal*, *Cyclical*, dan *Irregular* [6].

#### 2.2.1.1. *Trend*

*Trend* adalah istilah dari tendensi data yang menunjukkan kenaikan, penurunan, dan tetapnya nilai data pada sebuah jangka waktu yang panjang. *Trend* juga dapat dikatakan sebagai pergerakan data pada jangka waktu *long term* [3]. Contoh dari *trend* adalah tendensi naiknya nilai properti di wilayah ibu kota atau *trend* turunnya penurunan penggunaan telepon rumah.

#### 2.2.1.2. *Seasonal*

*Seasonal* adalah variasi nilai pada data *time series* yang terjadi dalam satu tahun atau dapat dikatakan per musim. Contoh dari *seasonal* adalah peningkatan penjualan jaket tebal di saat musim dingin atau peningkatan jumlah wisatawan di pantai saat libur musim panas. Variasi

*seasonal* sangat penting bagi pebisnis dan juga pialang saham untuk membuat rencana transaksi ke depan.

### **2.2.1.3. Cyclical**

Variasi *cyclical* berarti variasi nilai data *time series* yang terjadi pada jangka menengah dan perubahannya bersifat berulang dengan tahapan tertentu. Durasi dari sebuah tahapan atau *cycle* dapat berjalan dalam waktu yang cukup lama seperti dua tahun atau lebih, tetapi durasi *cycle* yang sedang berjalan tidak dapat diketahui berapa lama akan berlangsung. Contoh dari data *cyclical* adalah kepadatan kendaraan di jalanan ketika jam berangkat dan pulang kerja serta kembali sepinya kendaraan ketika di luar jam tersebut.

### **2.2.1.4. Irregular**

Variasi *irregular* atau *random* adalah variasi data yang disebabkan oleh faktor yang tidak dapat diprediksi dan tidak biasa sehingga variasi ini tidak memiliki pola yang jelas. Contoh dari faktor tidak biasa tersebut adalah bencana alam, perang, fenomena tertentu, dsb. Variasi *irregular* dapat menimpa seluruh jenis variasi yang sudah dijelaskan sebelumnya menyebabkan masing-masing menjadi contoh variasi *irregular*.

## **2.2.2. Lag Function**

*Lag* atau *delay* atau jeda adalah komponen *time series* yang digunakan untuk mencari korelasi antara sebuah data kepada data-data sebelumnya. Pada analisa data *time series*, umumnya hanya terdapat satu fitur yaitu data itu sendiri. Oleh karena itu, digunakan *lag* untuk mencari korelasi yang biasa disebut *autocorrelation* untuk mencari pola dari data dan menggunakannya untuk memprediksi data selanjutnya.

## **2.3. Deep Learning**

*Deep Learning* adalah kelanjutan dari *machine learning* yang menggunakan struktur saraf buatan atau *neural network*. *Neural network* terdiri dari banyak *node* yang saling terhubung satu sama lain yang disusun secara berlapis yang disebut

*layer*. *Deep learning* dapat dikatakan sebagai sebuah cara untuk melakukan analisa prediksi secara otomatis.

*Deep learning* mempelajari data dengan cara memasukkan data ke dalam algoritma *backpropagation* yang di dalamnya terdapat layer yang harus dilewati data. algoritma ini akan merubah *weight* dari setiap *node* bergantung dari data input dengan cara menghitung gradien dari *loss function*. Setelah melewati seluruh layer tersebut maka data akan menjadi *output* berupa model statistik. Jumlah dari layer yang digunakan pada algoritma ini yang akan membedakan *deep learning* dengan *machine learning* biasa. Proses ini akan diulang dalam beberapa *iteration* sampai *output* dinilai memiliki akurasi yang cukup. Perulangan iterasi ini juga sebaiknya tidak dilakukan secara berlebihan karena dapat terjadi *overfitting* yaitu prediksi menjadi tidak akurat saat menggunakan *dataset* selain data *training* karena terjadinya *oversaturation*. Dalam implementasi *deep learning*, terdapat beberapa parameter yang bisa disesuaikan untuk mencapai hasil yang maksimal. Parameter tersebut adalah jumlah perulangan (*epoch*), *learning rate*, *batch size*, jumlah *layer*, dan *hyperparameter tuning*.

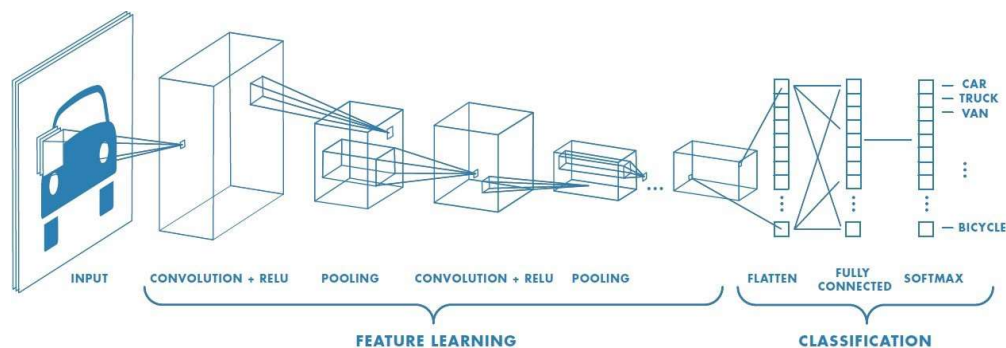
*Deep learning* dipilih sebagai metode prediksi pada penelitian kali ini karena kemampuannya yang luas serta fleksibilitasnya dalam mempelajari berbagai hal termasuk penggunaan listrik pada penelitian ini. *Deep learning* juga memiliki kurva pembelajaran yang linear yang berarti semakin banyak data yang dikonsumsi untuk dipelajari, semakin baik juga performa *deep learning* yang dimiliki. Penelitian ini dapat dikatakan menggunakan kombinasi *machine learning* dan *deep learning* karena mengkombinasikan penggunaan seleksi fitur secara manual dengan PCC dan juga ekstraksi fitur secara otomatis menggunakan CNN.

### 2.3.1. Recurrent Neural Network

Recurrent Neural Network (RNN) didasari oleh temuan David Rumelhart pada tahun 1986 [7]. adalah sebuah jenis *artificial neural network* yang setiap *nodenya* terhubung membentuk sebuah graf berarah maupun tidak berarah dalam satu *sequence*. RNN ditujukan untuk memproses data *sequential* seperti data *time series* yaitu data ramalan cuaca, harga saham, nilai tukar uang, dan sebagainya.

RNN bekerja dengan baik untuk fungsi data *sequential* karena RNN akan menyimpan informasi dari input sebelumnya dan dijadikan pembelajaran untuk ke depannya. RNN memiliki *feedback loop* atau ingatan yang dapat mengenali sebuah pola yang muncul dalam data. RNN bekerja untuk menentukan bagaimana data saat waktu ini ( $t$ ) mempengaruhi data setelahnya ( $t + 1$ ) [8]. Metode *training* yang digunakan pada RNN adalah algoritma *gradient descent*. *Gradient descent* adalah sebuah metode untuk mencari nilai minimum pada sebuah fungsi diferensial dengan cara melawan arah dari gradien hingga mencapai nilai minimum. Pada *neural network gradient descent* dapat digunakan untuk mengganti nilai bobot *node* untuk meminimalisir terjadinya error. Terdapat banyak sekali jenis arsitektur dari RNN, salah satunya adalah Long short-term memory (LSTM) yang ditemukan oleh Hochreiter dan Schmidhuber pada tahun 1997 yang mencetak rekor akurasi tertinggi pada beberapa aplikasi [9].

### 2.3.2. Convolutional Neural Network



Gambar 2.1 Struktur lapisan CNN [10]

Convolutional Neural Network (CNN) adalah sebuah artificial neural network yang dapat digunakan untuk klasifikasi. Biasanya CNN digunakan untuk menganalisa gambar, tetapi juga bisa digunakan untuk data time series. CNN terdiri dari *convolution layer*, *pooling layer*, dan *fully connected layer* seperti yang terlihat pada gambar 2.1 [10]. *Convolution layer* pada CNN berperan untuk melakukan konvolusi pada matriks input dengan mengalikannya dengan filter yang ditentukan [11]. Setelah itu input akan masuk ke *pooling layer* yang menerima *feature map* dari lapisan sebelumnya. Terdapat dua metode *pooling* pada lapisan ini yaitu *max pooling* dan *average pooling*. Lapisan terakhir dari CNN adalah *fully connected*

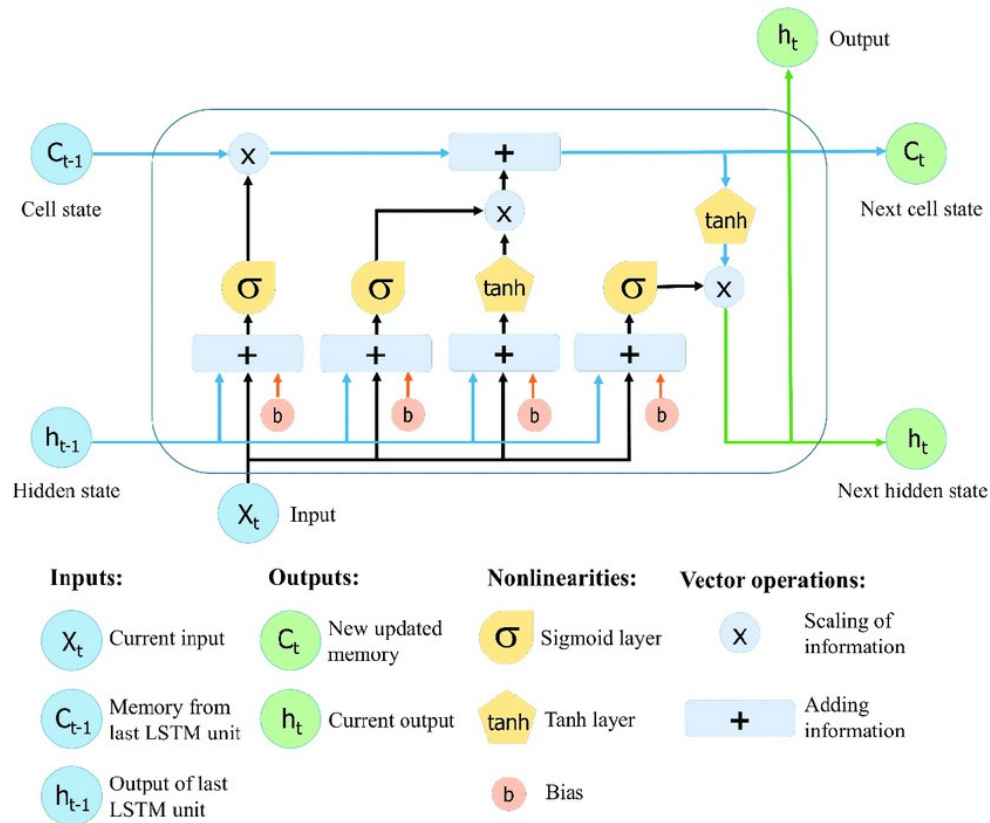
*layer* yang mengubah *feature map* menjadi vektor dan diklasifikasi. Aktivasi yang dapat digunakan di *layer* ini adalah *tanh*, *sigmoid*, *relu*, dan sebagainya. Pemilihan aktivasi tergantung dengan jumlah klasifikasi yang ingin dilakukan.

CNN akan digunakan dengan cara digabungkan dengan Long Short-term Memory (LSTM) yang bertujuan untuk menyeleksi fitur dan melakukan filtering terhadap data untuk meningkatkan akurasi. Berdasarkan penelitian [12], penggabungan LSTM dengan CNN memiliki *mean absolute percentage error* (MAPE) sebesar 3,71%, lebih rendah 0,37% dibandingkan dengan LSTM dan jauh lebih rendah daripada metode lain seperti Auto Regressive Integrated Moving Average (ARIMA) dan Support Vector Regression (SVR) untuk data yang jumlahnya satu tahun 11 bulan. Pada penelitian [12] juga digunakan Pearson Correlation Coefficient (PCC) untuk mengeliminasi fitur yang tidak diperlukan yang bertujuan untuk meningkatkan akurasi.

#### 2.4. Long Short-term Memory

Long Short-term Memory (LSTM) adalah sebuah jenis *artificial neural network* yang merupakan pengembangan dari RNN [9]. LSTM memiliki koneksi ke belakang yang disebut *feedback connection* yang dapat menunjuk ke sebuah poin data. LSTM biasa digunakan untuk mempelajari data tulisan tangan, pengenalan ucapan, dan deteksi anomali pada *traffic* data. LSTM dinilai cocok untuk melakukan klasifikasi, *processing*, dan melakukan prediksi terhadap data *time series* karena data *time series* terkadang memiliki *lag* atau jeda pada data yang dapat ditangani oleh LSTM menggunakan algoritma *vanishing gradient problem*. LSTM memiliki empat komponen utama yaitu *cell* sebagai pengingat atau *memory*, *input gate*, *output gate*, serta *forget gate*. *Training* pada LSTM dilakukan secara *supervised* dalam sebuah *sequence* menggunakan algoritma *gradient descent* yang dikombinasikan dengan *backpropagation through time* (BPTT) untuk menghitung gradien dan mengganti bobot *node* pada *neural network*. Namun pada penggunaan *gradient descent* terdapat sebuah kelemahan yaitu ketika data memiliki jarak waktu yang kosong dan dapat menyebabkan *exploding gradient*. Oleh karena itu LSTM dapat menangani galat ini dengan cara melakukan *backpropagation* dari *output layer* ke dalam *cell*. LSTM akan mempelajari *error* ini dan jika ada *error* baru akan

dikirimkan ke tempat yang sama sehingga LSTM tahu kapan untuk membuang *error* karena sudah dikenali sebagai *error*.



Gambar 2.2 Struktur lapisan LSTM [13]

Pada gambar 2.2 yang diambil dari [13] dapat terlihat struktur dari sebuah lapisan LSTM seperti yang dijelaskan sebelumnya. LSTM menggabungkan input poin data  $X_t$  dengan *hidden state* dari *layer* sebelumnya lalu diaktivasi menggunakan lapisan sigmoid. Hasil dari *cell state* atau *memory layer* sebelumnya dikalikan dengan penggabungan tadi dan ditambahkan dengan hasil perkalian *hidden state* yang diaktivasi dengan sigmoid dan tanh untuk menjadi *cell state* selanjutnya. *Output* didapatkan dari hasil perkalian *cell state* yang diaktivasi dengan tanh dan *hidden state* ditambahkan dengan *input* yang diaktivasi dengan sigmoid.

LSTM menerima input dalam bentuk numpy array tiga dimensi. Misalkan sebuah model memiliki input  $(x, y, z)$ . Input  $x$  merupakan poin data,  $y$  merupakan *timestep* atau lag dari fungsi *time series*, dan  $z$  adalah fitur yang digunakan pada *training*. Output dari LSTM dapat diatur menjadi empat jenis tergantung dari

parameter yang diberikan. Jika dibiarkan *default*, maka output dari LSTM adalah *hidden state* terakhir. Jika diberikan parameter *return sequences = True*, maka outputnya adalah semua *hidden state* dari semua *time step*. Jika diberikan parameter *return\_state = True*, output akan berupa *last hidden state + last hidden state + last cell state*. Dengan variasi output ini, LSTM dapat digabungkan dengan berbagai metode lain termasuk dengan menumpuk LSTM dengan lapisan LSTM lainnya.

*Stacked* LSTM adalah ketika sebuah model LSTM memiliki lebih dari satu lapisan LSTM. Tambahan lapisan LSTM ini bertujuan untuk merubah input data menjadi data sekuensial kepada lapisan LSTM selanjutnya. Metode ini bisa dilakukan dengan pmemberikan parameter *return sequence* kepada lapisan-lapisan LSTM sebelum lapisan LSTM terakhir. [14] menunjukkan bagaimana menambah kedalaman jaringan RNN dengan cara menyusun beberapa lapis *hidden layer* dan dapat memberikan hasil yang lebih baik dan lebih efisien secara eksponensial dalam beberapa fungsi tertentu.

## 2.5. Structured Query Language

Structured query language atau SQL adalah sebuah bahasa pemrograman standar yang digunakan untuk mengatur dan melakukan operasi terhadap data dalam *relational database*. SQL memiliki perintah-perintah standar seperti *select*, *add*, *insert*, *update*, *delete*, *create*, *alter*, dan *truncate*. SQL sudah menjadi standar industri dalam bahasa pemrograman *database* relasional yang berisikan tabel yang terdiri dari baris dan kolom. SQL juga dapat digunakan untuk analisa data dengan menjalankan *query* yang sesuai.

MySQL adalah sebuah aplikasi *relational database management system* (RDBMS) yang bersifat *open source*. MySQL ditemukan oleh Michael Widenius dan David Axmark, namun sekarang MySQL sudah dibeli dan dioperasikan oleh Oracle. MySQL ditulis dalam bahasa C dan C++ dengan *parser* yang ditulis dalam *yacc* dengan *lexical analyzer*. MySQL dapat berjalan di banyak sistem operasi umum sehingga banyak diadopsi oleh pengguna. MySQL juga memiliki keunggulan karena mudahnya menghubungkannya dengan bahasa pemrograman lain dengan menggunakan program pendukung yang bernama MySQL Connector.



## 2.6. Fungsi Koefisien Korelasi

Koefisien korelasi adalah fungsi yang digunakan untuk mencari tahu apakah sebuah data memiliki hubungan atau korelasi dengan data lainnya. Koefisien korelasi dapat bernilai di antara -1 hingga 1. Misalkan terdapat dua data yang dicari hubungannya dengan koefisien korelasi yaitu data A dan B. Jika nilai koefisien korelasinya -1, maka setiap data A bertambah positif, maka data B akan berkurang, begitu juga sebaliknya. Jika nilai koefisien korelasinya 0, maka data A dan B tidak memiliki hubungan sama sekali. Jika nilai koefisien korelasinya 1, maka setiap kenaikan positif data A akan diikuti kenaikan positif di data B, begitu juga dengan penurunan. Semakin nilai koefisien mendekati 1 atau -1, semakin kuat hubungan antara kedua data tersebut.

Pearson product-moment correlation coefficient (PPMCC) atau Pearson Correlation Coefficient (PCC) adalah salah satu jenis fungsi koefisien korelasi[15]. PCC dikembangkan oleh Karl Pearson yang merupakan pengembangan dari ide korelasi yang diperkenalkan oleh Francis Galton pada tahun 1880an. PCC mengukur korelasi linear antara dua data. Nilai koefisien merupakan rasio *covariance* dari dua variabel dan hasil dari *standard deviation*. Koefisien korelasi Pearson direpresentasikan oleh huruf Yunani  $\rho$  (rho). Misalkan terdapat dua data (X, Y), maka formula  $\rho$  adalah[16]:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \quad (1)$$

di mana cov adalah *covariance*,  $\sigma_x$  dan  $\sigma_y$  adalah *standard deviation* dari X dan Y.

Untuk menggunakan fungsi koefisien korelasi Pearson, sebuah data perlu memiliki syarat dan karakteristik seperti:

1. Memiliki dua atau lebih variabel
2. Tidak memiliki nilai kosong atau *null*
3. Relasi linear antara variabel
4. Tidak memiliki *outlier*.

## 2.7. *Mean Absolute Error*

*Mean absolute error* (MAE) adalah salah satu matriks penilaian yang menghitung nilai absolut rata-rata error pada hasil prediksi. MAE dapat memberikan ukuran kinerja model dengan mengetahui seberapa besar nilai error rata-rata dari model sehingga menggambarkan seberapa dekat model dengan nilai aslinya. Skor terbaik MAE adalah 0 dan semakin besar nilai MAE maka semakin buruk kinerja model. Rumus MAE adalah

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |X_i - Y_i| \quad (2)$$

di mana  $X_i$  adalah nilai prediksi dan  $Y_i$  adalah nilai asli. Nilai MAE dari sebuah model tidak bisa dibandingkan dengan model lainnya atau tidak standar dikarenakan perbedaan dari apa yang diprediksi. Jika *range* nilai prediksi sebuah model besar, maka nilai MAE kemungkinan juga akan besar, begitu pula sebaliknya. Jika nilai MAE suatu model lebih besar daripada model lainnya, bukan berarti model tersebut lebih tidak akurat karena bergantung kepada data validasinya.

## 2.8. *Coefficient of Determination*

Koefisien determinasi atau skor  $R^2$  adalah nilai di antara 0 dan 1 yang mengukur seberapa baik hasil prediksi dari sebuah model statistik [17].  $R^2$  dapat didefinisikan dengan

$$R^2 = 1 - \frac{\text{MSE}}{\text{MST}} \quad (3)$$

di mana MSE adalah *mean square error* dan MST adalah *mean square total*. Sehingga dari definisi tersebut  $R^2$  dapat diuraikan sebagai berikut:

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2 \quad (4)$$

$$\text{MST} = \frac{1}{m} \sum_{i=1}^m (Y_i - \bar{Y})^2 \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (Y' - Y_i)^2} \quad (6)$$

di mana  $X_i$  adalah nilai prediksi dan  $Y_i$  adalah nilai asli. Berbeda dengan MAE, skor  $R^2$  merupakan nilai yang standar untuk semua model karena nilai  $R^2$  merupakan rasio atau presentase. Oleh karena itu  $R^2$  dapat digunakan untuk membandingkan performa suatu model dengan model lainnya meskipun dengan data validasi yang berbeda.

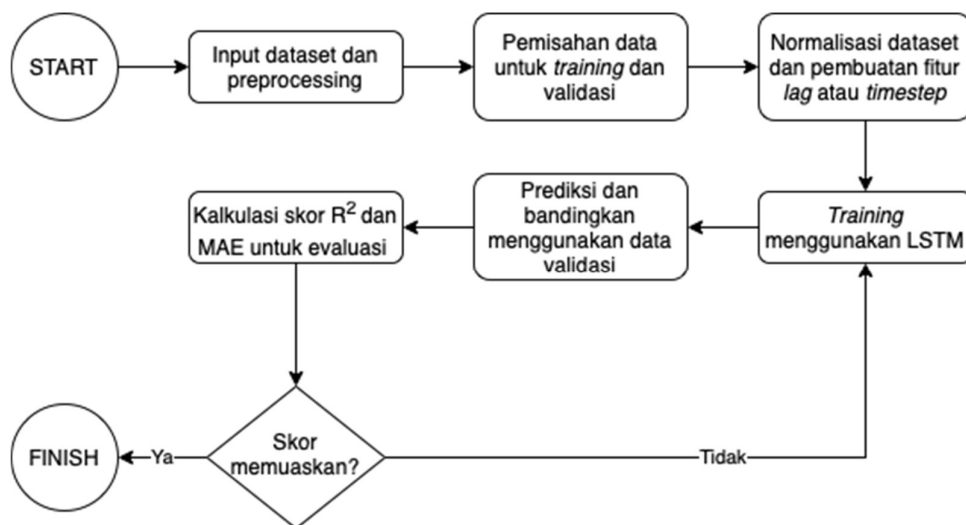
## BAB 3

### PERANCANGAN SISTEM

#### 3.1. Diagram alur sistem

Berikut ini adalah diagram alur tiga metode yang akan digunakan pada penelitian ini:

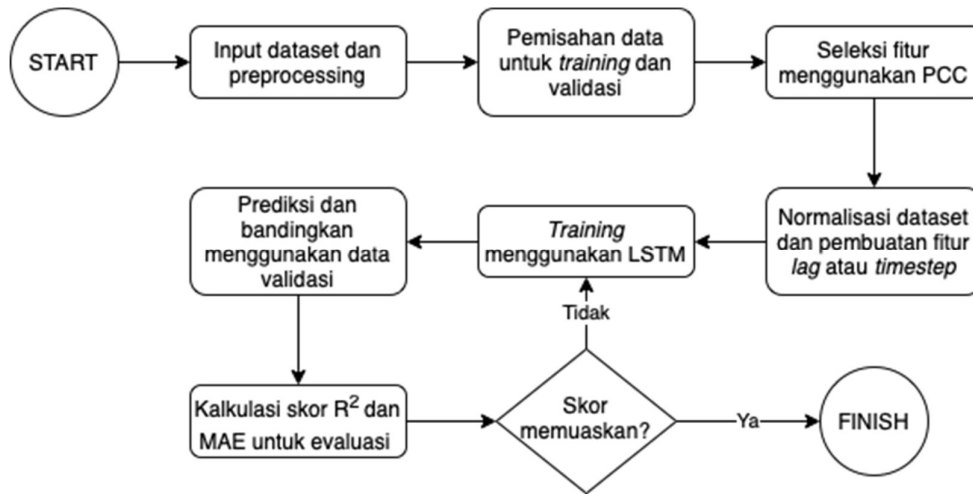
##### 3.1.1. Model LSTM Tanpa Data Cuaca



Gambar 3.1 Diagram alur metode LSTM tanpa data cuaca

Gambar 3.1 merupakan diagram alur skenario pengujian pertama pada penelitian ini. Skenario dimulai dari pengolahan data dan *preprocessing* untuk mengganti satuan data. Selanjutnya data dipisahkan menjadi data *training* dan validasi. Kemudian dilakukan normalisasi dan dibuat fitur *lag*. Selanjutnya data yang sudah siap digunakan untuk training dan dilakukan prediksi menggunakan model hasil training dan hasilnya divalidasi menggunakan data validasi. Dari hasil prediksi tersebut dikalkulasi skor matriks penilaiannya, jika skor sudah maksimal maka akan diambil sebagai pembanding dengan skenario lainnya.

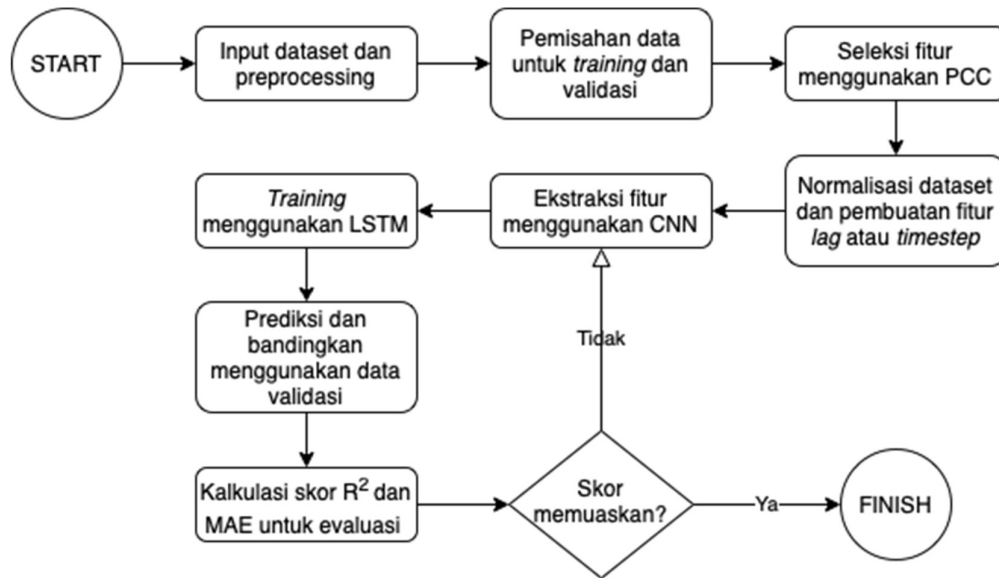
### 3.1.2. Model LSTM dengan Data Cuaca



Gambar 3.2 Diagram alur model LSTM

Gambar 3.2 merupakan diagram alur skenario pengujian kedua pada penelitian ini. Skenario dimulai dari pengolahan data dan *preprocessing* untuk mengganti satuan data. Selanjutnya data dipisahkan menjadi data *training* dan validasi. Berbeda dengan skenario sebelumnya pada skenario ini digunakan PCC untuk menyeleksi fitur yang akan digunakan pada *training*. Kemudian dilakukan normalisasi dan dibuat fitur *lag*. Selanjutnya data yang sudah siap digunakan untuk training dan dilakukan prediksi menggunakan model hasil training dan hasilnya divalidasi menggunakan data validasi. Dari hasil prediksi tersebut dikalkulasi skor matriks penilaiannya, jika skor sudah maksimal maka akan diambil sebagai pembanding dengan skenario lainnya.

### 3.1.3. Model CNN-LSTM



Gambar 3.3 Diagram alur model CNN-LSTM

Gambar 3.2 merupakan diagram alur skenario pengujian kedua pada penelitian ini. Skenario dimulai dari pengolahan data dan *preprocessing* untuk mengganti satuan data. Selanjutnya data dipisahkan menjadi data *training* dan validasi. Selanjutnya fitur akan diseleksi untuk digunakan pada *training* menggunakan PCC. Kemudian dilakukan normalisasi dan dibuat fitur *lag*. Selanjutnya data yang sudah siap digunakan untuk training akan dilakukan ekstraksi fitur menggunakan lapisan CNN lalu dilakukan *training*. Prediksi dilakukan menggunakan model hasil training dan hasilnya divalidasi menggunakan data validasi. Dari hasil prediksi tersebut dikalkulasi skor matriks penilaiannya, jika skor sudah maksimal maka akan diambil sebagai pembanding dengan skenario lainnya.

### 3.2. Dataset

Dataset yang digunakan pada penelitian ini adalah data penggunaan listrik Gedung S Fakultas Teknik Universitas Indonesia (FTUI) pada periode 1 November 2013 hingga 11 November 2014. Dataset ini dinilai termasuk sebagai data *time series* karena data ini memiliki poin penunjuk waktu. Dataset ini juga memiliki komponen yang sesuai dengan data *time series* seperti komponen *seasonal* yaitu periode di mana saat perkuliahan berlangsung dan periode di mana perkuliahan sedang ditiadakan. Komponen *cyclical* juga terdapat pada dataset ini yaitu adanya perbedaan intensitas penggunaan listrik di siang hari dan malam hari. Data tersebut tersimpan dalam database SQL dengan ukuran SQL *dump* sebesar 576 megabytes. Data tersebut berisikan empat tabel yaitu tabel *power* yang berisikan frekuensi, tabel *powernodes* yang berisikan nama gedung dan *nodeId*-nya, tabel *powergroup* yang berisikan kode *groupId* dan nama kelompok gedung, dan tabel *power3p* yang berisikan catatan penggunaan listrik. Pada penelitian ini akan digunakan tabel *power3p* yang berisikan 2783433 baris dan 38 kolom yang berisikan *id*, *timestamp*, *nodeId*, dan data listrik. Pada periode ini tercatat hanya ada dua *nodeId* yang diambil datanya yaitu *nodeId* satu dan dua yaitu data penerangan Gedung S dan data pendingin ruangan Gedung S. Pencatatan data dilakukan setiap 19 detik sekali untuk masing-masing node dengan periode berbeda sehingga *timeframe* dari data kedua *node* ini tidak sinkron. Gedung S sendiri memiliki lima lantai dan 25 ruangan. Gedung S merupakan gedung kuliah bersama yang digunakan oleh semua departemen di FTUI.

Selain itu digunakan juga dataset eksternal berupa pengukuran cuaca yang berasal dari Badan Meteorologi Klimatologi dan Geofisika (BMKG). Data berasal dari stasiun pengukuran BMKG yang terdekat dengan Gedung S yaitu Stasiun Cuaca Kemayoran, DKI Jakarta. Data tersebut merupakan data per hari yang terdiri dari temperatur minimum, temperatur maksimum, temperatur rata-rata, kelembaban rata-rata, curah hujan, lamanya penyinaran matahari, kecepatan angin maksimum, arah angin saat kecepatan maksimum, dan kecepatan angin rata-rata. Seluruh data tersebut tersedia untuk publik melalui halaman web [dataonline.bmkg.go.id](http://dataonline.bmkg.go.id).

### 3.3. Preprocessing dan Pembuatan Fitur

Dataset penggunaan listrik yang berbentuk database SQL memiliki banyak kolom yang tidak terpakai untuk penelitian kali ini. Maka dari itu penulis menggunakan *query select* terlebih dahulu untuk memilih kolom mana yang akan dipakai dari kolom *power3p*. Hasil *query* tersebut lalu di *export* menjadi *file* CSV. *File* CSV tersebut lalu di *import* menjadi Pandas Dataframe dan dilakukan pengecekan nilai kosong atau *null*. Setelah dipastikan tidak ada, data yang sebelumnya diambil per 19 detik diolah menjadi data per jam dengan satuan *kilo watt hour* (kWH) dengan cara menambahkan semua data salah satu jam lalu dikalikan 19 lalu dibagi 3600. Setelah data menjadi per jam maka cukup menjumlahkan data dengan tanggal yang sama untuk dijadikan data per hari. Setelah itu dilakukan pengecekan terhadap data untuk mencari apakah ada data yang kosong atau apakah ada data *outlier*. Ternyata terdapat beberapa tanggal yang tidak ada catatan pengambilan data atau pengambilannya tidak penuh selama 24 Jam. Berikut ini tabel penjelasan tanggal data kosong:

Tabel 3.1 Daftar tanggal data kosong pada dataset

Tanggal	Keterangan
1 November 2013	Awal data (data dimulai pukul 06.00 WIB)
2 – 3 November 2013	Akhir pekan
14 – 15 Desember 2013	Akhir pekan
31 Desember 2013 – 1 Januari 2014	Akhir pekan
3 – 4 Mei 2014	Akhir pekan
5 – 6 Juli 2014	Akhir pekan
24 Juli 2014 – 3 Agustus 2014	Libur kuliah pergantian tahun ajaran
10 Agustus 2014	Akhir pekan
12 Oktober 2014	Akhir pekan
18 – 19 Oktober 2014	Akhir pekan
1 November 2014	Akhir pekan
11 November 2014	Akhir data (data berakhir pukul 21.00 WIB)



Berdasarkan keterangan data tersebut, data yang kosong tidak akan dilakukan *preprocessing* atau pengisian nilai kosong dikarenakan pada saat akhir pekan dan pergantian tahun ajaran memang tidak ada penggunaan pendingin ruangan karena tidak adanya perkuliahan. Untuk data awal yaitu tanggal 1 November 2013 serta data terakhir yaitu 11 November 2014, data dibiarkan apa adanya karena pada data pertama pengukuran dimulai pukul 06.00 WIB dan pada data terakhir pengukuran berakhir pukul 21.00 sehingga data untuk tanggal tersebut dianggap representatif untuk tanggal tersebut mengingat jadwal perkuliahan yang biasanya dimulai pukul 07.00 WIB.

Data cuaca yang didapatkan dari BMKG berbentuk *file* XLSX yang berisi data setiap bulannya. Data tersebut diambil dan digabungkan menjadi satu *file* lalu disimpan menjadi CSV lalu dibuka menggunakan Python Notebook untuk menjadi Dataframe Pandas. Setelah itu data dilakukan pengecekan dan ditemukan beberapa data kosong. Data kosong yang terdapat di data penyinaran matahari dan curah hujan kemudian diisi dengan nilai *mode* atau nilai yang paling sering muncul dalam kolom tersebut. Setelah data kosong diisi kemudian data digabungkan dengan data penggunaan listrik menjadi satu *dataframe*.

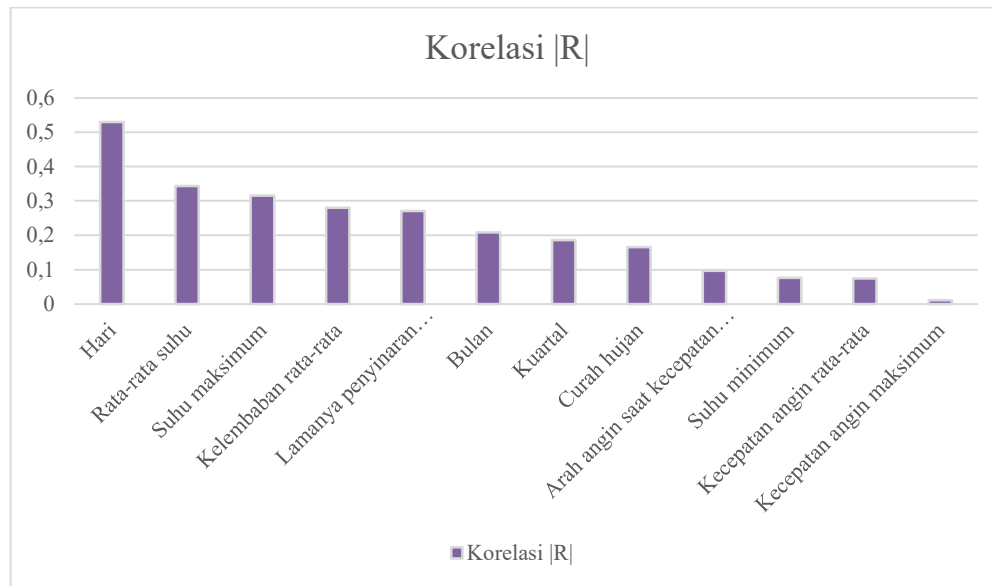
Dataset yang berbasis *time series* berarti fitur dapat diambil dari satuan waktu poin data tersebut. Dari data satuan waktu dapat dibuat fitur seperti tanggal, hari dalam pekan, pekan dalam tahun, bulan, hari dalam tahun, kuartal tahun, dan tahun. Pembuatan fitur waktu dilakukan dengan tujuan untuk melihat korelasi penggunaan listrik pada periode tertentu dan juga membantu model untuk mempelajari pola penggunaan listrik. Penggunaan listrik pendingin ruangan dalam satuan kWh per harinya akan dijadikan target. Dari data cuaca dapat diambil fitur berupa temperatur minimum, temperatur maksimum, temperatur rata-rata, kelembaban rata-rata, curah hujan, lamanya penyinaran matahari, kecepatan angin maksimum, arah angin saat kecepatan maksimum, dan kecepatan angin rata-rata.

Setelah fitur dibuat lalu dilakukan pengecekan korelasi fitur dan target dengan menggunakan fungsi koefisien korelasi PCC. Pengecekan ini bertujuan untuk meningkatkan akurasi dengan menyeleksi fitur yang tidak diperlukan atau

fitur yang tidak berkorelasi baik dengan target. Berikut ini adalah tabel korelasi PCC dari fitur yang dibuat:

Tabel 3.2 Tabel korelasi fitur berdasarkan PCC

<b>Fitur</b>	<b>Korelasi  R </b>
Penggunaan listrik	1.000000
Suhu minimum	0.076154
Suhu maksimum	0.314729
Rata-rata suhu	0.342783
Kelembaban rata-rata	0.280477
Curah hujan	0.165655
Lamanya penyinaran matahari	0.270747
Kecepatan angin maksimum	0.011398
Arah angin saat kecepatan maksimum	0.096679
Kecepatan angin rata-rata	0.074332
Hari	0.529650
Bulan	0.207942
Kuartal	0.186077
Tahun	0.029042
Hari dalam tahun	0.201774
Minggu	0.230045
Tanggal	0.073197



Gambar 3.4 Grafik batang perbandingan nilai PCC setiap fitur

Dari hasil penilaian korelasi tersebut, dapat terlihat pada tabel 3.2 dan gambar 3.4 bahwa nilai korelasi tertinggi terhadap penggunaan listrik adalah hari dengan nilai 0.529650. Fitur yang memiliki korelasi di bawah 0.1 tidak akan digunakan untuk pembuatan model karena bisa mempengaruhi akurasi model. Pengaruh tersebut akan dilakukan pengujian dan akan dibahas pada Bab 4. Meskipun fitur waktu seperti hari dalam tahun dan minggu yang memiliki nilai korelasi data diatas 0.1, kedua fitur tersebut tidak diikutsertakan pada penelitian kali ini karena dianggap tidak signifikan, kurang cocok, dan dikhawatirkan akan membuat model yang dibuat kurang praktis jika digunakan untuk kegunaan lain yang lebih umum. Maka dari itu pada gambar 3.4 kedua fitur tersebut tidak diikut sertakan.

Selanjutnya dilakukan normalisasi terhadap data dengan menggunakan fungsi *standard scaler* dari library *sklearn*. Normalisasi ini bertujuan untuk mengubah data angka yang sebelumnya memiliki jarak tidak menentu menjadi berskala di antara -1 dan 1. Pada kasus ini, data akan berskala antara nol dan satu karena tidak ada data negatif.

Selanjutnya data akan dimanipulasi agar bentuknya menjadi tiga dimensi untuk menyesuaikan *input* yang diterima CNN dan LSTM. Data *time series* ini dibentuk menjadi tiga dimensi dengan menggunakan *lag function*. Pada penelitian

kali ini digunakan jeda atau *lag* sebanyak tujuh, yang pada *dataset* ini berarti tujuh hari. Jeda yang digunakan sebanyak tujuh hari atau satu minggu dikarenakan pola pada dataset per hari ini terbentuk untuk setiap minggunya, di mana pada hari kerja penggunaan pendingin ruangan meningkat, dan pada akhir pekan penggunaan menurun hingga mencapai nol. Maka dari itu penggunaan jeda hingga tujuh hari diperkirakan akan menguntungkan pada model ini dan meningkatkan akurasi karena model akan mempelajari pola kenaikan dan penurunan data. Sebelumnya juga telah dilakukan percobaan untuk menggunakan lag sebanyak dua hari dan 14 hari, akan tetapi, hasil yang terbaik didapatkan dari penggunaan lag sebanyak tujuh hari dan juga data yang hilang akibat dari pembuatan fitur *lag* dapat seminimal mungkin

Setelah seluruh *preprocessing* selesai, data akan dilakukan pemisahan untuk *training* dan validasi. Untuk menjaga integritas pola data, maka data yang dipakai untuk *training* akan berjumlah selama satu tahun penuh atau 365 hari. Data yang akan digunakan untuk validasi adalah data 11 hari sisa data untuk dilakukan uji prediksi.

### 3.4. Skenario Pengujian

Pengujian dilakukan tiga kali dengan membuat dua model berbeda. Model pertama yang dibuat adalah model yang menggunakan dua layer LSTM atau stacked LSTM, masing-masing memiliki 128 hidden layer. Pada layer pertama LSTM digunakan parameter *return sequence = True* untuk mengubah output dari LSTM menjadi semua *hidden layer*. Pengubahan ini bertujuan agar data berubah menjadi data sekuensial dari semua *hidden layer* dan bentuk data tidak berubah yaitu tiga dimensi. Setelah dua lapis dari LSTM, maka akan dilanjutkan dengan dua lapisan *Dense layer*. Lapisan pertama dari *dense layer* akan berjumlah sama dengan jumlah fitur dan lapisan kedua dari *dense layer* berjumlah satu lapisan. *Dense layer* adalah lapisan *fully connected layer* yang bertujuan untuk mengubah dimensionalitas dari output layer sebelumnya agar model lebih mudah untuk mendefinisikan hubungan antar data. Pada model pertama ini dilakukan dua kali pengujian yaitu satu kali tanpa menggunakan data cuaca dan satu kali menggunakan

data cuaca. Hal ini dilakukan untuk mengetahui apakah cuaca memiliki pengaruh terhadap penggunaan pendingin ruangan.

Model kedua dibuat dengan gabungan dari dua metode yaitu CNN dan LSTM. Pada lapisan pertama akan digunakan lapisan CNN dengan jumlah filter sebanyak 128, *kernel size* sebesar dua, dan *stride* sebanyak satu. Fungsi dari lapisan CNN ini bertujuan untuk menyaring dan melakukan seleksi fitur terhadap data yang terbatas ini untuk meningkatkan akurasi prediksi. Lapisan kedua lapisan LSTM dengan jumlah *hidden layer* sebanyak 128 buah seperti pada model pertama. Kemudian lapisan ketiga dan keempat digunakan lapisan *dense layer* masing-masing sebanyak jumlah fitur dan satu buah seperti pada model pertama.

Metode pertama akan dilakukan percobaan sebanyak dua kali, pertama dengan data penggunaan listrik saja, dan kedua ditambah dengan data cuaca. Hal ini dilakukan untuk melakukan pengujian apakah data cuaca memiliki pengaruh terhadap akurasi model. Seluruh percobaan akan dilakukan dengan *training* parameter yang sama yaitu dengan *epoch* sebanyak 50 dan *batch size* sebesar 16. Penyetaraan parameter dilakukan agar perbandingan antar metode lebih terlihat perbedaannya. Selain itu juga akan ada skenario tambahan yaitu penggunaan metode CNN-LSTM tetapi tanpa dilakukan seleksi fitur sebelumnya menggunakan PCC. Hal ini dilakukan untuk memastikan bahwa PCC memiliki peran pada penelitian ini.

### 3.5. Rencana Evaluasi

Evaluasi dari hasil *training* kedua model tersebut akan dilakukan dengan membandingkan nilai *coefficient of determination* atau  $R^2$  score. Nilai ini akan merepresentasikan seberapa dekat garis hasil prediksi dengan nilai aslinya. Selain itu juga akan digunakan perbandingan dari nilai *mean absolute error* (MAE) untuk mengetahui rata-rata selisih prediksi dengan nilai aslinya. Nilai *mean absolute percentage error* (MAPE) tidak bisa digunakan pada penelitian kali ini karena adanya data yang bernilai nol. Hasil dari prediksi setiap model juga akan dibandingkan dengan nilai aslinya dalam sebuah grafik sehingga terlihat juga apakah pola data sudah sesuai atau belum.

## BAB 4

### HASIL PENELITIAN

#### 4.1. Implementasi Model

Pembuatan model LSTM dan CNN-LSTM dilakukan menggunakan Jupyter Notebook dalam aplikasi Jupyter Lab dengan bahasa Python. *Platform* ini digunakan karena Python menyediakan *library* yang siap digunakan untuk membuat model *neural network*. Berikut ini adalah daftar *software* dan *library* yang digunakan beserta versinya:

1. Python 3.8.9
2. Ipykernel 6.4.1
3. Ipython 7.28.0
4. Jupyterlab 3.4.2
5. Keras 2.9.0
6. Pandas 1.4.2
7. Plotly 5.8.0
8. Tensorflow 2.9.0
9. Numpy 1.22.3

Dataset penggunaan pendingin ruangan serta dataset cuaca dari BMKG yang sudah digabung disimpan dalam *file* CSV. Model akan disimpan dalam *file* h5 dan kode dalam bentuk Python Notebook atau IPYNB.

## 4.2. Hasil Pengujian

Berikut ini adalah hasil pengujian metode-metode yang telah dirancang. Hasil ini merupakan hasil terbaik yang diambil dalam 10 kali melakukan training

Tabel 4.1 Hasil Penilaian Model dan Error

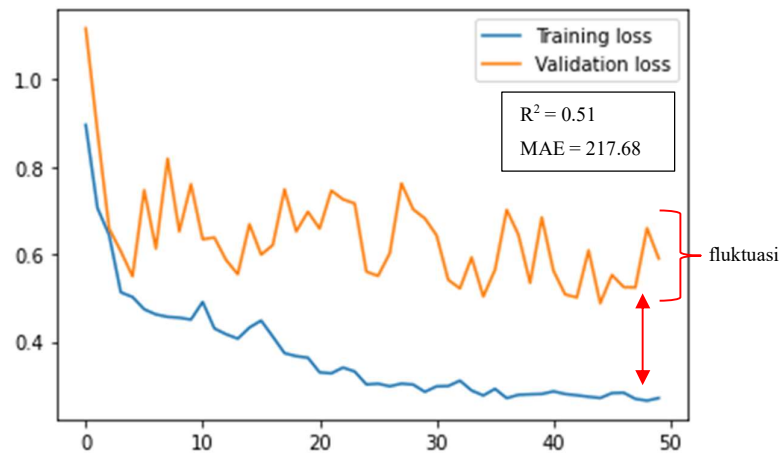
Model	Skor $R^2$	MAE
Stacked LSTM dengan data penggunaan listrik	0.51	217.68
Stacked LSTM dengan data cuaca	0.69	180.72
CNN-LSTM dengan data cuaca	0.82	156.94

Berdasarkan Tabel 4.1 dapat terlihat pengaruh data cuaca terhadap akurasi dan konfidensi model. Dengan metode yang sama, pengaruh data cuaca mampu meningkatkan konfidensi dan menurunkan *error* dari model yang sama yaitu peningkatan 0,18 dari skor  $R^2$  dan penurunan nilai MAE sebanyak 36,96. Hal ini menunjukkan bahwa faktor cuaca seperti lamanya penyinaran matahari, kelembaban, dan curah hujan berpengaruh terhadap penggunaan pendingin ruangan. Hasil ini juga menunjukkan bahwa prediksi *time series single variative* kurang cocok untuk dilakukan dengan jumlah data yang sedikit. Jumlah data yang diperlukan untuk prediksi *time series* satu fitur setidaknya tiga tahun agar model dapat mempelajari pola dari tahun yang berbeda.

Skor  $R^2$  juga terus mengalami peningkatan seiring optimalisasi model dan data. Model CNN-LSTM dengan data cuaca terbukti menjadi yang paling akurat dengan skor  $R^2$  tertinggi yaitu 0,82 disertakan dengan nilai MAE terendah masing-masing di angka 156,94. Hal ini membuktikan jika penggunaan CNN untuk melakukan filtering dan seleksi fitur sehingga meningkatkan konfidensi dan akurasi, mengingat jumlah data yang tidak terlalu banyak. Nilai MAE juga memiliki tren menurun seperti skor  $R^2$ . Seluruh hasil penelitian ini diambil sebanyak 10 kali dikarenakan adanya fluktuasi dari hasil skor  $R^2$  pada penelitian kali ini. Hal ini dapat terjadi dikarenakan kurang representatifnya data validasi yang hanya

berjumlah 11 hari dibandingkan dengan data *training* sebanyak 365 hari. Meskipun nilai MAE dapat lebih stabil, tetapi konfidensi model akan terpengaruh akibat dari posisi nilai prediksi.

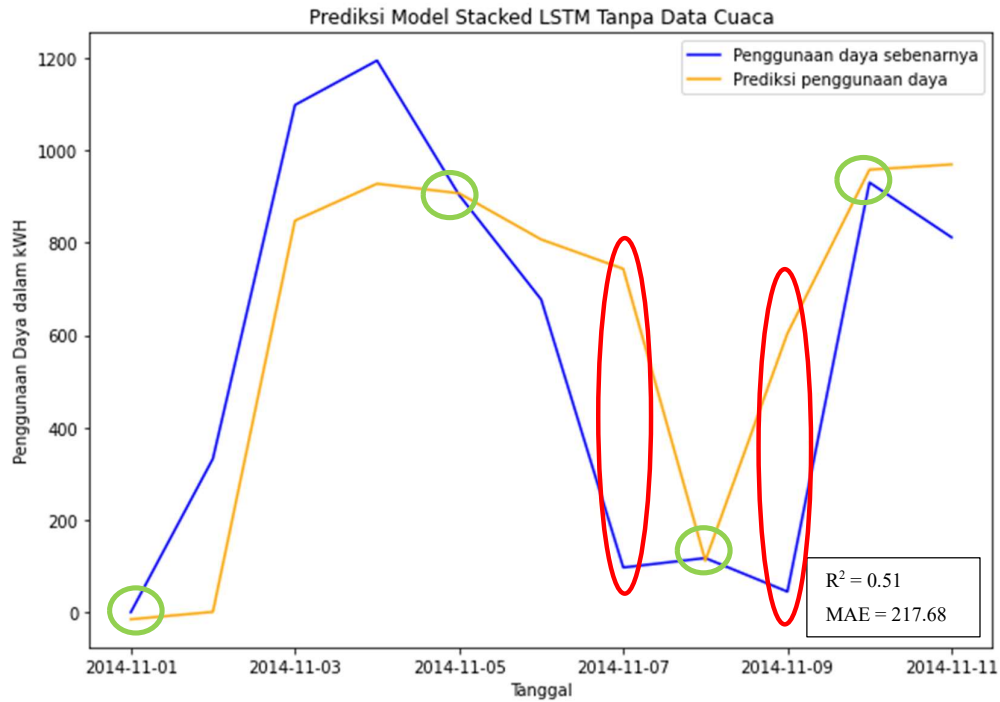
#### 4.2.1. Metode Stacked LSTM dengan Data Penggunaan Listrik



Gambar 4.1 Grafik *Validation Loss* dari model Stacked LSTM tanpa data cuaca

Pada Gambar 4.1 terlihat *training loss* yang memiliki tren menurun dan menuju kestabilan. Hal ini membuktikan bahwa model yang digunakan sudah tepat untuk data yang diberikan[18]. Namun *loss* validasi memiliki fluktuasi hingga akhir training. Hal ini dapat disebabkan kurang representatifnya data validasi terhadap data *training* mengingat data validasi yang hanya berjumlah 11 hari dibandingkan dengan data *training* yang berjumlah 365. Jarak antara dua grafik pada Gambar 4.1 yang cukup jauh mengindikasikan bahwa data *training* kurang representatif terhadap masalah yang ingin dipelajari oleh model sehingga mengkonfirmasi bahwa data penggunaan daya listrik selama satu tahun saja tidak cukup untuk melakukan *learning* dan prediksi penggunaan listrik. Idealnya kedua grafik memiliki tren menurun dan tidak ada fluktuasi. Jarak antara *loss* validasi dan *training* juga idealnya semakin mendekat.

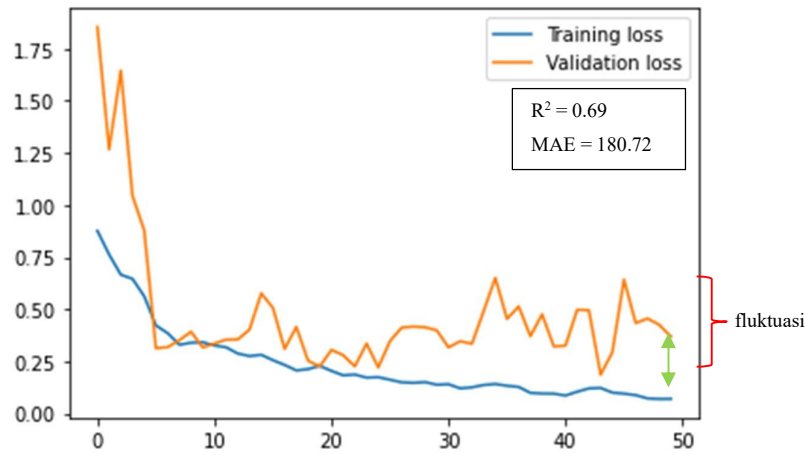




Gambar 4.2 Grafik penggunaan daya dan hasil prediksi

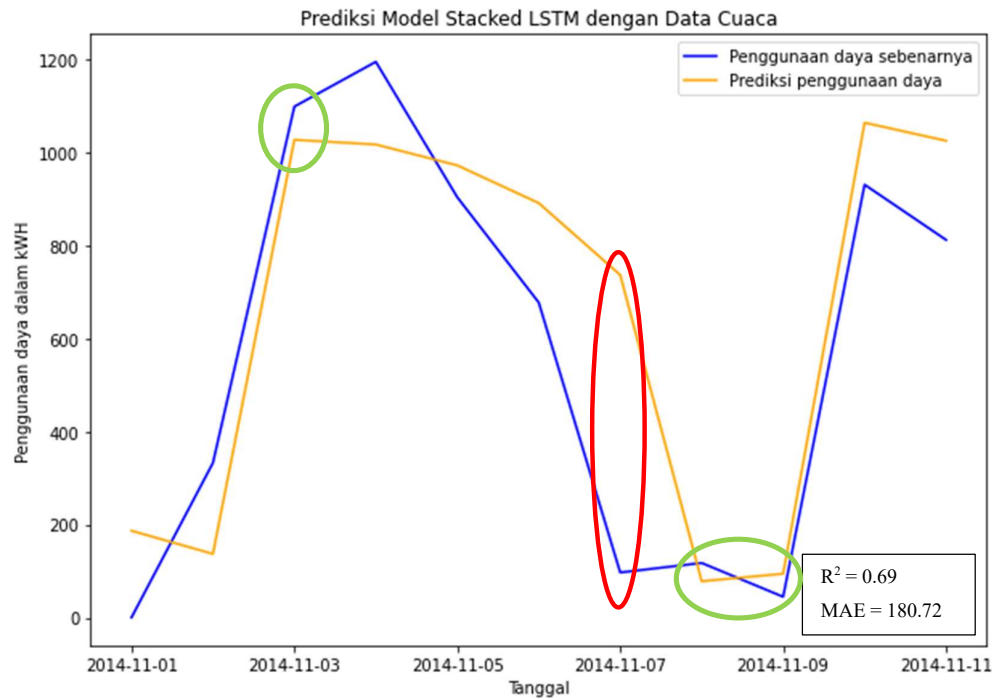
Pada Gambar 4.2 dapat terlihat bahwa hasil prediksi tidak terlalu akurat. Akan tetapi, terdapat beberapa poin data seperti tanggal 1, 5, 8, dan 10 yang ditandai dengan lingkaran hijau di mana prediksi model sudah mendekati aslinya. Pola kenaikan dan penurunan juga sudah terbentuk pada hasil prediksi, namun model masih kesulitan untuk memprediksi pola penggunaan listrik yang meningkat pada hari kerja dan menurun pada akhir pekan. Masih terdapat *error* yang cukup signifikan seperti pada tanggal 7 dan 9 yang ditandai dengan lingkaran merah yang bernilai sekitar 700 kWh turut berkontribusi terhadap rendahnya nilai  $R^2$  dan tingginya MAE. Kesalahan prediksi pada kedua tanggal tersebut disebabkan karena gagalnya model untuk mempelajari pola penggunaan listrik pada gedung kampus yang meningkat pada hari kerja dan menurun drastis pada akhir pekan. Perlu diketahui bahwa tanggal 7 November 2014 adalah hari Jumat, sehingga memang seharusnya sulit bagi model untuk memprediksi penggunaan yang mendekati 0 yang sebetulnya pada hari tersebut adalah hari kerja.

#### 4.2.2. Metode Stacked LSTM Dengan Data Cuaca



Gambar 4.3 Grafik *validation loss* model Stacked LSTM dengan data cuaca

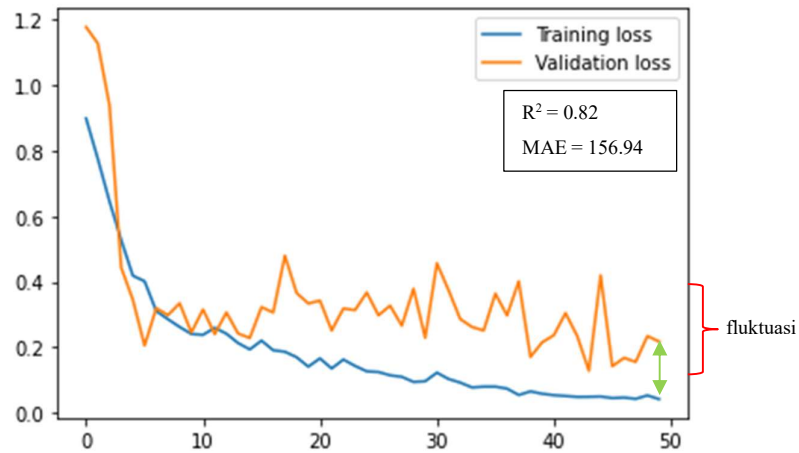
Pada Gambar 4.3 dapat terlihat bahwa *training loss* dan *validation loss* memiliki tren menurun setiap *epoch*-nya dan memiliki nilai yang lebih kecil secara keseluruhan dibandingkan dengan model sebelumnya yang tidak menggunakan data cuaca. Jarak antara *training loss* dan *validation loss* juga semakin dekat yang menunjukkan bahwa data *training* lebih representatif dengan tambahan data cuaca yang terpilih menggunakan PCC jika digunakan oleh model untuk mempelajari masalah. Akan tetapi *validation loss* pada model ini juga masih mengalami fluktuasi akibat dari kurang representatifnya data validasi yang hanya berjumlah 11 hari dibandingkan dengan data *training* yang berjumlah 365 hari. Oleh karena itu maka dengan epoch berapapun nilai *loss* validasi tidak akan mencapai kestabilan yang dinilai lebih ideal.



Gambar 4.4 Grafik penggunaan daya sebenarnya serta hasil prediksi model

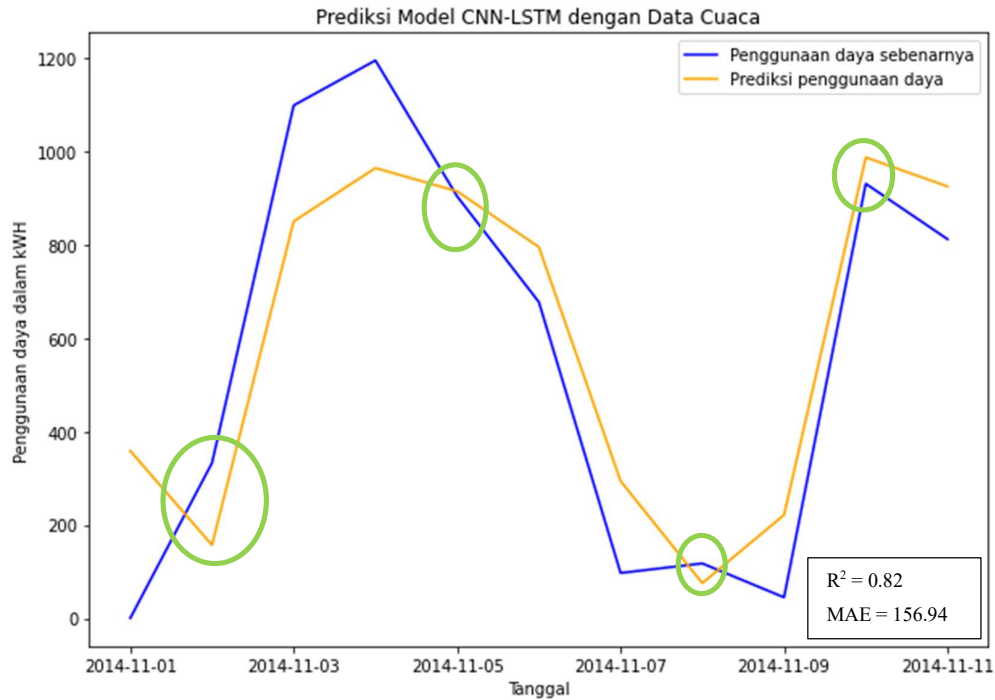
Terlihat pada Gambar 4.4 bahwa hasil prediksi lebih representatif terhadap data aslinya. Dibandingkan dengan model sebelumnya, *error* yang besar hanya terjadi pada tanggal 7 November dengan jarak sekitar 700 kWh. Ada kemungkinan bahwa model mengalami kesulitan memprediksi yang sama seperti pada skenario sebelumnya yaitu pada tanggal 7 November 2014 adalah hari Jumat yang seharusnya merupakan hari kerja, tetapi penggunaan listriknya tidak sesuai dengan hari kerja pada biasanya. Pola dari grafik hasil prediksi sudah mirip dengan data aslinya meskipun puncak data masih berada di tanggal yang salah. Dibandingkan dengan model sebelumnya, prediksi tanggal 1 November yang sebelumnya sudah sangat dekat kembali menjauh. Hal ini dapat disebabkan oleh hadirnya data cuaca pada metode ini yang dapat mempengaruhi hasil prediksi.

### 4.2.3. Metode CNN-LSTM



Gambar 4.5 Grafik *validation loss* model CNN-LSTM

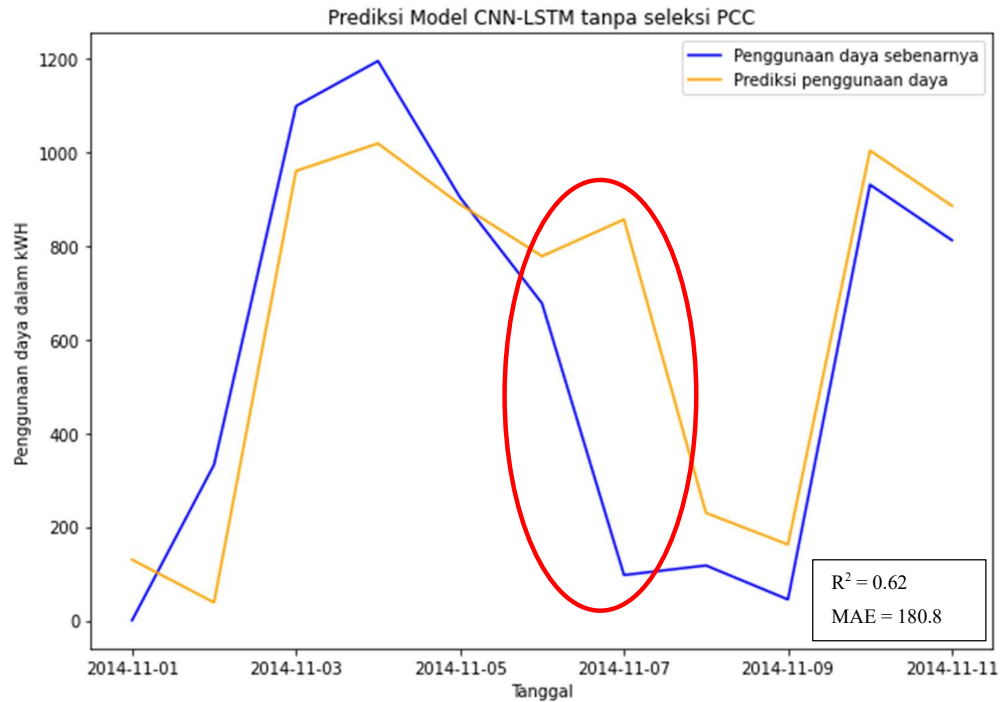
Pada Gambar 4.5 terlihat bahwa tren *training loss* dan *validation loss* menurun dan nilainya lebih rendah dibandingkan dengan model-model sebelumnya. Akan tetapi masih terlihat ciri-ciri yang sama seperti model sebelumnya yaitu fluktuasi nilai *validation loss* akibat dari kurang representatifnya data validasi yang hanya berjumlah 11 hari dibandingkan dengan data *training* yang berjumlah sebanyak 365 hari. Jarak antara grafik *loss* validasi dan *training* semakin mendekat ditunjukkan dengan panah berwarna hijau yang berarti data training sudah lebih representatif dibandingkan dengan pengujian pertama.



Gambar 4.6 Grafik penggunaan daya sebenarnya dengan hasil prediksi model CNN-LSTM

Pada Gambar 4.6 terbukti bahwa model ini terlihat sebagai metode yang paling mendekati dengan penggunaan daya listrik asli. Kesalahan prediksi signifikan yang sebelumnya terlihat pada tanggal 7 dan 9 pada percobaan sebelumnya tidak terlihat lagi pada model ini. Akurasi prediksi pada beberapa tanggal yang ditandai dengan lingkaran berwarna hijau sudah sangat dekat dengan data aslinya. Model ini juga berhasil untuk melakukan hal yang kedua model sebelumnya belum bisa lakukan yaitu memprediksi penggunaan listrik pada tanggal 7 November yang dapat dikatakan diluar normal.

Hasil ini merupakan hasil terbaik dari seluruh skenario pengujian termasuk dengan pengujian dengan tanpa seleksi fitur menggunakan PCC. Jika semua fitur cuaca yang memiliki korelasi rendah dimasukkan ke dalam data *training*, hasil prediksi terbukti lebih buruk dengan skor  $R^2$  sebesar 0.62 dan MAE sebesar 180.8. Hasil prediksi pada pengujian tersebut dapat terlihat pada gambar 4.7 sebagai berikut:



Gambar 4.7 Grafik prediksi CNN-LSTM tanpa seleksi PCC

Dapat terlihat pada gambar 4.7 bahwa terdapat kesalahan yang cukup besar pada tanggal 7 November meskipun pada tanggal lainnya tidak terlalu besar kesalahannya. Kesalahan tersebut cukup untuk membuat model ini menghasilkan skor yang lebih rendah dibandingkan dengan setelah seleksi fitur menggunakan PCC dikarenakan kesalahan pada pola penggunaan dimana seharusnya pada tanggal 7 penggunaan listrik menurun dikarenakan. Menuju akhir pekan namun diprediksi meningkat. Seperti pada dua skenario pertama, tanggal 7 November menjadi tanggal yang paling sulit diprediksi akibat dari penggunaan listrik yang diluar normal pada hari kerja.

#### 4.3. Percobaan Tambahan

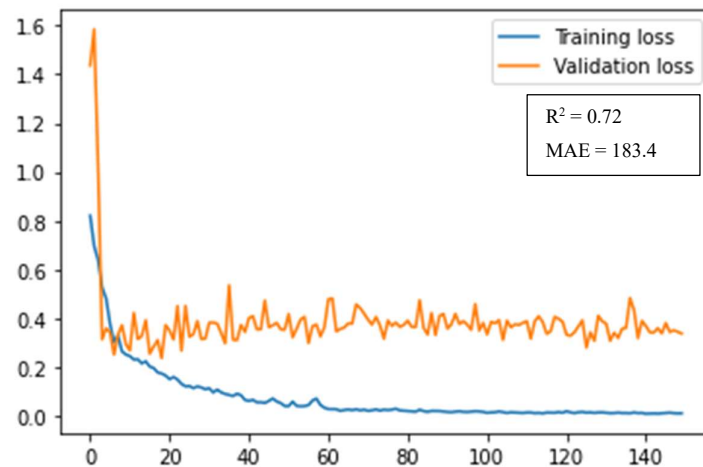
Dengan hasil yang didapatkan dari percobaan sebelumnya, penelitian dilanjutkan dengan mencoba beberapa parameter untuk mencari tahu apakah hasil dari metode CNN-LSTM sebelumnya sudah maksimal. Berikut ini adalah tabel hasil percobaan tambahan yang dilakukan.

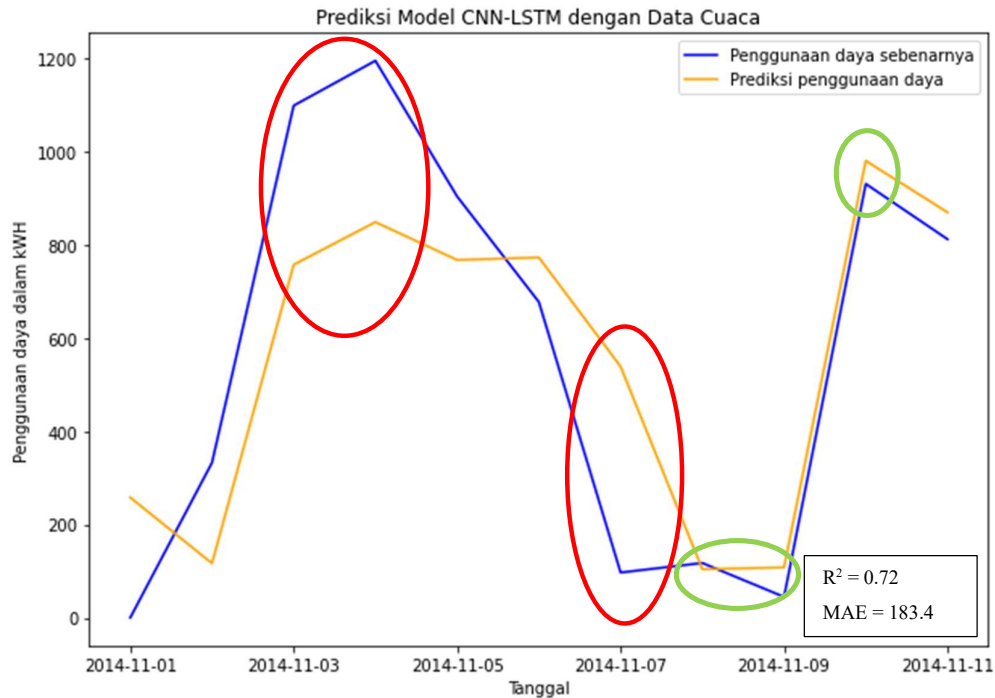
Tabel 4.2 Tabel Skor Hasil Percobaan Tambahan

Model	Skor $R^2$	MAE
CNN-LSTM dengan <i>epoch</i> 150	0,72	183,5
CNN-LSTM dengan <i>lag</i> 14 hari dan <i>epoch</i> 100	0,83	183,72

#### 4.3.1. Pengubahan Jumlah *Epoch*

Pada skenario sebelumnya jumlah *epoch* yang digunakan adalah 50 untuk seluruh skenario. Hal tersebut dilakukan untuk menjaga konsistensi tes dan menjadikan *epoch* sebagai variabel tetap. Berikut ini adalah hasil percobaan metode CNN-LSTM dengan jumlah *epoch* sebanyak 150:

Gambar 4.8 Grafik *loss* CNN-LSTM dengan epoch 150



Gambar 4.9 Grafik hasil prediksi CNN-LSTM dengan epoch 150

Pada gambar 4.8 terlihat bahwa grafik *loss* masih memiliki kriteria yang mirip dengan percobaan sebelumnya. Terdapat jarak antara grafik *loss* validasi dan training yang tidak cenderung mengecil yang menunjukkan bahwa data *training* masih kurang representatif bagi model untuk mempelajari pola penggunaan listrik serta penambahan jumlah *epoch* tidak terlalu berpengaruh terhadap hasil prediksi dan akurasi dari model tersebut. Fluktuasi pada grafik *loss* validasi masih konsisten terjadi pada percobaan ini karena data validasi yang masih berjumlah 11 hari dan dianggap kurang representatif terhadap data *training* yang berjumlah 365 hari.

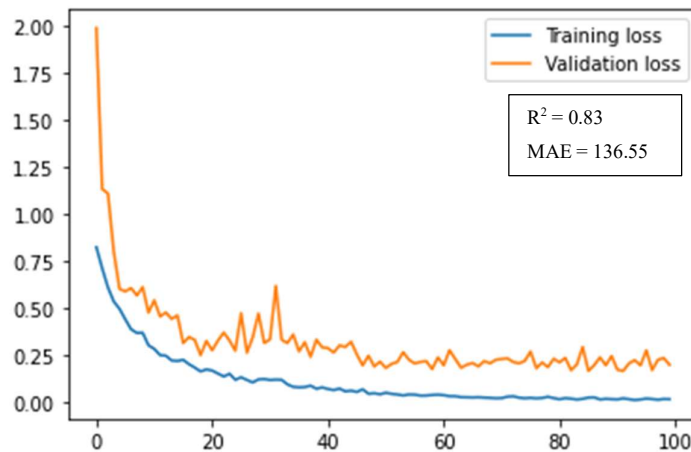
Pada gambar 4.9 juga terdapat beberapa kesalahan prediksi yang cukup besar terlihat pada bagian yang diberikan lingkaran merah. Namun secara umum, model sudah berhasil mengenali pola penggunaan listrik yang naik pada hari kerja dan menurun pada akhir pekan. Pada beberapa tanggal model sudah dapat memprediksi dengan nilai yang sangat dekat dengan aslinya yang ditandai dengan lingkaran hijau. Skor  $R^2$  dari percobaan ini memang lebih kecil daripada percobaan yang menggunakan *epoch* sebanyak



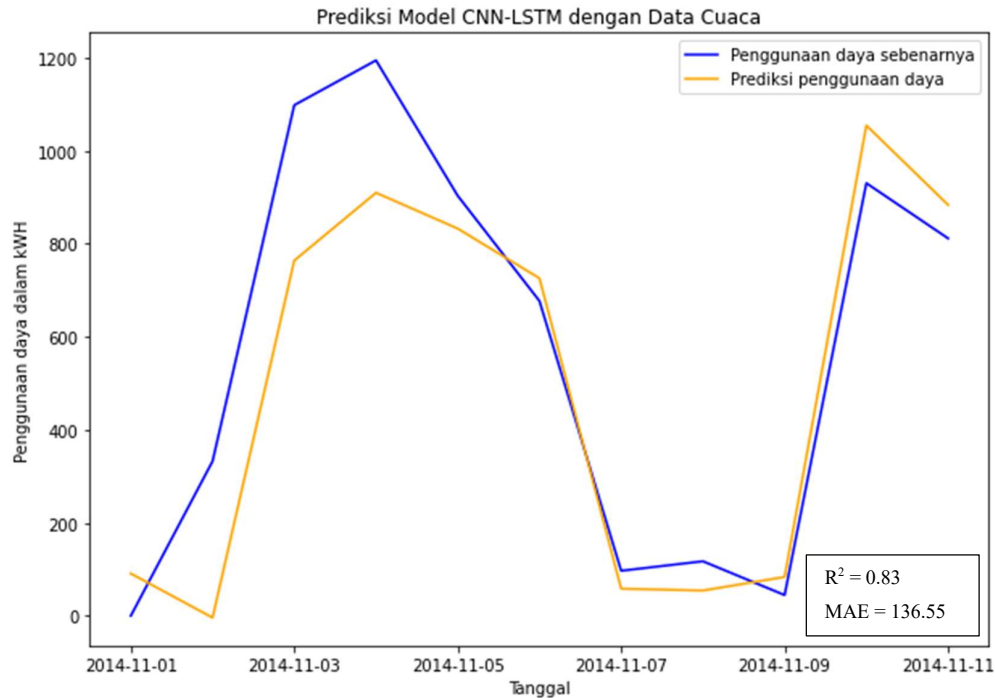
50 yaitu sebesar 0.72. Akan tetapi, skor MAE dari percobaan ini tidak terlalu berbeda jauh dengan sebelumnya yaitu sebesar 183.4, hanya berbeda tiga poin dengan percobaan sebelumnya.

#### 4.3.2. Pengubahan Jumlah *Lag* dan *Epoch*

Percobaan berikutnya adalah percobaan dengan mengganti *lag function* dari sebelumnya tujuh hari menjadi 14 hari dan menambah jumlah *epoch* menjadi 100. Penambahan *lag* menjadi 14 hari bertujuan untuk mengganti pola yang sebelumnya berdasarkan data satu minggu menjadi dua minggu dengan harapan pola dua minggu lebih cocok dengan target data validasi yang berjumlah 11 hari. Berikut ini adalah hasilnya:



Gambar 4.10 Grafik *loss* CNN-LSTM dengan *lag* 14 dan *epoch* 100



Gambar 4.11 Grafik Hasil Prediksi CNN-LSTM dengan *lag* 14 dan *epoch* 100

Hasil dari percobaan terakhir ini menjadi hasil terbaik pada seluruh penelitian ini. Dengan skor  $R^2$  sebesar 0,83 dan MAE sebesar 136,55 terbukti bahwa jumlah *lag* sebanyak dua minggu lebih cocok untuk model ini. Terlihat dari gambar 4.10 bahwa jarak antara grafik *loss* validasi dan training masih ada yang menandakan data training yang masih kurang representatif seperti percobaan sebelumnya. Fluktuasi pada grafik *loss* validasi juga masih terjadi akibat dari kurang representatifnya data validasi yang hanya berjumlah 11 hari dibandingkan dengan data training yang berjumlah 365 hari, meskipun fluktuasinya mengecil daripada sebelumnya akibat dari penggantian fitur *lag* yang menambah dimensi pada data. Pada gambar 4.11 terlihat bahwa pola grafik hasil prediksi pada percobaan ini sudah sangat bagus dan mendekati aslinya, meskipun masih ada beberapa nilai yang kurang tepat seperti pada tanggal 2, 3, dan 4. Meskipun begitu hasil ini merupakan hasil yang sangat baik karena tidak ada error yang signifikan.

## BAB 5

### PENUTUP

#### 5.1. Kesimpulan

Setelah melakukan penelitian dan implementasi beberapa metode serta mengevaluasi hasil dari penelitian ini, dapat ditarik kesimpulan sebagai berikut:

1. CNN-LSTM merupakan metode yang paling akurat dari metode lain yang diuji pada penelitian ini untuk digunakan pada prediksi penggunaan daya listrik gedung institusi pendidikan menggunakan data berbasis *time series* dengan nilai  $R^2$  sebesar 0,83 dan MAE sebesar 136.55
2. Data cuaca seperti data suhu maksimum, rata-rata suhu, kelembaban rata-rata, dan lamanya penyinaran matahari memiliki pengaruh terhadap penggunaan daya listrik khususnya pendingin ruangan pada gedung institusi pendidikan di wilayah tropis.
3. Dengan tambahan data cuaca, Model LSTM dapat meningkatkan akurasi yang terlihat pada skor  $R^2$  yang meningkat dari 0,51 menjadi 0,69 dan juga skor MAE yang menurun dari 217,68 menjadi 180,72.
4. LSTM memerlukan data *time series* yang cukup banyak untuk mempelajari komponen *time series* pada data dan mempelajari polanya.
5. Penggunaan lapisan CNN pada model LSTM dapat meningkatkan akurasi model dengan data *training* yang berjumlah sedikit.
6. Data penggunaan daya listrik selama satu tahun 11 hari belum cukup untuk menciptakan metode prediksi yang stabil.

## 5.2. Saran

Terdapat beberapa saran yang dapat diberikan setelah penelitian ini dilaksanakan, yakni:

1. Melakukan pencatatan penggunaan daya listrik lanjutan selama setidaknya tiga tahun untuk menciptakan model prediksi penggunaan daya listrik yang lebih akurat.
2. Melakukan pengambilan data cuaca secara mandiri pada lokasi yang lebih representatif terhadap cuaca sekitar gedung institusi pendidikan.
3. Melakukan pengujian model menggunakan data baru yang lebih terkini untuk membuktikan hasil penelitian lebih lanjut.
4. Melakukan pengujian dengan data yang lebih divariasikan jumlahnya untuk menilai performa model lebih lanjut.

## DAFTAR PUSTAKA

- [1] E. Mele, C. Elias, and A. Ktena, “Electricity use profiling and forecasting at microgrid level,” 2018, doi: 10.1109/RTUCON.2018.8659866.
- [2] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, “ARIMA models to predict next-day electricity prices,” *IEEE Trans. Power Syst.*, vol. 18, no. 3, pp. 1014–1020, Aug. 2003, doi: 10.1109/TPWRS.2002.804943.
- [3] X. Liao, N. Cao, M. Li, and X. Kang, “Research on Short-Term Load Forecasting Using XGBoost Based on Similar Days,” *Proc. - 2019 Int. Conf. Intell. Transp. Big Data Smart City, ICITBS 2019*, pp. 675–678, 2019, doi: 10.1109/ICITBS.2019.00167.
- [4] X. Song *et al.*, “Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model,” *J. Pet. Sci. Eng.*, vol. 186, p. 106682, Mar. 2020, doi: 10.1016/j.petrol.2019.106682.
- [5] J. Hwang, D. Suh, and M. O. Otto, “Forecasting electricity consumption in commercial buildings using a machine learning approach,” *Energies*, vol. 13, no. 22, Nov. 2020, doi: 10.3390/en13225885.
- [6] R. Adhikari and R. K. Agrawal, *An Introductory Study on Timeseries Modeling and Forecasting*. LAP Lambert Academic Publishing, 2013.
- [7] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986, doi: 10.1038/323533a0.
- [8] aditya.yanuar.r, “Recurrent Neural Network (RNN),” <https://machinelearning.mipa.ugm.ac.id/2018/07/01/recurrent-neural-network-rnn/>, Jul. 01, 2018. <https://machinelearning.mipa.ugm.ac.id/2018/07/01/recurrent-neural-network-rnn/> (accessed Dec. 23, 2021).
- [9] 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.
- [10] S. Saha, “A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way,” *Towards Data Science*, 2018. <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53> (accessed Jul. 15, 2022).
- [11] N. Faustine, “Pengembangan Algoritma GAN Untuk Generasi Citra Wajah Masyarakat Indonesia Dari Citra Sketsa Wajah,” University of Indonesia, 2021.
- [12] X. Deng, C. Zhang, H. Peng, G. Bi, T. Cheng, and C. Zhang, “Short-Term Load Forecasting for Regional Power Grids Based on Correlation Analysis and Feature Extraction,” in *2022 7th Asia Conference on Power and Electrical Engineering (ACPEE)*, Apr. 2022, pp. 1081–1085, doi: 10.1109/ACPEE53904.2022.9783694.
- [13] Le, Ho, Lee, and Jung, “Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting,” *Water*, vol. 11, no. 7, p. 1387, Jul. 2019, doi: 10.3390/w11071387.
- [14] R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio, “How to Construct Deep Recurrent Neural Networks,” Dec. 2013, [Online]. Available: <http://arxiv.org/abs/1312.6026>.
- [15] “SPSS Tutorials: Pearson Correlation,” *Kent State University Library*. <http://libguides.library.kent.edu/SPSS/PearsonCorr> (accessed Jun. 06, 2022).
- [16] “Statistical Correlation,” *Wolfram Mathworld*. <https://mathworld.wolfram.com/StatisticalCorrelation.html>.
- [17] D. Chicco, M. J. Warrens, and G. Jurman, “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation,” *PeerJ Comput. Sci.*, vol. 7, p. e623, Jul. 2021, doi: 10.7717/peerj-cs.623.

- [18] J. Brownlee, *Better Deep Learning*, V1.8. Jason Brownlee, 2020.